JuLIA
Justice, Fundamental Rights and Artificial Intelligence

HANDBOOK
JuLIA (101046631)
DG Just – JUST-2021JTRA

# ARTIFICIAL INTELLIGENCE, JUDICIAL DECISION-MAKING AND FUNDAMENTAL RIGHTS

SSM
Scuola Superiore della Magistratura
Università di Trento
upf. Universitat Pompeu Fabra Barcelona
Funded by the European Union

*Published within the framework of project:*
JuLIA (101046631) "Justice, fundamentaL rIghts and Artificial intelligence"
Co-funded by the European Commission (DG Just – JUST-2021-JTRA)

*Coordinating Partner:*
University of Pompeu Fabra

*Partners:*
University of Trento
Italian School for the Judiciary
Dutch National School of Magistracy
Consejo General del Poder Judicial
University of Paris Saclay (Versailles St Quintin)
University of Groningen
University of Coimbra
Polish Institute of Legal Studies of the Academy of Science
Bruno Kessler Foundation
LIBRe Foundation



First (draft) Edition: September 2023
Second Edition (published): October 2024


Scuola Superiore della Magistratura
ISBN 9791280600592

*Scientific Coordinator of the JuLIA Project:*

Mireia Artigot Golobardes

*Coordinators of the team of legal experts on AI and Justice:*

Gianluca Grasso - Paola Iamiceli

*Project Manager:*

Bruno Castro Coma

*Co-editors and Co-authors of this Casebook:*

Co-editors: *Mireia Artigot i Golobardes, Gianluca Grasso, Paola Iamiceli, Maria Rosaria Maugeri*

Co-authors: *Mireia Artigot Golobardes, Filiberto Brozzetti, Federica Casarosa, Mariavittoria Catanzariti, Maria Giuliana Civinini, Giuseppe Contissa, Antonio Corbo, Tommaso De Mari Casareto dal Verme, Filippo Donati, Paolo Ferragina, Alessia Fidelangeli, Benedetta Galgani, Federico Galli, Andrea Maria Garofalo, Marco Gioia, Gianluca Grasso, Paola Iamiceli, Giulia Lasagni, Marisaria Maugeri, Yannick Meneceur, Gaetana Natale, Nilcan Ozalp, Piera Santin, Francesco Perrone, Laura Piva, Silvio Ranise, Amedeo Santosuosso, Galileo Sartor, Etienne Vèrges, Géraldine Vial*

*Assistant editors:*

Beatrice Marone
Tommaso De Mari Casareto dal Verme

*Note on national experts and contributors:*

The JuLIA team would like to thank all the judges, experts, and collaborators who contributed to the project and to this Handbook by suggesting national and European case law.

# Table of Contents

# Section 1

# Artificial intelligence and justice: what challenges for courts and judicial decision-making?

## 1.1. A methodological introduction[1]

*Mireia Artigot i Golobardes – Universitat Pompeu Fabra*
*Gianluca Grasso – Corte di Cassazione*
*Paola Iamiceli – Università degli Studi di Trento*
*Maria Rosaria Maugeri – Università di Catania*

### 1. Handbook objectives and methodological approach

This Handbook seeks to identify questions more than it seeks to provide answers, to inspire judges and other legal experts rather than to deliver a complete set of information or knowledge on the use of AI in the realm of justice. Merely presenting the *status quo* of the AI debate, as it has currently developed in several communities of scholars, practitioners, and policy makers, stands well beyond the scope of this Handbook.

Building on previous experience in the field of judicial training, this training tool is presented to actual and prospective trainees as guidance and a source of stimuli for participation in training events as well as a starting point for future consideration in their daily practice.

Although the use of AI in the field of justice has not (yet) entered the ordinary practice of judicial decision-making, its impact is real and bears consideration well before its actualization. In fact, critical challenges are posed by existing experiments that seek to empower search engines through AI and machine learning and by testing tools for online dispute resolution based on the use of AI technologies. How are these tools designed, and by whom? What data, i.a., past judicial decisions, feed machine learning? Who should select this data and be responsible for feeding the system at time of development and, later, during its use? How and to what extent could this data be anonymized and by whom? What is an adequate amount of data needed for machine learning? How are biases prevented? Does AI reduce the limits of human rationality in judicial decision-making, its errors and its biases, or does it generate new errors and biases? How can due process and a fair trial be ensured when AI is involved? To what extent can the algorithm be "explained" and support a sufficiently motivated decision? What is the expected output of a given AI-based tool due that is to be used in the field of judicial decision-making? To what extent can we expect judges to preserve their autonomy in decision making once legal support is provided through a

[1] We are very grateful to Beatrice Marone (IUSS Pavia) for her invaluable contribution editing the chapters of this handbook and preparing the Annexes, and to Tommaso De Mari Casareto Dal Verme (University of Trento) for his dedication to complementing this work during a second phase of the project.

(highly "accurate") AI-based tool? Does it reduce their autonomy and their independence? In a nutshell, is the fundamental right to access to justice improved or hampered by AI?

These are some of the questions discussed in this Handbook. Their complexity is further intensified in the context of an ever-evolving legal framework, whose references are still being developed both at the national and supranational levels. Although the focus here is mainly on the EU legal framework and the possible impact of the new 'Artificial Intelligence Act' (hereinafter the AI Act), attention will also be paid to the soft, and more recently, hard law instruments developed by the Council of Europe (the most recent being the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, also signed by the EU on the 5th of September 2024), as well as to national approaches that have been or are currently emerging (e.g. see the reform proposals at the national level in Spain, "*Proyecto de Ley 121/116 de medidas de eficiencia digital del servicio pùblic digital de la justicia*," and the Italian legislative proposal on "*Norme di principio in materia di intelligenza artificiale,*" whose article 15 specifically regulates the use of AI in the field of justice - AS 1146, pending). Are such regulatory interventions due to stimulate or challenge the further development of AI in the field of justice? In particular, will qualifying certain AI systems as high-risk (such as those used in the administration of justice, under Annex III of the AI Act) hamper or steer their design and possible use? Which type of human oversight should be expected by deployers of AI systems in the field of justice in light of Art. 14 of the AI Act?

The methodological approach of this Handbook is necessarily interdisciplinary. Legal scholars, judges, and IT experts contribute to a multidisciplinary dialogue that aims to make judges aware of the core technological features of AI instruments so as to better understand both the strengths and limitations of AI applied to judicial decision-making, i.e., the emergence of biases, their impact on decision-making, and their possible defeat may be fully considered only by dealing with the technical design of AI tools.

As it is directed at judges and other legal practitioners – including lawyers – this Handbook focuses on the practices and general principles capable of steering such practices towards a more effective protection of fundamental rights. A practical approach is favoured over a more theoretical outlook. A list of essential bibliographic references is however provided in Section 8 to facilitate further analysis.

The presentation of a selection of experimental practices testing AI-tools in the field of justice provides a basis of discussion for judges and legal practitioners regardless of whether they've already had personal experience with the use of such technologies. In some instances, this "hands-on" approach should enable trainers to concretely design and trainees to more actively participate in training events, raising their awareness about experimental projects and involving them in discussion of real and hypothetical cases. Both the strengths and limits of these techniques are considered together with a case-based examination of possible challenges generated by the use of AI in the field of justice. Lacking a material stream of litigation in this field in the EU – adjudicating cases of the misuse of AI by judges and legal practitioners – examined cases are mainly hypothetical. However, guidance for training is offered based upon existing case-law as it has emerged both at the EU and national level and in other contexts, such as public administrations or markets.

As a front runner within the JuLIA Project on "Justice, fundamental rights, and artificial intelligence," this Handbook is meant to establish connections and complementarities with the

forthcoming JuLIA Handbooks on AI and public administration, AI and health, and on AI and markets. Following a common methodological approach (interdisciplinary, practice-oriented, focused on the emerging EU legal framework in dialogue with national approaches), these Handbooks will address a few horizontal issues such as: the impact of AI on decision-making, both by individuals and organisations, and the consequences to fundamental rights. Among the latter, special attention will be paid to data protection and non-discrimination as the principal fundamental rights concerned by the use of AI in all relevant contexts. With regard to data protection, an *ad hoc* Handbook will build on the data protection modules of the sectoral Handbooks to provide horizontal and more comprehensive guidance to one of the pillars of fundamental rights protection related to AI.

As a tool for active training, this Handbook is the result of continual growth and updates along with the training experience. It benefited from the contributions of judges before and after their participation in training events, as well as from further integrations and complementary contributions: as in previous judicial training projects (see *Re-Jus, FRICoRe*), mutual learning is critical.

## 2. The Handbook structure

This Handbook is divided into eight sections, including this introduction (**Section 1**) and a set of annexes (**Section 8**) which encompass lists of relevant regulatory instruments, case law and opinions, along with essential bibliographic references.

**Section 2** seeks to provide a technical and interdisciplinary overview of the use of AI tools in the domain of justice. The first contribution (**2.1**) provides for a comprehensive analysis of the relationship between justice and new technologies, as well as the potential benefits and risks of the ways technological innovation operates in, or will integrate into, the domain of the judiciary in the near future. In doing so, it investigates the Council of Europe's perspective and activities related to the topic, together with a presentation of the structure and results of a research project carried out in Italy by a joint multi-stakeholder effort. The second contribution (**2.2**) instead introduces analysis of the use of AI in the field of justice from a technical perspective, providing a basic understanding of the main features of AI tools and machine learning algorithms as they might possibly be used in the context of judicial decision-making. Not only is the process underlying the functioning of AI tools concisely presented but the basis for risk analysis is provided in order to guide users towards the objective of what is often called "responsible AI." An example discussing AI-based guidance for unfair terms assessment will substantiate the analysis. Finally, the third contribution (**2.3**) further complements this analysis by looking at one of the first use-cases of AI in the field of justice: legal search and search engines. Search engine development based on the use of AI represents a vivid example of how AI may support and complement the activity of judges and legal practitioners and how legal and technical experts can jointly contribute to improving the accuracy and efficiency of research and legal materials analysis.

**Section 3** provides an overview of the evolving supranational and international legal framework at the European level. The analysis begins (**3.1**) from a soft law perspective and discusses the enormous work conducted by the Council of Europe which resulted in the "European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their

environment" adopted in 2018, the subsequent toolkit for supporting its implementation, as well as the implications stemming from the recent adoption of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law in 2024. It continues (**3.2**) with a brief overview of the recently adopted EU Regulation laying down harmonised rules on artificial intelligence (AI Act), by presenting its general architecture and highlighting the aspects of major interest when AI is used in the context of judicial decision-making.

**Section 4** explores the impact of the use of AI in judicial decision making upon fundamental rights and the general principles that should be respected to ensure effective protection of the same. The first contribution (**4.1**) looks specifically at several AI instruments designed to support legal research and judicial decision-making – advanced precedent search systems, chat boxes and tools that automatically indicate the outcome of a dispute, or that automatically decide on a specific point in a dispute – with the aim of assessing how they can be used in compliance with fundamental rights and fundamental principles of civil procedure such as the principles of non-discrimination; transparency; due process; certainty; diligence; independence and impartiality. The second analysis (**4.2**) specifically examines one of the fundamental rights particularly hampered by the use of AI: data protection. It sheds light on how the General Data Protection Regulation (GDPR) has been applied in the field of AI technology and on the possible implications of the AI Act. The third contribution (**4.3**) tackles the theme of algorithmic discrimination by focusing on criminal justice. It analyses the risks related to the use of AI in this field for human rights, such as the right to a fair trial, the presumption of innocence, and the principle of non-discrimination. It also analyses the existing remedies against possible unlawful outcomes of AI systems used in criminal justice that are provided by EU soft law and proposed by the new EU Regulation.

**Section 5** adopts a "hands-on" approach by offering a description of some transnational research projects on AI and justice, funded or co-funded by the European Union, aimed at developing experimental tools in this field. In the first contribution (**5.1**) the CrossJustice project is presented, which combines legal analysis of the shortcomings and obstacles to the exercise of procedural rights under EU and national law with a legal-informatics approach. The latter consists of the translation of the EU Procedural Directives, as well as samples of national legislation, into a computable language. This method has become the core of the CrossJustice online platform. In fact, the project resulted in the development of a Legal Database and an Advisory Module, both freely accessible online. The FACILEX project is then presented, which aims at strengthening the implementation and application of the *acquis* on judicial cooperation in criminal matters with the help of digital tools, not limited to mere fact-finding research. The second contribution (**5.2**) describes the ADELE project that has developed artificial intelligence and legal analytics (LA) methods to support legal research and decision-making processes in the judiciary. The project focused on two legal areas: value added tax (VAT) and trademarks and patents (T&P).

**Section 6** narrows the legal analysis from the supranational context to some selected national experiences dealing with the relationship between AI and justice, specifically in France and Italy. The first contribution (**6.1**) deals with the state of the art in the use of AI in the French legal framework and the results of several research projects carried out within the context of the French judicial system are explained. In particular, the contribution investigates the existing tools for assisting judges; the open data of court decisions, its impact on the notion of case-law and

the search for relevant decisions for judges or lawyers; metrics and predictive justice; and the impact of algorithms on judges' decisions in French law. The Italian experience is reported in a broader subsection (**6.2**) composed of four different contributions preceded by an introduction. It seeks to draw a path from the role of the digitalisation of justice in the Italian legal system, through the presentation of the recent Italian draft law on AI, to the frontier topic of the Metaverse when it comes to courtrooms. The first contribution (**6.2.1**) highlights how the Italian justice system is transitioning from merely digitizing documents to leveraging digital tools and AI for judicial decision-making, by moving from traditional documents to data as enabled by recent, but certainly not definitive, technological advancements. It also analyses the role of digitization within the latest Italian reform of civil procedure law. The second contribution (**6.2.2**) introduces the Electronic Documentation Centre (CED) established at the Italian Court of Cassation, which is exploring the use of AI in judicial activities, with particular regard to key elements such as the pre-determination of decision-making criteria and the analysis of documents using syntactic sequences related to text structure and the statistical frequency of linguistic expressions. The third contribution (**6.2.3**) provides a reconstruction of the recent Italian bill on Artificial Intelligence, by analysing its main provisions and their relevance to the field of justice. Finally, the fourth contribution (**6.2.4**) begins again from the Italian reform of criminal procedure examined from the lens of remote justice which leads to the description of a new digital paradigm, the so-called "Metaverse," and its entry into the field of justice.

**Section 7** follows a practical approach by offering a set of training materials. Although the use of AI in justice has rarely been directly addressed by case law, inspiration is found in a set of judgements by European and national courts concerning cases in which some principles on the limits of AI, based on fundamental rights considerations, have been elaborated in a way that could be relevant in the field of justice. However, this section is meant to foster discussion within training sessions without suggesting any immediate transplants of decisions across sectors: another issue worth debating. Bearing this in mind and beginning from a recent UK court's decision, the first contribution (**7.1**) treats the use of predictive algorithms by the parties and the judge during so-called e-discovery, addressing the extent to which errors in the use of AI tools risks impacting the parties' and/or the court's ability to provide documentational evidence. The second contribution (**7.2**) highlights important elements in the famous SCHUFA Holding and Others (C-634/21) case where the CJEU was asked to evaluate whether and to what extent an AI credit scoring system is qualified to be an automated decision-making mechanism, pursuant to Article 22 of the GDPR, even where the decision-maker retains some (abstract) power to deviate from the AI suggestion. This is an issue that could, at least in principle, mirror an equivalent mechanism for a court to use an AI-system to complement its own decision-making process. On the other hand, at the national level, courts have predominantly addressed AI-based cases related to the use of predictive or mass surveillance tools adopted by law enforcement authorities, among which is an interesting and recent example examined in the decision of the German Federal Constitutional Court on the use of Palantir surveillance software by the police. The third contribution (**7.3**) discusses a case decided by the Italian Supreme Court in 2021 regarding a reputational rating to investigate problems with data protection when it comes to the use of AI in decision-making, leading to relevant considerations on the right to consent, transparency, and dignity in automated decision making, which may be useful outside of the specific context examined. The fourth contribution (**7.4**) examines a decision by the Italian Data

Protection Authority that has urgently ordered OpenAI LLC to limit the processing of the personal data of subjects in the Italian territory through ChatGPT, pursuant to article 58, paragraph 2, letter f of the GDPR. In the fifth contribution (**7.5**) a case decided by an Italian court in 2020 is the starting point for an analysis of various uses that algorithms can perform in decision-making, with a particular focus on the right to equality and non-discrimination, and of the risks for its protection due to biases both in the system and within the context in which the algorithm is deployed.

# 1.2. The use of Artificial Intelligence in judicial systems: ethics and efficiency

*Filippo Donati – Università degli Studi di Firenze*

*Summary: 1. Introduction – 2. The ongoing development of AI applications in the field of justice – 3. The AI Act – 4. The use of AI tools in legal analysis and decision-making by judges – 5. Concluding remarks*

*Abstract: The subject "Artificial Intelligence (AI) and justice" can be addressed in two ways. First, the use of AI systems may be the object of judicial proceedings, when decisions under judicial scrutiny have been made by algorithms operating without guidance from humans. This is the case, for instance, in disputes on damages caused by driverless cars, drones, or automated disease diagnosis and treatment systems. In such cases, consolidated legal principles on the law of evidence, on the quantification of damages, and on liability may prove difficult to apply.*
*Second, AI systems may be used to assist the judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts. Several companies currently provide new automated tools for due diligence exercises, for drafting documents, and for technical assessments, including the calculation of maintenance allowances for spouses or children or damages in the event of personal injury. Predictive AI systems may also be used to determine the possible outcome of a current or potential dispute. Such AI systems, currently used by law firms or insurance companies, could provide new tools to increase the efficiency of the judicial system. A "robotic" justice, it has been noted, may better accomplish the need for legal certainty. However, the use of AI in the judicial field raises a set of open questions. In the intervention, this latter profile is addressed, with the intent of suggesting some thoughts on the opportunities and risks that may derive from the use of AI by the judiciary.*

## 1. Introduction

The subject "Artificial Intelligence (AI) and justice" can be addressed from two different perspectives: AI as the object of judicial proceedings, on one side, and AI as a support tool to judges in the exercise of jurisdiction, on the other.

It is not difficult to foresee how an increasing number of disputes will regard the use of AI systems in the future. This is the case, for instance, in claims for damages caused by driverless cars, drones, or automated disease diagnosis and treatment systems. In such cases, the primary issue is whether, and to what extent, consolidated legal principles on the law of evidence, for the quantification of damages and liability, which traditionally refer to human behaviour, can be extended to robotic behaviour.

The use of AI systems may also trigger a different set of issues, when used to assist judicial authorities in exercising their jurisdiction. New automated tools for the exercise of due diligence, for drafting documents and for technical assessments, including the calculation of maintenance allowances for spouses or children, or damages in the event of personal injury, are currently available on the market. Law firms and insurance companies increasingly use predictive AI systems to determine the possible outcome of a current or potential legal dispute. Why not use the same tools, then, to increase the efficiency of the judicial system? As a matter of fact, the use of AI systems may help to increase the quality and efficiency of justice. At the same time, however, the use of AI in the judicial field raises a set of new and open questions.

This second perspective will be addressed in this intervention as well as some thoughts on the opportunities and risks that derive from the use of AI in the domain of justice.

## 2. The ongoing development of AI applications in the field of justice

In Italy and in most member States of the European Union, the digitalization of justice is completed or is nearing completion. Digitalization regarding communication, filing, and the exchange of documents has resulted in greater simplification for users and a strong contribution to the increased efficiency of judicial offices. Furthermore, the possibility of holding online hearings allowed trials to be conducted in oral form even when the pandemic prevented physical access to courtrooms. Digitalization has also allowed the creation of large digital databases for collecting judicial decisions, an indispensable prerequisite for the development of AI systems.

Despite trials underway in some countries, including Estonia, Canada, China, the United Kingdom and the United States, justice structures in most countries still make little use of AI systems. In fact, the features of AI systems appear incompatible with a set of fundamental principles applied in the field of justice, including transparency and the justification of judicial measures, the right to defence and cross-examination. Furthermore, it has been established that AI systems can be biased and produce errors and discrimination, resulting in the infringement of human rights. The case of COMPAS, an AI program designed to assess the risk of potential recidivism, is well known. This program, used by certain US courts, was found to be discriminatory because it tended to attribute a greater risk of recidivism to certain people in relation to the colour of their skin and the social environment of reference.

However, AI could also contribute to solving the problems and inefficiencies that afflict justice today, especially in terms of reducing excessively lengthy judicial proceedings that undermine the right to a fair trial. Additionally, the lack of a sufficient uniformity and predictability of judicial decisions undermines legal certainty in many countries. AI could speed up the delivery of judgments and ensure more predictable trial outcomes.

The application of AI, in substance, entails both risks and advantages at the same time. AI must be considered not only a threat, but also a tool to improve people's lives and their enjoyment of fundamental rights. However, AI systems are developing at an increasingly rapid pace and so it cannot therefore be excluded that, over time, problems like those regarding the opacity of AI systems, i.e., the black box effect, might be mitigated or overcome thanks to technological progress. The option of whether to allow the use of AI systems in the judicial sector, therefore, must not reflect a choice between ethics and efficiency. To the contrary, a human-rights perspective on the development and use of AI is possible and desirable.

The Ethical Charter on the use of AI in judicial systems, adopted in 2018 by the European Commission for the Efficiency of Justice (CEPEJ), identified the core ethical principles to be respected in the field of AI and justice: respect of fundamental rights, non-discrimination, the quality and security of data processing, transparency, impartiality and fairness, and human control. The Ethical Charter is based on the idea that AI, if used as a tool not for replacing, but for assisting judges, can promote the efficiency and quality of justice. Judges' autonomy must be increased and not restricted by AI tools and services.

The European Union, in its policy on AI, has followed the same approach. The European Commission's proposal for a regulation laying down harmonised rules on AI (the AI Act), published as Regulation (EU) 2024/1689 on June 13, 2024 after more than two years of work, analysis, and discussions, clearly states that AI should not substitute human autonomy or limit

individual freedom. Additionally, the AI Act aims at introducing safeguards to ensure the development and use of ethically embedded AI that respects Union values and human rights.

## 3. The AI Act

The AI Act follows a risk-based approach that, in order to introduce a proportionate and effective set of binding rules for AI systems, tailors the type and content of such rules to the intensity and scope of the risks that AI systems can generate. It therefore prohibits AI systems that pose unacceptable risks for fundamental public interests as recognised and protected by Union law, including fundamental rights, democracy, the rule of law, or the environment. The prohibition covers practices that have significant potential to manipulate persons through subliminal techniques beyond their consciousness or that exploit vulnerabilities of specific groups, such as children or differently abled persons, in order to materially distort their behaviour in a manner that is likely to cause psychological or physical harm. AI-based social scoring for general purposes carried out by public authorities is also prohibited, as well as the use of "real time" remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement.

For systems that entail limited risk, such as chatbots, the AI Act requires transparency obligations aimed at making users aware they are interacting with a machine. Free use is permitted for minimal-risk AI systems, such as AI-enabled video games or spam filters. High-risk AI systems that may create a high risk to human rights are subject to strict regulation that requires conformity assessment, certifications, registration obligations, and ex post controls. The classification of an AI system as high-risk is based on its intended purpose. The AI Act classifies as high risk those systems that are "intended to be used by a judicial authority or administrative body or on their behalf to assist a judicial authority or administrative body in researching and interpreting facts and the law and in applying the law to a concrete set of facts or used in a similar way in alternative dispute resolution." Therefore, AI systems at the service of justice shall comply with the strict regulation imposed by the AI Act.

## 4. The use of AI tools in legal analysis and decision-making by judges

It is worth noting that, as underlined in recital 41 of the AI Act, the fact that an AI system is classified as high-risk does not indicate that the use of the system is necessarily lawful or unlawful under other acts of Union law or under national law compatible with Union law, such as on the protection of personal data. Any such use is permitted to the extent that it complies with the "applicable requirements resulting from the Charter and from the applicable acts of secondary Union law and national law."

Several fundamental principles enshrined in national constitutions, the European Convention on Human Rights (ECHR) and in the Charter of Fundamental Rights of the European Union (CFREU) prevent AI systems from replacing human judges. As a matter of fact, a robot judge would affect the constitutional guarantees related to jurisdiction, such as the right to a fair trial, the parties' right to a defence, and the obligation for judicial rulings to state the reasons upon which they are founded.

Although AI cannot fully "replace" a human judge at present, it may still be useful in the courtroom in various ways. AI systems could provide more powerful search engines for

improving the research for court decisions and other legal texts. Additionally, AI tools may also aid judges with technical evaluations, such as calculating indemnity against unfair dismissals or maintenance allowance in divorce cases. AI can be used to analyse evidence, translate languages, assess factual data, as well as prepare draft measures or deal with simple, serial, repetitive, entirely documentary cases. Finally, AI systems can be used in alternative dispute resolution procedures, particularly those involving small claims that are seldom asserted before a judge. In such cases, effective legal protection of fundamental rights requires the provision of online platforms which, through AI systems, can offer inexpensive, rapid, and reliable forms of dispute resolution, that do not exclude recourse to judicial protection. It is therefore no coincidence that the use of algorithms in the judicial field is spreading in many countries, particularly in the USA, China, Canada, and the UK.

However, many scholars today still seem highly sceptical about the use of AI tools by judges. The problem lies in the risk of the so called "*effet moutonnier*" (sheep effect), which may lead the judge to avoid responsibility and simply follow the algorithm's advice. As a matter of fact, the risk of a judge being captured by the algorithm cannot be underestimated. AI support may relieve the decision-maker from the burden of motivation and may help to qualify decisions with a veneer of "scientificity" and "neutrality" which today surrounds algorithmic evaluation and gives it a peculiar – yet unfounded – authority. The risk is that the advice provided by an AI system will be blindly followed by a judge, without further autonomous assessment of the peculiarities of the case and of the applicable law.

Such risks should be avoided. The autonomy of the judge, who is solely responsible for the interpretation of the applicable law and the evaluation of the peculiarity of the case in question, cannot be limited. It is therefore essential that, as the Wisconsin Supreme Court ruled in the Loomis case, a judge must maintain full autonomy of judgment and not base their decision exclusively on the indications derived from AI. It is therefore worth noting that, pursuant to the AI Act, high risk systems, such as those that may be used to support judicial authorities, must be designed and developed in such a way that natural persons can oversee their functioning. Human oversight shall aim to prevent or minimise the risks to fundamental rights that may emerge in the use of such systems.

## 5. Concluding remarks

The use of AI in the service of justice is possible and desirable, provided it is made in compliance with the applicable ethical and legal principles.

A fundamental role in the success of the AI Act will be played by the authorities entrusted with the power to enforce its provisions. High-risk systems will be permitted after being subjected to an ex-ante conformity assessment carried out by conformity assessment bodies designated and monitored by national authorities. Ex-post supervision of the function of such systems by competent authorities will then follow. To this end, the AI Act sets up a dedicated governance system at the Union and national levels. At the Union level, a European Artificial Intelligence Board, composed of representatives from the member States and the Commission, will be established. At the national level, member States will have to designate one or more national competent authorities and, among them, a national supervisory authority, for the purpose of ensuring the application and implementation of the AI Act. Such national competent

authorities "shall have a sufficient number of personnel permanently available whose competences and expertise shall include an in-depth understanding of artificial intelligence technologies, data and data computing, fundamental rights, health and safety risks, and knowledge of existing standards and legal requirements" (Article 59(4)).

In this respect, the difference between the high-risk systems listed in Annex III of the AI Act cannot be underestimated. The requirements of AI systems intended to be used for the recruitment or selection of natural persons, for example, may not be identical to those intended to assist judges in the exercise of jurisdiction. In addition, independence of the judiciary from undue external interference is a prerequisite of the rule of law, which is one of the founding values of the European Union (Article 2 TEU).

In the domain of justice, sound technical knowledge of ethical and legal principles applicable to jurisdiction, along with the need to avoid undue interference by economic or political power, is therefore necessary. This means the judiciary should be involved and have a voice in the assessment and monitoring procedures of those AI systems intended to be used in support of jurisdiction.

The judiciary cannot miss the opportunity to make use of new technologies available today and in the future. AI may help promote the quality and efficiency of justice. When using AI systems, however, human control remains necessary. Judges' autonomy cannot be restricted by AI systems. In addition, issues regarding opacity, complexity, bias, unpredictability and partially autonomous behaviour of certain AI systems must be duly addressed, in order to ensure their compatibility with fundamental rights.

The judiciary may well contribute to the assessment and monitoring of AI systems to be used in support of jurisdiction. AI, therefore, is a great opportunity and, at the same time, a great responsibility for the judiciary.

# Section 2

# The perks of using AI tools in the justice domain

## 2.1. New technologies and justice

*Maria Giuliana Civinini – Council of Europe*

*Summary: 1. Introduction – 2. Judges coming to terms with technology – 3. Opportunities, challenges, and risks – 3.1. Opportunities – 3.2. Challenges – 3.3. Risks – 3.3.1. "Do we want to use these systems?" – 4. Applying new technologies in the Justice sector – 5. Decision Support Systems – 5.1. DSSs in the Justice sector: many projects and few applications – 5.2. "Question zero" and the foreseeability of judicial decisions – 5.3. Judging today with an eye to the future – 5.3.1. Tools tailored to judges' specific needs – 5.3.2. The participation of justice actors. A virtuous example: the Pisa pre-totype – 6. Before concluding – 7. (Initial and partial) Conclusions*

*Abstract: The theme of this contribution is to investigate the relation between justice and new technologies and the consequent effort to discover how innovation in technology currently operates in the domain of the judiciary or will integrate into it in the future. The aim is to carry out a comprehensive analysis of both potential benefits and possible threats, along with a suggestion of the most efficient way of dealing with them. After a thorough explanation of the Council of Europe's perspective and activities related to the topic, a particularly relevant research project carried out in Italy by a joint multi-stakeholder effort is examined in which both the structure and results are highlighted.*

### 1. Introduction

Judges can confront technological change and Artificial Intelligence on two different levels. They can judge cases in which the use of AI is at stake, i.e., a medical liability case for AI-supported diagnosis, automated drive liability cases, or cases on the use of algorithms and AI by public authorities in decision making processes. Owing to a delay in the adoption of significant legislation, AI-related cases are increasing in number and relevance, and courts will play a pivotal role in designing the framework around AI as well as the limits and requirements for its legitimate use.[2]

[2] O. POLLICINO, *Judicial protection of Fundamental Rights on the Internet*, Oxford, 2021; M. SUKKSI (ed.), *The Rule of Law and authomated decision-making*, Springer, 2023. One of the first judicial decisions in the world on the use of AI by the Public Administration is the District Court of the Hague, 6 March 2020, ECLI:NL:RBDHA:2020:865, available in English at: uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878 , on which see: A RACHOVITSA, N. JOHANN, *The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case,* Oxford, *Human Rights Law Review*, 2022, 22, 1–15, https://doi.org/10.1093/hrlr/ngac010; A. MEUWESE, *Regulating algorithmic decision-making one case at the time. Case note on: District Court of The Hague*, ECLI:NL:RBDHA:2020:865 (NJCM vs the Netherlands (SyRI)) in *European Review of Digital Administration & Law*, 2020, 1, 1, 209; M. VAN BEKKUM, F. Z. BORGESIUS, *Digital welfare fraud detection and the Dutch SyRI judgment* in *European Journal of Social Security*, 2021, 23, 4, 323. https://doi.org/10.1177/13882627211031257; S. BEKKER, *Fundamental Rights in Digital Welfare States: The Case of*

At a stage prior to, but profoundly related to, the exercise of jurisdiction, judges are coming to terms with technology at the organizational level, in court and case management, as well as the creation and management of case-law archives. For some time now, the digitalization of justice has been identified as one of the key elements for ensuring access to justice, speedy judgements, uniformity, and the quality of decisions. While judges are familiar with making decisions on new situations and new rights emerging from social, scientific, and technological progress (in fact, it can be said that the ability to adapt existing legislation to new cases through interpretation, that is not yet regulated by the legislature, is at the core of the judicial function), the impact of innovation on judges' work and decision-making has still not been sufficiently investigated.[3]

The relationship between justice and new technologies will be focused on in this section, attempting to explore – or at least to break ground on – the areas of penetration of technological innovation in the judiciary, its benefits, possible risks and risk mitigation tools. However, topics of evidence generated by AI tools and the use of predictive policing will not be covered.

This research, still in its infancy, is based upon several assumptions, further developed below:

- the introduction of technology in the administration of justice seeking to improve quality and efficiency cannot be reasonably opposed;
- judges cannot seclude themselves in their ivory tower and live remotely from a world in which professionals and young people are using technology. The legitimacy and perception of the judiciary would be severely impacted as well as the role of justice. Digitization is not neutral and permeates both the judicial organization and the way of doing justice and being a judge;
- a technological transformation of justice in compliance with the rule of law and fundamental rights is only possible if judges take a leading role in the process.

## 2. Judges coming to terms with technology

The swirling evolution of technology and the splashy debut of Generative AI have only rippled the waters of the sea of Justice despite the efforts of Judicial Training Institutes and the European Judicial Training Network (EJTN) which now regularly devote training actions to the topics of digitalization and AI. However, there is no real debate within European judiciaries on the impact of innovation and the possible use of AI tools in the exercise of jurisdiction. The discussion, if any, lingers on anecdotal profiles – i.e., the US lawyer who turned into a cat during

*SyRI in the Netherlands,* in O. SPIJKERS, W.G. WERNER, R.A. WESSEL (eds) *Netherlands Yearbook of International Law 2019*, T.M.C. The Hague, 2021, 289. DOI: 10.1007/978-94-6265-403-7_24

[3] M. ZALNIERIUTE, F. BELL, *Technology and the judicial role*, *forthcoming* in G. APPLEBY, A. LYNCH (eds.), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia*, Cambridge, 2020; P. W. GRIMM, M. R. GROSSMAN, S. GLESS, M. HILDEBRANDT, *Artificial Justice: the Quandary of AI in the courtroom*, Judicature International, 2022; C. AGUZZI, *Le juge et l'intelligence artificielle : la perspective d'une justice rendue par la machine,* in *Annuaire international de justice constitutionnelle 2019*, 2020, 35, 621, doi: https://doi.org/10.3406/aijc.2020.2794; https://www.persee.fr/doc/aijc_0995-3817_2020_num_35_2019_2794; Y. MENECEUR, C. BARBARO, *Artificial intelligence and the judicial memory: the great misunderstanding*, in *AI and Ethics*, 2022, 2, 269; https://doi.org/10.1007/s43681-021-00101-z; F. DE STEFANO, *L'intelligenza artificiale nel processo*, in *Giustizia Insieme*, March 6, 2020; EUROPEAN COMMISSION, *Study on the use of innovative technologies in the justice field – Final Report*, September 2020.

a videoconference; the Colombian judge who used ChatGPT and issued a wrong decision; the US lawyer who quoted non-existent case law suggested by ChatGPT – only to arrive at a reassuring confirmation of traditional rites and ways of judging. The weakness of policies for the digitalization of justice, the non-allocation of adequate funds, and the lack of integration into the overall strategy for the digitalization of the public administration result in a slow and unimpressive digital transformation of justice while the application of AI remains in its infancy.[4]

This situation and the tolerance often accorded to judges' rejection of digital systems have resulted in a reinforcement of the idea that justice can remain a "non-technological island." A desire which stems from the concern that automation could seriously jeopardize the independence and impartiality of the judge, rendering the judge's work to mere case management and the systematizing of decisions.

However, even in areas as strongly shaped by tradition as justice, new technologies can provide crucial support to improving quality, effectiveness, efficiency, accessibility, and transparency.

The whirlwind development of technology and fast-growing AI – perhaps one of the most important breakthrough technologies of our times – will bring about important changes to many aspects of life, and the legal sector will not be immune. A clear understanding of the benefits and risks of technological innovation as well as the areas to which it could be applied in the justice sector would put judges and judicial institutions in a position to govern the process.

## 3. Opportunities, challenges and risks

The European Commission and the Council of Europe have been devoting great attention to the topic of innovation in the justice sector for some years now.

[4] A recent EJFRI - e-Justice & Fundamental Rights International - *Survey on the digitalization of justice and the use of artificial intelligence in the Judiciary in the EU Member States* (2022) confirms what is stated in the text. The survey target group were persons competent at the Ministries of Justice, the high courts, court administrations or distinguished legal scholars. The report is based on answers to the questionnaire from 19 EU member States. All participants responded positively to the question of whether digital tools were being used in the justice sector in their country and, in 75% of cases, that there was a digitalization strategy for justice at different stages of implementation, mostly in the planning stage. When examining the digital tools actually used, it emerged that "a quite heterogenous picture, though electronic tools for communication (authority/authority or authority/party) are predominant, while … case management systems are still, to a considerable extent, paper based. Electronic registers are quite broadly used, while solutions like judicial app-stores or microservices are more the exception than the rule. Even at first sight, the answers in the field of Artificial Intelligence disclose that, on the one hand, EU member state judicial systems seem to be quite reluctant when it comes to using AI solutions in their systems. …Concerns in terms of the reduction of the human factor and the lack of transparency of the functioning of these models as well as data protection delay progress in this area. This background is reflected by answers received which, at the moment, indicate very little use of these solutions. Supportive technologies in administrative-related issues seem to be more accepted than technologies intervening in the judicial decision-making process. Developments related to big data and respective storage solutions does not seem to be at the top of the agenda of EU member states, though the experts we talked with were aware of the importance of a transition from mainly server- to cloud-based technologies with respect to scalability and costs. Concerns about data protection and IT-security are predominant in this field. … The majority of the EU member states have discovered the power of statistical data and their visualization on dashboards to optimize the efficacy of their systems and to base their respective policies on a strong data pool" (*Survey Executive Summary*, 6).

Since 2016, the European Commission for the Efficiency of Justice (CEPEJ) has developed important reference texts concerning the implementation of digital justice in the judicial systems of the member States of the Council of Europe.[5]

Acceleration towards further development has also been fueled by the health emergency. The Covid-19 pandemic caused significant damage to the functioning of justice in all countries affected by the virus. The justice system was reduced and even stopped its activity across Europe putting its credibility and the trust of people at serious risk. The Covid-19 crisis showed that only judicial systems equipped with the technological tools of *e*-justice were able to guarantee the handling of cases, particularly in the fields of Civil and Commercial law. Thanks to the dematerialization of the file, intelligent case management systems, and telematic notifications, it was possible to deal with almost all litigation. From the crisis a lesson for the future has been raised: the dematerialization of cases, digital procedures, and (at least partial) on-line handling of cases are effective tools for the ordinary management of justice and for backlog reduction strategies.

In 2020 the CEPEJ established the Cyberjust Working Group, and, in its Action Plan 2022-2025[6], it has made it its priority for the next few years to accompany States and courts in a successful transition towards the digitalization of justice in line with European standards and Article 6 of the European Convention of Human Rights (ECHR) in particular. Several Guidelines have thus far been adopted on e-filing and the digitalization of courts, on videoconferencing, and on e-auctions.[7] Others on online forms of Alternative Dispute Resolution (ADR) and online data-bases are being elaborated. The CEPEJ has also established a Resource Center on Cyberjustice and AI, that "serves as a publicly accessible focal point for reliable information on AI systems and other key cyberjustice tools applied in the digital transformation of the judiciary" that "shall help to gain an overview of such systems and tools, providing a starting point for further examination on their risks and benefits for professional and end-users in line with the 'European ethical Charter on the use of AI in judicial systems and their environment."[8]

---

[5] CEPEJ, Guidelines on how to drive change towards cyberjustice, CEPEJ (2016)13, December 7, 2016, available at: https://rm.coe.int/16807482de; CEPEJ, European Ethical Charter for the use of artificial intelligence in judicial systems and their environment, December 3-4, 2018, available at: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c; CEPEJ, Toolkit for the implementation of the Guidelines on Cyberjustice, CEPEJ(2019)7, June 14, 2019, available at: https://rm.coe.int/cepej-toolkit-cyberjustice-en-cepej-2019-7/168094ef3e

[6] CEPEJ, *2022 – 2025 CEPEJ Action plan: "Digitalisation for a better justice,"* CEPEJ(2021)12Final, December 9, 2021, available at: https://rm.coe.int/cepej-2021-12-en-cepej-action-plan-2022-2025-digitalisation-justice/1680a4cf2c

[7] CEPEJ, *Guide on judicial e-auctions*, CEPEJ(2023)11, June 16, 2023, available at: https://rm.coe.int/cepej-2023-11-en-guide-on-judicial-e-auctions-1-/1680abb674; CEPEJ, *Comparative Study on the use of judicial e-auctions in the Council of Europe Member States*, CEPEJ-GT-CYBERJUST(2023)1, June 16, 2023, available at: https://rm.coe.int/cepej-gt-cyberjust-2023-1-en-judicial-e-auctions-comparative-study-/1680abb7b4; CEPEJ, *Guidelines on electronic court filing (e-filing) and digitalisation of courts*, CEPEJ(2021)15, December 9, 2021, available at: https://rm.coe.int/cepej-2021-15-en-e-filing-guidelines-digitalisation-courts/1680a4cf87; CEPEJ, *Guidelines on videoconferencing in judicial proceedings*, June 2021, available at: https://rm.coe.int/cepej-2021-4-guidelines-videoconference-en/1680a2c2f4

[8] https://www.coe.int/en/web/cepej/resource-centre-on-cyberjustice-and-ai

The documents mentioned already contain accurate analyses of the benefits, risks, and ways of containing them. It is nevertheless useful to assess them concisely for the convenience of further reasoning.

### *3.1. Opportunities*

The digitalization of justice is about the way justice can be transformed by digital tools, AI instruments, networks and media. It could potentially provide faster, cheaper, and better service, whilst fully implementing the rule of law and guaranteeing the substantial and procedural rights of all citizens.

It primarily refers to solutions that facilitate access to justice by establishing a digital channel that enables the interaction and exchange of data and *e*-documents between courts and court users, and that includes electronic registers, digital file and workflow management, *e*-document, *e*-signature, *e*-service of decisions, and *e*-archives. But there are many other opportunities as well: the effective, timely, secure archiving of documents and acts (both judicial and administrative); secure and efficient file management; the transparency and complete traceability of acts; the processing of all data contained in documents; creation of navigable databases (especially for caselaw) through ML and NLP techniques and Big Data analysis; the acquisition and management of digital evidence; automatic allocation of cases, by algorithm, ensuring transparency and fairness in work distribution and protection from undue pressure; the distribution of workloads according to suitable criteria in order to reduce disposition time; remotely available procedures that are secure and user friendly; detailed statistical elaboration and analysis of flow and critical nodes that would allow for the detection and enhancement of procedures and practices. The outcomes are an improvement in accessibility, openness, and transparency, the promotion of consistent caselaw and ethical practices, a reduced risk of corruption, and a decrease in costs and inconveniences.

### *3.2. Challenges*

While the core values of justice remain unchanged, delivering justice in a digital world can be radically different from traditional methods.

The Court system thus faces major challenges: adapting very traditionalist practitioners to new methods; ensuring the impartiality and independence of the judiciary; protecting fundamental procedural rights – first and foremost the right to a fair trial – and human rights; managing a huge amount of data while protecting personal data; and avoiding discrimination.

The work approach of legal practitioners, judges, lawyers, and clerks would be revolutionized: paper files replaced by virtual files; trial documents, i.e., introductory proceedings and pleadings, minutes, taking of evidence and decisions, would be created in digital form, and as far as judges are concerned, on the basis of pre-set templates partially filled in with data registered by the system, i.e. names of parties, the register number, type of case, structure of the decision, elements of reasoning, via placeholders. Suits, acts and decisions could be filed remotely; the entire file, with documents ordered by provenance, type, date and organized into sub-folders, could be available in digital format so that judges have a complete overview of the workload and can monitor it, intervene promptly on urgent matters, and avoid unjustified delays.

Digitalization processes and the introduction of case management systems for civil and commercial proceedings already began in some European countries, i.e., Italy, Estonia, some twenty or thirty years ago. One might imagine that after such a long period of time information systems, related legislation, and application practices must have evolved to high levels of quality and effectiveness. The reality, however, is disappointing. Resistance from judges (especially in those countries where they are provided technical and legal assistance by specialized personnel which frees them from repetitive or simple tasks), maintenance of the paper-digital dual track, and inconveniences due to the technical quality of instruments have all had a negative impact on the spread of a digital-by-default and no-paper-by-default process. The perceived risk of a conformation of judicial work, the belief that an all-digital path could limit the procedural options available to judges, and subjugate procedural rights (Article 6 ECHR) to the rules of technology has also contributed to mistrust and disinterest. Digital judicial training, dissemination of knowledge in both institutional and associative (magistrate associations) fora, the participation of judges in the design, planning, and implementation of digital strategy and digital tools, as well as the involvement of all justice actors (lawyers, registrars, clerks) in designing and orchestrating the digitalization process are essential to overcoming these obstacles.

A new idea of court management is affirmed in countries such as France or Italy, where court presidents have significant managerial powers extended to the overall organization and functioning of the judicial office, and in countries like the UK, where the presidents' powers are limited to the exercise and organization of judicial functions. Court management is based on advanced statistics and graphs representing the flow of proceedings helping to identify critical nodes; the use of human and material resources is optimized and linked to specific objectives. Electronic registers and case management systems allow data to be collected and processed for statistical purposes. Advanced statistical tools allow articulated analyses of procedural flows, backlog formation, and their causes at all levels, including that of the individual judge. The judge's work can be analyzed statistically, ascertaining the time and manner of processing, the presence of anomalies, delays, as well as reasons for them. The range of reform of decisions by higher instances can be calculated. The concern for a distorted use of these instruments for the purpose of undue pressure and interference in the delicate process of assessing a judge's professionalism may also arise. Assessment, monitoring, and control of statistical systems by self-governing bodies (Councils for the Judiciary) or court administration is essential, as is the participation of judges in the development and testing phase.

Electronic registers allow moving from the "logic of the register" and the basic information contained therein to the *processing of all data* contained in documents, which also includes the purpose of office orientation, knowledge of unreported cases, the foreseeability of decisions, and litigation studies. The problems related to the right of individuals to privacy and to know, understand, and control the processing of their personal data by others are evident.

### *3.3. Risks*

When we move from the field of digital transition to that of algorithms and Artificial Intelligence tools in the field of justice, we should not only consider the challenges of modernity but also the risks that these technologies entail.

The unpredictable, swirling, enormous development of General AI tools, such as ChatGPT, Microsoft's Bing chatbot, and Google's Bard, has further polarized the debate between rejection and embrace, between opponents and supporters of AI. On one side, there is widespread alarm about the problems these tools may pose in the short term – biased responses, "hallucinations," the oversimplification of complex issues – and the serious risks they pose in the long term – from toxic texts, social-scale disruption through misinformation, propaganda, and false and harmful contents, to dreadful scenarios related to the manufacture of viruses or theft of nuclear codes. On the other side, technological evolution and AI are embraced without criticism.

The right solution may be to recognize the revolutionary contribution AI can bring to the way knowledge is created and synthetized, and to focus on risk neutralization through its responsible management (i.e., cooperation, research, regulation, governing authority), quality of data, model training, reinforcement learning, truthfulness (i.e., transparency, verifiability, the explicability of answers, reliability, a neutral point of view, accuracy), and alignment with human values.

### *3.3.1. "Do we want to use these systems?"*

Looking at these different approaches we must first ask what is called the "zero question:" "do we want to use these systems?" The answer must be a "conditional yes." Algorithms and AI are a revolutionary possibility for innovation and improvement but with some conditions. Systems, tools, and their practical applications must be in compliance with judicial values, such as independence and impartiality, equality, transparency and accountability; they must grant respect for human rights and the protection of all members of society from the risk of discrimination and misuse of personal data; they must respond to ethical, legal imperatives and fundamental guiding principles as laid down in the European Ethical Charter on the use of AI in judicial systems;[9] they must prefer an approach that values human control and the expansion of human capacity and that operates in an inclusive manner, overcoming the digital divide and protecting the digitally unskilled or disconnected.

As emphasized in the "CEPEJ Guidelines on e-filing" and in the CEPEJ tools on the transition towards digitalization, the basis for a fair and human rights-compliant digitalization of justice are the continuous involvement of stakeholders and the adoption of strategic choices in line with fundamental principles. This implies that changes in the field of cyberjustice shall be

[9] In 2018 the CEPEJ adopted the first European text setting out ethical principles on the use of AI in judicial systems, the Ethical Charter, which identified the following core principles to be respected in the field of AI and justice:

- respect for fundamental rights: ensuring that the design and implementation of AI tools and services are compatible with fundamental rights;
- non-discrimination: specifically preventing the development or intensification of any discrimination between individuals or groups of individuals;
- quality and security: regarding the processing of judicial decisions and data, using certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment;
- transparency, impartiality and fairness: making data processing methods accessible and understandable, authorizing external audits;
- "under user control:" precluding a prescriptive approach and ensuring that users are informed actors and in control of their choices.

court-driven and not technology-driven; each judicial information system shall be deployed with core judicial values in mind; judges and other justice actors must learn how to contribute to the development of such tools and systems and how to monitor and assess their quality and human rights compliance.

Shaping this technical revolution while adhering to ethical and legal values should be the "overarching mission," or the "background music," as was the building of independence through the Councils of Justice and the implementation of the Constitution for generations of judges in the previous century. The CEPEJ Guidelines should underpin this challenging transition. The European Union "AI Act" and the Council of Europe "[Framework] Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law" shall be our guide.

## 4. Applying new technologies in the Justice sector

The AI applications that most concern practical jurists are the following:

a) Components of more complex digitalized procedures. Primary examples concern Case Management Information Systems (CMIS) with complex technical architecture that may include algorithmic software systems for automated case registration, automated case assignment, automated recognition classification and filing acts and documents, automated production of partially filled forms, as well as the translation and transcription of speech-to-text. To date, this type of algorithmic software primarily operates on a predetermined and specific set of rules and data, i.e., automatic allocation, which provides for a distribution of cases based on rules normally set by the president of the court and conventional systems of assigning a weighting to the case. However AI is essential to achieving excellent performance and effective techniques.

b) Digital/scientific evidence. When speaking of evidence in this context, one can refer both to digital evidence in the strict sense (data and information in computer systems) and to scientific evidence generally introduced into the process through expert opinion, the conclusions of which may be based on the interpretation of complex statistical data (i.e. correlation between exposure factors and disease occurrence in the field of epidemiology), or on Decision Support Systems (i.e. diagnoses based on DSSs in the field of health or predictions of environmental damage based on computational calculation systems or ascertainment of the causal link between a disaster and human work based on DSSs in the engineering sector). In the latter cases especially, the judge may decide based on knowledge produced by AI (sometimes without knowing it).

c) Judicial training. Digital tools and interactive *e*-learning systems have long been used in judicial training, but a revolutionary opportunity is now offered by the metaverse, where young judges and prosecutors or older magistrates who change functions can practice innumerable activities, i.e., presenting evidence in court, listening to vulnerable witnesses, handling difficult situations in court, interacting with experts, and communicating with the media. To understand how technology can contribute to strengthening and reshaping (judicial) training, it is instructive to look at experiences in other highly complex sectors. I refer, for example, to the development of virtual ophthalmic surgical skills training.[10]

[10] E. GROSS MUNOZ, R. FABREGAT, J. BACCA-ACOSTA, N. DUQUE-MENDEZ, C. AVILA-GARZON, *Augmented Reality, Virtual Reality, and Game Technologies in Ophthalmology Training,* 2022, 13, 5, 222,

d) Caselaw database. Digital caselaw archives have existed in some European countries, Italy for example with its once revolutionary Documentation Electronic Center (CED) of the Court of Cassation, since the 1960s. In the drive towards the transparency and accessibility of judicial decisions, it is essential today to build digital archives of decisions that are complete and browsable in a simple and effective manner on the merits and legitimacy of the various types of courts. AI tools support the acquisition and categorization of decisions, as well as their anonymization or pseudo-anonymization with the aim of protecting privacy and personal data. AI is also essential for the introduction of new methods of intelligent archive search.
e) Decision Support Systems. One of the great opportunities offered by AI, along with significant risks, is the creation of Decision Support Systems for judges, grafted onto complete and structured caselaw archives. The following section is devoted to such systems.

## 5. Decision Support Systems

DSSs can be found in everyday life and are increasingly pervasive in day-to-day examples (from online shopping to choosing a TV series) as well as in relevant decision-making circumstances in the health, administrative, educational, judicial or police fields.

DSSs are computerized programs used to support determinations, judgments, and courses of action in an organization, a business, or an activity. They can be based on a set of predefined rules or on forms of Machine Learning (ML) and Natural Language Models (NLM). These work by collecting and storing a quantitatively large amount of data (in a data warehouse or data lake) and "labeling" it to extract significant syntagma/information. New inferences are then formulated using a mix of techniques, i.e. text mining (the extraction of information and, hopefully, knowledge from raw text with the help of AI and machine learning) and Big Data.

Their use is pervasive and growing exponentially in the health sector. Here DSSs are applied[11] in a variety of ways: drug prescription support; management of patient health records; diagnosis and referral to a specialist by processing imaging data (X-rays, ultrasound scans, mammograms, photographs); applying an algorithm trained to recognize images;[12] the diagnosis of cancer diseases (deep learning algorithm fed with data from health records, clinical studies, publications, recommendations of national agencies);[13] the identification of people at risk for certain diseases

https://doi.org/10.3390/info13050222; G. PELLEGRINO, M. C. BARBA, G. D'ERRICO, M. Y. KÜÇÜKKARA, L. T. DE PAOLIS, *eXtended Reality & Artificial Intelligence-Based Surgical Training: A Review of Reviews,* in *International Conference on Extended Reality*, 2023, 345, doi:10.1007/978-3-031-43401-3_22; C. GUPTA, *Development of virtual ophthalmic surgical skills training*, in *Eye,* 2023, 37, 290, https://doi.org/10.1038/s41433-021-01896-1

[11] See DESMOULIN-CANSELIER, LE METAYER, *Décider avec les algorithms*, Paris, 2020.

[12] See Y. ZHOU, M. A. CHIA, S.K. WAGNER ET AL., *A foundation model for generalizable disease detection from retinal images*, in *Nature,* 2023, https://doi.org/10.1038/s41586-023-06555-x

[13] Among others: W. T. TRAN, A. SADEGHI-NAINI, F-I. LU ET AL., *Computational Radiology in Breast Cancer Screening and Diagnosis Using Artificial Intelligence,* in *Canadian Association of Radiologists Journal*. 2021, 72, 1, 98, doi:10.1177/0846537120949974; A. C. S. TALARI, S. REHMAN, I. U. REHMAN, *Advancing cancer diagnostics with artificial intelligence and spectroscopy: identifying chemical changes associated with breast cancer* in *Expert Review of Molecular Diagnostics*, 2019, 19, 10, 929, DOI: 10.1080/14737159.2019.1659727 ; D.J. VAN BOOVEN, M. KUCHAKULLA, R. PAI, F.S. FRECH, R. RAMASAHAYAM, P. REDDY, M. PARMAR, R. RAMASAMY, H. ARORA, *A Systematic Review of Artificial Intelligence in Prostate Cancer,* in *Research and Reports in Urology*, 2021, 31, DOI: 10.2147/RRU.S268596; A. A. RABAAN, M. A. BAKHREBAH, H. ALSAIHATI, S. ALHUMAID, R. A. ALSUBKI, S. A. TURKISTANI, S. AL-

(breast or bowel cancer) and automatic scheduling of tests; and the management of rare resources (death risk assessment to direct pneumonia patients to hospitalization or hospital treatment; classification of patients awaiting transplants in order of priority, i.e. ScoreCoeur, an automated system with operating methods set by the French *Agence de la biomédicine*). The scientific community has rigorously debated the opportunities and risks associated with the use of such AI tools in research and applications,[14] but – in the opinion of an outside observer trying to follow the tumultuous development of the sector as a citizen – in the face of the exceptional relevance of the results achieved, it is committed to exploring new horizons, to scientifically verifying the correctness of results and to mitigate the risks.[15]

Examples from the medical world, concerning both the possibilities of use and the attitude of the relevant scientific community, are of relevance to the justice sector. Indeed, the problems posed by DSSs in the medical-health field, i.e. the legitimacy of the use of the tool, force, discrimination, and bias, data quality, the quality of results, control, explicability, and responsibility, are all very similar to those encountered in the judicial field.[16]

### *5.1. DSSs in the Justice sector: many projects and few applications*

There are no reports of AI application currently in use in judicial systems to support the decision-making process of judges.[17] Various pioneering projects, research, and even commercial products in different areas of the world have focused on the implementation of AI-based systems,[18] claiming that it is possible to "predict," among other things, the outcome of a

---

ABDULHADI, Y. ALDAWOOD, A. A. ALSALEH, Y. N. ALHASHEM ET AL., *Artificial Intelligence for Clinical Diagnosis and Treatment of Prostate Cancer*, in *Cancers*, 2022, 14, 5595, https://doi.org/10.3390/ cancers14225595.

[14] See also, EUROPEAN PARLIAMENT, *Artificial Intelligence in Healthcare*, June 2022, available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf.

[15] E. A. M. VAN DIS, J. BOLLEN, R. VAN ROOIJ, W. ZUIDEMA, C. L. BOCKTING, *ChatGPT: five priorities for research*, in *Nature*, 2023, 614; S. B. JOHNSON, A. J. KING, E. L. WARNER, S. ANEJA, B. H. KANN, C. L. BYLUND, *Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information*, in *JNCI Cancer Spectrum*, 2023, 7, 2, pkad015 https://doi.org/10.1093/jncics/pkad015; A. BLANCO-GONZÁLEZ, A. CABEZÓN, A. O. SECO-GONZÁLEZ, D. CONDE-TORRES, P. ANTELO-RIVEIRO, A. PIÑEIRO, R. GARCIA-FANDIN, *The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies*, in Pharmaceuticals, 2023, 16, 6, 891, https://doi.org/10.3390/ph16060891; C. BIEVIER, *The easy intelligence tests that AI Chatboats failed*, in *Nature*, 619, 2023.

[16] The problems of health DSSs and judicial DSSs are compounded when the former are used in the context of expertise for diagnostic purposes, for ascertaining causal links and assessing damage, in the context of a technical consultation to bring scientific knowledge into the process.

[17] I refer here to the process of deciding (questions of fact and law) a case presented before a judge. It is well known that in the 2010s, software was produced by private companies for predicting the risk of recidivism, to be used in imposing coercive measures or in sentencing (see the infamous Loomis case, available at: https://harvardlawreview.org/print/vol-130/state-v-loomis/). Such systems, when applied, gave rise to serious problems of discrimination. They have not been implemented in European judicial systems.

[18] - Mexico: the Expertius system, based on deep learning algorithmic architecture, which seeks to automate alimony rulings, a type of trial with a high volume of annual cases and a high degree of structuring and homogeneity in the process of evidence assessment and resolutions.

- UK: the introduction of an automatic online conviction procedure has been the subject of a government proposal and consultation which would allow some defendants to resolve their cases entirely online in appropriate cases. Under this proposal, defendants who opt in to the online procedure and plead guilty will be offered the option of accepting a pre-determined penalty (including the payment of any appropriate compensation and costs),

civil case, possible solutions to given questions of fact and law, or the amount of damages and costs of litigation. However, this research is either not yet at an advanced stage or deals with systems that are simply based on statistical inferences or that at best manage to identify certain frequencies in a limited number of past decisions.[19] We are still a long way from proven prediction and even simple support systems. In general, the purpose of this type of research or projects is not to introduce automatic dispute resolution systems or automatic decision-making systems but to provide judges with "expert suggestion," based on an in-depth knowledge of precedents, on how to solve specific cases (or, better, on how similar or identical cases were solved). Nevertheless, judges (and scholars and policy makers) spontaneously ask themselves many of the same questions. Could judicial independence be jeopardized? Might judges rely too much on AI recommendations and be reluctant to depart from them, thereby raising issues with their impartiality? Is there a risk that the independence of judges may be undermined by the combined actions of software engineers, AI researchers, and information technology companies participating in the design of automated judicial decision-making processes? Can a DSS based

---

be convicted, and pay the amount immediately. The testing of the system was proposed with the following summary for non-imprisonable offences: railway fare evasion, tram fare evasion, and possession of an unlicensed rod and line. The proposal has not been implemented at present.

- Estonia: Much publicity has been given to an Estonian small claims' decision automation project in recent years. The Ministry of Justice published the following note in February this year: "As there have been a lot of questions relating to the topic of AI Judges, we have to explain that the article about an Estonian project for designing a "Robot/Judge" in Wired from 25th of March 2019, is misleading. There hasn't been such a project or even an ambition in the Estonian public sector. The Estonian Ministry of Justice has not developed an AI robot judge for small claims procedures nor for general court procedures to replace human judges. We are still searching for ICT means to make the court's workload, including administrative burdens, more bearable. More precisely, the Ministry of Justice is looking for opportunities for optimization and automatization of the court's procedural steps in every type of procedure, including procedural decisions where possible.."

- France: "Travaux de recherche sur l'intelligence artificielle,"," based on an agreement reached among the Court of Cassation, HEC Paris and the Ecole Polytechnique, in partnership with the BAR Association, with the goal of studying the flux of cases in front of the CoC, the CoC made procedural documents and judgments, previously pseudonymized, available to researchers in order to identify arguments and legal issues, connections, and attempt to objectify the notion of the complexity of a case.

The project *www.lawdataworkshop.eu* is also very interesting, conceived and managed by M. Clément, Président de Chambre au tribunal administrative de Lyon. The program uses open data case-law, especially from the JADE database (Decisions of the Council of State and the Administrative Courts of Appeal) counting 504,194 decisions as of 30 September 2022, along with the maximum amount of information available (files, implicit or explicit references to other case-law), linking them together in order to better visualize the network nature of case-law, identify implicit citations of case law thanks to the recitals of principle.

- Italy: "LIBER-Lab of the Dirpolis Institute" of the Scuola Superiore Sant'Anna of Pisa analyzed case law material through machine learning techniques and Big Data analysis, creating a navigable archive with semantic modalities. It is based on agreements between the *SSSA* and the Courts of Genoa and Pisa, which have made the decisions stored in their IT systems available to researchers.

[19] This is also the case with the well-known research on ECHR caselaw; see: N. ALETRAS, D. TSARAPATSANIS, D. PREOTIUC-PIETRO, V. LAMPOS, *Pre- dicting judicial decisions of the European court of human rights: A natural language processing perspective,* in *PeerJ Computer Science*, 2016, 2, 93; on subsequent developments, see: M. MEDVEDEVA, M. VOLS, M. WIELING, *Judicial Decisions of the European Court of Human Rights: Looking into the Crystal Ball*, available at: https://www.jus.uio.no/pluricourts/english/news-and-events/events/2018/1a_medvedeva-european-court-of-human-rights.pdf ; M. MEDVEDEVA, X. XU, M. WIELING, M. VOLS, *JURI SAYS: An Automatic Judgement Prediction System for the European Court of Human Rights,* in S. VILLATA ET AL. (eds.), *Legal Knowledge and Information Systems*, 2020, 277.

on the similarity or identity of cases embody the specificities of a judge's reasoning? Might AI fail to consider crucial parts of judgments? Is there a risk that the solution to all cases is carried over from that of the average or perhaps a majority of cases? Is there room for customized decision-making?

Finally, returning to "question zero," do we really need support in the decision-making process?

### *5.2. "Question zero" and the foreseeability of judicial decisions*

The answer to "question zero" is again positive. Contemporary judges cannot decide alone, as a monad separate from the whole, to use electronic databases as they once used paper directories of case-law. They must take into account the enormous quantitative and qualitative development of case-law; the exponential growth of open data available and the development of related policies and legislation; the limited capacity of judges to have an effective and deep knowledge of case-law and precedent when faced with enormous amounts of data, particularly regarding difficulties in timing, analysis, systematization, and so-called "choice overload bias;"[20] the increased thirst of the public for transparency and equality accompanied by a relevant variety of sources of case-law knowledge that includes social media; the spread of projects and research financed by the private sector with the occasional participation of large law firms to "predict" judicial decisions which can result in cognitive discrepancies between judges and lawyers.

These challenges cannot be met by limiting the number of decisions published in archives, on the web, or in official collections, to those selected either by the study offices of the courts or by research systems that focus on abstracts rather than full texts; two methods that have historically had the function of guiding the formation of consistent case-law, especially at the supreme court level. The ease of online dissemination of any decision (by a lawyer, an individual judge, or by anyone with access to the decision) and its contradiction to open knowledge and open data policies makes it unviable.

AI-based instruments can help judges make decisions with a full awareness of the actual state of case-law, applying or disapplying it based on reasons, while always keeping the principle of equality in mind.

From this perspective, it seems necessary to abandon the suggestive binomial "predictive justice" first, which would seem to refer more to a crystal ball or a game of dice than it would exercise of the judicial function, and then replace it with the more appropriate notion of a "foreseeability" of decisions. The concept of foreseeability is linked to that of "legal certainty" – a value for the community and one of the main pillars of trust in the judicial system as it relies on the basis of free choice for the individual – and to a guarantee of equality.

While being fully aware of the complexity of the theoretical debate that, since the second half of the 20th century, has seen the intertwining of the themes of certainty, legal interpretation, and

[20] We are referring to a psychological phenomenon analyzed in the field of Economics, market and consumer choice; the expression is used in the text to highlight the interpreter's difficulties in identifying the right precedent (or set of precedents) when the available data are vast. See B. SCHWARTZ, *The paradox of choice: Why more is less*, New York, 2004; G. LOEWENSTEIN, *Is more choice always better*, in *Social Security Brief*, 1999, 7, 1, 7; L. CARMINATI, *Behavioural economics and human decision making: Instances from the health care system*, in *Health Policy*, 2020, 124, 6, 659.

the activity of judging as based on law or on rationalized intuition motivated ex post facto, this is not the place for an extended exploration of the subject.

The purpose of adopting the notion of foreseeability as an expression of the principles of legal certainty and equality is to emphasize that foreseeability is coessential to access to justice. Expressed in very simple terms, those persons claiming a violation of their rights must be able to know: i) whether their situation is recognized as a right; ii) whether the conduct attributed to the other party is considered unlawful; iii) how the right will be restored and the wrong redressed. Similarly, a person must be able to know whether one's conduct may constitute a criminal offence and what the consequences of the conduct may be in terms of punishment and a restriction of one's rights.

This also implies that different decisions for the same or very similar situations without sufficient reasoning being provided by the judge is intolerable. However, this happens when judicial decisions (even by judges from the same office) are not known or knowable – due to the overly large number of decisions that must be consulted as well as to the existence of "hidden" case-law – because they are serial, they do not relate to complex legal issues or are not suitable for publication. This is especially the case with judgments from trial court judges who deal with factual issues – such as determining spousal and child maintenance, child custody arrangements, sentencing for minor offences, small claims, and compensation for personal injuries – situations in which it becomes difficult to ensure the homogeneity of decisions, legal certainty and equality before the law.

In the case-law of the European Court of Human Rights (ECtHR), legal certainty is a fundamental aspect of the Rule of Law,[21] even though it is not an absolute value. The requirements of legal certainty and the protection of the legitimate confidence of the public do not confer an acquired right to consistency and case-law development is not contrary to the proper administration of justice, since a failure to maintain a dynamic and evolutive approach would risk hindering reform, the improvement of safeguards, and the protection of new rights.[22]

---

[21] One of the fundamental aspects of the rule of law is the principle of legal certainty (see ECtHr, *Brumărescu v. Romania* [GC], October 28, 1998, n. 28342/95), which, *inter alia*, guarantees a certain stability in legal situations and contributes to public confidence in the courts (see *Nejdet Şahin and Perihan Şahin v. Turkey* [GC], October 20, 2011, n. 13279/05). The persistence of conflicting court decisions, on the other hand, can create a state of legal uncertainty likely to reduce public confidence in the judicial system, whereas such confidence is clearly one of the essential components of a State based on the rule of law (see *Vinčić and Others v. Serbia*, December 1, 2009, n. 44698/06).

[22] ECtHR (*Lupeni Greek Catholic Parish and Others v. Romania* [GC], November 29, 2016, n. 76943/11, § 116; *Guðmundur Andri Ástráðsson v. Iceland* [GC], December 1, 2020, n. 26374/18, § 238; *Beian v. Romania (no. 1)*, December 6, 2007, n. 30658/05, § 39; *Nejdet Şahin and Perihan Şahin v. Turkey* [GC], 2011, § 58). In *Atanasovski v. the former Yugoslav Republic of Macedonia*, January 14, 2010, n. 36815/03; *Borg v. Malta*, June 26, 2012, n. 38590/10: "The Court has been called upon a number of times to examine cases concerning conflicting court decisions and has thus had an opportunity to pronounce judgment on the conditions in which conflicting decisions of domestic supreme courts were in breach of the fair trial requirement enshrined in Article 6 § 1 of the Convention (see *Paduraru v. Romania*, December 1, 2005, n. 63252/00,; *Iordan Iordanov and Others v. Bulgaria*, July 2, 2009, n. 23530/02; *Pérez Arias v. Spain*, June 28, 2007, n. 32978/03; *Ştefan and Ştef v. Romania*, January 27, 2009, n. 24428/03-26977/03; *Taussik v. the Czech Republic* (dec.), September 15, 2015, n. 9842/13; and *Tudor Tudor v. Romania*, March 24, 2009, n. 21911/03). In so doing it has explained the criteria that guided its assessment, which consist in establishing whether "profound and long-standing differences" exist in the case-law of a supreme court, whether the domestic law provides for machinery for overcoming these inconsistencies, whether that

Courts set out the values to be pursued and the guiding principles for balancing them. Social transformation can guide the evolution of case-law and changes, often slow and prepared by minor reforms, must be clear, understandable and foreseeable, especially in the criminal sector.[23] The motivations of the judge therefore acquires a fundamental role.

Systems using AI could help judges as well as lawyers to: i) obtain better and faster knowledge of cases; ii) improve and speed up the drafting of documents; iii) select the most significant parts of previous cases; iv) reuse part of previous documents in new documents; v) make better and more conscious use of precedent; vi) improve the readability of legal documents and court decisions; vii) foresee a possible decision on the basis of precedent; viii) understand how a possible decision would be positioned within the framework of precedent.

### *5.3. Judging today with an eye to the future*

In reality, to decide a case even today a judge must search archives that are generally well developed and organized for the supreme courts, but fragmentary and largely incomplete for the jurisprudence of trial courts. One wonders why intelligent systems have not been created to help the judge in the "diagnosis" of the concrete case and in the knowledge of how that case (or a similar case) was decided in the past, allowing them to consciously decide whether to stay in the groove of certainty or to depart from it for good reason.

However, the situation can change quickly. The acceleration imparted by the emergence and evolution of AGIs, along with the rapid shift from mistrust to its use in key sectors such as in science and education,[24] give us reason to believe that decision support tools will soon also be available in the field of justice.[25] Europe's commitment not only to regulating the use of AI but to research in the field[26] and funding for the digitalization of justice from the Recovery Fund and the creation of open source digital caselaw archives[27] have created the conditions for an

machinery has been applied, and if appropriate to what effect (see *Iordan Iordanov and Others*, cited above, §§ 49-50)."

[23] See: *Del Rio Prada v. Spain* [GC], October 21, 2013, n. 42750/09; *Rohlena v. Czech Republic* [GC], January 27, 2015, n. 59552/08; *Coëme and Achour v. France* [GC], March 29, 2006, n. 67335/01, *Contrada v Italy (n.3)*, April 14, 2015, n. 66655/13.

[24] K. ROOSE, *How Schools Can Survive (and Maybe Even Thrive) With A.I. This Fall*, available at: *https://www.nytimes.com/2023/08/24/technology/how-schools-can-survive-and-maybe-even-thrive-with-ai-this-fall.html?name=styln-artificial-intelligence&region=TOP_BANNER&block=storyline_menu_recirc&action=click&pgtype=Article&variant=undefined*; MagicSChool at: https://www.magicschool.ai

[25] Interesting tools dedicated to lawyers are beginning to appear on the market. Such is the case with CoCounsel, an AI legal assistant built on OpenAI's GPT-4 large language model. CoCounsel reviews documents, prepares for a deposition, searches a database, prepares legal research memos providing sources, summarizes interpreting and condensing critical information, extracts contract data, and checks contract policy compliance. For more see https://casetext.com/cocounsel/. There is a specific allegation that the tool is not hallucinating.

[26] See the research project COHUBICOL (Counting as a Human Being in the Era of Computational Law) funded from 2019-2024 by the European Research Council (ERC) under the HORIZON2020 Excellence of Science program ERC-2017-ADG No 788734 COHUBICOL to "investigate how the prominence of counting and computation transforms many of the assumptions, operations, and outcomes of the law." For more see https://www.cohubicol.com

[27] For example, one of the goals of Next Generation EU (Recovery Funds) for the Justice Sector in Italy is the "creation of a free, fully accessible, and searchable database of civil decision according to legislation;" we should

unprecedented growth of research at the intersection of Law and Information Technology. Concrete and operational results can soon be expected.

### *5.3.1. Tools tailored to the specific needs of judges*

The concept of "instruments tailored to the specific needs of judges" includes reference to the essential characteristics of the function (independence and impartiality) and its exercise (respect for human rights, the guarantee of equality, adherence to the concrete case, quality of the decision, reasoning).

The essential elements are summarized as follows:

- such tools will not aim to issue fully automated decisions;[28]
- they are automated (or automatic) support to the decision making process;
- the relevant facts of the case continue to be attached, proved, and established according to the rules and by the ordinary means of proof.

AI tools can support judges in synthesizing and systematizing statements of evidence by producing memos correlating the statement with evidence. The judge would retain control over the result and the correspondence to the actual results of the investigation carried out.

These tools can perform searches in public archives of first instance case-law, and appellate and supreme courts, providing the legal solution in text form that appears most appropriate and closest to the case with an indication of relevant sources. Judges, thanks to their high level of professionalism and knowledge of the Law, can verify, refine, extend, or circumscribe the result; they can also decide to follow past case-law or innovate by giving reasons. They are never bound by tool-suggested results.

To guarantee a judge's knowledge of how the system operates, what data it uses, and how the "suggested" solution was constructed, it is essential that: i) the software is produced by a public party (the Council of Justice, Ministry of Justice, or Ministry of Innovation according to national systems) in-house by hiring computer scientists or by public procurement where the contractor is bound to follow instructions from the contracting authority; ii) the values, principles, and rules inherent in the exercise of jurisdiction are incorporated into the programming and creation phase of the system; iii) the source code is made publicly available.

Within this framework, Decision Support Systems are not operational to date due to three problematic nodes (present at different levels in most countries): a) the quality and size of databases, b) the language of judicial decisions and the lack of a comprehensive thesaurus for languages other than English, c) the pre-processing of data.

a) The main database related issues are:

1. There are no comprehensive databases available, especially for judgments of first instance and appellate courts. This depends on three factors: i) judgments are not in a digital

also mention the EU funded Governance Project "Innovation and efficiency of Courts" whose goal, among others, is the "analysis, study, implementation and improvement of 'knowledge' tools available to the judiciary by contributing to the development of systems for analyzing case law, highlighting regulatory references, identifying specialized knowledge sources, and the semantic analysis of judgments."

[28] A fully automated decision making process (examples of which can be found in the activities of public administrations) is conceivable within the jurisdiction only for extremely simple and serial cases whose solution is based on technical elements, which exclude the exercise of discretion, and rules that are all predetermined in advance; an ordinary remedy should always be granted before an independent court.

format; ii) there are no platforms or systems where judgments can be stored securely; iii) there are no automatic anonymization systems in place.

2. Unsuitable criteria are often adopted when creating databases, i.e. the database is not fed with the full text of judgments, but instead by maxims or excerpts edited by the author or specialized staff; judgments are chosen for the archive on the basis of specific criteria (e.g. the quality of the text or compliance with supreme court judgments) that do not favor a real and comprehensive knowledge.
3. There is widespread resistance among judicial stakeholders: it is feared that erroneous judgments will be disseminated with a discrediting effect; a spirit of cooperation among judges to build shared legal certainty is missing; judges fear an impact on their independence.

b) A decision support system uses, depending on the domain, a Natural Language Processing (NLP) technique, which takes written text, interprets and transforms it into a form that the computer can understand, performs an intelligent analysis of large amounts of written text and generates insights from it. All types of NLPs need a *thesaurus* – a terminological language resource.

Legal/Judicial language is very peculiar since words that appear common can have an independent meaning. Moreover, the styles of judges can be very different, and the texts of judgments can lack homogeneity and standardization even when they resolve substantially identical cases. These conditions impair language processing.

c) The creation of a decision support system, capable of "suggesting" solutions to a concrete case, requires a series of operations. In the simplest terms, this involves collecting judgments, inserting legal markers, labelling them, designing an algorithm, and comparing the solutions obtained by the AI model with those of the human expert.

Labelling is an operation that requires a deep knowledge of the legal system in which the decision was rendered, of the rules of judgement that apply to a certain type of proceeding, of the relevance of the factors that may be contained in the document, and of the possible hidden bias.

Those with judicial experience, judges, or legal officers under the supervision of a judge, should play a primary role in labelling.

These problems must all be considered when designing new systems.

### *5.3.2. The participation of justice actors. A virtuous example: the Pisa pre-totype*

The project launched by the University of Pisa – in the Departments of Law, Informatics and Computer Science, and Linguistics – and by the Tribunal of Pisa under the PON Governance "Innovations and efficiency of courts – *Giustizia Agile*" is a virtuous example of collaboration between the judiciary and researchers in the development of new systems.

As explained in the final report of the project, "the work carried out by the IT and linguistic research group of the University of Pisa made it possible to create a pre-totype of a search engine for Italian-language judgments that synergistically uses NLP, AI, and information retrieval algorithms." The activities essentially involved the implementation of the pre-totype, its experimentation and evaluation according to standard metrics, and a series of usability tests

carried out by scholars, professors, and staff of the Supporting Staff Office (*Ufficio per il Processo* – UPP) of the Tribunal of Pisa.

The main contributions of the research group included:

- The creation of language resources: three gold-standard datasets for the legal domain (see below);
- A pipeline for the automatic extraction of metadata: various methods (from pattern-matching to neural model training) for the recognition of metadata within judgments were designed, implemented, and tested;
- A pre-totype: the realization of a pre-totype search engine based on ElasticSearch offered a series of basic and advanced functionalities to search for information within a corpus of judgments provided by the Court of Pisa and (semi-)automatically annotated, also thanks to the NLP/AI tools implemented in the project;
- An interface study: an in-depth study on the design of an interface for the efficient and forceful visualization of results produced by a legal search engine.
- A software deliverable: the implementation of a software package in "docker format" enabling installation of the search engine pre-totype on various hardware/OS.

The pre-totype allowed for the automatic extraction of two classes of metadata from the documents uploaded in PDF format: (i) low-level (i.e. judgment number, general role, publication date); and (ii) high-level (i.e. named entities, keywords, subject classes). This metadata was then indexed by the open-source search engine Elasticsearch, which can be queried through a graphical interface suitably developed for the project. The search engine not only allowed judgments to be searched "by metadata," but also enabled full-text searches, with results sorted by relevance, auto-completion functions exploited, and customized searches refined by users" (ToA).[29]

## 6. Before concluding

Before closing, three issues concerning contraindications to using AI and their ability to support judicial work should be mentioned.

The first can be described as the "risk of a lazy judge," i.e. a judge who would tend to be satisfied with a solution proposed by AI without looking any further, thus drying up the evolution of case-law and failing to pay attention to the details of the concrete case. Unfortunately, laziness is not related to AI and the image of a contemporary judge who uses decision models without adapting them to the specifics of a case, who does not study precedent, and perhaps does not listen to witnesses, is certainly no better than a colleague from the future. These are problems that are solved by training and professionalism evaluations and have little to do with technological development.

The second can be called the "irrational judgement model." This is the idea that a judge's reasoning and its hidden components cannot be reproduced by a machine by nature.

[29] As former president of the Court of Pisa, I would like to thank the University of Pisa and especially: Paolo Ferragina, Professor, PhD Dept. of Computer Science; Benedetta Galgani, Professor of Criminal Procedural Law; Giuseppe Campanelli, Professor of Constitutional Law and Prorector; Alessandro Lenci, Professor Dept. of Philology and Linguistics. Thanks also to the full research team for the opportunity to participate in the research, the shared passion, the chance to look at legal research, and the discovery of knowledge in a new and fruitful way.

An entire book could be written dealing with the never-defined diatribe between legal formalism and legal realism's thesis about the nature of the Law and judicial decision-making. In the footsteps of Dworkin and considering modern constitutionalism as well as the European conventional context, my belief is that judges decide by applying the law as interpreted in light of constitutional and conventional principles (EU Charter of Fundamental Rights, ECHR). This is not the place to explain my thoughts in detail and I limit myself to putting forty years of judicial work at the national and international level, thousands of judgments drafted, and thousands of hours in chambers on the scale. The conviction that what is considered *a posteriori* rationalization of a decision-making intuition is in truth an expression of the judge's "mirrored knowledge" ("*conoscenza riflessa*"),[30] acquired over years of cases, hearings, the reading of acts and writing of decisions, which leads to a simplification and shortening of the decision-making process. My conclusion is that Generative AI can generate legal reasoning.

The third can be defined "the screen of bias." There are decisions from the past, which are part of the database and which are the result of bias (e.g. cases of domestic violence and sexual assault). The problem of recognizing them and avoiding new discrimination is the same in a digital environment as it is in a paper environment. Data processing and the proper categorization and labelling of decisions can help to detect and remove them.

## 7. (Initial and partial) conclusion

Much still needs to be deepened and elaborated. At the conclusion of this initial survey of challenges, problems, and solutions, some very brief conclusions become clear.

Access to justice, equality before the Law and legal certainty are crucial values and fundamental components of the Rule of Law. New Information Technologies and AI can support the judiciary to deliver justice in a proper, open, transparent, efficient way but serious risks are associated with the embedding of AI in judicial tools.

European judges must be ready to take up the challenge of building advanced and innovative systems of justice management and decision-making. Judges cannot enclose themselves in a fortress with the risk that citizens will turn to the sirens of private justice governed by AI. Judges must build bridges to new solutions and be the primary interpreters of any technological reform. In this process training will be crucial. Only their presence and participation, rich in the unique experience of judging, can ensure that new systems are inspired by the principles of human rights protection, transparency, knowability, explicability, and democratic control.

[30] This concept is used by Franchi to explain the acquisition of extrajudicial knowledge by the judge, but it can also be used with reference to the creation of legal knowledge. In *La perizia civile*, Padua, 1959 Franchi says that "the organized experience that the judge has in the matter that does not relate to his professional education due to the fact of repeatedly becoming aware, in the act of judging, of events and phenomena that belong to the said matter;" in fact, "as a rule, repeated technical integration leads to the formulation of decisions containing technical evaluations of the same tenor, and knowledge of previous decisions... leads to knowledge of the technical criteria ...of evaluation, that is, to the absorption by the judge of the particular experience and... to the transformation of a particular experience, through the decisions of those who are not technical, in common experience (vulgarization)."

# 2.2. Technical & risk analysis

*Silvio Ranise – Fondazione Bruno Kessler & Università degli Studi di Trento*

*Summary: 1. Introduction to AI and ML – 2. ML algorithms: traditional and modern views – 3. Usage of ML in decision-making systems – 3.1. Usage of ML in judicial decision-making systems – 4. Responsible AI: technical aspects and risk analysis*

*Abstract: After defining AI and ML as subjects of investigation, an operational characterization of them is provided (Section 2). It is then explained how ML techniques can automate decision making (Section 3) with a focus on the judicial system. Further, the trustworthiness of ML techniques is discussed (Section 4).*

## 1. Introduction to AI and ML

Artificial Intelligence (AI) encompasses a wide range of disciplines including Robotics, Symbolic Reasoning, and Machine Learning. These disciplines use a disparate collection of techniques to enable machines to act intelligently by performing cognitive tasks like learning, reasoning, and analysis that were previously regarded to be the exclusive domain of humans – according to Prof. Maura R. Grossman of the University of Waterloo, Canada – or operate with varying levels of autonomy and can, for explicit or implicit objectives, generate output such as predictions, recommendations, or decisions influencing physical or virtual environments – as stated in the European Union's AI Act.

Because of its broad scope and the heterogeneous methodologies and techniques, giving a precise definition of AI turns out to be a daunting task. We seek to avoid this problem here by taking a pragmatic approach towards one discipline of AI, namely Machine Learning (ML). This choice is due to the fact that ML techniques have obtained spectacular results in several domains of application, including image and natural language processing with varying degrees of trustworthiness.

## 2. AI algorithms: traditional and modern views

An operational perspective on AI can be gained by considering the notion of an algorithm that identifies a finite sequence of instructions used to solve a class of specific problems or to perform a computation such as data processing. We now consider two ways of constructing algorithms according to a traditional and modern view as depicted in Figure 1.

- In the traditional view, humans create a model of the class of problems to be solved or the computations to be performed and then define a sequence of instructions that take as input an instance of the problem or a data set capable of producing in output the solution to the problem instance or the transformed data set. Because the model is created by humans, the set of features characterizing a problem instance is typically small and the sequence of instructions encodes a fixed set of logical rules. An example is classical planning, whose goal is the realization of action sequences, typically for the execution by autonomous robots or unmanned vehicles. In this context, planning problems use symbolic representations of

the set of possible states such as all the legal positions of chess, and actions, like all the legal moves of chess, and then transform one state, the initial position for example, into another such as white checkmate using a certain number of pieces in given positions. Forward or backward chaining procedures can search for a sequence of actions that transform an initial state into the goal state. This approach to planning is not flexible as it is defined on a predefined set of rules and suffers from the space explosion problem state, e.g., it is estimated that there are 10^40 possible positions when considering only legal moves.

To avoid such problems, heuristics must be designed, i.e. strategies that focus on particular ways of forming sequences of actions that are more likely to reach the desired goal state, although they do not explore all possibilities and may miss a solution even if it exists. In that sense they are incomplete.

An advantage of the traditional view of constructing algorithms is that a few problem instances with solutions are sufficient for human beings to define an algorithm that can be understood and proved correct, i.e., to find a solution when it exists.

- In the modern view, characterized by Machine Learning, a sufficiently large dataset – each data point representing a problem instance – is used to define a set of rules capable of producing some output, representing the solution to the problem instance encoded by the data point. This process is called training and uses optimization techniques to select the best element, with regard to one or more criteria, from a set of available alternatives. A concrete example is the definition of an algorithm for classifying images of pets: a sufficiently large number of pictures should be available that are labelled by one among, say, two possible alternatives, such as dog and cat.

  Using optimization techniques, such as the descent gradient method, training is able to define a sequence of instructions to classify the subject of a picture as a dog or a cat. Afterwards, the resulting set of instructions is used to classify images that do not belong to the training set, which is called deployment. A clear advantage of the modern view of building algorithms is that it does not require a high degree of ingenuity to solve complex problems and instead uses off the shelf powerful optimization techniques.

  A disadvantage is that the algorithm produced by training may be imprecise, e.g. classifying the subject of a picture as a dog when it is a cat or vice versa, especially when the training data set is not large enough or contains bias. Another disadvantage is that the algorithms produced by training are difficult to understand by humans as they consider too many features in order to produce a solution. There are approaches to mitigate these disadvantages such as monitoring performance and adjusting the algorithm if necessary, filtering the dataset used for training to eliminate biases and improve precision, or devising techniques to help humans predict the results produced by algorithms that result from training.

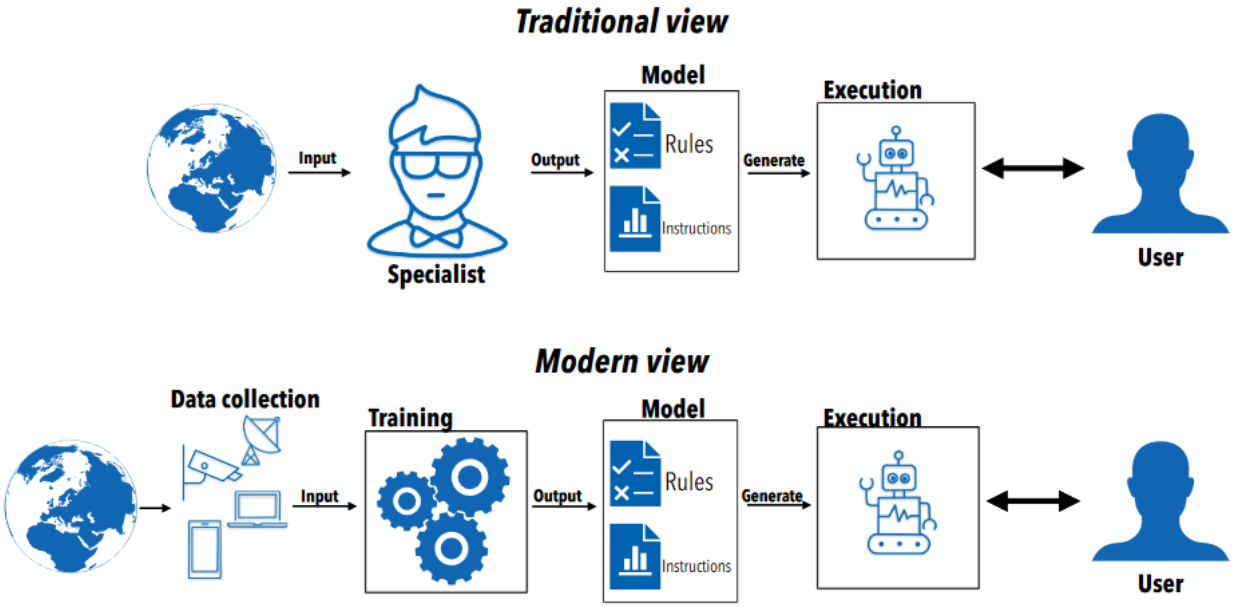



Figure 1. Constructing algorithms: traditional vs modern views

The choice of using the traditional or modern view to build AI algorithms depends on the requirements of the use case scenario in which the algorithm is going to be deployed. An adequate risk analysis with respect to several dimensions including security and trustworthiness must be performed in order to understand whether the risk of using an algorithm produced by AI techniques is acceptable or not.

## 3. Usage of ML in decision-making systems

A decision-making system supports business or organizational decision-making activities by helping people make decisions about problems that may be rapidly changing and not easily specified in advance. They can be either fully computerized or human-powered, or a combination of both. Examples of applications range from clinical decision support systems for medical diagnosis, to public administration tools for automating administrative procedures, to criminal judicial systems for providing investigative assistance and automating decision-making processes. The challenges and opportunities of harnessing ML in judicial systems and their implications for human rights should be carefully evaluated from the different perspectives of the stakeholders, including judges, involved in judicial ecosystems. A first step towards informed debate is to help stakeholders deepen their understanding of how ML algorithms can be used to build decision-making systems that are trustworthy enough to be used in the various stages of judicial processes. The use of ML in decision-making systems offers the opportunity to enhance and improve the quality of the process. By analysing large amounts of data, ML can automate many of the phases identified, thereby making the process more efficient and effective. Decision-making is a

complex process that involves four phases:[31]

1. Information acquisition. This phase involves gathering, filtering, prioritizing and understanding the relevant information about the problem or the decision. This information can come from a variety of sources, such as expert opinion and past results or experiences. ML can be used in this phase to automate the process of gathering the information, such as by collecting data from documents or the internet. For instance, with Natural Language Processing techniques to process and analyse large amounts of natural language data, it is possible to automatically extract product reviews from social media.

2. Information analysis. This phase involves analysing, interpreting, and making inferences and predictions on previously gathered information. This relates to identifying key elements and drawing conclusions from the data. ML can be used to automate this process by identifying recurring patterns that arise in a large dataset which allows a decision to be made. For instance, with Pattern Recognition techniques to distinguish and create emergent patterns, it is possible to understand what product customers are likely to buy based on their interests, including past purchases.

3. Decision selection. This phase involves prioritizing/ranking decision alternatives. The objective is to evaluate different options and choose the best one. ML can be used to automate this process by using algorithms to calculate an expected result from the different options. For instance, with Reinforcement Learning techniques enabling an agent, e.g., a computer, to take actions in unknown environments to maximize its reward, it is possible to set prices of products to maximize incomes.

4. Decision implementation. This phase involves the execution of choices. ML can be used to automate this process by creating a plan and monitoring the results. For instance, after choosing the best price option, ML algorithms can be used to monitor the market and dynamically adjust it based on different factors, such as demand, competition, and/or inventory levels.

ML can improve the efficiency of the decision-making process but depending on the phase and the use case scenario, it is advisable to foresee different levels of human involvement ranging from full automation to complete supervision of the results produced by the decision-making system. More precisely, we can consider five levels of automation intended as the sharing of responsibility in order to complete a certain task between humans and machines:32

1. Manual, which involves only human activities;
2. Low, where the human is the main actor, but the computer can be used for assistance;
3. Intermediate, where the computer is the main actor offering highlights and results with humans shadowing for contingencies;
4. High, where the computer is the main actor and informs humans only if required by context;
5. Full, where the computer executes automatically without displaying any information and result or allowing for any human intervention.

Depending on the use case scenario, it is possible to apply a different level of automation

[31] R. PARASURAMAN, T. B. SHERIDAN, C. D. WICKENS, *A model for types and levels of human interaction with automation*, in *IEEE Transactions on Systems, Man, and Cybernetics - Part A:Systems and Humans*, 2000, 30, 3, 286, doi: 10.1109/3468.844354, available at: https://ieeexplore.ieee.org/document/844354

[32] D. PETKEVIČIŪTĖ-BARYSIENĖ, *Human-Automation interaction in law*, May 2021, available at: http://inpact-psychologyconference.org/wp-content/uploads/2021/05/2021inpact070.pdf

to each one of the four phases discussed above. For instance, an e-commerce site could apply high automation to information acquisition and analysis but use low automation for decision selection and intermediate automation for decision implementation. Besides providing flexibility, choosing different levels of automation allows for increasing the trustworthiness of the entire decision-making process by supervising the results produced by the various ML techniques.

### *3.1. Usage of ML in judicial decision-making systems*

In the judicial decision-making system, the starting point of a case, from which the parameters of the debate are set, is the complaint filed by an applicant. A judge analyses the case, from the arguments of the respondent to the relevant case law, based on the complaint. In some cases, the judge should do the work "ex-officio," taking the part of the weaker position and, if necessary, modifying or rephrasing the complaint to address a specific issue. For example, the applicant could ask for the nullification of a contract, but the judge cannot find anything to justify such an action. By analysing the document for the proposed case, the judge may discover an unfair clause, which allows the complaint to be modified and the judge to act accordingly. Such an analysis, outside the specific complaint and made for the protection of the weaker party, will increase the workload of judges, who will be overloaded. By introducing the usage of ML to justice systems, it is possible to help courts better manage their workload.

An example of a helpful application can be found in situations of unfair clauses, where even legal experts have difficulty due to their complicated language and numerous types. In the following we provide a running example using the case of car financing. We will then present a characterization of the decision-making process phases for the judicial decision-making system[33] (note how these phases can be seen as a specialization of the general ones introduced above). In this scenario, the Applicant, who asks for a car loan, does not have enough money to buy a car, so he decides to apply for a loan from a Bank. Several months after signing the contracts, the Applicant realizes something is wrong and decides to complain to the court. The judge assigned to the case begins the judicial decision-making process and, as the request from the user is unclear, decides to get help from a Legal Tech application, which can be used by judges to evaluate sentencing options and promote consistency in the decision-making process.

1. Information acquisition. In this phase, the judge provides details about the scenario, such as documents, e.g. the contract and complaint, and filters to narrow the scope to cases related to finance. By uploading these documents, the judge should also take privacy concerns into consideration through data anonymization. The judge should always be aware of the information selected by the tool, to choose which is most appropriate and which should be ignored, as mistakes in this part of the process will spread to subsequent phases, e.g. ignoring unfair contract clauses will not contribute to the intended outcome.
2. Information analysis. In this phase, the analysis begins based on the information provided. The ML algorithm can identify patterns that could be difficult for humans to

[33] D. BARYSĖ, R. SAREL, *Algorithms in the court: does it matter which part of the judicial decision-making is automated?*, in *Artificial Intelligence and Law,* 2023, available at: https://doi.org/10.1007/s10506-022-09343-6

notice. For example, it can identify that certain types of contracts often contain unfair clauses and even in which provision. The judge should be aware of how the algorithm generates the outcome to identify and correct possible errors, e.g. if the clause is not unfair the report should be used to improve the algorithm.

3. Decision selection. In this phase, the ML algorithm, after acquiring and analysing the information, should explain its decisions and motivations. For example, it should produce a report, with one or more options, such as the presence of one or more unfair clauses, where they are located in the contract, and why they are qualified in such a way, e.g. which laws were violated, which terms were used, along with which precedent cases can be cited. This decision should be consistent and transparent, allowing stakeholders, i.e. judges, lawyers, the parties involved, to understand and, if needed, oppose the underlying reasoning.

4. Decision implementation. In this phase, the judge must implement the decision by writing up a verdict. The judge could benefit from the usage of the ML algorithm in communicating the decisions in a clear way and monitoring its implementation. In our example, when the presence of an unfair clause is verified, the judge will write up the verdict communicating it to the parties. The application could be available for further explanation and by monitoring the situation with the help of the police, could notify the judge of any anomaly.

As should be evident from this overview, judges may generally benefit from this technology, but they should be aware of possible limitations and problems. Most existing Legal Tech applications are centred around information acquisition – with intermediate and high level automation – and information analysis – with low automation. In these phases, ML algorithms can efficiently analyse vast amounts of legal texts, statutes, case laws and precedents to provide relevant information for judges' consideration, saving considerable time and effort, and improving the comprehensiveness of the research phase. This could also, however, introduce problems that will be reflected in the next phases. For example, if the algorithm used to identify unfair clauses contains an overrepresentation of women in the dataset, it is more likely to identify unfair clauses for this group even if the clauses are not unfair. There are also instances of Legal Tech applications related to decision selection, that, even with a low level of automation, did not produce outcomes with the desired level of trustworthiness. It is important to note that, while ML can provide assistance to judges, the final decision remains with them. No Legal Tech applications should provide full automation in any of the phases, especially the latter ones.

To ensure that the ethical aspects are duly considered when judges are supported by ML algorithms, judges should be made aware of the potential inaccuracies and problems that may arise. Even though AI generates value, it is not infallible and brings significant risks and potential adverse impacts if steps are not taken to address unfairness and inaccuracies.[34]

## 4. Responsible AI: technical aspects and risk analysis

To address the challenges presented in the previous section, a new field – Responsible AI – has emerged. Responsible AI encompasses a set of principles and practices aimed at ensuring that AI technologies are designed, implemented, and used in a manner that aligns

[34] EUROPEAN COMMISSION, *Shaping Europe's Digital Future*, April 2021, available at: https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

with societal values, respects human rights and minimizes potential risks and biases. It emphasizes the need for human oversight, interpretability of AI decisions, and clear guidelines for the responsible use of AI in sensitive domains.[35]

The non-implementation of principles can expose structures to the risks presented before. Thus, when assessing a system, it is important to consider the following five principles:

- Robustness, defined as the ability of a system to maintain its level of performance under a variety of circumstances.[36] The basic idea is that systems should be reliable, safe, and secure, not vulnerable to tampering or compromising the data they are trained on. The adversary
-
- purpose is to influence the learning process to achieve specific outcomes or to manipulate the system's behaviour in a desired way by, for example, raising ethical concerns.

  Consider an AI system trained on a vast amount of legal text, to assist in the analysis and classification of legal documents. It demonstrates high accuracy, providing positive answers in most of the tested cases. A robustness or reliability problem arises when the AI system encounters ambiguities, minor changes between models or the inclusion of a certain clause, or complex legal language, such as sophisticated phrasing and intricate sentence structures, that can lead to misinterpretations.

  AI systems should have robustness measures for detecting and mitigating attempts at algorithm tampering, thereby ensuring the integrity and trustworthiness of the system in legal contexts. Furthermore, by acknowledging the limitations of AI systems in understanding complex legal language and promoting human-AI collaboration, we can enhance the robustness and reliability of AI systems in the legal domain.
- Efficacy is the effectiveness or capability of an AI system to achieve its intended goals or objectives. It evaluates how well the AI system performs in achieving the desired outcome or solving the problem it was designed for. For these reasons, efficacy considers different factors such as accuracy, efficiency, usability, and each of them has a different evaluation system. For example, to reach an extremely accurate result, a lot of material is required which increases the time and resources needed to analyse it, which affects the system's overall effectiveness. Despite producing accurate answers, the system's effectiveness is limited by real-world constraints, which can cause delays and inefficiencies in the judicial process.

  Privacy generally refers to the norms and practices that help safeguard human autonomy, identity, and dignity. It is defined as the right to control over personal information, which refers to each piece of data that can be linked to a person, i.e. the "data subject," such as date of birth, social security number, or fingerprints. Privacy related risks may influence security, bias, and transparency and come with trade-offs with these other characteristics.[37] It is important to remember that the more precise the ML algorithm, the less privacy an individual may have.

---

[35] D. W. TIGARD, *Responsible AI and moral responsibility: a common appreciation*, in *AI Ethics*, 2021, 1, 113, available at: https://doi.org/10.1007/s43681-020-00009-0

[36] Source: ISO/IEC TS 5723:2022.

[37] NIST, *AI Risk Management Framework*, 2023, available at: https://www.nist.gov/itl/ai-risk-management-framework

A software example is Palantir Gotham, used by police departments around the world to predict crimes. In 2017, the state of Hesse, Germany adopted it. Gotham permits the police to collect and store the personal data of people who are not suspected of any crime. In 2020, the German Constitutional Court ruled that the use of this software was unconstitutional because it violated the right to privacy.[38]

- Bias refers to a systematic deviation from a true value or an objective standard, resulting in a consistent favour or prejudice towards certain outcomes.
  An example of software with a bias problem is COMPAS, analysed by a famous article from ProPublica.[39] The purpose of this software is to estimate the likelihood that a defendant will be arrested again, based on mostly demographic data. The defendant's race is not directly used in the calculation, however, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways: the formula was particularly likely to falsely flag black defendants as future criminals, wrongly labelling them such at almost twice the rate as white defendants; on the other hand, white defendants were mislabelled as being low risk more often than black defendants. This happened because it took previous cases into consideration and also sought data related to arrested relatives. In the US system, in which it was applied, there is an important correlation between the ethnicity of people and number of arrests, so this produced even further discrimination. Moreover, the algorithm used to create risk scores is the product of a for-profit company which, being a trade secret, did not to publicly disclose the calculations used to produce them, so it was not possible for either defendants or the public to see what might be driving the disparity.
  However, we have to consider that if AI could accurately predict which defendants were likely to commit new crimes without human bias, the criminal justice system could be fairer and more selective about who is incarcerated and for how long.
- Explainability refers to a representation of the mechanisms underlying the operation of AI systems, it is connected with interpretability, which refers to the meaning of output from AI systems in the context of their designed functional purposes. For instance, if we consider an AI system that is expected to predict with high accuracy the results of legal cases involving job discrimination, the system can emphasize surface-level details, such as job titles or the firm sizes involved in the instances, rather than understanding the legal considerations that establish discrimination. The AI system may produce biased or incorrect predictions based on the wrong assumptions.
  The lack of explainability and interpretability can be problematic in legal cases, as it becomes challenging to understand and justify the AI system recommendations. Ensuring explainability in AI systems used for legal analysis promotes transparency, accountability and trustworthiness. For these reasons judges should exercise caution when relying on AI technology by making themselves aware of potential limitations and giving thoughtful consideration to AI system outputs in the legal decision-making process.

To ensure that AI systems are developed and used in a responsible way both a technical and risk analysis should be conducted. Technical analysis focuses on the technical aspects

[38] M. MEAKER, *Germany Raises Red Flags About Palantir's Big Data Dragnet*, in *Wired*, February 2023, available at: https://www.wired.co.uk/article/palantir-germany-gotham-dragnet

[39] J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, *Machine Bias*, ProPublica, 2016, available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

of a system, such as the quality of the data used to train the system, the algorithms used to make decisions, and the security of the system. Risk analysis focuses on the potential risks posed by the system, such as biases, misuse, and privacy violations. Technical and risk analysis involve ongoing monitoring, evaluation, and mitigation, ensuring the system is not vulnerable. The main steps to performing technical and risk analyses are:

1. Risk Identification: identifying the potential threats and vulnerabilities of the system, along with their potential negative impacts;
2. Risk Assessment: estimating the probability and the consequences of each threat;
3. Risk Mitigation: identifying countermeasures to reduce the risk of each threat;
4. Implementation and Testing: applying those countermeasures and conducting tests;
5. Monitoring and Review: regularly checking the situation.

Some examples of how technical and risk analysis can be used to support Responsible AI are, in the case of the former, the possibility of identifying potential biases in training data, such as an over-representation of certain groups and, concerning the latter, the identification of risks associated with usage, by considering the potential impact on people and the possibility that it could be misused.

It may be evident that judges are often considered non-technical individuals when it comes to evaluating the trustworthiness of AI systems. This highlights the necessity for external support. To address this, it may be extremely helpful to provide tool-based support to enhance judges' awareness and understanding of AI systems. The primary goal is to develop an approach that empowers AI tool users and allows them to understand the inherent risks associated with these systems, as well as the ways these risks could compromise the principles of trustworthy AI. By adopting a risk-aware approach, users (judges in this case) can ensure that AI systems are developed and deployed in a manner that is responsible, ethical, and respectful of fundamental rights and values.

# 2.3. The evolution of search engines and their application to Justice: opportunities and challenges. A technical perspective

*Paolo Ferragina – Università degli Studi di Pisa & Scuola Superiore Sant'Anna*

*Summary: 1. Premise – 2. Algorithm versus AI definitions – 3. The history of Web Search Engines – 4. The role of AI in modern Search Engines – 5. Applications of search engine technology in Justice*

*Abstract: Inspired by Lessig, this contribution considers that, in order to understand the "age of cyberspace" we currently live in it is necessary to dig into the nature of all the impressive advancements that Computer Science has achieved in the last twenty years: mostly based upon search engines and AI. This chapter is divided into four main parts that deal with the definition of Algorithms and AI (often considered synonyms!), the history of web search engines (it is argued this "storytelling" will allow the reader to better understand the "opportunities and challenges" in their application to Justice), the recent role of AI in their design (which is a mix of ML/AI-based techniques for Natural Language Processing, Knowledge Graphs, and Generative AI), and conclude with a discussion on the value "that could be added" to legal search engines with all these algorithmic and AI technologies. Overall, this discussion will argue that more research and software development is still necessary in order to make the searching and mining of legal document collections easier, faster, more accurate, more "intelligent," and serendipitous in offering hints and views on legal arguments.*

## 1. Premise

Writing about Search Engines and their long journey from Digital Libraries to Web giants like Google and Bing is much too long a story to be presented in only a few pages. Nonetheless, a glimpse of their main underlying technologies and evolution is warranted because this "storytelling" allows the interested reader to better understand the "opportunities and challenges" in their application to Justice, particularly in light of the bursting developments that have recently affected Artificial Intelligence (AI) and Machine Learning (ML), which are two new key ingredients of modern search technology.

Motivation for the content and structure of this chapter is taken from the following excerpt of the book "Code and other Laws of Cyberspace," written in 1999 by Prof. Lawrence Lessig,[40] who wrote: "Ours is the age of cyberspace. It, too, has a regulator. […] This regulator is code—the software and hardware that make cyberspace as it is. […] In a host of ways that one cannot begin to see unless one begins to understand the nature of this code, the code of cyberspace regulates." Inspired by Lessig, I also believe that, with the aim of understanding the "age of cyberspace" we currently live in, even more now than in 1999, we must to dig into the nature of "code" and begin to acquaint ourselves with two ubiquitous terms: Algorithms and AI.

## 2. Algorithm versus AI definitions

The Oxford English Dictionary states that an Algorithm is, informally, "a process, or set of rules, usually one expressed in algebraic notation, now used especially in computing, machine translation, and linguistics." The modern meaning for Algorithm is quite like that

[40] Roy L. Furman, Professor of Law at Harvard Law School and the former director of the Edmond J. Safra Center for Ethics at Harvard University.

of the terms "method," "procedure," "routine," except that the word Algorithm in Computer Science connotes something more precisely described. The recognized definition worldwide for the word Algorithm is due to Donald E. Knuth, Professor Emeritus at Stanford, who stated at the end of the 1960s that "an Algorithm is a finite, definite, effective procedure, with some output."[41] Although these five features may be intuitively clear, their significance is so dense that we need to look at some of them in more detail, as this investigation will lead us to better understanding the difference the terms algorithm and AI. We restrict our attention to:

- Definite: "each step of an algorithm must be precisely defined; the actions to be carried out must be rigorously and unambiguously specified for each case." This means that anyone reading the algorithm's description will interpret it in a precise way and nothing will be left to personal choice. This unambiguity is currently guaranteed by using one of many programming languages such as C/C++, Java, or Python.
- Input-Output: the behavior of the algorithm is not unique, but depends on the data given as input to be processed, which produces an output that constitutes the answer returned by the algorithm for those inputs. The mapping between inputs and outputs is precisely defined by the problem the algorithm must solve and must be "guaranteed" for all possible inputs. This is the so-called correctness of the algorithm.

On the other hand, there are many definitions of Artificial Intelligence that, according to some statements published by the European Parliament, either "refer to systems that display intelligent behaviour by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals"[42] or "is the ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity."[43]

Among others, I prefer instead the definition given by David L. Parnas[44] in 2017, who described the AI approach as "heuristic programming." This definition is different from that given above for Algorithm because a heuristic program is one that "does not always get the right answer." Heuristic programs are based on rules that hang on experience, but are not supported by hand-written code or theory. Typically, "heuristic" is not a desirable attribute of software, but has been used effectively in recent years in more and more contexts, i.e. natural language understanding, audio and video processing, chat, text generation and translation, where finding a mathematically precise definition of the problem to be solved is difficult (if not impossible!). This approach gained popularity thanks to impressive advancements in the field of Machine Learning, which is another approach to the creation of Artificial Intelligence by constructing programs that "learn" from examples.

---

[41] D. E. KNUTH, *The Art of Computer Programming*, vol. 1-4, Addison-Wesley, 2023.

[42] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS, *Coordinated Plan on Artificial Intelligence*, COM(2018) 795 final, December 7, 2018, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0795&rid=3#:~:text=Artificial%20Intelligence%20refers%20to%20systems,autonomy%20%E2%80%94%20to%20achieve%20specific%20goals

[43] EUROPEAN PARLIAMENT, *What is Artificial Intelligence and how is it used?* in *News*, September 4, 2020, available at: https://www.europarl.europa.eu/news/en/headlines/society/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used

[44] D. LORGE PARNAS, *The Real Risks of Artificial Intelligence*, Communications of the ACM, October 2017, 60, 10, 27, doi: 10.1145/3132724.

This seems easier than designing an efficient algorithm that solves the actual problem and then coding for it. Nevertheless, these ML-based programs may be untrustworthy because they often exhibit weaknesses like incomplete or biased experiences, which are plugged into the system as "examples" and then turned into wrong answers by the learning process. In fact, ML solutions may fail when given unusual or untrained inputs, i.e. situations.

Let's now focus our attention on one of the most ubiquitous and sophisticated software tools currently available on every personal device, deploying the most advanced algorithms and AI/ML techniques ever designed by academia and industry: the Search Engine. Its simplicity of use has driven most users to consider it a trivial software to build.

Let's now briefly summarize the history of search engines over the last 30 years, and then "re-map" it onto the case of search tools for Justice. Sometimes a parallel example helps understand what can and cannot be done.

## 3. The history of Web Search Engines

The history of search engines, as we know now, is strongly linked to the history of the Web, which was born in 1991. The context at that time was very different from the present: there were only a few internauts and the Web consisted of only a few million well-maintained and reliable documents, i.e. pages, belonging to government or university sites. Those search engines, with names like Wanderer and Aliweb which are all now but forgotten, were based on extremely elementary algorithms for searching for user-specified keywords through meta-information that the authors of the pages manually associated to them. The search proceeded on that meta-information by using a linear scan, which was efficient only because of the limited number of existing Web pages.

The sudden growth in size of the Web made these approaches completely inefficient and new search engines were born, with perhaps more familiar names such as AltaVista, Lycos, Excite, and Yahoo!. These search engines introduced a set of criteria that could be used for sorting the results of a search for the first time, since they were growing more and more numerous given the growth of the Web. The concept of the relevance of results emerged which was addressed by means of two primary approaches: first, the Boolean retrieval model and, then, the more powerful Vector-space model. The former was primarily derived from the DataBase setting: a query consisted of a set of logical criteria for retrieving documents. The criteria specified the presence, and possibly the proximity, of indicated terms for documents to be responsive. The relevance measure ranked results in terms of how completely the Boolean criteria for the query were satisfied. However, terms were given "the same" weight, although this did not reflect their frequency or discriminative power, i.e. articles *versus* nouns. In the Vector-space approach to relevance, documents and queries were represented as vectors in a multi-dimensional space in which dimensions corresponded to terms and each vector component denoted the frequency of the corresponding term within the document and across the document collection (the so-called TF-IDF score). This representation surpassed the term-agnostic limitations of the Boolean retrieval model, and allowed for computing the similarity of documents/queries in purely mathematical terms by means of the scalar product of vectors. The final retrieval results achieved by Altavista and its competitors, at that time, were excellent and depended heavily on the fact that documents available on the Web were of high quality.

Around the year 1997, use of the Web in the business sphere and knowledge of how search engines worked paved the way for malicious practices aimed at influencing the ranking

of search results, a practice now known as "spamming." This heavily penalized the performance of search engines, making them often unusable for queries that contained frequent terms of interest to Web users. Countermeasures therefore became necessary as it soon became clear that Web page content alone was insufficient for determining their relevance to users' queries.

The subsequent biennium marked the beginning of the third generation of search engines and coincided with Google's birth, with its famous PageRank algorithm crucially based on the interconnections between Web pages, i.e. their hyperlinks. This generation of search engines, to which Ask Jeeves, Yahoo!, and Bing also belong, dominated the Web search scenario during the following decade. In the initial version of Google, the relevance of a page depended on its content, as in AltaVista, but also on the relevance of other pages pointing to it as well as which text, so called "anchor-text," surrounded those hyperlinks. This "centrality" measure was named PageRank, it was recursive in nature, and has proven to be one of the most important and persistent measures used for determining the relevance of a "node" in a network, whether it be on the Web, a social network, or a set of posts on Instagram, Facebook, or Twitter. Even now, whenever there is a network to analyze, a descendant algorithm of PageRank is typically one of the first options to consider. The third generation of search engines thus combined the textual information contained in Web pages and in anchor-texts with general information on the structure of the Web graph. This approach was so effective at answering user queries that second-generation search engines soon disappeared within a short time.

But, as is often the case in the world of Web search engines, the mechanisms for determining page relevance were quickly threatened by new spamming techniques, the most famous of them termed Google bombing.[45]

We are now living in the age of fourth generation search engines, in which there is world-class engagement between the two giant protagonists, Bing and Google, plus a multitude of others at the national level – such as Baidu in China and Yandex in Russia –providing specific contents (i.e. products, publications, users, maps), or claims of "semantic" searches, i.e. DuckDuckGo, and the most recent AI-based versions of Bing and Google, that interpret users' questions and carry out an in-depth analysis of document content. This latest generation is marked by an improvement in the efficiency and effectiveness of the search technology to "understand" the user query and document collection.

## 4. The role of AI in modern Search Engines

The above-mentioned capacity to "understand" hinges on ML/AI-based techniques for Natural Language Processing and Understanding (NLP/U), Knowledge Graphs (KG), and the latest advancements in Generative AI, such as ChaptGPT. The first techniques are used to process input texts, identify keywords or entities (possibly formed by multi-words) and then perhaps assign roles to those tokens and sentences, with their eventual "meaning." Here, "meaning" can refer to their Part of Speech – the so-called PoS, i.e. subject, verb – or, more interestingly, the corresponding concept in a Knowledge Base, such as Wikipedia or DBpedia. The former case fits into the classical realm of Computational Linguistics, which dates back to the 1960s, but with a revamped interest and more effective algorithms thanks

[45] https://en.wikipedia.org/wiki/Google_bombing

to the recent progress of AI/ML-tools. The latter was born in 2012 with Google's introduction of the first very-large Knowledge Graph, also known as a Semantic Network,[46] which represents a network of real-world entities – i.e. persons, events, objects, or concepts – and models the relationship between them thanks to links with (possibly many) associated types. There are now many known and freely available Knowledge Graphs. The key idea of their use in Search Engines is to disambiguate terms in indexed pages or query keywords by linking these terms to the proper corresponding nodes in the KG, i.e. Wikipedia pages. This represents not only a new way of mapping terms to concepts, but also a manner of empowering machines to extract interrelated concepts by percolating the KG starting from those nodes. By way of example, the query "Leonardo painted the Mona Lisa" clearly refers to the scientist and artist Leonardo da Vinci. So, the search engine, with the help of a KG, i.e. Wikipedia, connects "Leonardo" to the node representing "Leonardo da Vinci," i.e. https://en.wikipedia.org/wiki/Leonardo_da_Vinci. After which point, traversing the adjacent nodes in the KG, the search engine could discover the cities of Vinci and Florence, their nation Italy, or other information related to the famous scientist, such as the fact that he lived during the Renaissance. In some sense the KG expands the notion of ontologies to other kinds of entities and allows software engineers to develop highly sophisticated techniques for semantically annotating texts to support more intelligent and concept-based searches.

It goes without saying that the size and quality of the KG is crucial for the concreteness and completeness of these concept-based searches and reasoning, which requires very sophisticated AI techniques and algorithms to process and digest large volumes of (often unstructured) texts from which that Knowledge is extracted and interconnected to form these Graphs.[47] This approach gives algorithms the power to reason about the significance of terms and texts, along with finding similarities that go much beyond the (syntactic) sharing of terms.

More recently, this "understanding capacity" has been further extended with the advent of Transformers[48] and other sophisticated ML tools that, by processing billions of texts, extract mathematical representations of keywords capturing, in a sense, their "semantics" from their co-occurrence with other words in those large textual collections. The most notable ML-techniques in this setting are the original GPT (Generative Pre-trained Transformer)[49] and BERT (Bidirectional Encoder Representations from Transformers).[50] The former has been mainly used to generate human-like text beginning from questions or phrases, so-called prompts; the latter has been mainly used to build tools for some important end-to-end applications, such as entity recognition and document classification, thanks to novel and effective vector-based representations of token/words or sentences, which bring with themselves useful context-based (semantic?) information. Note that these approaches are orthogonal to previous ones and thus can be – and indeed have been – used to mine KGs

[46] https://en.wikipedia.org/wiki/Knowledge_graph

[47] See X. L. DONG, E. GABRILOVICH, G. HEITZ ET AL., *Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion* in *Proceedings of* the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2014, New York, USA, 601.

[48] https://en.wikipedia.org/wiki/Transformer_(machine_learning_model)

[49] T. WOLF, L. DEBUT ET AL., *Transformers: State-of-the-Art Natural Language Processing*, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020, 38, doi:10.18653/v1/2020.emnlp-demos.6.

[50] *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*, in *Google AI Blog*, November 2, 2018.

or reason more efficiently than them (though it is too technical to discuss the algorithmic details here).

## 5. Applications of search engine technology in Justice

We are now ready to dig into the "value-added" or "value that could be added" to legal search engines with respect to Web search engines, in light of the technological achievements just outlined above.[51] The two most known and notable commercial search systems for legal documents in the United States are Westlaw Next (WN) and Lexis Advance (LA), which offer many of the features commented on below and that should be compared with the features offered by, for example, the Italian *Italgiure* system, currently available for the personnel of Italian Courts.[52] All those systems, in one way or the other, have followed the evolution of Web search engines, though they have not yet achieved the same level of sophistication and efficacy, for many reasons that are also intrinsic to the nature of legal documents and user needs.

The first issue to deal with here is the composition of a user query. In the legal context, more than in the classic Web search, syntactic searches are not enough to match the needs of (legal) users. Concept-based retrieval is essential, and is becoming more and more mandatory as the size of digitalized legal document collections has increased in the last few years. Additionally, the use of techniques such as query auto-completion or query expansion could turn out to be very effective in empowering legal users to design queries that are better composed. In the former, a "dictionary" of potential queries is fundamental upon which a user's query is matched for its completion. Web search engines use dictionaries drawn from many sources, the most important of which are query-logs, built by the search engine during its use. In the latter, legal ontologies or, better, legally-centered Knowledge Graphs should be built to properly interpret and then expand the "syntactic" queries posed by users with additional meaningful terms that solve polysemic or synonym issues present in them.

In this context, a crucial role could also be played by so-called user relevance feedback, which would have a key role in "personalizing" the ranked list of search results via expert-generated annotations. The key algorithmic idea would be to flag some relevant results, in a sort of human-in-the-loop feedback system, that helps the machine learn a model embodying those judgments in a way applicable to new documents. Re-ranking might employ evidence derived from those expert-generated annotations, frequency information in the text of documents, citation networks and document popularity from previous queries. The (re-)ranking function could also be optimized by using ML to determine the weights to ascribe to those different features.

It is evident that the interpretation of user queries is strictly tied with the analysis of legal sentences. This action may occur at different levels of granularity: from classic and simple Part-of-Speech tagging to Entity linking discussed above, up to ultimately adding more semantic information about the role of the sentences in legal arguments. These "annotations" could be exploited by the search engine to rank the results depending on the "role" played by the searched terms into the indexed sentences, but also to reason about the concepts

[51] See K. ASHLEY, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge, 2017.

[52] https://www.italgiure.giustizia.it/

involved in those results or to visualize them better by deploying representation schemes that surpass the usual linear list.

In fact, the third key issue to be mentioned is related to the visualization of query results. Web search engines have habituated us to the linear list of so-called "ten blue links," with their associated snippets of a few lines and searched keywords in bold. This choice leaves it up to the users to read the list of resulting documents and decide which of them is relevant and refine the query based on what was found or not. Although poor from an effectiveness perspective, the linear list has become the de facto visualization standard, mainly dictated by the small screens of our mobile devices, the simple needs of Web users, i.e. navigational, transactional, and informational, and by their short patience (a few seconds…).

In the legal context, the scenario is much different: users may spend time looking carefully through the search results as their needs are more conceptual, they are also possibly open to serendipitous suggestions, and legal-document collections are not as big as the Web. These characteristics should lead search-engine designers to afford for a more sophisticated processing of documents and queries while, on the other hand, they should be advised that legal language is more "complicated," so that polysemy/synonym issues, as well as the role of words/sentences in those documents, are taken into account. These issues were already discussed in the previous section, with respect to their impact on the (re-)ranking of search results. Browsing a plain (linear) list of hundreds or thousands of search results (possibly sorted by date or other simplistic criteria) isn't humanly feasible and can't be performed with the right level of attention even for legal users. As such, it should be declined. In this respect, the legally-centered Knowledge Graphs mentioned above as well as the citation networks, based on articles, entities, or concepts mentioned in the resulting documents, could be deployed to arrange search results in the form of *graphs* that could be far more visually effective than the plain list and allow users to extract a fast and meaningful glimpse from them. Moreover, graphs are browsable in several conceptual dimensions, thus resulting in a greater flexibility than linear lists to adapt to the search needs of users and to accompany them more efficiently and effectively in retrieving what they were searching for or to discover "new" concepts or arguments that are useful for their work.

The fourth and last issue is dedicated to the impressive progress of generative AI, and ML in general, that is often succinctly summarized these days by the tool "ChatGPT." This family of tools is referred to this way because Search and Generative AI will progressively converge. Generative AI went mainstream in 2022 with ChatGPT and Dall-E, offered by OpenAI, after the seminal work done by Google with Transformers. However, Generative AI still needs a lot of effort to check the factuality and groundedness of its generated phrases, by preventing "hallucinations" (I prefer "rambling") in texts that look correct, but which are, in fact, not. Web search, on the other hand, could help with fact checking algorithms and with Web references provided together with AI-generated text.

But there is another way to "effectively merge" these two approaches, as some tools/apps are already pursuing in various contexts: namely, to use ChatGPT for reasoning about a large part of search results. That is, deploying the power of ChatGPT or similar tools to produce humanly-readable summaries of "significant parts" of search results, to answer specific questions regarding them, to extract argument-related information, or finally to order the results in a manner that is tailored to the problem a user seeks to address. This would avoid the fatigue of reading hundreds of results snippets and generate outputs that are currently not possible with current textual searches. In this scenario, it can be argued that search engines for legal documents could benefit the most by the mixing of Generative AI, classic

AI/ML, and algorithms, thus making the most advanced search assistant one could think of available to their users.

Let us conclude this chapter by mentioning that, in the last year, we have investigated some of these challenges and issues within the context of Italian legal documents, thanks to the support of a PNRR-PON project, named "*Giustizia Agile,*"[53] that has seen fruitful collaboration between the Court of Pisa and several Departments of the University of Pisa. Its goal was the study, design, implementation, and experimentation with a software platform for the analysis, indexing, search, and visualization of Italian legal documents, by following some of the methods suggested above. During this project, three main gold standards were constructed semi-automatically – for NER, document classification, and keyword extraction – that could be adopted for further studies with Italian legal documents, and a preliminary experiment on keyword extraction via Generative AI was proposed. Overall, the study demonstrated that a lot of research and software development is still necessary in order to bring Italian legal search engines up to the task of matching the (not so futuristic) vision discussed in the paragraphs above, in which the most advanced AI and algorithms blend together to make the searching and mining of legal document collections easier, faster, more accurate, more intelligent, and possibly serendipitous in offering hints and views to legal arguments for the daily work of court officials, judges, and lawyers.

In the end, the opportunities for applying "intelligent" search engines as well as interesting and powerful AI/ML and algorithmic techniques to the justice sector already abound these days both in industry and academia. Advancements in legal search engines are sure to surprise us in the not-so distant future.

[53] A special thanks to all colleagues and students who collaborated on implementing the software platform we designed for the PNRR-PON project, "Giustizia Agile:" Leonardo Calàmita, Piero Cossu, Matteo De Francesco, Chiara De Nigris, Alessandro Lenci, Giacomo Mariani, Giovanna Marotta, Lucia Pàssaro, Erika Pistolesi, Mattia Proietti, and Giacomo Vaiani. A warm thanks also goes to the colleagues (especially, PIs Benedetta Galgani and Giuseppe Campanelli), researchers, and fellows of the Departments of Law and Management Engineering, as well as to the officials and judges of the Courts of Pisa, Lucca, and Livorno, who contributed to shaping and driving the success of this project. My final warmest thanks and gratitude goes to Maria Giuliana Civinini, who was the real "[search] engine" of collaboration that accompanied the intense study and software design behind this project.

# Section 3

# The International Legal Framework

## 3.1. The Council of Europe's perspective on Artificial Intelligence

*Nilcan Ozalp – Council of Europe*
*Yannick Meneceur –Université de Strasbourg*

*Summary: 1. Premise – 2. The early work on algorithms, Big Data, and Artificial Intelligence – 2.1. Definition of concepts – 2.2. Work on algorithms, Big Data, and AI – 3. The transversal work of regulating AI – 3.1. The Ad-Hoc Committee on Artificial Intelligence (CAHAI) – 3.2. Committee on Artificial Intelligence (CAI) – 4. Work of specialized sectors relevant to AI and the field of justice – 4.1. Cyberjustice – 4.2. Data protection – 5. Conclusion*

*Abstract: The Council has long been at the forefront of technology regulation in Europe. Since the Convention on the Protection of Personal Data adopted in 1981, the Council has intervened to ensure that the development of information technology in its Member States is carried out with respect for human rights, democracy, and the rule of law. The revival of artificial intelligence applications since the early 2010s has led the Council to propose various types of instrument, including an ethical charter for use in judicial systems, and finally, and most importantly, the first-of-a-kind treaty: the Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law in 2024.*

### 1. Premise

The Council of Europe places human rights at the centre of its approach to Artificial Intelligence (AI), emphasizing that the design, development, and deployment of this technology must be in line with Council of Europe standards. The work of the Council of Europe in the field of AI seeks to guide member states in navigating the digital environment while safeguarding the principles of human rights, democracy, and the rule of law. In addition to its first-ever international, legally-binding treaty on AI, the Council of Europe has been addressing common issues related to AI governance, standards, and best practices in various related fields, including data protection and Cyberjustice, through different committees and working groups.

This article will explore the Council of Europe's early work on AI technology and the trajectory of the Framework Convention of AI. It will also cover the work of specialized sectors relevant to AI and its use in the judiciary field.

### 2. The early work on algorithms, Big Data, and Artificial Intelligence

#### *2.1. Definition of concepts*

Defining AI presented no small difficulty for regulators to agree on precise and consistent terms in the titles of their public policies. In some documents, the choice of words is often only an editorial choice, reflecting the consensus reached among experts who contributed to the work. In other documents the tendency of editors to not focus solely on the impact of

*machine learning* led to a search for the most appropriate terms in order to cover the subject more broadly and in which the terms algorithm, Big Data, and AI were used equivalently in the normative output or studies of many intergovernmental organisations as a result. These algorithmic processing methods have not always been identified as a central element in some of the early IT-related documents, such as texts on the prioritisation of online content or mass surveillance.

The Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law[54] opted for the definition taken from the OECD defining AI systems as "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments." The definition not only captures the dynamic and multidisciplinary nature of AI technologies but also considers the various contextual interpretations in the governance of AI at the global level. It should be emphasized that the European Union has also aligned itself with the definition set out by the OECD in its AI Act, adopted in 2024. From a legal point of view, this convergence brings some clarity to the notion of AI, even though the field remains in continuous evolution from a technological point of view.

### *2.2. Work on algorithms, Big Data, and AI*

The Council of Europe has been exploring issues related to the *information society* since the early 2000s. The Council has addressed the issue primarily from the standpoint of freedom of expression (Article 10 of the European Convention on Human Rights), media, and Internet governance.

In 2010, the Committee of Ministers (CM) adopted a Recommendation on the protection of individuals regarding the automatic processing of personal data in the context of the profiling[55] of personal data processed by calculation, comparison, and statistical correlation software, and which addressed the need for extra safeguards to protect personal data and privacy. The Parliamentary Assembly of the Council of Europe (PACE) adopted a Resolution on mass surveillance[56] in 2015 which addressed deep concerns about the collection of large amounts of personal data by intelligence agencies, the use of personal data for unlawful purposes by State or non-State actors, and the need for adequate legal regulation and oversight. Although these documents do not explicitly mention algorithms, Big Data, or AI, they suggest that the widespread use of these technologies has effects far beyond the sole issue of data protection and a respect for private and family life.

---

[54] COUNCIL OF EUROPE, Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (CETS No: 225) available at: https://rm.coe.int/1680afae3c

[55] COUNCIL OF EUROPE, *Recommendation on the protection of individuals with regard to automatic processing of personal data in the context of profiling*, CM/Rec(2010)13, November 23, 2010, available at: https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805cdd00

[56] COUNCIL OF EUROPE, *Resolution on mass surveillance*, Resolution 2045(2015), April 21, 2015, available at: https://pace.coe.int/en/files/21692/html

In 2017, the PACE published a Recommendation on technological convergence, AI, and human rights[57] highlighting the need to strengthen existing regulation and which addressed the need for internet governance that was not dependent on private groups or States. An Expert Committee on Internet Intermediaries (MSI-NET) was established with a two-year mandate (2016-2017) under the Steering Committee on Media and Information Society (CDMSI). It produced a study entitled "Algorithms and Human Rights"[58] on "the human rights dimensions of automated data processing techniques and possible regulatory implications:" one of the first documents of the Council of Europe to comprehensively address the consequences of the widespread use of algorithms. The text set out several principles and actions, which included transparency, accountability, ethical frameworks and better risk assessment, regarding the possible impacts of algorithms on human rights.

In 2019, the CM adopted a Declaration on the manipulative capabilities of algorithmic processes,[59] warning against the risk of using algorithmic processes to manipulate individuals' social and political behaviour and which underlined the need for additional protective frameworks. It included key areas of action such as conducting human rights impact assessments, public consultations, data protection and privacy, and independent oversight. In the same year, the Council of Europe Commissioner for Human Rights issued a Recommendation entitled "Unboxing artificial intelligence: 10 steps to protect human rights,"[60] which built on work done in this area by the Council of Europe and other international organisations. The Recommendation listed a number of steps to be taken by national authorities to prevent the negative impact of AI systems on individuals while maximizing their potential. In the framework of the CDMSI, the Expert Committee on the human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT) published a study on Responsibility and AI. The study focused on the implications of advanced digital technologies, including AI, on human rights and fundamental freedoms and underlined the need for accountability for the adverse consequences of such technologies.

All this work has gradually converged to illustrate the increased need for regulation, both inside and outside new digital environments that goes well beyond the issue of personal data protection or privacy. This work consistently recognises the potential of this new socio-technical context to improve daily life, but also to seriously undermine human rights, democracy, and the rule of law. The Council of Europe's ambition to explore issues related to AI has gradually become a political priority, in order to support member States in their digital transformation, but also to guide them towards preventing actual or potential risks to individuals and society as a whole.

---

[57] COUNCIL OF EUROPE, *Recommendation on technological convergence, artificial intelligence, and human rights*, Recommendation 2102(2017), April 28, 2017, available at: https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=23726

[58] COUNCIL OF EUROPE MSI-NET, *Algorithms and Human Rights*, DGI(2017)12, 2018, available at: https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10

[59] COUNCIL OF EUROPE COMMITTEE OF MINISTERS, *Declaration on the manipulative capabilities of algorithmic processes*, CM Decl(13/02/2019)1, February 13, 2019, available at: https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b

[60] COUNCIL OF EUROPE COMMISSIONER FOR HUMAN RIGHTS, *Unboxing AI: 10 steps to protect human rights, Recommendation by the Commissioner for Human Rights*, May 2019, available at: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

## 3. The transversal work of regulating AI

### *3.1. The Ad Hoc Committee on Artificial Intelligence (CAHAI)*

In 2019, the CM set up an *ad hoc* Committee on Artificial Intelligence (CAHAI) to examine the feasibility of a legal framework for the development, design and application of AI based on the organisation's standards on human rights, democracy, and the rule of law. In parallel, the CM adopted a new Recommendation on the human rights impacts of algorithmic systems[61] in 2020 with guidelines for public and private sector actors on how to take these impacts into account in order to prevent human rights breaches. This document, prepared by the Committee of Experts on the Human Rights Dimensions of Automated Data Processing and Artificial Intelligence (MSI-AUT) calls on Member States to ensure that algorithmic systems incorporate safety, privacy, data protection and security safeguards by design. That same year, the PACE issued a report entitled "Need for democratic governance of AI"[62] focusing on the impact of AI on democracy and placing particular emphasis on the adoption of a legally binding treaty-like framework.

The CAHAI delivered its feasibility study, based on a broad consultation with academia, civil society, and the private sector[63] in 2020. This study proposed developing "an appropriate legal framework" which would consist of "a combination of binding and non-binding legal instruments, which complement each other." More specifically, the study mentions "that a binding instrument, a convention or a framework convention, of a horizontal nature, could consolidate the common general principles – contextualised to apply to the AI environment and using a risk-based approach – and include more granular provisions in line with the rights, principles, and obligations identified in this feasibility study." The study also identified "key substantive rights" and "key obligations" ranging from the need to protect human dignity and prevent violations of human rights, democracy and the rule of law, to the need to ensure human autonomy – for example through human oversight, agency, and ultimate decision making. In addition, it set out transparency rules recognising the need for individuals to know when and under which conditions AI systems were used and how it affected their lives and rights. Emphasis was also placed on the need to ensure the explicability of AI decisions, the auditability of AI systems, and the accountability and responsibility of those involved in the life cycle of an AI system. According to the feasibility study, the future legal framework should also be dynamic, applicable throughout the life cycle of an AI project, and adaptable to current and future technologies. In concrete terms, proposed mechanisms for the operationalisation of rights and principles were also mentioned. These included human rights impact assessments to identify, prevent, mitigate, and report on the negative impacts to human rights arising from these activities and the establishment of regulatory sandboxes and automated monitoring of the operation of AI systems. The CAHAI also simultaneously

---

[61] COUNCIL OF EUROPE COMMITTEE OF MINISTERS, *Recommendation on the human rights impacts of algorithmic systems*, CM/Rec(2020)1, April 8, 2020, available at: https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

[62] COUNCIL OF EUROPE COMMITTEE ON POLITICAL AFFAIRS AND DEVELOPMENT, *Need for democratic governance of artificial intelligence*, Doc.15150, September 24, 2020, available at: https://pace.coe.int/en/files/28742/html

[63] CAHAI, *Feasibility Study,* CAHAI(2020)23, December 17, 2020, available at: https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

published a study entitled "Towards regulation of AI systems,"[64] which compiled high-level academic contributions employed by working groups to produce the feasibility study by analysing the challenges arising from AI systems and possible regulatory responses. The study also contained national perspectives from observer States, i.e. Israel, Japan and Mexico, on AI regulation as well as an analysis of legally-binding, international instruments.

In 2021, the CAHAI completed its task, producing "Possible elements of a legal framework on Artificial Intelligence," based on the Council of Europe's standards.[65] The document stressed the need for a legally binding transversal instrument containing fundamental principles for the protection of human dignity and the respect for human rights, democracy, and the rule of law for the development, design and application of AI systems.

### *3.2. Committee on Artificial Intelligence (CAI)*

The Committee on Artificial Intelligence (CAI) was set up by the CM in 2021 and was inaugurated in 2022, bringing together 46 Council of Europe member states, the European Union and 11 non-member states (Argentina, Australia, Canada, Costa Rica, the Holy See, Israel, Japan, Mexico, Peru, the United States of America and Uruguay), and representatives from other international and regional organizations, the private sector, civil society, and research and academic institutions who contributed as observers.

In parallel to CAI's work, the Council of Europe Human Rights Commissioner issued a follow-up Recommendation[66] in 2023 entitled "Human rights by design - future-proofing human rights protection in the era of AI," reviewing the key challenges faced by member States in protecting and promoting human rights in the use of AI. The Commissioner pointed to a lack of comprehensive and human rights-based approaches, insufficient transparency and information sharing, and the lack of initiative on the part of member States to use AI to strengthen human rights. She highlighted the need to reinforce supervision and oversight by independent institutions and to proactively explore the potential of AI to boost, rather than harm, human rights protections.

In its 10th Plenary meeting in March 2024, the CAI adopted the draft Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law.[67] The Framework Convention was adopted by the CM in May 2024, making it the first-ever, legally-binding international treaty. The framework convention was opened for signature in Vilnius, Lithuania on September 5th for the occasion of a conference of Ministers of Justice where it was signed by Andorra, Georgia, Iceland, Norway, the Republic of Moldova, San Marino, the United Kingdom as well as Israel, the United States of America, and the European Union, on behalf of its 27 member States.

---

[64] CAHAI SECRETARIAT, *Towards regulation of AI systems*, DGI (2020)16, December 2020, available at: https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a

[65] CAHAI, *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law,* CAHAI(2021)09rev, December 3, 2021, available at: https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d

[66] COUNCIL OF EUROPE HUMAN RIGHTS COMMISSIONER, *Follow-up Recommendation to "Unboxing AI" (2019). Human rights by design, future-proofing human rights protection in the era of AI*, May 2023, available at: https://rm.coe.int/follow-up-recommendation-on-the-2019-report-human-rights-by-design-fut/1680ab2279

[67] COUNCIL OF EUROPE, *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law* (CETS No: 225) available at: https://rm.coe.int/1680afae3c

The Framework Convention introduced a legal framework that applied over the course of the design, development, and application of AI systems throughout their lifecycle, covering AI systems by public authorities – including private actors acting on their behalf – and private actors themselves. The Framework Convention offered Parties two modalities for compliance with its principles and obligations when regulating the private sector: Parties may opt to be directly obliged by the relevant Convention provisions or, as an alternative, take other measures to comply with the treaty's provisions while fully respecting their international obligations regarding human rights, democracy, and the rule of law.

The Framework Convention did not apply to national defence matters nor to research and development activities, except when the testing of AI systems may have the potential to interfere with human rights, democracy, or the rule of law.

The principles listed in relation to the development and use of AI systems included human dignity and individual autonomy, equality and non-discrimination, respect for privacy and personal data protection, transparency and oversight, accountability and responsibility, reliability and safe innovation. The Framework Convention is human-centred and adopts a risk-based approach to the design, development, and use of AI systems ensuring the prevention of harmful uses of AI systems and promoting the use of this digital technology for the good of society, including by allowing for safe innovation.

In terms of the risk and impact management framework, it stated that each Party shall take legislative and other measures for the identification, assessment, prevention and mitigation of risks and impacts to ensure that AI systems respect human rights, the functioning of democracy, and the rule of law. It also requires each Party to establish effective oversight mechanisms to oversee and supervise compliance with the obligations in their domestic legal system.

The Framework Convention establishes a follow-up mechanism for determining the extent to which its provisions are being implemented. It also serves as a forum for international cooperation for facilitating information on various aspects of artificial intelligence.

The Framework Convention will enter into force after five signatories, including at least three Council of Europe member states, have ratified it. This instrument will potentially become the global legal standard for the protection of human rights, democracy, and the rule of law, with high-level principles and measures in the field of artificial intelligence.

## 4. Work of the specialized sectors relevant to AI and the field of justice

### *4.1. Cyberjustice*

The approach of the Council of Europe to cyberjustice focuses on ensuring the protection of human rights principles while promoting the responsible and ethical deployment of digital technologies in the field of justice. While the Council of Europe does not have a specific treaty or convention dedicated to cyberjustice, it has been developing various strategies and tools for promoting the use of such technologies and for improving access to justice by making legal processes more efficient and accessible to individuals. The origin of this term is attributed to the work of the "Cyberjustice Laboratory" in Montreal,[68] which specifically dealt

[68] See the website of the laboratory: https://www.cyberjustice.ca/en/laboratoire/presentation/

with both the socio-legal and techno-legal aspects of the use of technologies by actors involved with justice.

The European Commission for the Efficiency of Justice (CEPEJ) is the Council of Europe's body for improving the quality and efficiency of European judicial systems. One of the focuses of the CEPEJ is the cyberjustice field, where it documents details related to the information and communication technologies (ICTs) applied to the justice system as well as the risks and challenges it poses for both justice professionals and policymakers. Recall that the CEPEJ's mandate is to enable member States of the Council of Europe to guarantee access to their courts within a reasonable timeframe, within the meaning of Article 6 of the European Convention on Human Rights. To this end, the use of new technologies was clearly specified by the Commission's terms of reference, adopted in 2002.

One of the early works of CEPEJ in this field was the publication of the "Guidelines on how to drive change towards Cyberjustice"[69] in 2016. The guidelines document the existing ICT systems deployed in the courts and assess their ability to improve the quality and effectiveness of judicial systems. Moreover, the guidelines put the general principles for development and implementation of European digital justice policies into perspective. They list four main categories to be promoted: access to justice; communication between the courts and professionals; court administration; and direct assistance for the work of the judge and the registrar. The document provides detailed information including benefits, long term developments, and essential aspects for the success and potential risks of ICT applications. The document also contains a compilation of good practices based on data provided by the member States.

The most notable work of the CEPEJ in this field is the "European Ethical Charter on the use of artificial intelligence in judicial systems and their environment,"[70] adopted in 2018. The Charter was accompanied by an in-depth study on the use of AI in judicial systems which sought to guide policy makers, legislators, and justice professionals on the rapid development and use of AI in this field. The Charter is the first European text that sets out ethical principles relating to the use of AI in the judicial system. It sets out five fundamental principles that should govern the development of tools using AI in the interest of the proper administration of justice: respect for fundamental rights; non-discrimination; quality and safety; transparency, impartiality and fairness; and user control, so that users are informed actors and retain control of their choices.

In 2020, the CEPEJ adopted a feasibility study on the possible establishment of a certification mechanism for AI tools and services[71] based on the Ethical Charter. The study addressed the need for the creation of an objective and neutral certification for AI applications in a high-risk sector, such as the judicial sphere, taking both ethical concerns and fundamental rights and freedoms into account. The study also identified the risks and opportunities relating to the CEPEJ certification.

---

[69] CEPEJ, *Guidelines on how to drive change towards Cyberjustice*, CEPEJ (2016)13, December 7, 2016, available at: https://rm.coe.int/16807482de

[70] CEPEJ, *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*, December 3-4, 2018, available at: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

[71] CEPEJ, *Possible introduction of a mechanism for certifying artificial intelligence tools and services in the sphere of justice and the judiciary: Feasibility Study*, CEPEJ(2020)15Rev, December 8, 2020, available at: https://rm.coe.int/feasability-study-en-cepej-2020-15/1680a0adf4

In parallel, the European Committee on Legal Co-operation (CDCJ), the standard-setting body in the field of public and private law, conducted a technical study[72] in 2016 for online dispute resolution (ODR) mechanisms, seeking to analyse compatibility with Article 6 (Right to a fair trial) and Article 13 (Right to an effective remedy) of the European Convention on Human Rights (ECHR). The study contained three main elements: a literature review on ODR, interviews with experts in the field of ODR, and responses to a Questionnaire by experts from member States. The study also covered specific ODR techniques including the use of AI and big data analysis currently used or that is planned for use by the courts. In September 2023, the CDCJ also began working on the consequences of AI use by administrations. The handbook "The Administration and You" is to be reviewed in light of the use of AI and non-AI algorithmic systems and further revision will be based on the comparative study "Artificial Intelligence and Administrative Law"[73] consisting of reports received from twenty-four member States.

In 2019 the CEPEJ established the Working Group on Cyberjustice and Artificial Intelligence (CEPEJ-GT-CYBERJUST) entrusted with the task of developing tools and offering a framework and guarantees to member States and legal professionals willing to create ICT and AI systems in judicial systems in an effort to improve the efficiency and quality of justice. The Working Group operates in co-ordination with other CEPEJ working groups and relevant bodies of the Council of Europe dealing with this field. The Working Group deals with topics such as quality criteria for videoconferencing, AI used in alternative methods of dispute resolution, enforcement of court decisions or court proceedings in a digital context, as well as developing training programmes for CEPEJ tools.

In 2019 the CEPEJ also adopted a "Toolkit for supporting the implementation of the Guidelines on how to drive change towards Cyberjustice"[74] which sought to guide national authorities in carrying out cyberjustice change management processes. The guidelines set out principles for the implementation of ICT in European judicial systems and provide concrete examples as well as the most common challenges encountered across Europe. They contained a checklist of the different steps and actions for the design, development, and implementation of ICT projects in judicial systems as well as a grid for the evaluation of such projects.

The CEPEJ adopted Guidelines on videoconferencing in judicial proceedings in 2021. [75] The guidelines aimed to ensure that use of videoconferencing in judicial proceedings was in line with the provisions of Article 6 of the ECHR and with the Convention for the Protection

---

[72] COUNCIL OF EUROPE EUROPEAN COMMITTEE ON LEGAL COOPERATION, *Technical study on online dispute resolution mechanisms*, CDCJ(2018)5, November 14-16, 2018, available at: https://rm.coe.int/cdcj-2018-5e-technical-study-odr/1680913249

[73] PROF. DR. J. WOLSWINKE FOR THE COUNCIL OF EUROPE, *Artificial Intelligence and Administrative Law*, December 2022, available at: https://www.coe.int/documents/22298481/35097084/CDCJ%282022%2931E+-+FINAL+6+%281%29.pdf/787157dd-f386-3e51-5b93-9aac50da1489?t=1671009161932

[74] CEPEJ, *Toolkit for supporting the implementation of the Guidelines on how to drive change towards Cyberjustice*, CEPEJ(2019)7, June 14, 2019, available at: https://rm.coe.int/cepej-toolkit-cyberjustice-en-cepej-2019-7/168094ef3e

[75] CEPEJ, *Guidelines on videoconferencing in judicial proceedings*, June 2021, available at: https://rm.coe.int/cepej-2021-4-guidelines-videoconference-en/1680a2c2f4

of Individuals with regard to the Automatic Processing of Personal Data[76] during remote hearings. The document contained procedural issues concerning all types of judicial proceedings as well as technical and organisational requirements covering different techniques, including the use of AI, Big Data analysis techniques, and other forms of automation. It also contained a checklist of basic requirements for the implementation of videoconferencing in judicial practice. In terms of AI, the guidelines require that the court's autonomy should not be restricted by the use of technology, by AI tools in particular. It also required that the use of AI tools such as sound or video e-filters be under the control of the court. As a complementary source to the Guidelines on videoconferencing, the CEPEJ also publishes a regularly updated document entitled "Selected National Good Practices."[77] This document compiles good practices from member States with the details of both the legal and ICT aspects of videoconferencing during judicial proceedings in reference to the Guidelines.

In 2021, the CM adopted the "Guidelines on online dispute resolution mechanisms in civil and administrative court proceedings"[78] prepared by the CDCJ, which aimed to ensure that member States adapted ODR mechanisms in line with the provisions of Article 6 and 13 of the ECHR without compromising the benefits concerning the costs and speed of dispute resolution in particular. The guidelines serve as a practical tool for fair procedures, including: access to justice, equality of arms, evidence, effective proceedings, delivery of decisions, the right to a reasoned decision, enforcement of the decision, the right to judicial review in cases involving purely automated decisions, transparency in the use of ODR, and requirements for hearings. The guidelines also contain special issues related to transparency, cybersecurity, and personal data protection. That same year, the CEPEJ also adopted Guidelines on electronic court filing (e-filing) and digitalisation of Courts.[79] This document provided a framework based on fundamental legal principles for the design and implementation of an e-filing system to enable the interaction and exchange of data and e-documents between courts and their users. The guidelines contain a governance strategy, organizational and technical aspects, as well as a checklist for deploying an e-filing system in the courts.

In 2022, the CEPEJ Artificial Intelligence Advisory Body (AIAB) was set up to support the CEPEJ in monitoring the use of AI in the justice sector, to provide expert guidance on ethical considerations and fundamental rights, on the operationalization of principles of the CEPEJ Charter, and to advise the CEPEJ working groups on new strategies for using AI in the justice system. The information obtained via the regular monitoring by the AIAB is

---

[76] COUNCIL OF EUROPE, *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, ETS No. 108, October 1, 1985, available at: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=108

[77] CEPEJ WORKİNG GROUP ON CYBERJUSTİCE AND ARTİFİCİAL INTELLİGENCE, *Selected national good practices on videoconferencing in judicial proceedings*, CEPEJ-GT-CYBERJUST(2021)11, November 30, 2021, available at: https://rm.coe.int/cepej-gt-cyberjust-2021-11-video-conferencing-good-practices-15-12-202/1680a4e1d9

[78] COUNCIL OF EUROPE EUROPEAN COMMITTEE ON LEGAL CO-OPERATION, *Guidelines on online dispute resolution mechanisms in civil and administrative court proceedings*, CM(2021)36add4-final, June 16, 2021, available at: https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680a2cf96

[79] CEPEJ, *Guidelines on electronic court filing (e-filing) and digitalisation of Courts*, CEPEJ(2021)15, December 8-9, 2021, available at: https://rm.coe.int/cepej-2021-15-en-e-filing-guidelines-digitalisation-courts/1680a4cf87#:~:text=Thus%2C%20the%20Guidelines%20deliver%20a,or%20receive%20notifications%20and%20summons

published in the Resource Centre on Cyberjustice and Artificial Intelligence.[80] It serves as an online repository for information on AI systems and relevant tools used in the justice sector. The information is based on contributions by the relevant authorities or information available in the public domain. The data is collected through the CEPEJ's European Cyberjustice Network (ECN) from Europe as well as other regions and is updated on a quarterly basis. The Centre serves as an important and reliable source of information for the digital transformation of the judiciary.

In February 2024, the CEPEJ published an information note, setting out a series of practical suggestions and advice on the basis of its ethical charter for using generative artificial intelligence systems in the judicial context.[81]

The Council of Europe's approach to cyberjustice is thus particularly focused on ensuring that the use of information technologies allow for control over procedural timeframes, rather than contributing to an extension of them due to the complexity or immaturity of IT systems. The intergovernmental and multidisciplinary approach adopted along with the creation of the CEPEJ, and the Commission's commitment to producing concrete and operational tools in this field, places the Council of Europe's expertise at the heart of the development of many public policies.

### *4.2. Data Protection*

The Council of Europe has played an important role in shaping global data protection standards and principles through the development of legal frameworks to ensure the protection of human rights, privacy, and the responsible use of technology in this field. One of the most important outcomes of the Council of Europe in the data protection field was the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108).[82] Adopted in 1981, the Convention was the first legally-binding international instrument in this field. It set out principles concerning the fair and lawful collection and automatic processing of data, storage for specified legitimate purposes, the quality and accuracy of data, the confidentiality of sensitive data, as well as the information of the data subjects and their right of access and rectification. It gathered fifty-five State Parties and around forty observers from all over the world.

An amending protocol[83] was adopted in 2018, with the aim of modernising Convention 108 to account for new challenges that have emerged as a result of new information and communication technologies and strengthening the effective implementation of the Convention. The modernized Convention, referred as Convention 108+, is expected to enter into force in 2024.

---

[80] CEPEJ Resource Centre on Cyberjustice and AI, available at: https://www.coe.int/en/web/cepej/resource-centre-on-cyberjustice-and-ai

[81] CEPEJ information note on the Use of Generative Artificial Intelligence (AI) by judicial professionals in a work-related context, accessible at: https://www.coe.int/en/web/cepej/-/information-note-on-the-use-of-generative-artificial-intelligence-ai-by-judicial-professionals-in-a-work-related-context

[82] *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, cit.

[83] COUNCIL OF EUROPE, *Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, CETS No. 223, October 10, 2018, available at: https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=223

The Convention established a Consultative Committee (T-PD), consisting of representatives of Parties to the Convention complemented by observers from member and non-member States and international organisations. The tasks of the T-PD include interpreting the provisions and improving the implementation of the Convention as well as drafting reports and guidelines on a broad range of topics related to data protection, including the issues emerging from the increasing intersection between data protection and AI. The TP-D issued Guidelines on Protecting Individuals from Data Processing in a Big Data World[84] in 2017 to provide a framework for applying appropriate policies and measures in the context of Big Data on the basis of the principles of Convention 108. The guidelines point out to the nature of Big Data which may pose risks and challenges in the implementation of traditional data processing principles, such as data minimisation, purpose limitation, fairness and transparency, along with free, specific, and informed consent.

In 2018, the T-PD issued a Practical Guide on the use of personal data in the police sector.[85] The document highlights the most important issues in the use of personal data by law enforcement authorities, especially in public prosecutor services, and provides practical information and concrete examples from everyday operation, underlining the importance of the lawful collection and use of personal data for law enforcement purposes. In 2019, the T-PD adopted the Guidelines on Artificial Intelligence and Data Protection[86] designed to ensure that AI applications do not undermine the right to data protection. The document stressed the importance of innovation in the AI field to giving careful consideration to the potential risks of the processing of personal data. The document also contained a set of baseline measures for legislators, policy makers and developers, as well as manufacturers and service providers. The T-PD published its Guidelines on facial recognition in 2021.[87] This document provided a set of benchmarks that governments, facial recognition developers, manufacturers, service providers and entities using such technologies should follow to ensure they do not infringe on human rights, including the right to protection of personal data. The document covers the uses of facial recognition technologies, including live facial recognition technologies in the private and public sectors.

In 2021, the CM issued a Recommendation on the protection of individuals with regard to automatic processing of personal data, including machine learning, in the context of profiling.[88] The Recommendation acknowledged the recent technological developments in

---

[84] COUNCIL OF EUROPE CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING OF PERSONAL DATA, *Guidelines on Protecting Individuals from Data Processing in a Big Data World*, T-PD(2017)01, January 23, 2017, available at: https://rm.coe.int/16806ebe7a

[85] COUNCIL OF EUROPE CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING OF PERSONAL DATA, *Practical Guide on the use of personal data in the police sector*, T-PD(2018)01, February 15, 2018, available at: https://rm.coe.int/t-pd-201-01-practical-guide-on-the-use-of-personal-data-in-the-police-/16807927d5

[86] COUNCIL OF EUROPE CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING OF PERSONAL DATA, *Guidelines on Artificial Intelligence and Data Protection*, T-PD(2019)01, February 25, 2019, available at: https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8

[87] COUNCIL OF EUROPE CONSULTATIVE COMMITTEE OF THE CONVENTION FOR THE PROTECTION OF INDIVIDUALS WITH REGARD TO AUTOMATIC PROCESSING OF PERSONAL DATA, *Guidelines on facial recognition*, T-PD(2020)03rev4, January 28, 2021, available at: https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3

[88] *Recommendation on the protection of individuals with regard to automatic processing of personal data, including machine learning, in the context of profiling*, CM/Rec(2021)8, November 3, 2021, available at:

the context of profiling and affirmed the need for additional safeguards to protect the personal data and private life of individuals. The text also included reference to the use of automated decision-making systems based on AI and additional risks for data subjects due to possible errors and biases.

## 5. Conclusion

The Council of Europe has long been a frontrunner in Europe for the binding and multilateral regulation of digital issues. The Convention 108 on data protection, the Budapest Convention on Cybercrime, and the recent Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law demonstrate the organization's work to address a broad range of digital issues and its continuous concern for the protection of human rights, democracy, and the rule of law. In the last decade, the work of the Council of Europe on the production of soft law on digital issues has intensified, in particular through Recommendations, in order to offer member States a more flexible framework. Following the recent and more global shift towards a risk-based approach in the regulation of AI, in 2024 the Council of Europe adopted a similar approach in its legally binding instrument: the Framework Convention on AI.

The current instruments of AI regulation in the international context place an emphasis on risk management by seeking to prevent damage *ex ante* instead of creating new rights and exceptions in order to establish violations *ex post*. Considering the fact that it is not always possible or desirable to transform rights into technical standards, this risk-based approach has been criticised by some scholars for failing to provide effective models for assessing the impact on human rights[89] as well as for failing to counter adverse impacts on society, beyond individual and collective harm.[90] Even though the adoption of the AI Act and the Framework Convention seem to have settled the debate, international efforts for the regulation of AI are far from over.[91] It is safe to assume that it will be a combination of *ex ante* and *ex post* measures that will need to be examined in order to demonstrate the effectiveness of these regulatory models. The use of AI for judicial administration, considered high-risk in Annex III of the AI Act, will thus likely seek to guarantee its compliance before undergoing any mass dissemination and before assuming responsibility in the event of discrimination, a breach of privacy, or unlawful obstacles to judicial access.

https://search.coe.int/cm/pages/result_details.aspx?ObjectId=0900001680a46147

[89] A. MANTELERO, *Beyond data: Human rights, ethical and social impact assessment in AI,* Springer Nature, 2022, 200.

[90] N. A. SMUHA, *Beyond the Individual: Governing AI's Societal Harm*, in *Internet Policy Review*, 2021, 10, 3.

[91] F. G'SELL, *Regulating under Uncertainty: Governance Options for Generative AI, SSRN*, 2024.

# 3.2. Regulation by the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act): an analysis

*Federica Casarosa – Scuola Superiore Sant'Anna*[92]

*Summary: 1. An overview – 2. Provisions – 3. The Annexes – 4. Open issues – 4.1. Independence – 4.2. Selection of data – 4.3. Fair trial*

*Abstract: The 'Artificial Intelligence Act' answers the call for legislative action to ensure a well-functioning internal market for artificial intelligence systems where both the benefits and risks of AI are adequately addressed at the Union level. The final structure has been subject to several modifications during the legislative process. This analysis focuses mainly on the relevant aspects related to the use of AI in the judicial sector, highlighting a set of open issues that may emerge in the application of it to this sector.*

## 1. An overview

The "Artificial Intelligence Act"' (hereinafter the "AI Act") answers the call for legislative action to ensure a well-functioning internal market for artificial intelligence systems ("AI systems") where both the benefits and risks of AI are adequately addressed at the Union level. The AI Act is part of the Commission's agenda to make Europe fit for the digital age. Since 2018 the Commission has put forward a European strategic plan for AI,[93] which resulted in some relevant initiatives, such as the Ethics Guidelines for Trustworthy AI[94] and Policy Recommendations[95] published by the High-level Expert Group on Artificial Intelligence, followed by the White Paper on AI.[96]

The AI Act concluded the legislative procedure for approval in December 2023, but the final text was only released in June 2024. The initial text proposed by the European Commission on April 21, 2021[97] was subject to amendments by the European Parliament on

92 Research was supported by *SoBigData.it project (Prot. IR0000013 – Call n. 3264 of 12/28/2021) initiatives aimed at training new users and communities in the usage of the research infrastructure (SoBigData.eu).*

93 EUROPEAN COMMISSION, *Artificial Intelligence for Europe,* COM(2018) 327 final, April 25, 2018, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237

94 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Ethics Guidelines for Trustworthy AI*, April 8, 2019, available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html.

95 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *Policy and investment recommendations for trustworthy AI*, June 26, 2019, available at: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence

96 EUROPEAN COMMISSION, *White Paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, February 19, 2020, available at: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

97 EUROPEAN COMMISSION, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts*, (COM(2021) 206 final), April 21, 2021, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

June 14, 2023,[98] and then by intense debates in the Trilogue discussion in November and December 2023. The current analysis considers the adopted legislation, underlining the changes that occurred in the version by the European Parliament, and focuses mainly on the relevant aspects related to the use of AI in the judicial sector.

The AI Act is based on the need to establish the EU's technological leadership in AI to ensure that European citizens benefit from a technological development that is compliant with EU values and fundamental rights and principles. The objectives of the proposed regulatory framework include:

a) ensuring that AI systems placed on the Union market and used by European citizens are trustworthy, safe, and compliant with existing law on health, safety, fundamental rights, democracy and the rule of law, and environmental protection;

b) guaranteeing legal certainty in order to facilitate investment and innovation in AI;

c) enhancing governance and the effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;

d) facilitating the development of a single market for lawful, safe, and trustworthy AI applications, while at the same time preventing market fragmentation.

In order to achieve these objectives, the AI Act adopted a risk-based approach distinguishing the rules applicable to AI development and market placement and the use of AI systems in the EU market depending on the type of risk.

## 2. Provisions

Understanding the general framework of action is essential in order to assess the specific implications for the administration of Justice in all its nuances and applications.

In Chapter I (Articles 1-4), the subject matter and scope of application of the new rules are defined, covering AI systems market placement, their being put into service, and use. Note that the AI Act applies to providers, deployers, importers and distributors, as well as persons affected by AI systems. The revised definition of actors subject to legislation aims at extending application of the rules not only to initial developers of AI but also to subsequent stakeholders in the value chain who will be able to use AI systems in specific sectors or applications.

*Example*: The national Ministry of Justice may decide to use facial recognition systems developed by company A to authenticate the entry to prison of detained individuals. The same system may be adopted by a multinational company providing transport services in order to verify the identity of users.[99] The types of risks and issues associated with use of this AI system are different depending on the deployer.

The AI Act lists the principles that should be applied to the development of AI systems (or in general-purpose AI models), including human agency and oversight, technical

[98] EUROPEAN PARLIAMENT, *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*, (P9_TA(2023)0236), June 14, 2023, available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf

[99] L. EDWARDS, *Regulating AI in Europe: Four Problems and Four Solutions*, Ada Lovelace Institute, 2022, available at: www.adalovelaceinstitute.org/report/regulatingai-in-europe

robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, and social and environmental well-being.

In Chapter II (Article 5), following a risk-based approach, the AI Act differentiates between uses of AI that bear (i) an unacceptable risk, (ii) a high risk, and (iii) a low or minimal risk. It prohibits altogether the use of AI systems whose risks are considered unacceptable, as contraventions of EU values or a violation of fundamental rights, such as:

- subliminal techniques that aim or result in materially distorting the behaviour of a person or a group by appreciably impairing their ability to make informed decisions;
- manipulative or exploitative practices against vulnerable groups in order to distort their behaviour in a manner that is likely to lead to uninformed decisions and/or to cause psychological or physical harm;
- social scoring systems for general purposes;
- risk assessment for predicting the occurrence or reoccurrence of an actual or potential criminal or administrative offence;
- facial recognition databases from Internet and CCTV footage;
- emotion recognition in areas of law enforcement, border management, in workplace and educational institutions;
- "real-time" remote biometric identification systems in publicly accessible spaces as well as "post" remote biometric identification systems, except if subject to pre-judicial authorisation.

An extensive number of provisions in Chapter III (Articles 6-51) deal specifically with high-risk AI systems: a high-risk classification is based on the intended purpose of the AI system and its modality of use. Section 1 sets classification rules and identifies two main categories of high-risk AI systems: i) AI systems intended to be used as safety components of products that are subject to third-party ex-ante conformity assessment; ii) other stand-alone AI systems with mainly fundamental rights implications that are explicitly listed in Annex III, which also contains a predefined list of AI systems whose risks have already materialised or are likely to materialise in the near future. Section 2 sets out the legal requirements for high-risk AI systems. The requirements include a risk management system (Article 9), data and data governance (Article 10), documentation and record-keeping (Article 11 and 12), transparency and the provision of information to users (Article 13), human oversight (Article 14), and robustness, accuracy and cybersecurity (Article 15). The content of such requirements may be substantially determined by consulting the Guidelines (see Article 96), the harmonised standards and common specifications (respectively Article 40 and 41), or otherwise be developed in accordance with the state-of-the-art of scientific knowledge at the discretion of the AI system provider. Section 3 places clear obligations on providers and deployers of high-risk AI systems, as well as obligations on other parties involved, such as importers, distributors, and authorised representatives. Such obligations are primarily articulated in the adoption of a quality management system (Article 17), a duty of information across the value chain (Article 25), a cooperation obligation with national and European authorities (Article 21) and a fundamental rights impact assessment (Article 27). Sections 4 and 5 are specifically dedicated to the conformity assessment procedure: while the former sets the framework for involving notified bodies in the conformity assessment procedure, the latter explains the procedure for carrying out the conformity assessment for each type of high-risk AI system in detail. It is important to note that AI systems that comply with the relevant harmonised standards (Article 40) or other common specifications (Article

41) where such standards do not exist, are insufficient, or specific safety or fundamental rights concerns need to be addressed, are presumed to be in conformity (Article 42) with the requirements set out in Section 2. The conformity assessment procedure (Article 43) may be carried out either based on internal controls (Annex VI) or based on an assessment of the quality management system and the technical documentation with the involvement of a notified body (Annex VII). AI systems undergoing substantial modifications shall be subject to new ex-ante conformity assessment.

With regard to those AI systems that (i) interact with humans, (ii) are used to detect emotions or determine association with (social) categories based on biometric data, or (iii) generate or manipulate content ("deep fakes"), the AI Act imposes in Chapter IV (Article 50) transparency obligations to allow users to make informed choices starting from their first interaction or exposure to the content/system.

Chapter V (Articles 51-55) deals with the newly envisaged General-purpose AI models with systemic risk. The providers of General-purpose AI models are required to provide at least the information listed in Annex IX (used to inform the AI office and the national competent authorities) and documentation following the elements set out in Annex XII (used to inform providers of AI systems that will exploit such models).

Chapter VI envisages the possibility of national competent authorities setting up regulatory sandboxes in order to establish a controlled environment to test innovative technologies, while Chapter VII (Articles 64-70) proceeds to establish an Artificial Intelligence Office (Article 64), composed of representatives from Member States, the Commission, and a representative of the EDPS.[100] The AI Office is tasked with facilitating a smooth, effective and harmonized implementation of the AI Act (Article 66). At the national level, Member States are tasked with designating one or more national supervisory authorities. An EU-wide database for high-risk AI systems with fundamental rights implications, operated by the Commission, will be established for the registration of AI systems before their placement on the market, according to the provisions of Chapter VIII (Article 71). Finally, ex-post obligations concerning monitoring and reporting of AI-related incidents and malfunctioning are established. In fact, national supervisory authorities are entrusted with ex-post enforcement and compliance monitoring after the AI system is placed on the market or otherwise put into service according to Chapter IX (Article 72-94). The AI Act also includes remedies in case of violation of the legislation, introducing a right to lodge a complaint before the national supervisory authority (Article 85) and a right to explanation of individual decision-making (Article 86).

Chapter X (Articles 95-96) provides a framework for the creation of codes of conduct to encourage providers of non-high-risk AI systems to voluntarily comply with the mandatory requirements for high-risk AI systems and other voluntary commitments, such as environmental sustainability, accessibility for vulnerable people, and stakeholder participation.

To conclude, Chapter XI and XII (Articles 97-101) impose rules for the exercise of delegation and implementing powers by the Commission to ensure uniform application of the Regulation, and the penalties – and other enforcement rules – to be adopted at the national level in case of a lack of compliance with the regulation. Finally, they also lay down

[100] European Data Protection Supervisor, see role and functions at https://edps.europa.eu/_en.

provisions for the differentiated transitional period prior to the initial date of applicability of the Regulation.

## 3. The Annexes

The text of the AI Act is accompanied by a total of eight Annexes. Especially worthy of attention is Annex I, which includes a list of Union Harmonisation Legislation, and Annex III, devoted to the High-Risk AI Systems Referred to in Article 6(2).

This Annex pinpoints the examples and criteria to be used for identifying high-risk AI systems, in the case of law enforcement in particular, or AI systems intended to be used, by, or on behalf of law enforcement authorities, or by Union agencies, offices, or bodies in support of law enforcement authorities:

(a) to assess the risk of a natural person becoming the victim of criminal offences;

(b) as polygraphs and similar tools;

(c) to evaluate the reliability of evidence during the course of an investigation or prosecution of criminal offences;

(d) for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups;

(e) for the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 during the detection, investigation, or prosecution of criminal offences.

Besides, in the case of administration of justice and democratic processes, AI systems are intended to be used by a judicial authority or administrative body or on their behalf to assist a judicial authority or administrative body in researching and interpreting facts and the law and in applying the law to a concrete set of facts or used in a similar way in alternative dispute resolution. It is important to highlight that not all AI systems that can be adopted fall into the category of high risk, as the annex specifically focuses on the research and interpretation of facts and law. Therefore, AI systems used by courts for the allocation of cases to different judges (based on availability, expertise, or other factors) or for scheduling hearings within a judicial procedure may be excluded from this qualification. However, a thorough analysis of the expected activities carried out by the AI system is crucial in order to verify its classification.

## 4. Open issues

### *4.1. Independence*

According to point 8 of Annex III, any AI systems intended to be used by a judicial authority or administrative body in researching and interpreting facts and the law and in applying the law to a concrete set of facts is subject to the obligations set up in Chapter II of the AI Act.

These obligations apply to the provider that develops the AI system, which can be any "natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge" (Article 3(2)). The provider, however, is usually a private-sector company. Additionally, the

application to the specific sector is adopted by deployers that can adapt the general AI system to the needs of the subsequent user.[101]

It is important that, in order to safeguard the judicial independence of AI system users, i.e. judges, neither the provider nor the deployer are identified with the executive or legislative powers in order to guarantee the existence of a separation of powers. For example, if a national government is tasked with designing algorithms used in courts, this allows for interference with decision-making and the external aspect of judicial independence would be impeded. The fact that providers and deployers are entities free from the influence of public authorities is crucial. However, it is possible such private sector entities are also situated outside the EU, leading to risks of surveillance and control from foreign states.

### *4.2. Selection of data*

An AI system, in order to achieve a trustworthy level of accuracy, needs to be trained, tested and validated on an ongoing basis. It is possible, then, to distinguish three types that are relevant for the development of an AI system: training data, validating data, and testing data.

- Training data: Data that are selected and annotated in order to train an algorithm, the annotation (or labelling) will allow the algorithm to recognize the parts or features that are similar and create a pattern or those that are relevant to identify new patterns.
- Validating data: data used to validate the ability of the AI system to react to randomness, showing not only the strengths but also the potential weaknesses in recognizing such patterns.
- Testing data: new data that are used to assess the accuracy of AI systems on an ongoing basis, as well as the ability to react to untrained data and circumstances.

For example, an AI system used in the judicial sector for predictive justice will use case-law issued on a specific sector as training data. Validation data will include new cases that were decided but not used in the training dataset. Whereas unsolved cases can be used as testing data.

It is crucial that all the data used are "appropriate for the context of use as well as the intended purpose of the AI system" (Article 10(2)). If data are biased, old (stale data), few in number, not relevant for the application, or not sufficiently representative, they may undermine the accuracy of the AI system or embed bias in the AI system's application. Accordingly, in order to trust the AI system, the quantity and suitability of the data sets should be assessed in advance.

### *4.3. Fair trial*

According to Article 13(1), the AI system should comply with the principle of transparency. Specifically, "the user shall be enabled to understand and use the AI system appropriately by generally knowing how the AI system works and what data it processes, allowing the user to explain the decisions taken by the AI system to the affected person." In

[101] Note that the Proposed regulation specifies two types of AI systems: a "foundation model" as "an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks" (Article 3(1c)) and a "general purpose AI system" as "an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed" (Article 3(1d)).

the case of an AI system used in the judicial sector, it is extremely important that the user be informed about the degree to which the AI system can provide an explanation for the decisions it makes (Article 13 (3)(iv)). The lack of such information may hamper fair trial guarantees in the case of using AI systems in the judicial system. For instance, the judge (as a user) should receive information about the logic upon which the algorithm provides its decisions, i.e., an explanation of the result.

# Section 4

# The impact on fundamental rights: rules and principles concerning the use of AI tools

## 4.1. Fundamental principles in the use of "artificial justice" tools

*Andrea Maria Garofalo – Università degli Studi di Trento*

*Summary: 1. Introduction – 2. Protection of fundamental rights – 3. The principle of non-discrimination – 4. The principle of transparency – 5. The principle of due process – 6. The principle of certainty – 7. The principle of due diligence – 8. The principle of independence and the principle of impartiality*

*Abstract: Artificial Intelligence systems that can be used in the field of justice are very diverse and raise heterogeneous problems. This paper looks at some of these systems, namely advanced precedent search systems, chatboxes, and tools that automatically indicate the outcome of a dispute or decide on a specific point in a dispute, with the aim of assessing how they can be used in compliance with fundamental rights and fundamental principles of civil procedure (the principle of non-discrimination; the principle of transparency; the principle of due process; the principle of certainty; the principle of diligence; the principle of independence and the principle of impartiality).*

### 1. Introduction

The use of AI in the field of justice, although already experimented with in some jurisdictions, represents a realm that is still developing. In fact, the definition of the area of research is unclear. When referring to AI in the justice system, one may have in mind very heterogeneous software and many different applications and uses.[102] The variety is remarkable and ranges from advanced jurisprudential search engines to online platforms for automated dispute resolution outside the court system, to assistance in drafting contracts and court documents, to predictive analysis of the chances of winning a case in court and finally to the drafting of a judgment itself (a "robot judge"). And these are but a few examples.

Obviously, each of these systems raises very different and heterogeneous problems. It is therefore worth limiting these pages, written from the perspective of those who administer justice, to the use of AI for finding solutions to legal cases. This can be done by means of advanced case-law search systems, which automatically select the most relevant decisions, or by means of chat boxes, which directly indicate the solution (or possible solutions) to a given legal problem, or even by means of tools that automatically indicate the outcome of a dispute (i.e., by writing down the text of a decision) or, once more automatically, decide on a specific point in a dispute (i.e. the amount of compensation or maintenance allowance). We shall comprehensively call these systems, which at present are all based on processing large amounts of data mostly represented by court decisions (or legal texts or doctrinal opinions), "artificial intelligence applied to the judicial function" or "artificial justice" systems. Since

[102] EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Strasbourg, December 2018, 17.

these systems share some common features, we will discuss them in a unitary manner, pointing out certain peculiarities of one or the other instrument of "artificial justice" where necessary. Finally, the perspective applied here is that of Civil Law.[103]

## 2. Protection of fundamental rights

Artificial justice systems can put certain fundamental rights at serious risk.[104] The recent AI Act provides that several artificial justice systems fall into the high-risk system category: in fact, Annex III states that these are also the "AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution." Consequently, these instruments are also regulated by Art 8 of the AI Act.

The risks regarding fundamental rights posed by artificial justice systems are generally taken into account by means of the principles discussed below.[105] Nevertheless, such systems may directly pose critical issues, especially from a data protection perspective.[106] They feed on large masses of data and, in turn, create new data. Is the Big Data they use personal data or, because they have been anonymized for instance, are they no longer personal? If they remain personal, what precautions and measures must be taken when processing such data? Is the new data that artificial justice systems create, such as a draft decision for a certain case, personal data? If so, how can it be legitimately processed?

Legal systems have adopted heterogeneous measures for the anonymization of judicial precedents, in an attempt to balance the principle of the publicity of decisions and those underlying the protection of personal data. In the case of big data to be supplied to an AI system, the need for anonymization is even more pronounced and should be fully respected, above all, by virtue of the principle of risk prevention.

As for the data that AI systems create, if it can be linked to a specific case and or a specific person, it should be guarded with particular care, since it is sensitive data. What is more, except in cases (not conceivable at present) of actual robotic decisions, such data should be subject to the regime of all those documents that help judges form their opinions or that record an interlocutory opinion, such as research, drafts, and discussions in chambers: a regime that, unlike for judgments, usually excludes publicity. This is why such data should not be covered by the data subject's right of access; moreover, it should be deleted as soon as there is no longer any need for justice that would require its preservation.

## 3. The principle of non-discrimination

The use of artificial justice systems, like all AI, poses risks of discrimination.[107]

---

[103] On the fundamental principles of civil procedure, see for instance the ALI-UNIDROIT *Principles of transnational*
*civil procedure* drafted in 2006.

[104] CEPEJ, *European ethical Charter*, cit., 8.

[105] From a general point of view, see also the recent Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law of the Council of Europe.

[106] CEPEJ, *European ethical Charter*, cit., 25.

[107] CEPEJ, *European ethical Charter*, cit., 9.

To explain this, we must remember that AI systems do not reason like human beings. In fact, they merely aggregate large masses of data and find frequent correspondences or patterns; on the basis of this they are able to give a final result or output. Most of the time, it is not possible for those who use AI, and even for those who program it, to know how to arrive at a final result. The risk, therefore, is that the AI, by providing output, has exploited patterns that reproduce discrimination prohibited by law, for example based on race, ethnic origin, socio-economic background, religion, or political opinions.[108] For instance, it is conceivable that an AI, when questioned by a judge on the amount of divorce allowance a certain person is entitled to, would note that on most occasions allowances awarded to persons of a certain ethnic origin are lower and would directly link to this fact, rather than to the fact that in most cases the couples involved were low income couples. The indication of a low amount for the allowance to be awarded to the person involved in the case would be discriminatory.

To the contrary, however, it should be noted that the use of AI systems can also reduce the risk of human discrimination, such as that based on an individual judge's bias. Obviously, it cannot avoid those discriminations that have sedimented in the data used (which are often represented by case-law precedent), since artificial justice systems do not allow for evolution with respect to existing data, but rather base their outputs on these and, indeed, reiterate their solutions.[109]

The problem of discrimination must primarily be solved by the person who designs the artificial justice system (ethical-by-design): for instance, a programmer should ensure that the system does not base its results on discrimination.[110] The judge should, in any case and especially where there is no such assurance, at least verify on a sample basis, by means of more traditional systems of consulting legal texts, precedents, and doctrinal opinions, the absence of discrimination in the "reasoning" of the machine. This is especially true if the artificial justice system used is very incisive; that is, it is less true for systems aimed at selecting precedents and much more so for systems that prepare the draft text of a judicial decision.

It follows from this that artificial justice systems can be a valuable aid to judges (even helping them overcome biases), but they can neither replace judges themselves nor entirely substitute the traditional tools of their work.

## 4. The principle of transparency

Transparency, as applied to AI, first of all requires that the process by which the algorithm arrives at a certain result is made evident. Unfortunately, it is well known that such a high

[108] EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, *Bias in Algorithms. Artificial Intelligence and Discrimination*, Vienna, 2022.

[109] M. TEGMARK, *Life 3.0: Being Human in the Age of Artificial Intelligence online*, New York, 2017, chapter 3, § Law.

[9] To tell the truth, the recent AI Act fails to provide for a direct obligation in this direction, although several recitals stress the importance of non-discrimination. However, there are some tools that aim to reduce the risks of discrimination (see for example the fundamental rights impact assessment set out in Article 27 of the AI Act; in literature, see F. LÜTZ, *The AI Act, gender equality and non-discrimination: what role for the AI office?*, in *ERA Forum*, 2024, 6).

level of transparency is not easy to achieve due to technological limitations.[111] This further applies to artificial justice systems, where the problem is even more acute.[112] In fact, one of the cornerstones of due process is represented, in all jurisdictions, by the contestability of a judge's decision: contestability based on making known the arguments the judge used to decide a case, i.e. the duty to furnish motivations.

It is commonly known that judges often decide on the basis of their pre-understanding,[113] often without even intending to do so, and only seek arguments to ground the decision they have intuitively reached downstream. However, this does not exclude the existence of a rational process, based on the arguments typical of legal discourse, which may also lead judges to revise the decision they arrived at initially. Such a rational process, however, is not carried out by the machine, which merely – as already noted – identifies patterns within large masses of data and, consequently, draws up a list of precedents that are apparently the most relevant to the outcome of a dispute, though it may specify the arguments used to resolve similar disputes, or draft a decision or part of a decision.[114] The machine does not reason in the same way a human being does and, even in its way of reasoning, AI does not usually allow us to understand how it arrived at a certain result. In fact, algorithms are not programmed to reveal the patterns that have been found, so much so that not even computer experts are able to understand the pattern; even when they do, for trade secret reasons they would likely prevent the public from knowing the "reasoning" carried out by the machine.

For these reasons, the use of artificial justice systems may endanger the transparency of the decision-making process.[115] In the face of this, it must be assumed that judges cannot use artificial justice systems that reduce their motivational burden, even though this might lead to saving time and resources, along with faster and more effective decisions.

However, automated search systems of precedents do not pose particular problems. Consider for example chat boxes that usually render, as output, text where legal arguments appear: judges may use this text as an aid, though the decision-making process must remain their own. They will consequently have to examine the correctness of the arguments and integrate them, by means of traditional search systems and by recourse to traditional technical-legal reasoning.

To the contrary, systems that produce actual decisions or pieces of decision are not compatible with the duty to motivate. They cannot in any way be used to arrive at decisions; their results must be incorporated into traditional legal-technical reasoning.

In the future, for purely quantitative aspects of decisions – such as, for example, compensation for certain factual elements – the law may allow the court to entrust its decision in part to an algorithm, limiting its task to verifying that the automated decision was reasonable and, in any case, not discriminatory. In such cases, in fact, a machine's calculation capacity would likely seem even more reliable than a judge's, who would inevitably have great difficulty knowing all the precedents expressed in a certain matter (especially in the case of serial litigation).

---

[111] A duty of transparency, with reference to high-risk AI systems, is now set out – with several inevitable limitations – in Article 13 of the AI Act.

[112] CONSEIL DES BARREAUX EUROPÉENS, *Position paper on the proposal for a regulation laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act)*, October 8, 2021, 9.

[113] J. ESSER, *Vorverständnis und Methodenwahl in der Rechtsfindung*, Frankfurt am Main, 1970.

[114] CEPEJ, *European ethical Charter*, cit., p. 35. In general terms see E.J. LARSON, *The Myth of Artificial Intelligence, Why Computers Can't Think the Way We Do*, Cambridge, 2021.

[115] E. BATTELLI, *Giustizia predittiva, decisione robotica e ruolo del giudice*, in *Giustizia civile*, 2020, 293.

The law permitting, the judge should indicate the use of an AI tool, specify which one, and how it was used. The tool itself should at least specify which precedents were used as the basis for arriving at a given output. The decision, insofar as it incorporates the output, should always be appealable.

## 5. The principle of due process

The principle of due process is strongly affected by the use of artificial justice systems.[116]

It may be endangered in various ways: persons may be discouraged to seek protection for their rights, knowing that an algorithm used in court would lead to a negative outcome of the dispute; an error in the decision of a dispute, repeated in some precedent, may be reproduced countless times. At the same time, artificial justice systems can also facilitate the full realization of the principle of due process by reducing the costs of justice, avoiding unnecessary litigation, and speeding up trials.

It is therefore necessary to find a solution that would allow the benefits of AI to be enjoyed while limiting the associated risks. Once more, the best mechanism seems to be a "weak use" of artificial justice systems, so that they may aid judges, but not replace them. In particular, no judge should be limited to solely considering the response of an AI, without evaluating the arguments of the parties. Similarly, citizens must be able to include legal arguments in their pleadings in order to indicate what they believe to be the correct solution to a dispute, knowing these arguments will be taken into account by a judge.

Looking to the future, one may wonder whether, in disputes of modest or derisory value, the legislature may allow a more incisive use of artificial justice systems, based on the principle *de minimis non curat praetor*. However, the right to appeal the decision before a human judge should always be permitted, laying out real legal arguments.[117] These litigations in and of themselves pose minor risks, as recognized by all legal systems, which, for instance, assign their cognition to inferior or non-professional judges or which, in order to avoid excessive use of resources, otherwise compress the procedural guarantees of citizens.

Moreover, many jurisdictions are familiar with provisional measures that can be obtained *inaudita altera parte*, sometimes jurisdictional and sometimes non-jurisdictional, aimed at quickly obtaining an order, which the opposing party must then challenge. In these proceedings, greater use of artificial justice systems might be permissible in the future, provided that such use is disclosed in the motivation of the order, that there are no excessive costs or time limits to opposing the order, and that it does not concern litigation against consumers or affect constitutionally protected rights, i.e. the right to housing, the right to alimony, and the right to maintenance.[118]

---

[116] CONSEIL DES BARREAUX EUROPEENS, *Position paper*, cit., 3.

[117] A particular model (proposal for a decision by the machine; right of the parties to accept it, or to ask a human
judge to intervene) is imagined for small claims by M. R. MAUGERI, *I robot e la possibile "prognosi" delle decisioni giudiziali*, in A. CARLEO (ed.), *Decisione robotica,* Bologna, 2019, 163.

[118] As S. PENASA, *Giustizia e variabile algoritmica. Una prima valutazione di sostenibilità tecnica e costituzionale*, in L. ANTONIOLLI, M. CARDILLO, F. CORTESE, L. DE CARBONNIÈRES,F. MYNARD, C. PICCIOCCHI (eds.), *Numérique & environnement. Université d'été franco-italienne. Actes du colloque*. 6-8 Julliet 2022. Université de Limoges, Trento, 2024, 140, observes that there may be situations where, provided sufficient technical and legal safeguards are in place, replacing the judge could align with constitutional principles. At the same time,

Should a more penetrating use of AI be permitted, citizens should know which artificial justice systems are used by judges in order to arrive at actual decisions (or even draft decisions), so that they can assess the risks of litigation at an early stage and consciously decide, in the case of plaintiffs, whether to have recourse to a special procedure based on the extensive use of AI or not.

## 6. The principle of certainty

Each legal order lives in a compromise between the principle of certainty and the need to ensure a certain flexibility. This applies to all systems, albeit somewhat differently, due to the different balance between legal formants.

In Civil Law systems, the system "improves" – i.e. finds a more efficient or fairer solution – not only through new legislative measures, but also through new judicial precedents. What is more, the evolution of the legal system can also serve to adapt it to changing social reality, without the need for the legislature to intervene. That sometimes happens "quietly," in the sense that case-law slides towards adherence to different principles, and sometimes "noisily," by means of strong breaks with the past. This is ensured not only by the fact that judicial decisions are grounded in legal arguments, i.e. on logical argumentation based on a set of values given as an assumption in Civil Law systems, but also by the fact that judges are decision-makers who live in a society and are familiar with its value structure, including all the contrasting values that distinguish different ways of seeing things.

In contrast, algorithms do not reason in a human way and do not directly know values; they filter them through judicial decisions, which transpose them. A machine will not therefore be able to ensure any evolution of the legal system; an algorithm will tend to be conservative and tend to crystalize and repeat decisions already rendered forever.[119]

At the same time, it cannot be neglected that legal systems also highly value legal certainty. It constitutes the essential safeguard in terms of fairness, allowing similar situations to be treated identically, and also efficiency, allowing economic operators to know how their conduct and choices are likely to be judged. Having a digital tool that ensures high levels of legal certainty at one's disposal therefore seems, from this perspective, a much more important safeguard for legal certainty than what can be provided by a judge, who is inevitably subject to limited knowledge of precedent, as well as to their own tastes and prejudices. From such a perspective, a "weak use" of AI should be encouraged, as it enables judges to increase their knowledge, overcome any bias in their search for legal arguments and solutions, and thus permits an increase in the level of certainty.[120]

On the other hand, a stronger use of artificial justice should also be avoided from this perspective. The legislator could permit it to be used for serial litigation of modest value and with the guarantee that users could appeal digital decisions. The possibility of appealing also represents, from this view, a means of obtaining human decisions in subject matter where a stronger use of artificial justice is exceptionally approved. The fact that we are referring to

activities that are formally defined as assistance but effectively influence the judge's discretionary powers could undermine the balance of constitutional principles and safeguards governing the judicial function.

[119] M. LIBERTINI, M. R. MAUGERI, E. VINCENTI, *Giustizia predittiva e giurisdizione civile. Primi appunti*, in A. PAJNO, F. DONATI, A. PERRUCCI (eds.), *Intelligenza artificiale e diritto: una rivoluzione?, II, Amministrazione, responsabilità, giurisdizione*, Bologna, 2022, 523.

[120] R. BICHI, *Intelligenza Artificiale tra "calcolabilità" del diritto e tutela dei diritti*, in *Giurisprudenza italiana*, 2019, 1178.

serial litigation makes it possible to increase the number of recent and human precedents on which a digital decision can be based, thus making it more reliable.

## 7. The principle of due diligence

All jurisdictional systems require that the outcomes of litigation be "correct," in the sense that they be rationally justified on the basis of arguments that each system recognizes as worthy and that are most pertinent in the specific case.

An AI based on "wrong" decisions would, inevitably, lead to a repetition of such errors, as explained above. It must be said, however, that AI itself could, for instance due to a defect in its functioning, produce wrong outcomes. It is enough to consider a situation in which an AI bases the outcome of a dispute on an argument, used in numerous precedents, but that is no longer relevant due to recent legislative reform.

If AI could really decide disputes, the question of liability[121] for such an error would then have to be posed.[122] For instance: is the State or the programmer liable for it? Or perhaps the State is liable, even though it could demand compensation from the programmer? However, this scenario still seems far off today. It is preferable then to ask other questions such as whether it is possible to avoid errors produced by AI systems as much as possible?

First of all, judges who use artificial justice tools should take the precaution of checking their reliability. The legislature could also provide that such tools, when used officially, must be checked by special certifying bodies.[123] In any case, judges must monitor the results produced by AIs through traditional study and research systems. Even here, this duty of control is less stringent where tools are used to identify suitable precedents and should be much more penetrating where tools are allowed to draft actual decisions. In the context of this control, judges must remember that the results of AI systems are more reliable if they have a large body of precedents, as is the case in serial litigation; and much less reliable if there are few precedents, as is the case with sporadic litigation or litigation decided in a different manner by case law (especially where there are several opposing views). In such a case, it would be appropriate for the artificial justice system to indicate the different views in each case, so as to give a full account of them.

## 8. The principle of independence and the principle of impartiality

Independence refers to the principle recurrent in Civil Law legal systems according to which a judge is subject only to the law, and to nothing else, and must be a third party to the dispute. Can the use of artificial justice systems call this into question? It certainly would if the machine could decide certain disputes autonomously. The algorithm could not claim to be "subject only to the law:" first, because it decides on the basis of precedent and by means

[20] Generally speaking, the question of AI liability is addressed by many recent publications. See *inter alia* S. LOHSSE, R. SCHULZE, D. STAUDENMAYER (eds.), *Liability for AI*, Baden-Baden/Oxford, 2023.

[122] CONSEIL DES BARREAUX EUROPÉENS, Position paper, cit., 10.

[22] For the time being, a duty of accuracy, with reference to high-risk AI systems, is laid down in Article 15 of the AI Act. With regard to the responsibility and accountability of their providers see Article 8 and above all Article 16 of the AI Act (and for certificates see Article 44 of the AI Act).

of reasoning that differs from human reasoning; second, because it is the result of programming by experts, who might, for example, be subject to pressure from third parties.

It cannot be ruled out then, that in the future and within the (narrow) limits in which robotic decision-making is permitted, the aforementioned principle might be legitimately compressed. Nonetheless, provisions will have to be made for artificial justice systems to be checked and monitored by public authorities, for instance by means of certifying bodies.

For the time being, a "weak use" of AI does not breach the principles of judicial independence and impartiality, provided that judges use the results of artificial justice systems with caution and care. They must be even more cautious and careful if the artificial justice system does not offer guarantees of independence and impartiality.[124]

[124] See also, as for providers, Article 27 of the AI Act.

# 4.2. Privacy and its challenges in the era of AI

*Gaetana Natale –*
*Avvocato dello Stato & Legal Advisor at Italian Data Protection Authority*

*Summary: 1. The fundamental rights of the person on the digital horizon – 2. Technical hints: AI and Big Data – 3. Legal aspects: data protection in the system of international and EU law sources – 3.1. The GDPR: Regulation (EU) 2016/679 – 3.2. Regulation on AI – 4. Law and technology: a possible combination?*

*Abstract: There is consensus on the fact that privacy will be fundamentally changed by the advent of AI. This contribution seeks to conduct an all-embracing assessment of the state of the art with reference to the normative bodies currently or soon to be in force, while providing some useful insights on the potential cooperation between legal and technology experts with the aim of an enhancing the protection of privacy.*

*"Magna pars est profectus velle proficere"*
Lucius Anneus Seneca

## 1. The fundamental rights of the person on the digital horizon

The challenges posed by unstoppable scientific and technical progress engage every branch of knowledge, especially the Law, which is responsible for the inalienable functions of regulation and protection. Specifically, legal systems that give a central role to persons and their dignity are called upon to meet two opposing requirements: to ensure the protection of fundamental rights on the one hand, and to allow for the development of technology and science on the other.[125]

In this regard, the European Economic and Social Committee has identified privacy[126] as one of the eleven areas destined to be changed/eliminated due to Artificial Intelligence. The possibility of monitoring tastes, preferences or habits, a person's movements, and even learning about the most intimate aspects of their private life,[127] makes it imperative to devise instruments that give users the power to control their data. It is therefore undeniable that the legal horizon of the digital revolution opens new scenarios in terms of fundamental rights, which are destined to change depending on the frame of reference. Indeed, as far as the right to privacy is concerned, it is now anachronistic to identify it with the absolute notion of the

[125] Consider the Italian legal system, in which the rights of the human person are defined as inviolable by Article 2 of the Constitution, qualified as a general clause for the protection of humans and their interests. Given their universality, they find expression in important international and EU documents, such as the Universal Declaration of Human Rights (1950) and the Nice Charter (2000), in which the protection of humans and their dignity operates expressly as a limit both with regard to those who hold power and with regard to relations between private individuals. See P. STANZIONE, *Manuale di diritto privato*, Turin, 2021.

[126] CESE, Document C-288, August 31, 2017. "AI poses challenges for society": ethics; security; privacy; transparency and accountability; labour; education and skills; (dis)equality and inclusivity; legislative and regulatory arrangements; governance and democracy; warfare; superintelligence.

[127] Hence the gradual emergence of a tendency towards a surveillance society where all social relations that take place online are naturally traceable. See S. ZUBOFF, *Surveillance Capitalism. The future of humanity in the age of new powers*, Rome, 2019.

"right to be let alone,"[128] meant as a categorical prohibition on the collection of information, since data is now an unavoidable component of social life. To the contrary, it should be declined in terms of the right to procedural lordship, i.e. as the possibility of directly controlling the way information is collected and circulated, as well as the right to interrupt its processing if the data subject considers it to be damaging to their interests, i.e. the "right to exit."

Moreover, it goes without saying that the relationship between rights and new technologies is of constant complementarity and integration. In view of the multidisciplinary nature of the subject, after briefly outlining the state of the art, and analysing the main regulatory sources in the EU sphere, we will focus on the main critical issues related to the use of AI for the fundamental rights of the person and attempt to outline useful methods and tools in regulating the complex and constantly evolving relationship between man and machine.

## 2. Technical hints: AI and big data

Prodromal to all questions concerning the legal aspects of new technologies is the understanding – at least in broad terms – of the phenomenon.

Defining the technology looming on the horizon is already a difficult operation in itself for jurists, first, because they are typically *neophytes* in the technological field, and second, for the presence of a *mare magnum* of notions. Typically, the definition of AI refers to the idea of human intelligence, which includes the ability to learn and extract, to reason and use language, to predict, and to decide with varying degrees of autonomy.[129] In fact, in 1950, Alan Turing, considered the founding father of computer science, stated that "the idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer."[130] In other words, if the process is qualified as intelligent when performed by a human being, then it can also be qualified as intelligent when performed by a machine. At the regulatory level, however, we would like to point out the formula contained in Article 3 n. 1 of the European AI Act Regulation,[131] whereby an artificial intelligence system is defined as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."

With regard to operation, these machines are based on algorithms, i.e. ordered sequences of actions that, given certain input data (input), arrive at producing a desired end result

[128] S. WARREN, L. BRANDEIS, *The right to privacy*, in *Harvard Law Review*, 5, 1890.
[129] B. MARCHETTI, voce *Amministrazione digitale*, in *Enciclopedia del diritto*, Milan, 2022.
[130] A. TURING, *Computering Machinery and intelligence*, in *Mind*, 1950, 59, 436.
[131] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj. For a more precise definition see that formulated by the High Level Expert Group on Artificial Intelligence appointed by the European Commission in the document on *A definition of AI: Main Capabilities and Disciplines,* Brussels, April 2019, available at: https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

(output), which constitutes the solution to the problem for which the algorithm was constructed. While this statement can be applied to any intelligent system, the phenomenon, as mentioned earlier, must necessarily be understood in plural terms, moving from the simplest expert systems, to gradually more refined devices, even capable of autonomous learning. The following classifications are proposed in greater detail:[132]

a) model-based algorithms: work according to hard rules, i.e. those defined and unambiguous instructions provided by experts in a given field which, when executed, lead to a certain and definite result;

b) machine learning (ML) algorithms: starting from structured and categorised data, the systems learn how to classify new data according to type and are optimised by human feedback, which indicates incorrect and correct classifications;

c) deep learning (DL) algorithms: like the former are characterized by the ability to learn autonomously from experience and to develop their own logic to arrive at a final result, but by exploiting neural networks they are able to process unstructured data. Unlike the latter, training by a developer is not necessary.

Briefly, there are at least two critical issues that the most sophisticated algorithms present, which are relevant from both an engineering and a legal point of view. The first, located in the learning phase, concerns the large amount of data (Big Data) required for the machines to provide reliable results (at least 100 million data points for DL systems). The second relates to the defect of its lack of explainability, since it is not possible to know the process by which the system, given certain inputs, arrives at certain outputs, i.e. the "black box" problem.[133] In fact, once the training phase is over, the algorithm develops autonomous decision logics with experience, which the programmer is neither able to predetermine nor predict. It should not be forgotten that such results may be correct, incorrect, and even discriminatory (biased). Hence the well-known difficulty of using these intelligent systems to assist or even replace humans in public decision-making processes, considering the high level of guarantees provided for by national legal systems and European and international law.[134]

## 3. Legal aspects: data protection in the system of international and EU law sources

It emerges from this concise discussion that, in the age of AI, data are essential resources for economic, social, and technological development, representing the raw material upon which technology feeds.[135] In this regard, the Economist reported that "data will be (and perhaps already are) the oil of the future."[136] This statement aptly describes the phenomenon

[132]https://www.ionos.com/digitalguide/online-marketing/search-engine-marketing/deep-learning-vs-machine-learning/.

[133] For all, compare F. PASQUALE, *The black box society*, 2016.

[134] On this subject the literature is endless. *Ex plurimis*, E. PICOZZA, *Artificial intelligence and law. Politica, diritto amministrativo, and artificial intelligence,* in *Giurisprudenza italiana,* 2019, 7; C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *Biolaw Journal*, 2019, 2; F. DONATI, *Intelligenza artificiale e giustizia,* in *Rivista AIC*, 2020, 415.

[135] M. CASTELLS, *The rise of the Network society,* Oxford, 2000.

[136] *The world's most valuable resource is no longer oil but data. The data economy demands a new approach to antitrust rules,* in the *Economist,* May 6, 2017.

if one considers that the predictive capacity of algorithms, besides being used to pursue general interests, can also be employed to maximise the profit of private power. In concrete terms, these machines predict consumption and market trends, the wear and tear of infrastructure, diagnoses and cures, disasters and political decisions, and even electoral results. Of course, there is often a cost to this: economic exploitation and the commodification of personal data. This mechanism must be regulated, as data is not just an input, from which a machine proceeds to arrive at a certain result, but encompasses a universe of information about an individual's life, which it detects as an object to be protected.

The protection of personal data is first and foremost a principle that has multiple normative foundations in international, supranational, and domestic law. Proceeding by hierarchy, the wording of Article 8 of the European Convention on Human Rights (ECHR) is significant, which, by recognising the right of every person to respect for their private and family life, home and correspondence, represents the parameter on the basis of which the Strasbourg Court ascertains possible violations of the right to privacy.[137] Other normative references are Article 12 of the Universal Declaration of Human Rights, which states that "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks" and Article 17 of the International Covenant on Civil and Political Rights, which incorporates it verbatim.

The European Union, in addition to Article 16 TEU, adds the right to the protection of personal data in Article 8 of the Charter of Nice (CFREU), making it a fundamental right that binds not only EU institutions, but extends to all member States, pursuant to Article 51 of the same Charter. In particular, it represents a specific declination of the right to respect for private and family life referred to in Article 7 of the same document and already provided for by Article 8 of the ECHR. Specifically, the provision establishes that "1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data shall be processed fairly, for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Every person shall have the right of access to the data collected concerning him or her and the right to have them rectified. 3. Compliance with these rules shall be subject to control by an independent authority."

What emerges from all these sources is a common conception of privacy that does not coincide with the traditional concept of the right to anonymity or to be let alone, but rather with the idea that everyone should have the right to control their own personal information, as a prerequisite for the exercise of many other rights of freedom, especially of a cyber nature.[138]

---

[137] In *Sidabras v. Lithuania*, July 27, 2004, n. 55480/00-59330/00, the ECtHR gave a very broad interpretation of the right to privacy under Article 8 of the ECHR. The Strasbourg judges held, in fact, that the protection provided by this article extends to encompass the right of each person to develop social relations free from all forms of discrimination or social stigmatisation, thus also allowing him or her the full enjoyment of his or her private life. The Court, therefore, considered the overall place of the person in society, stating that full respect for privacy is a condition for equality and the enjoyment of fundamental rights, such as the right to work.

[138] Reference is made to the doctrine of cyber-freedom, a theory that was put forward in 1981 and had its ideological matrix in the conception of a new liberalism. It was originally distinguished by positive and negative freedom. Negative freedom of information technology expresses "the right not to place in the public domain certain information of a personal, private, confidential nature (qualifications that may not coincide with each other in certain cases); positive freedom of information technology, on the other hand, expresses the faculty to exercise a right of control over data concerning one's own person that has escaped the circle of privacy

The European Union, in its aim to assert a European "digital sovereignty," envisages the construction of a regulatory framework, essentially based on four pillars:

a) the protection and enhancement of personal data: the former covered by Regulation (EU) 2016/679 "on the protection of individuals with regard to the processing of personal data and on the free movement of such data" (better known as the GDPR); the latter by the *Data Act*, the *Data Governance Act,* and the proposed regulation on European health data space;

b) digital services and the digital market: the subject of the Digital *Services Act* and the *Digital Markets Act*;

c) digital identity: the 2014 E-IDAS regulation is to be revised in this respect;

d) AI: a Regulation laying down harmonised rules (the AI Act) and amending certain pieces of Union legislation has been adopted.

### *3.1. The GDPR: Regulation (EU) 2016/679*

In order to understand the transformation of privacy in the age of AI, it is necessary to start from an analysis of the GDPR Regulation.[139] According to Article 2, it also applies to the processing of personal data carried out in whole or in part by AI.[140] Article 1, in defining object and purpose, states that "This Regulation lays down rules relating to the protection of individuals with regard to the processing of personal data, and rules relating to the free movement of such data." The first provision already demonstrates how the right to privacy is not protected absolutely, but must be combined with the need for the free movement of data. Indeed, the purpose of the regulation is not only to guarantee the protection of personal data, but also to promote the development of the Digital Single Market.[141]

To increase citizens' trust in the use of new digital services, a trustworthy digital environment must be created, in which the identity of the data controller, the procedures, and the levels of protection are known. The regulation focuses on the principles of accountability and compliance, as set out in Article 5(2), i.e. "The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1" ("accountability"). These are primarily incumbent on the data controller, who is called upon

because they have become input elements of an electronic programme; and therefore positive freedom of information technology, or the recognised subjective right, to know, correct, remove or add data in an electronic personal file." See THUS V. FROSINI, *La protezione della riservatezza nella società informatica*, in N. MATTEUCCI (ed.), *Privacy and* data *banks,* Bologna, 1981, 37 (later included in vol. Id., *Informatica diritto e società,* II, Milan 1992, 173).

[139] https://protezionedatipersonali.it/privacy-by-design-e-by-default.

[140] Article 2 "Material scope:" "This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system." See G. FINOCCHIARO, *XVIII lezione: intelligenza artificiale, privacy e data protection*, in U. RUFFOLO (ed.), *XXVI Lezioni di diritto dell'intelligenza artificiale*, Turin, 2021, 331.

[141] Recital 7: "Those developments require a strong and more coherent data protection framework in the Union, backed by strong enforcement, given the importance of creating the trust that will allow the digital economy to develop across the internal market. Natural persons should have control of their own personal data. Legal and practical certainty for natural persons, economic operators, and public authorities should be enhanced."

to choose the most appropriate measures to prevent risks, to take the necessary decisions and to prove that they are adequate, on pain of liability under Article 24.

The GDPR's approach is based on risk assessment (*risk based*), a parameter against which the degree of accountability of the data controller or processor is measured. Obviously, the controller is bound by specific principles set out in the regulation: in particular, *privacy by design* and *by default* and a *Data Protection Impact Assessment*.

The principle of *privacy by design,* referred to in Article 25(1), provides that, when taking the state of the art and the costs of implementation into account, as well as the nature, scope, context and purposes of the processing, in addition to risks of differing probability and severity for the rights and freedoms of natural persons constituted by the processing, the controller must implement, both when determining the means of processing and at the time of the processing itself, "appropriate technical and organisational measures, such as pseudonymisation," referred to in Article 4(1)(5). The structure is designed to effectively implement the principles of data protection, such as data minimization, to incorporate necessary safeguards in the processing and to meet the requirements of the Regulation and protect the rights of data subjects.

Linked to this criterion is the principle of *privacy by default*, enshrined in the second paragraph of Article 25: the data controller must implement "appropriate technical and organisational measures to ensure that only the personal data necessary for each specific purpose of the processing are processed by default." The individual is thus protected in a strengthened way since the provision establishes access to an indefinite number of natural persons by machines (without the intervention of the natural person) and provides that the obligation is calibrated on aspects such as the amount of data, the scope of processing, the retention period, and accessibility.

Also of note is Article 35, concerning the so-called *Data Protection Impact Assessment*. It states that when a type of processing, involving the use of new technologies in particular – while taking the nature, subject matter, context and purpose of the processing into account – may present a high risk for the rights and freedoms of natural persons, the data controller shall carry out, before processing, "an assessment of the impact of the intended processing on the protection of personal data."[142]

Despite its complexity – one hundred and seventy-three *recitals* and ninety-nine articles – and its proactive and flexible approach to the subject of personal data protection, the GDPR cannot be considered a self-sufficient and immutable body of legislation. The drafters themselves were aware of these qualities: in Article 12(8) and Article 43(8) they empowered the European Commission to adopt delegated acts and implementing acts to lay down technical standards concerning the certification mechanisms and data protection seals and marks; they also delegated the adoption of more specific rules for adapting the application of the Regulation to member States. Furthermore, Article 97 provides for a review of the GDPR every four years, allowing the Commission to propose amendments to the Regulation, specifically "in particular developments in information technology and progress in the information society."

[142] Again, https://protezionedatipersonali.it/privacy-by-design-e-by-default.

### 3.2. Regulation on AI

As much as the tools and principles provided by the GDPR may lend themselves to extensive application in today's "data-driven society," there is an urgent need to develop models for regulating new technologies. Self-regulation? Homogenous or sector-specific regulation? Where there has been a move towards a self-regulatory model in the United States, and in China there has been specific and detailed regulation, the European legislator has opted for a horizontal approach, with rules applicable to each sector, including the health and financial sectors. The European Union AI Act aims to ensure that AI systems placed on the EU market are safe and ethical, comply with existing fundamental rights legislation, and respect EU values through a proportionate risk-based approach.

AI systems are classified into three categories according to the risk they present: a) AI with unacceptable risks; b) AI with high risks; c) AI with low or minimal risks. First of all, systems that present an unacceptable risk are banned. These include "real-time" remote biometric identification systems in publicly accessible spaces for example.[143] Instead, for low-risk AI systems, certain transparency obligations are laid down and codes of conduct are encouraged. For instance, in AI systems that are intended to interact with individuals, it is required that users be informed they are interacting with an AI system; for so-called "deep fake" systems that generate or manipulate images or audio or video content that closely resemble existing persons, objects, places, or other entities or events and that could appear falsely authentic or true, it is required that users disclose that the content has been artificially generated or manipulated. Finally, the obligations for the adoption of high-risk AI systems are listed in detail. In particular, it is stipulated that such systems are subject to an ex-ante conformity assessment procedure, which concludes by affixing the CE mark. Additionally, high-risk AI systems must be designed and developed in such a way as to guarantee, by means of automatic event logging, the traceability of their operations throughout their life cycle which must be sufficiently transparent to enable users to interpret their output and use it appropriately.

It should be evident that the proposed new regulation borrows its main axes from the GDPR: from the risk-based approach to the duties of transparency towards users, to certifications and codes of conduct. Furthermore, the unavoidable incidence point for both subjects is not marginal: the processing of personal data is functional to feeding AI systems with a view to their automatic learning. It is apparent then how errors or mistakes in the processing of data, functional to feeding the machine, are reflected in as many distortions of the algorithmic process.[144] There is then an objective need to avoid the emergence of antinomies between the different disciplines mentioned, in order to render the regulation of the matter more organic and effective as a whole.

[143] This is the only system in which there are exceptions to its prohibition, pursuant to Article 9 of the GDPR, in cases of searching for victims of crime, threats to life, or terrorist acts, or searching for persons guilty of serious criminal offences. In these cases, the use of the system may be permitted, subject to authorisation by a judicial authority, or independent administrative authority.

[144] C. UTZ ET AL., *(Un)informed Consent: Studying GDPR Consent Notices in the Field*, in *ACM SIGSAC Conference on Computer and Communications Security*, November 11-15, 2019, London.

## 4. Law and technology: a possible combination?

In a climate of general mistrust towards technological and scientific progress, the European Union's attempt to regulate the AI phenomenon is certainly welcome, although we are aware that the speed at which AI is progressing and the complexity of the issues carry the risk of rendering any regulation obsolete. If law and technology travel at two different speeds, perhaps it would be appropriate to adapt legal instruments to the speed of the latter? Perhaps this can be accomplished by opting for soft law instruments rather than hard ones? Or perhaps by preferring general rather than detailed legislation? One thing is certain: the transnationality of the phenomenon requires that all questions be answered at a global level.

# 4.3 AI risk assessment tools for criminal justice: risks to human rights and remedies

Marco Gioia – Tribunale di Cassino



*Abstract: The right to a fair trial has become a shrine for the protection of all the subjects involved in a judicial proceeding, bearing particular importance in the Criminal Law justice system. It has been evidenced how the introduction of algorithms into trials has begun to present risks for compliance with the abovementioned right. An analysis of a landmark case on the topic is provided, as well as a more all-embracing examination of current and future trends with regard to the maximization of benefits without the minimization of safeguards.*

## 1. AI in the criminal trial: risk assessment tools

Artificial Intelligence, which includes machine learning and other automated systems based on analytical algorithms, has become an important aspect of our lives. In recent years, this technology has gradually been used in the criminal justice systems of several countries, playing an important role in the decision-making process of some criminal cases.

One of the main areas where courts have used AI systems is in assessing the likelihood of recidivism. The most analysed and discussed examples come from the United States, where most of the software is currently being applied. The usage of risk assessment tools to classify offenders has a long history in the US criminal justice system. These tools are used to inform court decisions at various stages of the criminal justice process, from pretrial services to proceedings closely related to defendants' freedom. These tools are used for example to set bail amounts for suspects when they appear in court after arrest, in decisions related to parole or probation, and even in making judgements themselves in States such as Arizona, Colorado, Delaware, Indiana, Kentucky, Louisiana, Oklahoma, Virginia, Washington and Wisconsin.

The evaluation relies on two types of factors: static and dynamic. Static factors refer to unchangeable characteristics, like a person's ethnic group, while dynamic factors relate to characteristics that can evolve over time, such as family and marital status, employment, and educational level. The risk factors that are widely acknowledged as being the most crucial aspects to evaluate are: past antisocial conduct, trends of antisocial personality, antisocial thinking, antisocial associations, difficulties within the family or marriage, challenges in work or school, and substance misuse.

Risk assessment tools rely on the actuarial method, which is known for its mechanical and algorithmic approach. Actuarial methods assess the comparative risk presented by an individual in comparison to a benchmark group within a designated time frame. These instruments rely on the utilization of static risk factors obtained from empirical research.

These elements are given a numeric value and an algorithm calculates a sum, which is then utilized to predict the probability of someone committing another crime within a designated timeframe.

The latest generation of risk assessment instruments still includes static risk factors, but also integrates dynamic risk factors that are theoretically associated with reoffending and become known as "criminogenic needs." They could change over time, and incorporating them into risk evaluations enhances efficiency by allowing criminal justice authorities to focus on risk mitigation efforts and, contextually, supervise them.

One of the most famous risk assessment tools is COMPAS (short for Correctional Offender Management Profile for Alternative Sanctions), used in the United States. It is a decision support tool developed by a private company (originally named Northpointe and renamed Equivant in 2017) that makes predictions based on the defendant's criminal record and a questionnaire. More precisely, "this software predicts a defendant's risk of committing a misdemeanor or felony within 2 years of assessment from 137 features about an individual and the individual's past criminal record." These features include things like age, gender, and criminal history, but do not include ethnic group. Although originally designed for pretrial release decisions and post-judgement decisions, i.e. parole, the use of this tool has gradually expanded to sentencing decisions, as illustrated in the Loomis case.

COMPAS is a classic example of supervised machine learning. In the past decade, the manufacturer Northpointe implemented machine learning techniques to reliably assign new cases to appropriate perpetrator categories. According to Northpointe, random forests and support vector machines are used as machine learning methods. The random forest method is considered a black box method because the relationship between the predictor variables and the outcome variable is not explained. The algorithm learns from past data analysed using a decision tree and develops model relationships between independent and dependent variables. After sufficient training, the algorithm is applied to determine the defendant's recidivism rate. COMPAS generates risk scores, displayed as a histogram, consisting of three columns indicating the pretrial risk of recidivism, the general risk of recidivism, and the risk of violent recidivism. Each bar indicates the defendant's risk level on a scale of one to ten.

## 2. Case law regarding the use of AI risk assessment tools in sentencing: the *State v. Loomis*

In the Loomis case,[145] the result of the assessment by COMPAS was included in the file as part of the investigation report submitted to the sentencing court for the purpose of reaching a judgement.

Loomis was alleged to have been a driver in a drive-by shooting case who denied his involvement. Later, he waived his right to trial and entered a guilty plea to only two of the less severe charges. The plea agreement stated that the other counts would be dismissed, but read in. After accepting Loomis' plea, the circuit court ordered a pre-sentence investigation (PSI) which included, as an attachment, the COMPAS risk assessment. Loomis' COMPAS risk scores indicated that he presented a high risk of recidivism on all three bar charts. The circuit court referenced the COMPAS risk score along with other sentencing factors in ruling out probation and affirming the following: "you're identified, through the COMPAS

[145] Supreme Court of Wisconsin, State of Wisconsin v. Eric L. Loomis, Case n. 2015AP157-CR, 5 April – 13 July 2016, *881 N.W.2d 749 (Wis. 2016)*.

assessment, as an individual who is at high risk to the community. In terms of weighing the various factors, I'm ruling out probation because of the seriousness of the crime and because your history, your history of supervision, and the risk assessment tools that have been utilized, suggest that you're an extremely high risk to re-offend."

Loomis appealed and the court of second instance certified the appeal to the Wisconsin Supreme court. Loomis asserted that the use of a COMPAS risk assessment by the circuit court in the context of the judgement violated the defendant's right to due process on three grounds: i) violation of the right to be sentenced based upon accurate information, in part because the proprietary nature of COMPAS prevented him from assessing its accuracy; ii) violation of the right to an individualized sentence; iii) improper use of gendered assessments in sentencing.

The Wisconsin Supreme Court concluded that a COMPAS risk assessment can be used for sentencing, but such a use must be circumscribed, i.e. the limitations and cautions must be observed and addressed, in order to avoid potential due process violations.

According to the Court, any PSI Report containing a COMPAS risk assessment must inform the sentencing court about the following cautions regarding the COMPAS risk assessment's accuracy: "(1) the proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are to be determined; (2) risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed; (3) some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority renders as having a higher risk of recidivism; and (4) risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations. Providing information to sentencing courts on the limitations and cautions attendant with the use of COMPAS risk assessments will enable courts to better assess the accuracy of the assessment and the appropriate weight to be given to the risk score."

Regarding Loomis' argument that a circuit court's consideration of a COMPAS risk assessment amounted to sentencing based on group data, rather than an individualized sentence based on the charges and the unique character of the defendant, the Court stated: "If a COMPAS risk assessment were the determinative factor considered at sentencing this would raise due process challenges regarding whether a defendant received an individualized sentence." However, the Court explained that as the report was not the sole basis for a decision, sentencing that considered a COMPAS assessment would still be sufficiently individualized since courts have the discretion and information necessary to disagree with an assessment when appropriate. However, the Wisconsin Court warned: "Due process implications compel us to caution circuit courts that, because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders — not a particular high-risk individual. Accordingly, a circuit court is expected to consider this caution as it weighs all of the factors that are relevant to sentencing an individual defendant."

Finally, the court explained that risk scores may not be used "to determine whether an offender is incarcerated" or "to determine the severity of the sentence." Therefore, judges using risk assessments must explain the factors other than the assessment that support the sentence imposed. In her concurring opinion, Justice Abrahamson affirmed that while she agreed with the judgment, she was concerned that the court had difficulties in understanding

algorithmic risk assessments. She stated that she would have required a more extensive record from sentencing courts on "the strengths, weaknesses, and relevance to the individualized sentence being rendered of the evidence-based tool."

## 3. Advantages and disadvantages of using AI risk assessment tools in criminal proceedings

The use of AI in criminal case adjudication, as suggested by the private companies creating these tools, is intended to enhance the quickness, effectiveness, and seemingly unbiased nature of judicial decision-making. It has also been argued that the use of such tools would result in more reliable and consistent decisions that are not dependent on each judge's personal sensibilities or opinions or even positions influenced by undue pressure.

Despite the benefits mentioned, the utilization of AI in judicial proceedings also brings different and considerable disadvantages, which are more significant for Criminal Law compared to the domains of Civil or Administrative Law.

### *3.1. Problems related to data collection. Risk of incomplete data and biased decisions.*

Systems that rely on data necessitate the gathering and utilization of an extensive set of data. However, circumstances of poor quality, inaccurate, and/or poorly transcribed data also have an effect on outcomes. Additionally, mistakes and/or biases could be built into the system itself. Furthermore, the recommendations of a predictive algorithm could potentially hide the subjective opinions of the individuals who created the system regarding which data are used, incorporated or excluded, as well as how to assign importance to the them and which information is prioritized or minimized.

The afore-mentioned COMPAS software offers an example of these issues. This software was designed not to consider the ethnicity of individuals and so we should expect this type of data should not influence a recidivism risk assessment. Nevertheless, a study by ProPublica, an independent non-profit NGO, revealed that, ethnicity did matter indirectly and even outweighed other explicitly included factors due to the cross-referencing of different data such as place of residence or profession, but also because certain ethnicities were overrepresented in the data used to train the system. For example, African-American populations were assigned a high-risk recidivism rate twice that of other populations within two years of sentencing, though this effect was not sought by the designers, whereas the algorithm considered other ethnic populations as being less likely to repeat an offence. This demonstrates that the risk of biases is significant and not easy to address. AI tools trained with past data inevitably create models that replicate what has happened in the past.

The use of COMPAS in the *Loomis* case, however, also shows that the quality (and accuracy) of risk assessments is determined by the quality of data collected. In particular, the Court pointed out how the tool had been trained with data collected nation-wide, without cross-validation with the Wisconsin population. As a result, the specific characteristics such as demographics and social, economic, and legal status, of the particular State where the tool was used may not have been adequately reflected. The Wisconsin Supreme Court thus required judges and other parties of the trial where COMPAS was used to receive notification about the limitations of the system.

Furthermore, in some legal systems, there may be a lack of adequate and/or precise data that can be utilized as a recidivism prediction tool.

### *3.2. The opacity of AI systems or the "black box" problem*

A black box indicates that "the processes happening inside of them are difficult – and sometimes impossible – to fully understand." The issue of opacity is a significant concern, specifically in the context of machine learning and particularly for systems that rely on neural networks.

The issue of the black box is closely connected to the need for providing explanations and justification for judicial decisions in criminal cases, where the decision impacts a defendant's freedom and fundamental rights more deeply. It is extremely important that all parties involved understand the technical aspects of the tool, how the algorithm selects data, and how it generates a particular outcome. Without transparency, system predictions about the risk of recidivism may be ill-founded or even arbitrary.

## 4. AI risk assessment tools and human rights

### *4.1. The right to a fair trial and transparency*

The use of algorithms in criminal justice systems raises serious concerns with regard to Article 6 of the European Convention on Human Rights, on the right to a fair trial, as well as Article 47 and Article 48 of the Charter, dealing with the principle of the equality of arms and adversarial proceedings as established by the European Court of Human Rights.

The fair trial standards contained in Article 6 of the ECHR guarantee the right of the accused to participate effectively in the trial and include the presumption of innocence, the right to be informed promptly of the cause and nature of the accusation, the right to a fair hearing, and the right to defend oneself in person. Paragraph 3 specifically requires that all evidence against the accused be produced in his or her presence at a public hearing, thereby giving defendants an effective opportunity to challenge the evidence against them. In order to ensure the effective participation in a trial, the accused must be able to contest the algorithmic score on which their conviction is based.

Transparency is essential to the right to a fair trial. The defendants should have full access to case documents and should be allowed to provide their input on the evidence presented against them. A significant obstacle to ensuring transparency for example, could be a lack of notification of the use of AI tools. If a person is not informed about being subject to an automated decision by an AI system, such an individual will not have the possibility to contest the decision or the data the decision relies upon.

As mentioned earlier, another significant obstacle to the contestability of AI systems is of a technical nature. The "black box" problem can mostly be attributed to how AI systems are designed. It is therefore crucial to have rules ensuring these systems can be interpreted and understood when they are being used**.** It is essential that the "reasoning" of such systems is made known to suspects and accused persons, similarly to how judicial decisions must contain "sufficient reasoning and address specific features of a given case," especially where they concern a deprivation of liberty. Decision-making processes of AI systems and the way in which they have produced an outcome in a particular case should thus be disclosed to accused persons, in a form that is intelligible to a layperson.

The commercial interests of the owners of the AI software used, should not be regarded as a valid justification for non-disclosure. Furthermore, EU Law does not explicitly recognize any derogation from the right of access to materials that is essential to challenging the lawfulness of an arrest or detention. In order for member States to comply with these standards, any exception to the disclosure of information regarding AI systems must be applied very narrowly.

### *4.2. Presumption of innocence and the "in dubio pro reo" principle*

Although risk assessment tools like COMPAS do not determine a person's guilt or innocence, their evaluation can impact decisions that lead to a deprivation of their liberty or the severity of their conviction. These types of high-impact decisions cannot be delegated to automated processes that are based on identifying correlations rather than establishing causal links between human characteristics and likely behaviour.

An examination of HART, a risk assessment tool used by the Durham Constabulary in the United Kingdom, also reveals that the tool considers under-estimations of risk levels as a more serious error than over-estimations. In other words, HART is deliberately designed to underestimate who is eligible for entry into the diversion programme, so it is predisposed to over-criminalise. This approach conflicts with the notion that any doubt in a criminal case should be interpreted in favour of the defendant, i.e. the principle of "*in dubio reo.*" A human rights compliant approach to criminal justice decision-making would thus do the opposite of what HART does, because when doubts arise, they should be decided in the defendant's favor.

### *4.3. The principle of personal criminal liability*

The ECtHR, in the case of *G.I.E.M. S.R.L. and Others v. Italy*, § 251, pointed out that Article 7 precluded the imposition of a criminal sanction on an individual without their personal criminal liability being established and declared beforehand. Instead, the inclusion of algorithmic variables such as criminal history and family background means that the past behaviour of a certain group may decide the fate of an individual, who is, of course, a unique human being with a specific social background, education, skills, degree of guilt and even distinct motivations for committing a crime.

### *4.4. The principle of non-discrimination*

As mentioned previously, one of the most frequent criticisms of AI systems and their use in criminal justice systems is that they can lead to discriminatory outcomes, especially along racial and ethnic lines. The best-known example of this is the above-cited study made by the US media outlet ProPublica on COMPAS. ProPublica found that COMPAS was more likely to rate black defendants as "high-risk"" than white defendants and was almost twice as likely to mislabel white defendants as lower risk with respect to the assessment for black defendants.[146]

Removing visible biases from AI systems cannot be the sole or primary solution to their discriminatory impact, because AI systems can be biased even if they have not been

[146] J. ANGWIN ET AL., *Machine Bias*, in *ProPublica*, 2016.

deliberately designed that way. Bias is often unintentional, and even if the AI system appears on the surface to be neutral, its algorithms can lead to discriminatory assessments and outcomes. COMPAS, for example, does not include ethnicity as a variable, yet researchers found that it consistently gave black defendants higher risk scores than their white counterparts, rendering them less likely to be released from detention.

## 5. European regulation of risk assessment tools in criminal proceedings

To address some of the risks to fundamental rights just outlined, the Council of Europe's European Commission on the Efficiency of Justice (CEPEJ) has developed the "Ethical charter on the use of artificial intelligence in judicial systems and their environment" identifying five ethical principles to which the use of AI in the judiciary should respond:

- the principle of respect for fundamental rights: ensuring that the design and implementation of AI tools and services is compatible and in compliance with them;
- the principle of non-discrimination: specifically preventing the development or intensification of any discrimination between individuals or groups of individuals;
- the principle of quality and security: with regard to the processing of judicial decisions and data, using certified sources and intangible data with models elaborated in a multi-disciplinary manner and in a secure technological environment;
- the principle of transparency, impartiality, and fairness: making data processing methods accessible and understandable, while authorising external audits;
- the principle of being "under user control:" precluding a prescriptive approach and ensuring that users are informed actors and in control of the choices they make.

Regarding the use of AI in criminal proceedings, in the Annex to the Ethical Charter the CEPEJ highlights that "when algorithms are used in the context of a criminal trial it seems essential to fully guarantee respect for the principle of equality of arms and presumption of innocence established by Article 6 of the ECHR. The party concerned should have access to and be able to challenge the scientific validity of an algorithm, the weighting given to its various elements, and any erroneous conclusions it comes to whenever a judge suggests that he/she might use it before making his/her decision. Moreover, this right of access is also covered by the fundamental principle of personal data protection. All people have the right not to be subject to significant decisions affecting them, made solely on the basis of automated data processing, without their point of view being heard beforehand."

In this respect, there is a difference between Europe and the United States with regard to the right of access to algorithms: while in the United States judicial authorities are still reluctant to recognize this right fully and weigh private interests – particularly the protection and the enforcement of Intellectual Property rights – against the rights of the defence, in Europe the framework is more protective because of the GDPR, which expressly establishes a right to information on the underlying logic of decisions made using algorithms.[147]

[147] Article 15,1(h) of EU Regulation 2016/679 affirms that "The data subject shall have the right to obtain from the controller" ... "the following information:" ... "the existence of automated decision-making, including profiling, as referred to in Article 22, paragraphs 1 and 4, and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."

On the Criminal Law side, these principles are reiterated in the European Parliament "Resolution on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters."[148] The EU Parliament notes that "AI tools and applications are also used by the judiciary to support decisions on pre-trial detention, in several countries worldwide, including to support decisions on pre-trial detention, in sentencing, calculating probabilities for reoffending and in determining probation […]" and stresses "the potential for bias and discrimination arising from the use of AI applications such as machine learning, including the algorithms on which such applications are based." The EU Parliament also "highlighted the power asymmetry between those who employ AI technologies and those who are subject to them; stresses that it is imperative that use of AI tools by law enforcement and judicial authorities not become a factor of inequality, social fracture or exclusion; underlines the impact of the use of AI tools on the defence rights of suspects, the difficulty in obtaining meaningful information on their functioning and the consequent difficulty in challenging their results in court, in particular by individuals under investigation." For these reasons the EU Parliament "considers it essential, both for the effectiveness of the exercise of defence rights and for the transparency of national criminal justice systems, that a specific, clear and precise legal framework regulates the conditions, modalities, and consequences of the use of AI tools in the field of law enforcement and the judiciary, as well as the rights of targeted persons, and effective and easily available complaint and redress procedures, including judicial redress; underlines the right of the parties to a criminal proceeding to have access to the data collection process and the related assessments made by or obtained through the use of AI applications."

EU institutions have been progressively shaping the EU's AI agenda with several policy documents. In April 2021, the Commission unveiled a legislative proposal for an AI Act[149] which was definitively adopted in May 2024.[150] The Regulation provides for a risk-based classification of AI systems.

AI systems for justice and law enforcement are considered "high-risk" and should comply with strict requirements, such as conformity assessments, transparency, and human oversight. Particularly in 38) where the Commission states that "actions by law enforcement authorities involving certain uses of AI systems are characterised by a significant degree of power imbalance and may lead to surveillance, arrest, or deprivation of a natural person's liberty as well as other adverse impacts on fundamental rights guaranteed in the Charter. In particular, if the AI system is not trained with high quality data, does not meet adequate requirements in terms of its accuracy or robustness, or is not properly designed and tested before being put on the market or otherwise put into service, it may single out people in a discriminatory or otherwise incorrect or unjust manner. Furthermore, the exercise of important procedural fundamental rights, such as the right to an effective remedy and to a

[148] EUROPEAN PARLIAMENT, *Resolution on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters*, 2020/2016(INI), October 6, 2021, available at: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html

[149] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts – Com/2021/206 Final

[150] Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

fair trial as well as the right of defence and the presumption of innocence, could be hampered, in particular, where such AI systems are not sufficiently transparent, explainable and documented. It is therefore appropriate to classify as high-risk a number of AI systems intended to be used in the law enforcement context where accuracy, reliability, and transparency is particularly important to avoid adverse impacts, retain public trust and ensure accountability and effective redress. In view of the nature of the activities in question and the risks relating thereto, those high-risk AI systems should include in particular AI systems intended to be used by law enforcement authorities for individual risk assessments, polygraphs, and similar tools or to detect the emotional state of natural persons, to detect 'deep fakes,' for the evaluation of the reliability of evidence in criminal proceedings, for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on the profiling of natural persons, or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups, for profiling in the course of detection, investigation, or prosecution of criminal offences, as well as for crime analytics regarding natural persons."

However, the AI ACT proposed by the Commission only sets general rules of conformity assessment, transparency, and human oversight that should be applied to all AI activities classified as high-risk. It does not designate any specific regulations for the use of AI systems particularly designed for judicial decisions in criminal proceedings.

The AI act was presented to the European Parliament which approved a text with numerous radical amendments to the Commission proposal on June 14, 2023. The finalized text included a ban on predictive policing and criminal prediction systems used by law enforcement and criminal justice authorities in the EU. The concerns regarding the possible violation of human rights caused by the use of risk assessment tools induced the European Parliament to state, in 26a), that "AI systems used by law enforcement authorities or on their behalf to make predictions, profiles, or risk assessments based on the profiling of natural persons or data analysis based on personality traits and characteristics, including the person's location, or the past criminal behaviour of natural persons or groups of persons for the purpose of predicting the occurrence or reoccurrence of an actual or potential criminal offence(s) or other criminalised social behaviour or administrative offences, including fraud-prediction systems, hold a particular risk of discrimination against certain persons or groups of persons, as they violate human dignity as well as the key legal principle of a presumption of innocence. Such AI systems should therefore be prohibited." The Regulation thus sets out Article 5 – paragraph 1 – point d, that includes, among the prohibitions "the placing on the market, the putting into service for this specific purpose, or the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity."

# Section 5

# Hands-on: from theoretical thinking to experimental projects

## 5.1. Deploying AI technology to empower procedural safeguards and judicial cooperation

*Giuseppe Contissa, Giulia Lasagni – Università degli Studi di Bologna*

*Summary: 1. Introduction – 2. The CrossJustice project – 3. The FACILEX project – 4. The way forward – 5. Hands-on: Case Studies*

*Abstract: The contribution describes two research projects, funded by the EU Commission, in which AI tools are being developed by a multi-disciplinary team of jurists expert in criminal procedure, legal informatics, and computer scientists, with the purpose of improving and facilitating judicial cooperation in criminal matters in the European Union.*

### 1. Introduction

Criminal law in the EU currently represents a good example of how the interaction between Law and legal informatics can bring an important synergy that enhances the efficiency of the overall system as well as the protection of citizens' fundamental rights.
The protection of fundamental rights for persons accused or suspected of a crime is one of the main aims of EU policy in the area of criminal justice. However, the effective implementation of such rights throughout the EU is heavily affected by the highly varying legal frameworks which characterize member State regulation on procedural rights in criminal proceedings.[151] In this context, legal actors often struggle to identify which legislation and therefore which procedural rights are applicable to persons accused or suspected in specific cases, both due to linguistic barriers and the peculiarities of different national legal systems.

While we face such challenges in Europe, on a global level the domain of Law is also on the brink of a period of fundamental and irreversible change and transformation. Information technologies and computer systems are the main drivers of such change. Increasingly capable machines, operating in support of experts, are transforming the way legal activities and procedures are undertaken. New legal applications and platforms have appeared, changing the practice of Law, specifically by enriching legal texts with

[151] The problem, at the basis of EU initiatives since the Green Paper on *Procedural Safeguards for Suspects and Defendants in Criminal Proceedings throughout the European Union,* COM(2003) 75 final, February 19, 2003, also remains significant in the post-Stockholm framework, as highlighted by several studies. See, *ex multis*, S. ALLEGREZZA, V. COVOLO, *Effective Defence Rights in Criminal Proceedings. A European and Comparative Study on Judicial Remedies*, Padua, 2018; M. CAIANIELLO, G. LASAGNI, *Comparative Remarks*, in G. CONTISSA, G. LASAGNI, M. CAIANIELLO, G. SARTOR (eds.), *Effective Protection of the Rights of the Accused in the EU Directives. A Computable Approach to Criminal Procedure Law,* Leiden, 2022, 229.

computational models of legal reasoning, providing users with automated assessments and explanations of their outcomes with reasons and arguments.[152] New AI enhanced tools, such as large language models like Chat GPT, are being introduced to analyse and reason on case-law and legislation, although several hard challenges are still to be addressed to make such tools of true support to the legal professions.[153]

Against this background, we introduce the results of two research projects, one that has concluded and another ongoing, that try to address such problems from a multi-disciplinary perspective. They not only investigate the theoretical aspects, but also deliver concrete tools to support legal professionals and citizens in retrieving the necessary knowledge to face criminal proceedings at the EU level.

## 2. The CrossJustice project

The CrossJustice ("Knowledge, Advisory and Capacity Building Information Tool for Criminal Procedural Rights in Judicial Cooperation") project is a thirty-month international research project funded by the EU Commission (DG JUST) and conducted under the supervision of the University of Bologna.[154] The study counted seven international partners including eminent scholars and researchers in the field of European Law and Criminal Procedure, Legal Informatics and Computer Science, as well as private sector experts working in the field of online platform development.

The CrossJustice study tackled the domain of Criminal Procedure Law by critically examining the procedural safeguards of the suspect and the accused as recognised by the six EU Procedural Rights Directives: Directive 2010/64/EU of 20 October 2010 on the right to interpretation and translation; Directive 2012/13/EU of 22 May 2012 on the right to information; Directive 2013/48/EU of 22 October 2013 on the right of access to a lawyer and to have a third party informed; Directive (EU) 2016/343 of 9 March 2016 on the presumption of innocence and the right to be present at trial; Directive (EU) 2016/800 of 11 May 2016 on procedural safeguards for juvenile defendants; Directive (EU) 2016/1919 of 26 October 2016 on legal aid. All these Directives were analysed both with regard to the statutory legal framework and in their practical implementation in the case-law of the national courts of member States.

Where the existing EU acquis establishes common minimum standards on criminal procedural rights, the need to promote their effective and coherent application remains particularly pressing due to highly fragmented national enforcement. This fragmentation, in addition to linguistic diversity, makes it difficult for legal professionals to understand how and to which extent fundamental rights are safeguarded in a different legal order or in transnational proceedings. In particular, the project investigated the primary lacunae in the protection of defence rights in eleven EU jurisdictions (Bulgaria, Croatia, France, Germany, Italy, Portugal, Poland, Romania, Spain, Sweden, the Netherlands).

---

[152] M. BILLI, R. CALEGARI, G. CONTISSA, F. LAGIOIA, G. PISANO, G. SARTOR, G. SARTOR, *Argumentation and defeasible reasoning in the law*, in *J*, 2021, 4, 4, 897.

[153] T. DAL PONT, F. GALLI, A. LOREGGIA, G. PISANO, R. ROVATTI, G. SARTOR, *Legal Summarization through LLMs: The PRODIGIT Project*, 2023, available at: arXiv:2308.04416.

[154] https://site.unibo.it/crossjustice/en. Grant Agreement n. 847346; the project was funded by the European Union's Justice Programme (2014-2020).

The research proposes an innovative perspective on the topic in two ways. First, analysis of the shortcomings and obstacles that procedural rights of Directives meet in their national implementation was carried out through the lens of legal reasoning. Second, the research integrates a legal informatics approach, consisting of translation of the EU Procedural Directives, as well as samples of national legislation, into a computable language that became the core of the CrossJustice online platform. In particular, the project resulted in the development of a Legal Database and an Advisory Module, both freely accessible online.

The Legal Database aims at providing access to legislative, judicial, and expert documents that fall within the material scope of Criminal Procedure Law, including international treaties and standards, relevant EU legislation and case-law, as well as national legislation and case-law, all provided in an English translation.

The Advisory Module, in turn, is an interactive expert module that helps legal professionals identify and apply relevant rules of EU and national legislation concerning the procedural rights of persons suspected or accused of crimes, both at the domestic and cross-border level. The Module can be of assistance on two horizons: 1) it assesses the compliance of national instruments implementing the six Directives of the EU acquis, by highlighting potential gaps in the implementation process (e.g. is the legal framework in member State X in compliance with the right of access to a lawyer as recognized by the EU Directives?); 2) it horizontally assesses the compatibility between national frameworks resulting from the implementation of EU directives (e.g. is the legal framework for the presumption of innocence in member State X compatible with the legal framework on the same matter in member State Y?).

Against this background, the Advisory Module is composed of three computer tools, grounded on highly advanced juridical data, collected and uploaded by legal experts, and then processed in a semi-automated way to provide customized support to legal professionals.[155]

The first tool, called "Mass Testing," produces a comparative report displaying the legal framework (legislation and case-law) of the relevant national systems on a hypothetical legal case concerning a particular procedural right. This use modality seeks to support policy makers to evaluate the level of harmonisation of national legal frameworks within the EU acquis.

The second tool, termed a "Single Assessment," requires the user to provide concrete information concerning a specific legal situation, and, on that basis, produces a legal assessment concerning the potential critical legal issues. This use modality aims to support legal professionals in a first legal assessment of practical cases.

The third and last tool, the "Automated Reasoner Assessment," aims to assist users in developing and analysing the legal reasoning process of a given case. This tool represents a technological innovation in the application of computational models of legal reasoning to the domain of Criminal Law. The system is structured as a knowledge base containing legal norms and a reasoning module that applies rules to specific cases.

The user is presented with a list of questions, ranging from the personal information of the defendant to the stage and matter of proceedings, and all known answers should be given. The system takes into consideration the specific facts of the case and presents the solutions it infers. By giving a detailed assessment of the case, the tool automatically considers and analyses all the relevant European or national legal norms. Depending on certain answers to

[155] https://www.crossjustice.eu/en/index.html.

the form, the user may be presented with additional questions, related to the facts of the case already stated.

Internally, the system has a machine-readable model of the norms that are queried to give the automated assessment. The outcome provided by the tool includes all the steps of the legal reasoning, leading to a specific conclusion, as well as an automated assessment of the level of harmonisation of the national legislation with the relevant Directive.

The platform tools were tested by legal practitioners of several member States in the course of the research project and their feedback was used to refine the system. A User Manual was also created to propose a common methodology of legal training to better improve and implement the procedural rights enshrined in the EU Directives, combining traditional training techniques with potential use of the CrossJustice platform.[156]

## 3. The FACILEX project

Taking the lead from the previous project, the University of Bologna took the initiative to expand the research, adopting a similar approach to the matter of transnational cooperation among member States in the domain of criminal law. Starting in December 2022, the FACILEX project ("Facilitating mutual recognition: Analytics and Capacity building Information LEgal eXplainable tool to strengthen cooperation in the criminal matter") will last until December 2024 and is also funded by the EU Commission-DG JUST.[157]

The project aims to strengthen the implementation and application of the acquis on judicial cooperation in criminal matters through the help of digital tools, without being limited to fact-finding research. More specifically, FACILEX goes well beyond CrossJustice, in two ways.

On one hand, it changes the subject matter of the analysis, bringing the focus of the research to the dimension of horizontal cooperation and mutual recognition instruments that have heretofore not been examined previously. Namely, in light of the new Strategic Agenda 2019-2024,[158] the project focuses on the European Arrest Warrant (FD 2002/584/JHA), the European Investigation Order (Directive 2014/41/EU), and Regulation 1805/2018 on freezing and confiscation orders, that have been implemented or are operational in nine different member States: Bulgaria, Croatia, France, Germany, Italy, Poland, Portugal, Spain and the Netherlands.

On the other, the project also greatly improves the scientific methodology and approach, focusing on the innovative design and modelling of complex and multilayer legal norms in the field of Criminal Law cooperation and adopting novel legal informatics approaches for the development of online tool functionalities provided by the platform. Unlike other currently available legal information tools that only enable access to case law and legislation,

---

156 https://site.unibo.it/cross-justice/en/project-results/publications, the results of the project have been published in the collective volume G. CONTISSA, G. LASAGNI, M. CAIANIELLO, G. SARTOR (eds.), *Effective Protection of the Rights of the Accused in the EU Directives. A Computable Approach to Criminal Procedure Law,* Leiden, 2022.

157 https://site.unibo.it/facilex/en. Grant Agreement n. 101089634; the project was funded by the European Union's Justice Programme (2022).

158 EUROPEAN COUNCIL, *A new strategic agenda for the EU 2019-2024*, available at: https://www.consilium.europa.eu/en/eu-strategic-agenda-2019-2024/.

FACILEX will include an enhanced, AI-enabled advisory function providing accurate comparative analysis on the state-of-play of criminal cooperation across member States.

With regard to the online platform, FACILEX is expanding the CrossJustice project first by integrating the LegalDataBase Module, including the legal framework concerning the three chosen mutual recognition instruments, both at the EU and at the national level. Similarly to CrossJustice, all data will be available in English and also encompass the relevant EU and national case-law. Secondly, the project is developing a dedicated "Customized Single Test Advisory Module" focused on judicial cooperation, based on an explicit and computable representation of legal knowledge and reasoning, based on symbolic logic rules.

The tool is being designed to be user-friendly and accessible to legal experts. It is structured in such a way to provide a deeper level of explainability, thanks to the adoption of novel technological symbolic and sub-symbolic approaches. Building upon previous methodology, programmers aim to write the rules to be detailed and expressive, enabling a high degree of isomorphism in relation to source legal material. The goal is to provide legal experts, who do not have knowledge of any programming language, to understand, and even draft, the formal representation of legal rules.

An advanced user interface enables the inclusion and analysis of multiple information sources for the case under consideration, e.g. case law and opinions. Moreover, the system is able to engage in abductive reasoning (e.g. "what are the conditions for a specific legal outcome?"), helping legal professionals reason strategically about a legal case or procedure. The module provides legal advice in a straightforward way, i.e. by answering a step-by-step questionnaire designed to take users' inertia and a general lack of tech abilities into account.

Specifically, the system allows the user to define a case, with relevant information produced by the same user (e.g. the nationality of the defendant, issuing and executing States, grounds for refusal) and receive a customized legal assessment of potential cooperation flaws and available remedies in the specific legal system (e.g. if a specific ground for refusal can be invoked in the case at stake). The Module will draw upon all possible solutions depending on variable parameters regarding the basic information provided by the user. For instance, if the user indicates basic information such as: the nationality of the defendant, the issuing Member State (e.g. Bulgaria) and a ground for refusal, but not the executing member State, the Module will show all possible outcomes for each of the member States included in the project.

The system thus assesses the level of compliance of national legal systems with the relevant EU acquis. It also offers customized support to legal professionals in the first legal assessment of practical cross-border cases. Furthermore, the system enables the generation of mock cases, for academic purposes, including the possibility of stating an incomplete set of facts, and having the system fill in possible conditions, in order to show different possible solutions for research purposes.

Finally, the FACILEX integrated platform includes a dedicated Harmonization Mass Test Advisory Module, that provides a harmonization index for each member State. The harmonization index is a measure of proximity between the terminologies of European and national legislation. This enables legal professionals to assess the level of compliance of national law with the EU acquis in matters of judicial cooperation, as well as access to an automated evaluation of the level of harmonization of the terminology. As terms and definitions can assume different meanings in legal texts across the EU, this part of the research has been based on semantic analysis. Specifically, Natural Language Processing (NLP), an AI technique, has been employed to identify definitions in legal texts.

The system also contains relevant case-law from member States, used to assess the conformity between each State and the European Union. Case-based reasoning techniques are employed to evaluate the interpretation given by each State to relevant legal principles, in compliance with EU law. Additionally, the explanation will be enhanced with the application of Large Language Models (LLMs), used for translating the explanations produced by the rule-based system, from high-level programming language to natural language, allowing all stakeholders access to a fast, clear, and accessible interaction with the technology.

Such explanation involves verifying a set of conditions derived from the relevant legal text, whether it be European, national, or both. Human experts first translate these conditions into a programming language, which is then used to compute an outcome based on the information provided by the user. The result is passed to the LLM, which is prompted to generate a natural language explanation, including a summary of the case and a list of applicable legal conditions. Any LLM can be used at this stage (with the appropriate prompt), as the language used by the programmer is close enough to the language used by the legal text to enable the model to automatically recognize and translate the output of the system.

Testing sessions have been arranged to verify the development of the online platform, the quality of its legal knowledge, and advisory modules.

By accessing the FACILEX platform, legal professionals can receive a tailored legal assessment on potential cooperation flaws and available remedies, drawing upon all possible solutions relative to given parameters. The FACILEX project represents a valuable tool for streamlining EU judicial cooperation in criminal matters. By leveraging the full potential of digital technologies, the project fosters mutual knowledge, communication, and ultimately trust among the relevant actors in the Area of Freedom Security and Justice.

## 4. The way forward

The CrossJustice and FACILEX projects illustrate a way of rigorously and usefully integrating automated technology with the practice of criminal proceedings. These results, however, far from representing an end point, move the starting line forward for studies that are still to come. In this regard, there are several profiles that immediately reveal themselves as potentially relevant for further investigation.

First, the integrated analysis and reasoning methodology of the projects could find application in other aspects of Criminal Procedure Law, for instance concerning the rights of victims, also subject to EU harmonisation with Directive 2012/29/EU.

Second, the approach developed in these studies may be useful for the objective of evidence-based policy making, as it provides ways to analyse the impact of the law, supporting the legislator to understand gaps, inconsistencies, and interpretative issues that may hamper the effort of harmonisation. Such an approach may be extended to other areas of law influenced by EU harmonization and by a complex enforcement model, for instance in financial supervision.

Third, the symbolic logic approach to the development of technologies provides a decisive advantage over data driven approaches: the system grounds its reasoning on an explicit model of law and ensures transparency and the explainability of outcomes, enabling human users to integrate such outcomes into their decision-making activity. However, the integration of LLMs and other text analysis techniques with symbolic approaches shall be explored further.

Other relevant research lines have only been partially examined by present studies, but are acquiring a growing weight in shaping the effectiveness of procedural safeguards on a national and supranational dimension. For instance, the increasing deployment of digital technology requires a much deeper reflection on the integration between criminal procedure rules, remedies, and potential violations to the rights of privacy and data protection.

All these developments, however, only represent a part of what needs to be done to ensure swifter cooperation across the EU and a more substantial implementation of fundamental rights. The other side lays in practical, but essential aspects: financial support to the institutional framework and adequate training for all actors involved in the proceedings, from lawyers to judges, from prosecutors to law enforcement, and linguistic operators. Training programs, in particular, often struggle today to keep pace with technological development and the globalized context in which jurists must take their own steps. We hope this research, beyond its limits, can contribute to changing this educational paradigm, providing a methodological architype that permits an improved navigation of the increasingly complex, integrated, multidisciplinary, but also exciting environment that is our society.

## 5. Hands-on: Case Studies

In order to illustrate the functioning of the CrossJustice and FACILEX integrated platform, we highlight below the following case studies.

Case scenario 1. Focus on procedural safeguards (CrossJustice online platform)

An Italian citizen, Mr. X, is arrested in Bulgaria for fraud-related offences, on the basis of an order issued by Italian authorities. In Bulgaria, Mr. X is questioned by the police. This investigative act concerns several sensitive fundamental rights, such as the right to access to a lawyer, to silence, to be informed, and to an effective remedy.

For instance, an Italian lawyer defending Mr. X could aim to exclude the use of statements made by the arrestee to the Bulgarian police during an Italian trial for the fraud related offences, by claiming that:

1) in general, the Bulgarian legislation concerning the safeguards of the accused was not in compliance with the EU acquis (the six Directives on procedural safeguards) and thus a violation of such safeguards occurred in the specific case; or

2) although generally in compliance with the EU acquis, the procedural rights recognized in that specific case by the Bulgarian authorities were not compatible with those recognised by the Italian law in the same case.

To verify such assumptions, the lawyer needs to know the applicable Bulgarian legislation, as interpreted by the local courts, and evaluate both its compliance with the EU acquis on procedural safeguards and its compatibility with Italian procedural rights recognised in the same case.

To this end, the lawyer may first use the "Mass Testing" function, to obtain an overview – in English – of the applicable Bulgarian Law.

The lawyer may then refine the assessment by using the "Single Assessment" tool. Providing a limited set of information, such as the age of the person, the country involved, and the phase of the proceedings, the platform will generate a list of the applicable rights in the specific case.

Finally, the lawyer can obtain an automated and more refined case analysis by making use of the "Automated Reasoner Assessment" tool by providing more information concerning the specific case to the system. Such information includes, for instance, the national legal systems involved (Italy and Bulgaria); whether the accused is a minor or an adult; where the trial is taking place (Italy); what investigative act is under assessment (suspect interview); where the act took place (Bulgaria); the phase of the proceedings (European Arrest Warrant request); and the language spoken by the person (Italian). On the basis of the information provided, the system will automatically make a tailored assessment concerning the compatibility of Bulgarian law applied to the specific case with the EU acquis and the Italian procedural rights at stake.

The assessment report will include an explanation of the reasoning carried out by the platform, and references to relevant legislation that justify the answer.

Case scenario 2. Focus on cooperation mechanisms (FACILEX online platform)

A Polish prosecutor needs to obtain documents in the possession of a lawyer located in Italy, in order to produce evidence at a trial in Poland. To do so, a European Investigation Order is issued, requesting that Italian authorities collect the documents. Following the cooperation request, the Italian authorities will examine whether to execute the EIO or to refuse it, on the basis of client-attorney privilege.

To foresee the potential outcome of the cooperation request, the Polish prosecutor can assess whether there are specific grounds for refusal that could lead the request to be opposed by the Italian authorities. In order to do so, the "Harmonization Mass Testing" function may be first used, to obtain an overview – in English – of the applicable Italian and Polish Law and to obtain a harmonization index for each member State involved. This allows the prosecutor to generally assess the conformity between each involved State and the European Investigation Order Directive.

Then, the Polish prosecutor can obtain an automated and more refined analysis by making use of the "Customized Single Test Advisory Module" tool. More information concerning the specific case must be provided to the system. Such information includes, for instance, what are the involved national legal systems (Italy and Poland); what investigative act has been requested (confidential documents production); and what rights are at stake (attorney-client privilege). On the basis of the information provided, the system will automatically make a tailored assessment concerning the likelihood that grounds for refusal are invoked. The system will identify what grounds for refusal might be invoked by the executing authority, in light of the interpretation given in Italy to such grounds by case-law (e.g. the investigative act is not available in a similar domestic situation).

The assessment report includes an explanation of the reasoning carried out by the platform and references to relevant legislation that justify the answer, highlighting the potential weaknesses of a cooperation request that should be addressed (where possible) by the requesting authority.

# 5.2. Analytics for Deciding Legal Cases: The ADELE Project

*Federico Galli, Alessia Fidelangeli, Piera Santin, Galileo Sartor – Università degli Studi di Bologna*

*Summary: 1. The ADELE project – 2. Background – 3. Objectives – 4. Data collection and annotations – 5. Citation extraction and network analysis – 6. Ontology framework – 7. Summarization and keyword extraction – 8. Argument extraction – 9. Outcome prediction – 10. Validation and feedback – 11. Ethics*

*Abstract: This contribution describes the ADELE research project, funded by the EU, that aims to use legal analytics to perform various activities in relation to different corpora of legal decisions belonging to different domains and written in different languages. The purpose is the implementation of tools and a methodological framework capable of being further transferred in other relevant fields.*

## 1. The ADELE project

The ADELE Project ("Analytics for DEcision of LEgal cases") was co-funded by the European Union's Justice Programme (2014-2020). The project ran for 30 months from February 1, 2021 until July 31, 2023.

It was developed in cooperation with the Alma Mater University of Bologna, the Bulgarian legal informatics company APIS Europe, the University of Turin, the Centre for Judicial Cooperation of the European University Institute of Florence, the University of Luxembourg, the Bulgarian LIBRe Foundation, and the Bulgarian Union of Jurists.

The project developed Artificial intelligence (AI) and legal analytics (LA) methods to support legal research and decision-making processes in the judiciary. The project focused on two legal areas: value added tax (VAT) and trademarks and patents (T&P) and analysed both national and EU case law and legislation. It engaged in multiple tasks, such as the annotation of legal decisions, keyword and summary extraction, knowledge representation, network analysis, judicial argument extraction, and outcome prediction.

The project's main achievement was the development of an AI pilot tool that offered various functionalities and was embedded within an open platform available on the web.[159]

This contributions presents the different activities carried out during the project below.

## 2. Background

ADELE was built on a paradigm change in AI: techniques of knowledge representation and logical inference which are complemented by machine learning as applied to large datasets (Big Data). The application of ML techniques today enables the discovery of correlations, often yielding unexpected insights, and is giving rise to many new applications: conceptual retrieval, speech and image recognition, question-answering, translation, planning, etc. Such applications are transforming many aspects of life and delivering many benefits, i.e. in the fields of healthcare, commerce, and transportation.

[159] https://adele-tool.eu.

The way law is practiced is also being affected by AI as new techniques unfold for legal cognition and practice.[160] The emerging field of legal analytics (LA) for example develops applications in the legal domain to extract legal knowledge, infer undiscovered relations, and engage in data-driven predictions.[161]

So far, legal analytics applications have been developed in the following domains:

- the identification and representation of legal knowledge, defining types of annotations, concepts, links and their integration with conceptual ontologies;
- machine learning with legal texts, managing complex unstructured datasets, classifying paragraphs of texts, extracting legal rules, comparing rules across jurisdictions;
- the extraction of arguments from cases, including argument structure, claims and substantive legal factors;
- the connection of computational reasoning models and legal texts, annotating obligations in privacy policies, delivering Q&A services, making legal predictions in IPR cases, and providing visual maps linking cases and concepts.

Many applications have been developed as private-sector commercial initiatives for lawyers or legal organisations.[162]

However, in the past few years, we have seen an increasing interest in developing and using AI in the judiciary and the CEPEJ launched an initiative to collect data and information on current applications in Europe.[163] This initiative resulted in an open database where helpful information can be found, i.e. the year the system became functional, a short description including the underlying technology, a link to an official public source/reference of the system, and the status of the system, i.e. whether the system is currently functional or is still in the pilot phase. The database includes applications that are strictly dedicated not only to domestic courts, but also to lawyers and law firms. Different categories of users have been identified, such as court users (general public), court management, judges, lawyers and prosecutors.

Several useful applications are being developed in this field, for example:

- management tools in administration activities, from setting up calendars of meetings and hearings to the more sensitive issue of allocation of cases depending on the topic and relevance;
- information retrieval, for instance, in the recognition of patterns in text documents and files, as eDiscovery, or argument extraction from cases;
- a triage function, helping judges to quickly assess the relevance of previous cases to the present decision, such as court cases to verify litigation risk assessment, or simply gathering relevant information for claims, such as the solution explorer at the Civil Resolution Tribunal;
- citation retrieval and management;

---

[160] R. SUSSKIND, *Tomorrow's Lawyers*, II, Oxford, 2017.

[161] K. D. ASHLEY, *Artificial intelligence and legal analytics: new tools for law practice in the digital age*, Cambridge, 2017.

[162] Among the most famous applications, we can cite Casetext, LexisNexis (incl. Lex Machina, Ravel Law), Premonition, ROSS Intelligence in the US; Case Law Analytics, Doctrine, Predictice for France; CourtQuant, Justis, Solomonic in the UK.

[163] Resource Centre on Cyberjustice and AI, https://www.coe.int/en/web/cepej/resource-centre-on-cyberjustice-and-ai.

- document-assisted generation systems, where the software automatically generates text that can help judges write judicial documents;
- speech-to-text applications, allowing the judge to receive the transcript of the hearing or courtroom record;
- risk prediction systems, where the algorithm is able to forecast a possible outcome of the case or an aspect related to it, which include compensation and litigation fees.

Policy-makers worldwide are increasingly recognizing the potential of these solutions and have raised government funding for initiatives and projects aimed at integrating AI into the judiciary.

## 3. Objectives

Against this background, the objectives of the ADELE project were to design a methodological framework of legal analytics (LA) for court decisions and to implement it in a pilot tool designed for Italian and Bulgarian case law in the fields of Intellectual Property Rights (IPR) and Value Added Tax (VAT).

The first objective was accomplished by designing a LA methodology consisting of: (1) a data collection and annotation phase, whereby relevant knowledge was added as metadata or machine-readable information to a particular document and integrated with legal ontologies; (2) a data processing phase, whereby machine learning and NLP techniques were applied to annotated legal texts to extract relevant information, i.e. arguments, claims, citations, sections, keywords, and summaries.

The aforementioned LA framework allowed for the development of the following functionalities:

- ontology-based searches, which provide a search function through a comprehensive view of the domain-specific maps with a representation of the most relevant concepts and their connections;
- citation extraction and network analysis, which enables judges to collect useful insights from previous judgments that are conceptually or functionally similar and to get a complete overview of the operative construction of the law;
- summary and keywords extraction, enabling the visualization of legal information to quickly understand the main factual and legal issues discussed in the decision;
- argumentation extraction, which allows the extraction of argumentative patterns from a decision. This application enables judges to establish the rationale of previous judgments and offers a conceptual and argumentative toolkit to make the final decision in the case at hand;
- outcome prediction, that enables the anticipation of what could be the decision to a specific case according to past case law.

In the following paragraphs, the methodology and practical steps taken to create the above-mentioned functionalities are discussed, starting with data collection and annotation, and then specifying how the data was used for each task presented above.

## 4. Data collection and annotations

Several judicial decisions were selected to create the source corpora for processing activities. In particular, approximately 500 decisions were collected between Italy and Bulgaria, both in the field of VAT and Trademark and Patents.

As a general rule, judicial decisions containing personal data were collected anonymously. For instance, both Bulgarian corpora were prepared by selecting anonymized decisions from public data sources. When possible, Italian corpora were also annotated by selecting decisions that were already anonymized. In cases where judicial decisions included personal data, anonymisation techniques were employed, accounting for the risk of data breaches balanced against the purposes of processing. In particular, any identification data of natural persons involved in proceedings were immediately anonymised. The names of legal persons were also anonymized as a precautionary approach, especially when the company name referred to natural-person entrepreneurs. In the Trademarks and Patents dataset, trademarks and patents were not anonymized – even when they refer to natural persons who are entitled to the relevant protection – as their name is often crucial for acquiring information about the dispute, which can also be important for machine learning purposes.

During the collection processes, we encountered different problems related to data accessibility, especially regarding Italian case-law. In particular, issues pertained to: (1) the multiplicity of databases through which decisions are accessible and whether a specific database includes all decisions or just a number of them; (2) the limited access to subscription-based databases or those restricted only to judicial professionals; (3) the lack of search filters in the database, which hampers the granularity of the search and thus the homogeneity of the created corpus; (4) the non-digital-native nature of documents and the machine-unreadability of formats used.

After the collection of data, a preliminary analysis of case-law was carried out to understand the recurrent structure of the decision, the type of language used, the recurrent argumentative patterns, the citation styles and overall, the content of the decisions.

Based on the aforementioned analysis, a set of annotation guidelines was developed. These are a set of rules or instructions defining the criteria legal experts use to annotate the text. They define what should be included in each annotation and can also provide a structure for annotation, such as a template. In particular, the annotation guidelines were developed to annotate the structural elements of decisions and their relationships and to annotate judicial arguments found in the court's motivation. These annotations were used for the automated extraction of arguments and outcome prediction. Citations were not annotated in the text, as a more traditional and less time-consuming method was adopted through regex.

Based on annotation guidelines, a corpus annotation was carried out in an incremental fashion. Annotation is the process of adding metadata or additional machine-readable information to a particular data or document. The annotation was carried out by experts in the analysed legal fields and consisted of the insertion of additional information within the texts of the analysed judgments. A part of the documents was tagged through a double-blind method, where two annotators were asked to tag the same documents and required to agree on the use of tags. Inter-annotator agreements were also measured to ensure the soundness and coherence of the datasets, i.e. to avoid the same legal information being treated differently by different annotators.

The corpora were annotated using XML text editors, including Sublime Text, NotePad++ and Visual Studio Code. Different annotation software (such as Gloss and Inception) was also experimented with and some developing with SenTag was also begun. This provided an environment for annotating and editing case law using XML tags, with several customized functionalities such as a semi-automated guided annotation procedure, a visualization space for the argumentative structure of decisions, and an automated agreement procedure.[164]

The annotated case law was contained in the ADELE platform as searchable and browsable case-law in which annotations can be highlighted through information retrieval functions. In addition to that, decisions were used to train the machine learning models for outcome prediction and argument mining, as well as to extract case law citations.

## 5. Citation extraction and network analysis

Given the large availability of data, the automated analysis of legal texts has become increasingly relevant in recent times. Network analysis, in particular, is being employed to analyse intricate legal domains by representing legal documents (judgements, statutes, regulation) in terms of nodes, i.e. cases, legislative documents, and their corresponding relationships or edges, such as citations and normative referrals. A network is particularly useful in understanding the overarching structure of a domain, allowing for the identification of specific characteristics, including the case most frequently cited, clusters of similar cases, which contribute to predictive analysis.[165] While most experiments in this field have been applied to EU case law, there are still ongoing efforts to apply the same methodology to national courts.[166]

In the context of the ADELE Project, network analysis in the field of VAT and T&P cases was carried out in Italy and Bulgaria. The pipeline included the automated extraction of cases and then the development of networks.

The automated extraction was performed by using a regular expression (regex) algorithm, namely a computational procedure used for matching patterns in strings based on pre-defined language rules. To do that, the first step was to manually detect judicial citation patterns, i.e. how judges usually refer to and cite other documents within a case. It was noticed, for example, that most Italian cases did not follow a common style of citations, as the same document could be referred to in multiple ways. For instance, the court identifier could be written fully, shortened, or, in some cases, skipped completely. The same occurs for the date. Through trial and error, we attempted to build regular expressions as high-level as possible to accommodate the automated extraction of all possible citation styles. For the purpose of the project, only citations in specific parts of the original decision text were searched, to include only the citations that were relevant in decision making, in particular those made by the judge.

---

164 A. ZERBINATI ET AL., *SenTag 2.0: A Cooperative Annotation Tool*, in *Proceeding of the 21st International Semantic Web Conference*, CEUR Workshop Proceeding, 2022.

165 M. DERLÉN, J. LINDHOLM, *Goodbye van Gend en Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments*, in *European Law Journal*, 2014, 20.5, 667.

166 A. LOUIS, G. VAN DIJCK, G. SPANAKIS, *Finding the Law: Enhancing Statutory Article Retrieval via Graph Neural Networks*, January 2023, available at: arXiv:2301.1284; R. WINKELS, J. DE RUYTER, *Survival of the Fittest: Network Analysis of Dutch Supreme Court Cases,* in M. PALMIRANI ET AL. (eds.), *AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents*, Lecture Notes in Computer Science, Berlin, Heidelberg, Springer, 2012, 106.

Once extracted, the citations were converted into a pre-established format: [Court][Branch][Date][Case number][Text section]. This made it possible to maintain a standard structure, with different elements directly retrieved in the text of the parsed document. A similar structure was extracted for legislative citations, while keeping the different legal sources in consideration. Unfortunately, citations were extracted only at the document level. This was due to the fact that judges, at least in Italy and Bulgaria, rarely refer to specific paragraphs of cited decisions in their judgments. For this reason two classes of citations could be identified: complete well-formed citations and incomplete ambiguous citations. Only complete citations permitted searching for the document and extracting additional information on the case, such as additional cited documents. However, in the Italian dataset, cited cases were often not publicly available. Unlike in EU law datasets which had been worked on previously,[167] this grossly limited the possibility of extracting complex networks of citations.

To enhance the output of the system, existing solutions were implemented that could be tailored to our use. For the Italian dataset in particular, the Linkoln system was used, a piece of open-source software developed at IGSG-CNR (*Istituto di Informatica Giuridica e Sistemi Giudiziari del Consiglio Nazionale delle Ricerche*) for the automated detection and linking of legal references contained in legal texts written in Italian.[168] With the capabilities of the library, which was itself built on the same methodology used (common patterns, regular expressions), it was possible to identify most citations and fix those that were still incorrect by modifying its output. Another module was then developed to check online sources for the document, validate the metadata extracted, and add more if necessary.

At this point, the extracted citations were ready to be converted to a graph structure and analysed further. In particular, the structure was imported into a graph that could be visualised. The result, while not very deep, still showed recurring patterns and groups of cases with common citations.

[167] G. SARTOR ET AL., *Automated Extraction and Representation of Citation Network: A CJEU Case-Study*, in R. GUIZZARDI, B. NEUMAYR (eds.), *Advances in Conceptual Modeling*, Lecture Notes in Computer Science, Cham: Springer International Publishing, 2022, 102.

[168] The software can be found at: https://gitlab.com/IGSG/LINKOLN/linkoln/

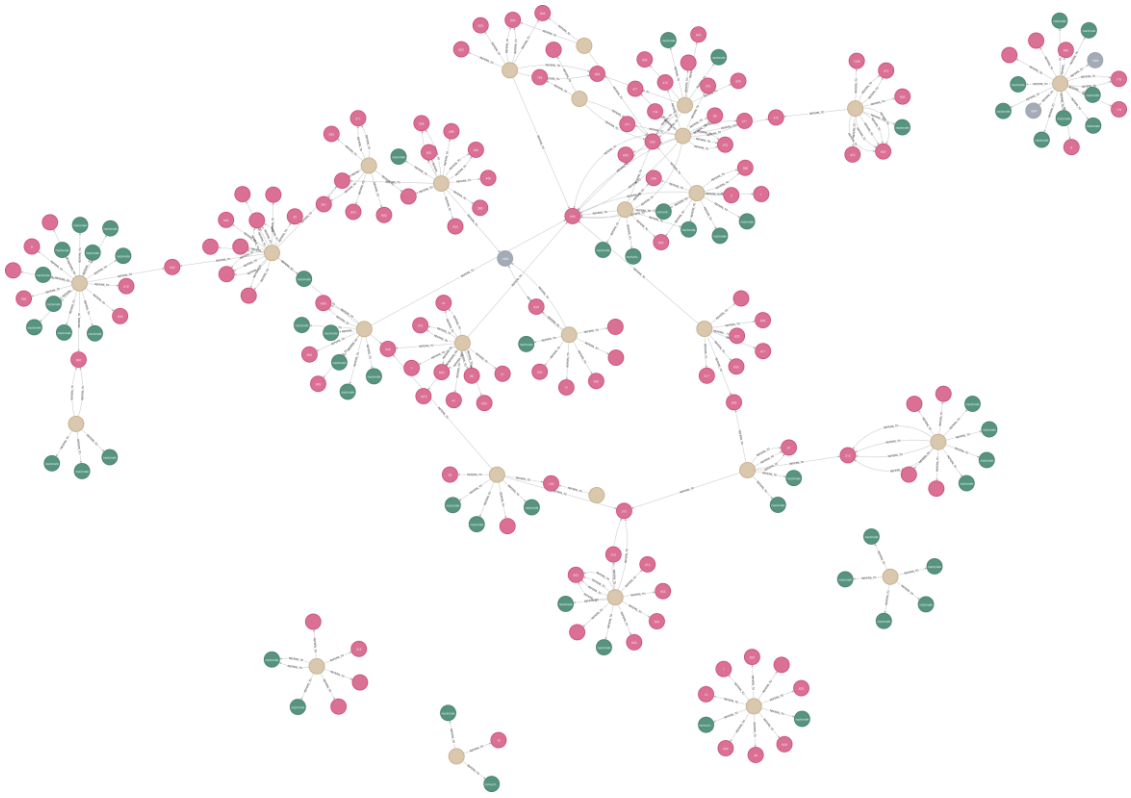


Figure 1. Section of the citation network

As we can see in

Figure *1*, the ADELE cases, shown in light brown, are linked to the cited case law (pink) and legislative documents (green). We can clearly see that patterns, or clusters, can be identified and, even with incomplete data available, this is already a good result. The clusters have citations in common, and the citations that have many incoming edges, i.e. that are cited by many different cases, could be more relevant than others.

The graph can be further analysed with degree centrality metrics, measuring the number of incoming relationships for the citation nodes in the graph. With this information, a list of the more relevant (cited) cases was compiled and divided into four categories: national case law, European case law, national legislation, and European legislation.

This analysis could be enhanced further, with more information available, but as it is now, it can help identify the important cases or legislative documents, particularly when integrated with the ontology, that will be described in the following section, by linking the concepts in the ontology to the relevant legal documents and adding this information into the citation network.

## 6. Ontology framework

Ontologies are formal representations of a specific domain's knowledge, which provide a structured and organised way of describing concepts and their relationships. They are used in AI, semantic web, natural language processing, and knowledge management systems to represent knowledge in a machine-readable and structured form. Such representations can then be used for information retrieval (enhancing the search and retrieval of legal

documents), facilitating the interoperability between different systems, to support automated reasoning and decision-making processes, and enabling the development of intelligent applications.[169]

In the ADELE project, two ontologies pertaining to the selected domain were built: (1) the OntoVAT, containing concepts relating to VAT taxable/exempt transactions 170 and (2) the PaTrOnto, integrating primary trademark and patent concepts, among which owner, validity, and transfer.[171]

The aim of the ontologies was twofold: (1) organizing and visualizing domain knowledge and (2) automatically linking the judgements in the dataset to the most relevant domain concepts.

The first of those aims was achieved by embedding a tool that provides the user with a comprehensive overview of a conceptual map of the respective legal areas in the ADELE platform. Each of the two ontologies offered a structure of terms with the respective legal definition provided in the legislation or case-law. Links between concepts are also built to allow the identification of synonyms and related terms between national and European legal concepts in the relevant field. To build the ontology, the most important concepts based on a twofold approach were selected: (i) top-down, starting from the pertinent legal sources, i.e. Directive 2006/ 112/CE for VAT, and (ii) bottom-up, based on the most recurrent concepts contained in the collected Italian and Bulgarian VAT decisions. Based on this analysis, a team comprised of lawyers and computer scientists provided a seminal representation of the domain's notions for each element of the ontology. Next, legal experts identified definitions, examples, synonyms and related concepts, as well as the relevant European and national legislation in which the concepts were mentioned or defined and the most common examples instantiating that concept. For this reason, a multilingual OWL ontology enriched with a SKOS lexicalization was built and implemented in English, Italian, and Bulgarian. Finally, the gathered results were validated by the legal team that returned them to the technical team, who implemented the new information in the ontology. These last three steps were then iterated several times to refine the ontologies.

Different approaches were used, instead, to achieve the second aim, namely, to link judgements in the dataset to the most relevant domain concepts. The first layer was represented by the link between the ontological concepts and the citation automatically extracted from the datasets. The pertinent legislation associated with each ontological concept was then benchmarked with the citation extracted from cases. If the decision in the dataset cited one or more legal provisions contained in the ontology, it was connected to the related ontological concept. This means that if a decision cited a legal provision, and that legal provision contained a definition of a concept, then the case was marked as "related to" a certain concept. This link was based on the assumption that if a decision cited one or more

---

169 P. CASANOVAS ET AL., *Introduction: Theory and methodology in legal ontology engineering: Experiences and future directions*, Springer, 2011.

170 D. LIGA, A. FIDELANGELI, R. MARKOVICH, *OntoVAT, an ontology for knowledge extraction in VAT-related judgments*, in *New Frontiers in Artificial Intelligence: JSAI-isAI* 2023 Workshops, AI-Biz, EmSemi, SCIDOCA, JURISIN 2023 Workshops, Hybrid Event, June 5–6, 2023, Revised Selected Papers, Springer, 2024.

171 D. LIGA, D. AMITRANO, R. MARKOVICH, *PaTrOnto, an ontology for patents and trademarks*, in *New Frontiers in Artificial Intelligence: JSAI-isAI 2023 Workshops,* AI-Biz, EmSemi, SCIDOCA, JURISIN 2023 Workshops, Hybrid Event, June 5–6, 2023, Revised Selected Papers. Springer. 2024.

legal provisions containing the definition of a legal concept, then the decision was also relevant to that concept.

A more refined NLP pipeline is also in the process of being implemented which uses the two ontologies to determine whether an ontological concept is relevant for a specific judgement, i.e. if a specific decision deals with one or more of the ontological concepts. For this reason, mixed techniques were used, relying on: (1) supervised learning based on expert annotations and (2) unsupervised learning based on the ontology content itself. Regarding the first, legal experts were asked to select concepts from the ontology they considered more relevant to the cases collected in the dataset. Then, legal experts were asked to manually annotate nearly 70% of the judgements by including the information of whether each selected concept was relevant in each judgement by associating a binary value, where "0" was "non-relevant" and "1" was "relevant." The concept was considered relevant if the decision of the court concerned the concept from a substantial point of view. A designed algorithm then encoded the information contained in the ontology to predict whether or not a was relevant, comparing the results with the gold standard defined in the previous step.

Regarding the second step, the algorithm used synonyms and related concepts in the ontology to provide an expanded knowledge base to run NLP techniques which were able to link the ontological concept to case-law. For instance, the presence of definitions, examples, and related terms permitted ontological concepts like "supply of goods" to be linked to judgements in which the "transfer of ownership of a specific object" was mentioned without some explicit reference to the term "supply of goods."

The ontologies built in the ADELE project can also be used for multiple purposes. For the organisation of judgements according to structured taxonomies that are more precise than those currently in use for example, which could help database searches for work and study purposes; or the creation of intelligent mind maps that allow for easy access to specific topics on a subject matter with conspicuous references to other areas of Law such as taxation; or for the extraction of keywords in judgements that are particularly homogeneous from a lexical point of view; or as a navigation and visualisation tool in intelligent databases.

## 7. Summarization and keyword extraction

Providing a set of relevant key terms (keywords) and summaries facilitates the retrieval of legal information by quickly assisting users to understand the main factual and legal issues discussed in a particular case without reading its text. In the context of ADELE, only extractive, not abstractive, summaries were worked with. Extractive summarisation selects the most meaningful sentences in the input text and combines them to form a summary. No change was made to the textual content of the extracted sentences.[172]

The automated key term extraction was created based on a trained spaCy model based on Named Entity Recognition (NER). Two specialised dictionaries were used for the training of the dataset, which contained legal terms in the relevant fields of Law – one in the field of VAT and one in the field of TM&P. Both were compiled by legal experts. In addition, a few more general law-related dictionaries, such as EuroVoc, were exploited. The applied model labels the key terms recognized in the text of a decision in two categories: the terms from the specialised dictionaries were labelled as priority key terms, whereas those contained in other

[172] Abstractive summarisation on the other hand generates new text which aims to provide a synoptic statement of the content of input documents, without reproducing their wording.

dictionaries were labelled as non-priority key terms. The terms that the trained model classified as priority key terms were directly included in the final selection, whereas the KeyBERT library[173] was used for the non-priority keywords to extract only those keywords that were most relevant to the processed document. However, before using KeyBERT, additional filtering was applied for non-priority key terms. For instance, only non-priority key terms that had at least two occurrences within the text of the processed document were included.

Similar to automatically extracted key terms, summaries were extracted following a supervised learning approach. The experiments with text summarisation began by using the TextRank model,[174] but the results were unsatisfactory. It was therefore decided to train a spaCy model[175] on a dataset of about 40,000 Bulgarian court decisions summarized manually by legal experts. The model was based on sentence text categorization, whereas the aforementioned manually annotated data were used to generate a training dataset containing two categories of sentences: summary and not_summary. The applied approach led to a significant improvement in results for the Bulgarian decisions. For Italian decisions, initially, the TextRank model was also attempted. Since the results were unsatisfactory as well, the trained "Bulgarian" spaCy model was also used, as it was based on a multilingual transformer. In this way, far better results were achieved.

In the ADELE tool, key terms and sentences are displayed under the title of judicial decisions, in the lists of documents corresponding to the results of a user's search, as well as in an opened document. Depending on the length of the decision, between 5 and 15 key terms are usually shown.

## 8. Argument extraction

The argument extraction module sought to identify arguments and their features. For this purpose, we deployed and refined state-of-the-art techniques in argument mining, i.e. the process of automatically detecting arguments from natural language texts to identify and analyse their structure and content.

Experiments were initially developed for and applied to CJEU decisions in the field of European Tax Law, specifically on fiscal State aids appeal decisions, and then were tested on national VAT and T&P datasets.[176]

Within CJEU decisions, sections describing the Court's reasoning which led to the final ruling were focused on in particular.[177] For the annotation process, sentences independent of one another were considered as tagging units. If a sentence contained more than one argument, usually separated by semicolons, each argument was considered and labelled separately. Two main elements were identified in particular: 1) the premise(s) and 2)

[173] M. GROOTENDORST, *KeyBERT: Minimal keyword extraction with BERT*, in *Zenodo*, 2020.
[174] https://cran.r-project.org/web/packages/textrank/vignettes/textrank.html.
[175] https://spacy.io/usage/models.
[176] Given the complexity of the task at hand, the decisions of the CJEU provided two advantages: (1) the standard (although not fixed) structure of CJEU decisions, in which complex and highly variable arguments are embedded; and (2) the English language, which prevented possible issues deriving from linguistic differences between Bulgarian and Italian judgments, possibly resulting in different argumentation structures (which may hinder a harmonised annotation process).
[177] These sections are usually denoted as "Findings of the Court."

conclusion. When an argument consisted of multiple steps, the conclusion of each step was the premise for the next, so each intermediate step was labelled as a premise and only the last step as a conclusion.

In the annotation, some mandatory or optional properties were also attributed to each part of the arguments. For example, a distinction was made between legal and factual premises, the former on applicable law or relevant case-law and the latter related to statements of facts of the case. Legal premises were also characterised by argumentation schemes, which identified the stereotypical structure of an argument.[178] The following classes of argument were included: authoritative, verbal classification, interpretation, literal interpretation, precedent, principle, intention of the legislator, rule, systematic interpretation, and teleological argument.

An additional layer of complexity pertained to linking each premise to other premises or the conclusions through a logical connection. These were: (1) support, indicating a support relationship between a set of contiguous premises or between a premise and its conclusion; (2) attack, indicating either a rebuttal or an undercutting relation; (3) rephrasing, indicating that a particular premise was entirely rephrased by another premise, both having the same semantic meaning, even though they were differently formulated.

From a machine learning point of view five tasks were defined: (1) argument detection, i.e. classifying a sentence as premise, conclusion, or neither; (2) argument classification, i.e. classifying a sentence that is known to be argumentative as premise or conclusion; (3) type classification, a multi-label classification problem where a sentence that is known to be a premise is classified as legal and/or factual; (4) scheme classification, i.e. a multi-label classification task where a sentence, known to be a legal premise, is classified according to its argumentative scheme; and (5) link prediction, where two inputs, i.e. the source and the target that belong to the same document and are known to be argumentative, predict whether there is a link from the source to the target.

For tasks 1-4, we adopted three different representations of the input text: TF-IDF, Sentence-BERT (SBERT),[179] and Legal-BERT.[180] A set of traditional machine learning techniques that have a low computational footprint were chosen as classifiers, i.e. Linear SVC, Random Forest, and K Neighbors. For task 5, the neural architecture comprising stacks of dense and Long Short-Term Memory (LSTM) layers , a co-attention module, and residual connections were experimented with.[181] Moreover, 300-dimensional GloVe pre-trained embeddings were used for text representation. The model was also designed to encode the distance between the source and the target as a 10-bit array and a complete description of techniques and results was presented in two previously published papers.[182]

---

[178] D. WALTON, C. REED, F. MACAGNO, *Argumentation schemes*, Cambridge, 2008; D. WALTON, F. MACAGNO, G. SARTOR, *Statutory interpretation: pragmatics and argumentation*, Cambridge, 2021.

[179] N. REIMERS, I. GUREVYCH, Sentence-bert: Sentence embeddings using siamese bert-networks, 2019, available at: arXiv preprint arXiv:1908.10084.

[180] I. CHALKIDIS ET AL., *LEGAL-BERT: The muppets straight out of law school*, 2020, in *arXiv preprint* arXiv:2010.02559.

[181] A. GALASSI, M. LIPPI, P. TORRONI, *Multi-task attentive residual networks for argument mining*, in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[182] G. GRUNDLER ET AL., *Detecting Arguments in CJEU Decisions on Fiscal State Aid*, in *ArgMining 2022: 9th Workshop on Argument Mining*, 2022, 143; P. Santin et al., *Argumentation Structure Prediction in CJEU Decisions Fiscal State Aids*, in *Proceeding of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023)*, June 19–23, 2023, Braga, Portugal. ACM, 2023.

## 9. Outcome prediction

The number of studies where machine learning and NLP techniques have been applied to judicial decisions to predict the outcome of cases has greatly increased in the past several years. Most of these studies adopt judicial decisions – or features extracted therefrom – and their related outcomes as examples for training machine learning classifiers. Roughly three kinds of approaches have been identified:[183] (1) approaches based on features unrelated to the merits of the case;[184] (2) approaches based on legally relevant factors;[185] and (3) approaches based on the textual description of a case.[186]

In the ADELE project, the third approach was followed by applying machine learning and NLP techniques only to a portion of cases, i.e. the requests of parties (request, claims, and arguments). Based on a request, the task was to predict the likely outcome of a decision on that specific request. Questions were raised about whether such a task could actually be described as a "prediction"[187] since the system was trained to "predict" the outcome of decisions that were already made. If correctly trained, the system should be able to classify decisions based on outcomes. This task was mainly useful for identifying predictors like facts, arguments, and judges of court decisions within the text of judgements, i.e. possible correlations between the requests of parties and the court's decision to uphold or reject such requests. However, whether such detected predictors could be used to anticipate future cases based on new requests was controversial and subject to discussion.[188] The aforementioned task was realised in ADELE using supervised learning, in which the system was provided with the relevant annotated text together with the corresponding outcomes.

The annotation guidelines reflected the structure of the decisions and their portions. The guidelines focused on the identification of the following elements in partiuclar: (i) the parties, (ii) the related requests, claims, and arguments; (iii) the Courts' motivation, and (iv) the final decisions. Such information could be of different lengths and details and was often enclosed within the same portion of text. For this reason, hierarchical levels of annotation were identified. For example, the section containing the parties' submissions (denoted with a <partreq> label) included (a) specific requests (denoted with a <req> label), i.e. the judicial measures sought by the party or the counter-party; (b) the claims (denoted with a <claim> label), i.e. the ultimate reasons for grounding a request, usually supported by premises and which may have concerned procedural or substantive facts; (c) arguments (denoted with a <arg> label), i.e. statements that supported or attacked a claim and which could be legal or factual. The annotation guidelines were developed with the aim of making them applicable

---

[183] F. GALLI, G. SARTOR, *AI Approaches to Predictive Justice: A Critical Assessment*, in *Humanities and Rights Global Network Journal* 5.2, 2023

[184] D. M. KATZ, M. J. BOMMARITO, J. BLACKMAN, *A general approach for predicting the behaviour of the Supreme Court of the United States*, in *PloS one*, 2017, 12.4, e0174698.

[185] K. D. ASHLEY, S. BRÜNINGHAUS, *Automatically classifying case texts and predicting outcomes*, in *Artificial Intelligence and Law*, 2009, 17.2, 125.

[186] N. ALETRAS ET AL., *Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective*, in *PeerJ Computer Science*, 2016, 2, e93.

[187] M. MEDVEDEVA, M. WIELING, M. VOLS, *Rethinking the field of automatic prediction of court decisions*, in *Artificial Intelligence and Law*, 2022, 1.

[188] M. MEDVEDEVA, P. MCBRIDE, *Legal Judgment Prediction: If You Are Going to Do It, Do It Right*, in D. PREOTIUC-PIETRO ET AL. (eds.), *Proceedings of the Natural Legal Language Processing Workshop* 2023, Singapore: Association for Computational Linguistics, December 2023, 73.

not only to the specific domains of he projects, but possibly to all legal domains, with minimal adjustments. Moreover, they were designed to be applicable to different legal systems.

Afterwards, NLP and machine learning techniques were applied to the texts of decisions. The aim was to establish whether, based on the requests and arguments of the parties, the system was able to predict the outcome of individual requests and with what margin of error. Two representations of the input text were adopted: TF-IDF and Sentence-BERT (SBERT). As classifiers, Linear SVC, SVC, Random Forest, Gaussian Naïve Bayes, and K-Neighbours were chosen. The F1 score obtained for each class and their macro-average were also measured. The task of determining the decision outcome based only on the claims and arguments of the parties reached a maximum score of 0.68 with Linear SVC and SBERT. The introduction of the motivation and decision sections gave conflicting results, improving some classifiers but worsening others. A complete description of techniques and results was presented and discussed in a previously published paper.[189]

The results obtained were in line with those of other projects, but insufficient to imagine a real use in application hypotheses. Moreover, the experiments concerned judgments that had already been written and, thus, in order to assess any further application, a first assessment of the appropriateness of the use of the methodology adopted for the acts of the parties would be necessary.

## 10. Validation and feedback

Besides relying on common validation techniques for machine learning, testing and validation events with judges were planned. Two series of events were organised in particular to test the beta version of the platform and to validate the final version. The events were organised in parallel in Italy and in Bulgaria, with representative judges from the two selected domains. The two national validation events were organized to assess the technical and user-operative capacity of the final version of the tool. The Italian validation event was held on May 12, 2023, as a one-day event, at the European University Institute in Florence. The 15 participants were all Tax Law judges working in different jurisdictions and at different levels of proceedings. The Bulgarian validation event took place on May 17, 2023, in a hybrid mode. 50 participants took part in it, of which 16 were judges and judicial assistants. Lawyers and tax consultants also showed great interest in the event.

During the events, the features of the ADELE pilot tool were demonstrated using examples from the fields of VAT and TM&P Law. After that, a moot case was presented and participants were required to solve it by using the different platform functionalities. Finally, a discussion was held with participants in which they shared their thoughts and experiences regarding the use of AI systems in the practice of law. At the end of the event, feedback surveys on the platform modules[190] were collected by questionnaire. In total, 31 feedback questionnaires were collected.

The results showed a high level of satisfaction with the tool on average (see Figure ) and an overall positive stance towards certain uses of AI by the judiciary (see Figure ). Judges seem to have a highly positive estimation of AI-powered functionalities in particular that supports the easy and effective retrieval and summarisation of information in texts. This

[189] F. GALLI ET AL., *Predicting outcomes of Italian VAT decisions*, in *Jurix 2022, Proceeding of the 35th International Conference on Legal Knowledge and Information Systems*, 2022, 188.

[190] The "Database Module" included the legal database, visualization of annotated judgements, automatically extracted keywords and summaries and extracted citation and analysis.

included automated summaries and key term extraction, citation analysis, ontology visualization and search, and the partially automated extraction of arguments. The largest doubts pertained to the outcome prediction.


Figure 2. Effectiveness of the ADELE modules

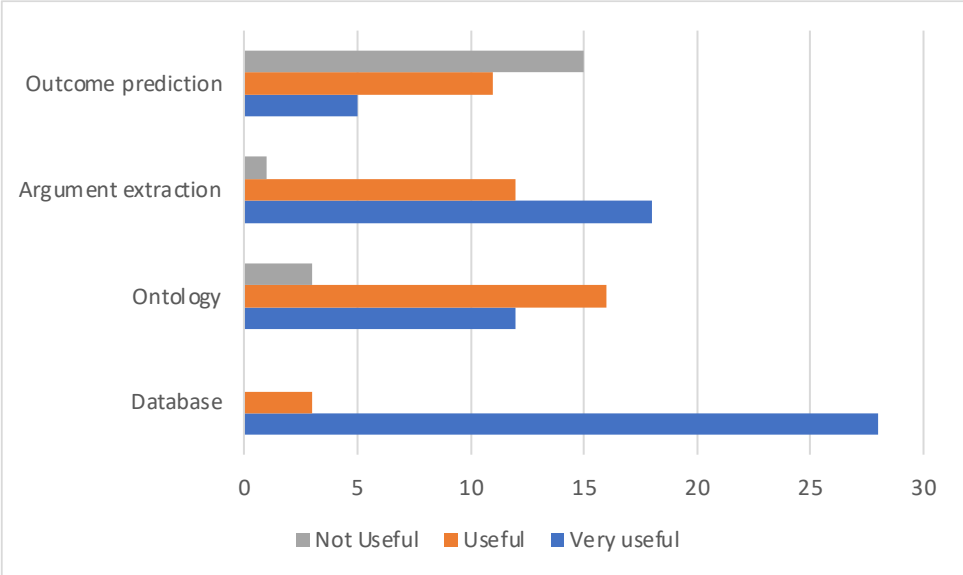



Figure 3. Usefulness of the ADELE modules

## 11. Ethics

During the project, an ethical self-assessment was carried out by the Consortium regarding the development of the AI-powered pilot tool. The tool was developed in compliance with the EU Ethical Guidelines on AI, adopted in 2019 by the European Commission through

the High-Level Expert Group on Trustworthy AI[191] and the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment, adopted in 2018 by the Council of Europe acting through the European Commission for the Efficiency of Justice.[192]

The tool was designed with human oversight, traceability, and auditability in mind, serving as a support instrument for legal professionals rather than a stand-alone decision-making tool. Its use was triggered by professionals for exploratory purposes, with outcomes intentionally initiated by users for specific domains and languages. The tool's user manual, drafted for the testing events, provided clear guidance on functionalities and machine learning model assumptions. Guidelines for annotating case-law were published and tested by legal professionals. Moreover, the tool prioritised compliance with fundamental rights, emphasising adherence to the European Convention on Human Rights (ECHR) and data protection laws. It incorporated privacy by design and default principles, by way of restricting processing to anonymized data. While designed to align with the ECHR, further investigation is still required to assess its compatibility with rights such as access to a fair trial. The tool's voluntary and exploratory nature, along with judicial responsibility for its use, mitigates concerns about fundamental rights. The independence of judges is highlighted, with measures taken to inform judges about the tool's nature and limitations, leaving the decision to use it or not to their judgment.

Furthermore, the tool adhered to fairness principles in AI design and usage, ensuring unbiased input data and algorithm design. Legal domains chosen for piloting (VAT and TM&P) also avoided sensitive personal data. Training data was properly anonymized, and rigorous testing revealed no discrimination tendencies. Efforts were also made to inform judges and legal professionals about the model limitations and training specifics.

Finally, users of the tool were informed of its use of AI, with clear explanations of its abilities, limitations, risks and benefits. The tool, designed for scientific debate in the legal use of AI, emphasised purposeful limitations. While the "Outcome Prediction" Module was not fully explainable, users received information about its AI algorithms, training data, and the non-binding nature of its predictions. The disclaimer specified that neither the ADELE consortium partners, nor the European Commission were responsible for the tool's provision.

[191] EUROPEAN COMMISSION HIGH-LEVEL EXPERT GROUP, *Ethics guidelines for trustworthy AI*, April 2019, available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[192] CEPEJ, *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*, 2018.

# Section 6

# National experiences: the cross-border dynamic between France and Italy

## 6.1. The use of Artificial Intelligence in predictive justice, present and future: the French experience

*Etienne Vergès, Géraldine Vial – Université Grenoble Alpes*

*Summary: 1. Tools for assisting judges – 2. Open data and searching court decisions – 2.1. Anonymizing court decisions to publish them – 2.2. Finding the right decision in a database – 3. Metrics and predictive justice – 3.1. Metrics – 3.2. Predictive justice – 4. Predictive justice and modeling judicial decision processes: how does it work? – 5. "To predict is to prescribe"? The impact of algorithms on judicial decisions*

*Abstract: In this discussion, the state of progress of AI in the French legal field is presented and the results of several research projects conducted in the context of the French justice system are described. The remarks focus on four main topics: (1) tools for assisting judges; (2) open data of court decisions and its impact on the notion of case-law and searches for relevant decisions for judges or lawyers; (3) metrics and predictive justice and, finally, (4) the impact of algorithms on judicial decisions.*

### 1. Tools for assisting judges

There are many software programs used in the French justice system. In criminal justice, police investigators and, more generally, fact finders use AI to identify suspects. For example, the scientific police, *gendarmerie,* and customs have been using serial analysis software – Anacrim or Salvac among them – for several years, which makes it possible to cross-reference data relating to an offense from several databases. Currently, the most innovative systems are facial or voice recognition. For example, a file called "criminal history processing" stores 8 million photos of people who have been involved in criminal proceedings. When investigators have an image of a suspect from a video surveillance camera, they can use facial recognition software to identify the profiles of people who most closely resemble the image of the suspect in the image database.

AI is also used to improve work tasks. For example, a company has developed an AI system to automatically classify documents in a criminal file and to detect procedural irregularities. For police officers or judges, this software makes it possible to correct files that contain visual irregularities, when a document has not been signed for instance. To carry out this task, the AI system is programmed in advance to identify the formalities that must appear on an investigation document and visually verifies whether this formality has been carried out. Another more specific example concerns the French Court of Cassation. This court receives 20,000 appeals each year. These appeals must be examined by one of the six chambers of the Court depending on the legal area concerned, among which are Contract

Law, Liability Law, and Family Law. This task is carried out automatically by AI which detects the legal issue at stake by reading the text of the appeal.

## 2. Open data and searching court decisions

### *2.1. Anonymizing court decisions to publish them*

Since 2016, the French government has had to release all court rulings handed down by judges. This represents 3.9 million decisions per year. Since the statute law also requires the anonymization of all sets of decisions – which applies to items such as names, addresses, wedding dates, and bank accounts – the data scientists of the Supreme Court have developed an AI engine to automatically identify and anonymize these rulings. The accuracy of the engine is quite high, identifying and anonymizing the names of parties in about 95% of cases. But the law requires total anonymization, which is why the same Supreme Court has hired twenty civil servants to control and manually correct the process.

### *2.2. Finding the right decision in a database*

Open data raises the problem of access by professionals to large numbers of court decisions. There are two types of responses to this problem.

From an academic point of view, an official report[193] proposed the creation of a classification of court decisions that would envision each judge classifying decisions according to their legal value.

From a practical point of view, lawyers seek to find decisions that would allow them to win cases for which AI solutions are quite useful. Publishers are also looking for solutions to help lawyers find the best decisions that apply to their case. For example, a legal publisher named *Lexbase* has developed an application using vectorization technology to associate words or expressions called "Similar facts."[194] A lawyer can describe the facts of a case in a few lines and the AI system will search for court decisions with similar facts. The lawyer can then study each decision and find the one to be communicated to the judge as precedent.

These two approaches to jurisprudence are very different. In the official approach, not all decisions rendered by judges have the same legal value. It is therefore the judge who is best placed to assess the legal interest of the decision rendered. This is a hierarchical conception of jurisprudence. Conversely, for lawyers, the right decision is one which concerns a case similar to that of their client and was rendered favorably. Among the 4 million decisions issued each year, lawyers seek to identify the most relevant. New AI technologies in the field of NLP can therefore help perform this task precisely and automatically. As the publisher states on its website: "only the facts count."

[193] L. CADIET, C. CHAINAIS, J.-M. SOMMER (eds.), S. JOBERT, E. JOND-NECAND (rapp.), Dissemination of decision-making data and case law, June 2022.
[194] https://www.lexbase.fr/lexia/acces-restreint.

## 3. Metrics and predictive justice

### *3.1. Metrics*

Metrics are the first application of AI to case-law. Metrics are also a new way of exploring the mass of case-law rulings by focusing on new categories of information provided by legal decisions.

For an illustration of a digital service created by a legal tech startup, consider the example of the company *Predictice* which analyzes large quantities of court decisions to extract metrics. For example, their AI can analyze a set of eighteen thousand decisions on the chances of losing a case and provide quantitative information such as the percentage of the chances of success or the length of a procedure. Their AI is also capable of identifying decisions based on keywords and extracting quantitative data such as an evaluation of the expectation of winning a case. For example, for the chance of losing a case, the AI is able to calculate that around 12,000 applications are won out of 18,000 and qualify the acceptance rate as "high."


Figure 1.

The AI can also calculate that the average duration of a procedure from the first instance to appeal is 2 years and 4 months.


Figure 2.

*Predictice* is not a predictive justice tool. It is possible to know average durations, but the AI is not capable of analyzing a particular case. On the other hand, it is capable of providing metrics on several hundred legal subjects.

To go further into legal analysis, the *Predictice* company uses AI such as ChatGPT to analyze legal information sites. It has, thus, created a legal assistant capable of interacting with users and providing answers to their questions. The results of this legal assistant are mixed. Sometimes the answers are specific and accurate, other times they are general, and other times they are wrong. Despite these contrasting results, *Predictice* has paved the way for the use of Large Language Models (LLM) to make knowledge of the law more accessible.

### *3.2. Predictive justice*

Predictive justice is an advanced way of using metrics. Predicting the future decision of a judge is not a trivial task. It takes hours and hours to study hundreds or thousands of rulings and to extract the relevant facts and evidence that will be decisive in a judicial solution. For instance, if you want to predict divorce compensation, you must evaluate, among other things, the influence of the spousal incomes, each person's property, and the age of the parties. After a learning process however, the AI engines are able to calculate the probability of compensation and an amount.

It is necessary to understand that justice prediction is, in fact, a quantification of judicial risk. In other words, you can't predict the future decision of a single judge, but you can anticipate the probability of different possible solutions.

In France, several legaltech firms have developed prediction tool techniques based on AI, such as the company *Case law analytics* or *Legal quantum*. The figures below show a probability distribution regarding the compensations that an employee could receive for unfair dismissal. Figure 3. shows that 17 percent of judges award 8 months' salary in a particular case. It is also possible to show that, in this case, the employee has only a 21 percent chance of obtaining more than 12 months' salary.

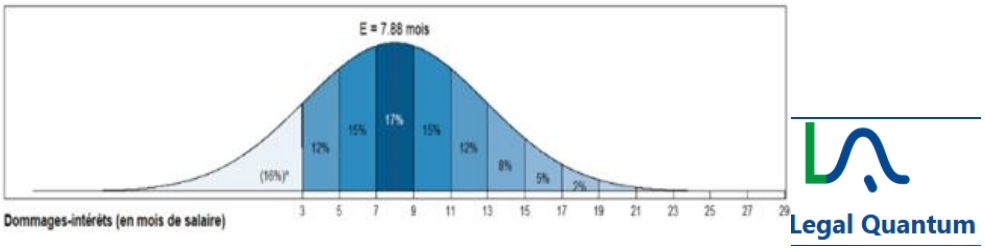



Figure 3.

Predictive justice is actually probabilistic reasoning. This involves calculating the chances of success in obtaining a favorable decision in court, but also calculating the chances of winning. This requires knowing the different criteria that are used by judges to make their decisions.

There are several companies in France developing predictive justice solutions, but this activity is rare and still only emerging. Transformations are underway as major legal firms and editors are currently absorbing start-ups to develop predictive justice solutions.

## 4. Predictive justice and modelling judicial decision processes: how does it work?

Research on predictions is growing at many universities around the world. At the legal research center of Grenoble Alpes University, research in Family Law is being developed, and more particularly on determining a child's residence in post-separation litigation.

Various criteria which influence a judge's decision in a set of a thousand appeals court decisions have been identified, like the parent's housing, the age of the child, and the results of the social survey. Court decisions are manually annotated and these criteria (input) are associated with a solution (output). So the influence of each factor can be quantified individually.

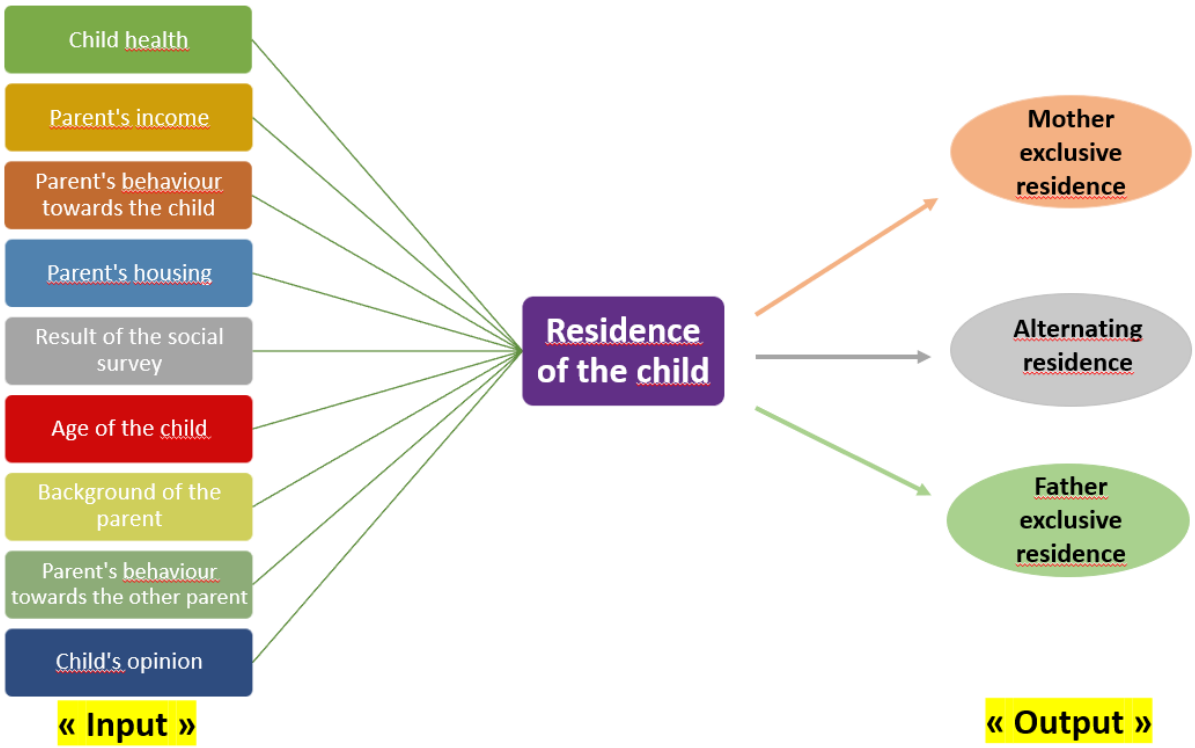



Figure 4.

For example, we have been able to establish a statistical link between the quality of a parents' housing and the chances their child will be placed with them at home. If their housing is unstable, the parent has no chance of having the child placed with them. To the contrary, if the housing has been specially adapted for the child, the chances they will be placed with the parent increase. The link is also very strong between a child's wishes and the judge's decision. The judge follows the child's wishes in more than eighty percent of cases.

Modeling a judges' decision consists of creating decision trees or neural networks to combine all possible solutions and finding the most likely decision.

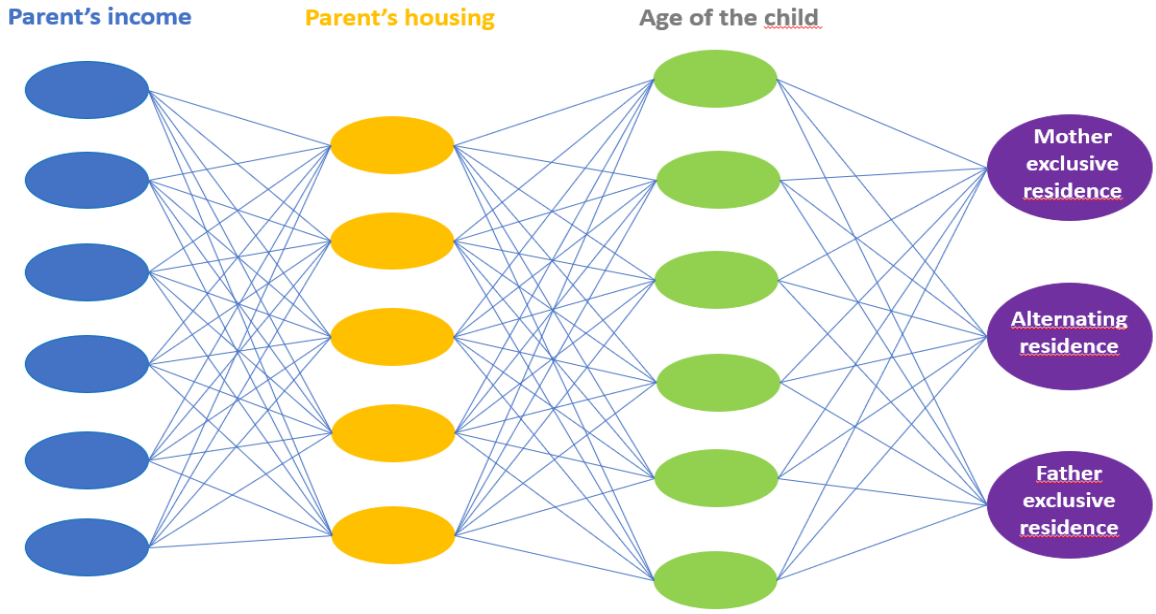



Figure 5.

Manual annotation of court decisions is very time-consuming which is why it was decided to test other technologies to find out whether it was possible to automate text analysis with natural language processing. This technique is based on the zoning of decisions. In French decisions, the zoning operation must be done manually because relevant information is spread throughout many parts of the text, as can be seen below from the display of manual zoning decisions.

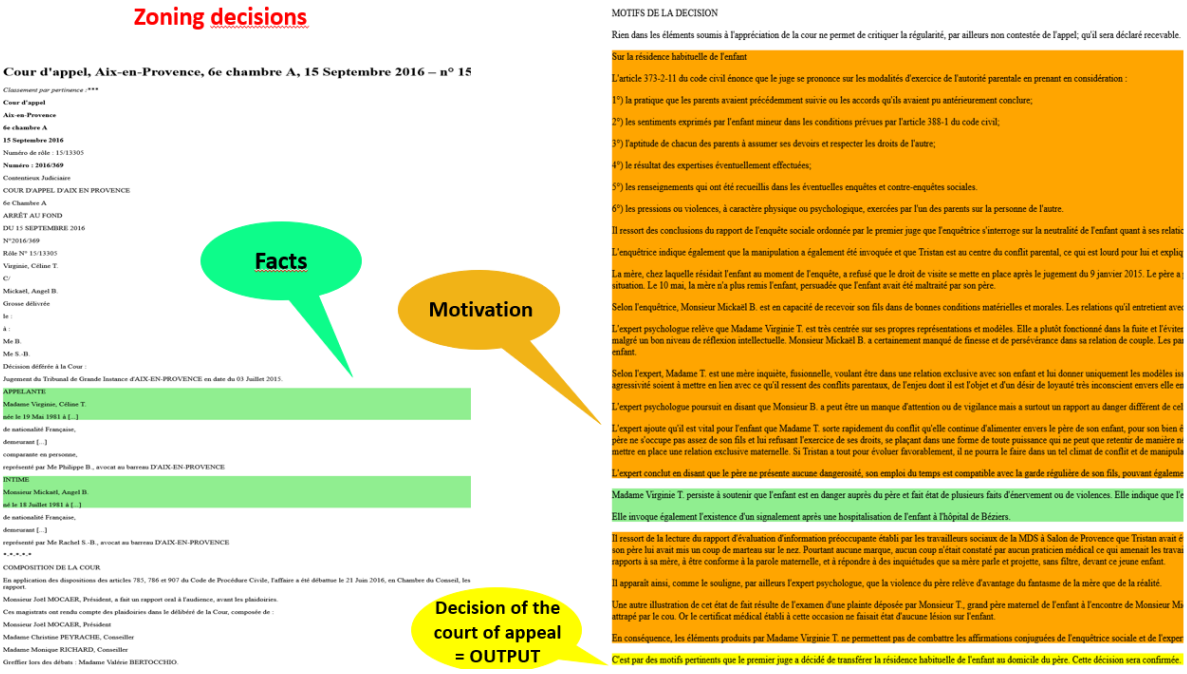



Figure 6.

After the zoning operation, AI systems are used to automatically analyze the content of text and link words and sentences with the output, which is the solution. Theoretically, NLP

solutions are very effective for analyzing texts such as Google search engines or ChatGPT. Unfortunately, with judicial language, it is more complicated.

In order to test the effectiveness of these different techniques, a competition between different AI solutions and natural intelligence was organized. To this end, five members of our team (lawyers and computer scientists) read one hundred decisions in order to predict the solution after reading the text. The accuracy of the NLP was low. In comparison, the accuracy of the models using annotated data was high to very high. The best human result was very close to the AI annotation model. However, humans do not work with the same information. They have access to the context and, specifically, can understand the position of a judge by the style of writing sentences. Much subjective information comes from the text.

We can learn several lessons from this experience. Compared to studies on the European Court of Human Rights,[195] the results of the NLP models were very low for predicting decisions of the Court of Appeals in France. This seems explained by the fact that the writing of ECtHR decisions is rather homogeneous while, in France, daily decisions handed down by judges are not. Despite these differences between judges, the annotation model is very accurate. When criteria are worked on precisely and each decision is carefully annotated, i.e. taking about 40 minutes per decision, Big Data is not really relevant for predicting judicial decisions. To the contrary, with a small set from a well annotated dataset, a high level of accuracy can be achieved. Without manual annotation, AI systems are not yet good enough to automatically extract the facts of a case from court decisions.

Models trained on very precise annotated datasets are the most efficient at predicting judicial decisions. One question is to examine the impact of these predictions on a judge's behavior in the future, the famous issue of whether "*to predict is to prescribe*."

## 5. "To predict is to prescribe"? The impact of algorithms on judicial decisions

In order to try and answer this question, an experiment was conducted with judges from 2019 and 2021[196] in which two tricky cases were chosen, involving different algorithmic skills.

The civil case was related to compensatory allowance in divorce litigation. The file contained many types of evidence and the facts were ambiguous. It was possible to award or to refuse the compensatory allowance for the ex-wife as well as to award a small or a large amount. We divided the judges into two groups. One with algorithmic expertise and the other without this particular skill. According to the group with algorithmic expertise, ninety-seven percent of judges would award compensatory allowance in this case, the likely amount ranging from €32,000 to €44,000.

In the criminal case, the issue was about determination of guilt in a case involving shaken baby syndrome. The context incriminated the father, but the fact that the child was a victim of SBS was up for debate. One portion of the judges had medical expertise and concluded a SBS diagnosis was "probable." In addition, the other portion of judges had algorithmic

[195] M. MEDVEDEVA, M. VOLS, M. WIELING, *Using Machine Learning to Predict Decisions of the European Court of Human Rights*, in *AI and law*, 2020, 237

[196] E. VERGES, G. VIAL, *Justice et intelligence artificielle, l'impact des algorithmes sur la décision de justice*, Recueil Dalloz, 2022, 1919.

expertise and evaluated the likelihood of SBS at fifty-seven percent. The determination of guilt was therefore balanced.

In the civil case, a significant influence of algorithmic expertise on respondents' decisions was observed. The average amount awarded by judges with knowledge of the results of algorithmic modelling was in the middle range (€36,380). By comparison, the average amount awarded by other respondents was outside the range (€30,945).

In the criminal case, the results were totally different. The percentage of guilty decisions was observed to be very close from one group to the next (10.3% and 6.7%). Algorithmic expertise had an influence on a very small number of judges.

These results were correlated with judges' perception of AI tools. In criminal cases, judges considered that the algorithmic evidence was not relevant and undermined the judge's authority (79,3% "AI is not relevant" vs. 20,7% "AI is relevant"). To the contrary, in civil cases algorithmic evidence was seen as helpful and capable of ensuring the fair treatment of litigants (89,3% "AI is helpful" 10,7% "AI is not helpful").

This brings us to conclude that the use of AI is still emerging in France, as it is in many other countries. However, many researchers and companies are investing in this field, and professional tools are already on the market. Not all technologies are efficient, but the question today is whether these technologies will be accepted and adopted by the legal community.

## 6.2. The Italian experience: an overview

The Italian experience with AI and justice is presented in the following section with several contributions, illustrating different elements of the still pending process that aims to enable the use of AI in the field of justice in Italy. Although the Italian experience may resemble, in certain respects, the experience of other EU Member States, the present analysis is not meant to be representative of any specific cluster or even less so, the EU as a whole.

The starting point begins with a description of how the Italian judiciary has been transitioning from merely digitizing documents to leveraging digital tools and AI systems for judicial decision-making, by moving from traditional documents to data as enabled by recent but surely unremitting technological advancements. This is further explored in an analysis of the latest Italian reform of procedural law (see paragraph 6.2.1).

Following that, the role played by the Electronic Documentation Centre (CED), established at the Italian Court of Cassation, to explore the use of AI in judicial activities is discussed and described, with particular regard to key elements such as the pre-determination of decision-making criteria and the analysis of documents using syntactic sequences related to text structure as well as the statistical frequency of linguistic expressions.

Next, this section could not omit analysing the recent Italian bill on Artificial Intelligence submitted on May 20, 2024 by the national government. This proposal represents the first legally-binding intervention on AI from a Member State aiming to ensure enforcement of the newly adopted AI Regulation Act. On the other hand, the proposal has also already raised doubts as to its relationship with the Regulation as a general framework binding all Member States which may be complemented, but not departed from, at the national level. Against this background, it remains to be seen to what extent national regulatory measures may enhance or potentially limit the opportunities that certain advancements, such as predictive justice algorithms, could bring to improving the judiciary.

Finally, beginning from the Italian reform of criminal procedure, the section ends by examining remote justice and laying the foundations for a discussion on a pioneering topic leading to a new digital paradigm, the so-called "Metaverse," from the perspective of its entry into the field of justice: a challenge whose link with AI goes beyond the purpose of this contribution.

# 6.2.1. From "documents" to "data:" an epochal shift in Italian justice is possible

*Amedeo Santosuosso – IUSS Pavia*

*Summary: 1. Premise – 2. Three apparently minor legislative innovations (Cartabia Reform) – 3. Assumptions and implications – 4. The reasons for judgement? A temporary data aggregate – 5. The August 2023 Ministerial decree – 6. Everything easy and within reach? Certainly not*

*Abstract: The Italian justice system is about to make a historic transition to an activity inherently shaped by the possibilities that the digital environment offers by using artificial intelligence (AI) techniques. Until now, the content of documents was essentially unchanged from the past, the only difference being the transfer into digital format (digital "documents") of information originally expressed as signs imprinted on paper. In the present phase, information is produced in a way that can be directly explored by AI, which operates on "data" and not "documents."*
*Is the Decree of the Italian Ministry of Justice (August 7, 2023) on the way to moving from "documents" to "data"? If one considers the provisions indicating the sections that each document must have, Yes. However, the most innovative part (dealing with computer schemas) is very poor. This is the road, but it seems to be a long one.*

## 1. Premise

The Italian justice system is (or, to be prudent, could be) on the verge of making a historic transition from the mere digitization of its documents, that is, from the mere use of non-analog (i.e. digital) tools, to an activity that is, first, inherently modeled on the possibilities that the digital environment offers and, second, that also makes use, for the first time operationally in decision-making, of Artificial Intelligence techniques. In the previous phase, the content and mode of document acquisition, as well as the proceedings, were essentially unchanged from the past, the only difference being the transfer of information into digital format originally expressed as signs printed on paper (whether by pen, typewriter, or computer makes little difference), producing digital "documents." In the phase that is now opening (or may open), information is directly produced and organized in such a way that it is machine-readable and has characteristics that make it directly explorable through AI techniques (legal analytics), which operate on "data" and not on "documents" by definition.

It is the shift from documents to data, long known and investigated at the theoretical level,[197] that is now becoming the cornerstone of a major innovation (i.e. implementation) in the judicial field.

## 2. Three apparently minor legislative innovations (Cartabia Reform)

Three minor legislative changes, which do not affect the Civil Procedure Code *per se*, but only its Implementing Provisions (*Disposizioni di attuazione)* mark this shift from documents to data. First we must have the patience to follow this legislative path, which is part of the so-called Cartabia Reform of Civil Procedure in Italy.

[197] M. A. LIVERMORE, D. N. ROCKMORE (eds.), *Law as Data: Computation, Text, & the Future of Legal Analysis*, Santa Fe Institute Press, 2018.

Article 196-quater of the Implementing Provisions of the Civil Procedure Code established that in civil proceedings before a Tribunal, Court of Appeals, the Court of Cassation and the Justice of the Peace, all procedural documents and records shall be *compulsorily* and *exclusively* filed through a computer system (i.e. the PCT[198]), while (and this is the second step) the subsequent article 196-quinquies provides that documents formed in this way shall be digitally signed and filed in the computer filing system.[199]

The third change concerns the way these documents must be written and structured, of which the articles mentioned above provide for their filing through a computer system and their destination in that system. Article 46 of the Implementing Provisions of the Civil Procedure Code[200] states that judicial documents must be written in a clear and easily readable font (an ancient request, it must be said!), but above all (and here is the new fact) that the Minister of Justice, having consulted the Superior Council of the Judiciary (CSM) and the National Council of the Bar (CNF), must define by decree the computerized schemes of court documents and the structuring of fields necessary for the entry of information into trial records. The same decree shall establish the limits of court documents, taking into account their type, value, complexity of the dispute, the number of parties and the nature of the interests involved. The documents must contain an index and a summary of the contents of the document itself. In addition, in a separate rule, it was established that these rules shall apply to all proceedings instituted after June 30, 2023.[201]

In brief, from now on (a) all documents in civil trials will exclusively be in digital format; (b) they will be filed by computer system; (c) the documents of lawyers and civil judges will have to be structured according to digital schemes, the fields of which are intended to receive

---

[198] https://pst.giustizia.it/PST/it/homepage.page

[199] Article 196-quater of the Provisions for the Implementation of the Code of Civil Procedure establishes the mandatory use of electronic filing for procedural documents and decisions in civil cases before the tribunal, court of appeal, Court of Cassation, and justice of the peace. Such filings must be carried out exclusively through electronic means, in compliance with legal and regulatory provisions governing the signing, transmission, and receipt of digital documents.
Article 196-quinquies addresses procedural acts prepared in electronic format, specifying that such acts, when drafted by judges or judicial office personnel (including notification, enforcement, and protest offices), must be duly signed. Updated texts can be accessed via "Normattiva".

[200] Article 46 of the Provisions for the Implementation of the Code of Civil Procedure requires that minutes and other judicial documents be written in a clear and legible manner. When drafted in electronic format, they must comply with the legal and regulatory standards governing the drafting, signing, transmission, and receipt of digital documents. The Minister of Justice, in consultation with the High Council of the Judiciary and the National Bar Council, issues a decree defining the electronic templates for judicial acts, specifying the necessary fields for entering case register information. The same decree establishes limits for procedural documents based on factors such as the type, value, and complexity of the dispute, the number of parties involved, and the nature of the interests at stake. These limits exclude formal elements such as headings, indexes, and summaries. The decree is updated at least every two years. Non-compliance with the technical specifications or formal requirements does not invalidate the document but may influence the court's decision on legal costs. Judges must draft acts and rulings in accordance with these criteria (the whole text of Article 46 Disp.Att. Cpc is available through Normattiva).

[201] Legislative Decree No. 149 of October 10, 2022, established (through Article 35, paragraph 1) that, unless otherwise specified, its provisions take effect from June 30, 2023, and apply to proceedings initiated after that date. For proceedings already pending as of June 30, 2023, the previous provisions remain applicable. The same article also specified that, unless otherwise provided, the decree's provisions take effect from February 28, 2023, and apply to proceedings initiated after that date. For proceedings pending as of February 28, 2023, the earlier provisions remain in force.

the information contained in the proceeding records; (d) the documents will have to contain a table of contents and a summary of the document itself; (e) the computer schemes and further indications of the breadth of the documents are defined by the Minister of Justice, after consultation with the CSM and CNF.

Apart from the prejudicial resistance that may exist from parts of the legal profession and the judiciary, there is ample material to reflect on as we move from some theoretical perspectives advanced by the academy to innovation, that is, to a practical change in the context and manner of administering civil justice in Italy.

## 3. Assumptions and implications

The starting point is as follows: if all the acts of parties and judges are in digital format and converge in the computer system, each act of the lawyer or of the judge, each part of the minutes of proceedings change their nature: it ceases to be an isolated entity (a document) and becomes a provisional aggregate of information shared with other parties during the proceedings and recomposed in various ways according to the position of the person drafting the act and the stage of the proceedings.

As pointed out in a contribution a couple years ago,[202] the term "justice" is used to refer to a complex set of institutions and activities, all related and, at the same time, each having their own function and distinct characteristics in which digitization plays a different role. "Justice" is the ministerial apparatus with its peripheral articulations. In this context, the digitization of assets, of personnel management and of administrative activities, on the one hand, can improve the functioning of the system and, on the other hand, has limited effects on the reduction of time and, mostly, on the quality of decisions.

If, then, by "justice" we mean judicial activity, certainly here, too, the digitization of the management of administrative and office staff, as well as of technical means (computers, connections, networks and others) can give judicial activity a more solid foundation, which will enable it to proceed more expeditiously, however, the impact of this, albeit useful, digitization will only be indirect on the time and quality of decisions.

The issue takes on a different aspect if by "justice" we refer to the rules governing proceedings, and thus the rules of procedure. The relationship between procedural rules and digitization is a delicate one: in other words, procedural rules should not be constructed as if they were in the analog era and then checked for digital compatibility, but should be *digital by design*, that is, conceived of from the outset as immersed in the context of a digital process.

All of this, as important as it is, only touches on the activity of judging; it informs its context but does not touch on the cognitive aspects of judging, which concern judges and lawyers, the sources from which they draw information on case-law and legislation, and by what technological means: the way their documents are written, how they relate to each other, the technologies that Judge's offices use, etc. In fact, it is only at this last level that digitization ceases to be a mere technological context for the legal professions and becomes a transformative factor in the Law itself in the way it lives in social relations, opening up to a profound and unpostponable juridical reflection, in which Legal Analytics (LA) tools come into play, and a technical environment in which various disciplines, such as data science,

[202] A. SANTOSUOSSO, G. PINOTTI, *Una Giustizia "digital by design": ecco come realizzarla*, in *Agenda Digitale,* June 21, 2021, available at: https://www.agendadigitale.eu/documenti/giustizia-digitale/giustizia-digital-by-design-ecco-come-realizzarla/

artificial intelligence (AI), machine learning (ML), natural language processing (NLP) and statistics converge.

At this point, with the regulatory changes mentioned above, it could be said we are there, i.e. that we are at the point of making this transition. However, we only need figure out where and how to begin.

## 4. The reasons for judgement? A temporary data aggregate

It may be useful to start with the nature of reasons for judgment, i.e. where judges explain the reasons behind their decisions (*motivazione* in Italian). It is a decisive concept from a constitutional point of view and an important indicator of the relations between judges and lawyers, as well as a decisive aspect of the formal regulation of their respective acts regulated by the Ministry, the CSM, and the CNF (see above Article 46 Disp. Att. Civil Procedure Code).

We could begin by asking what the components of a judicial opinion are, what the internal organization of the opinion is, and what it should contain with respect to the totality of the materials that have merged into proceedings.[203] The most convincing answer is that a judicial opinion explains the reasons why each element of materials from proceedings, that have been discussed by the lawyers of all parties according to the adversarial principle, were accepted or rejected, provided that judges place all these materials in a logical-legal order to support their decision. This does not appear to be a reductive definition. It responds to the rule that prohibits judges from ruling beyond or outside the requests of parties[204] and to the adversarial principle, according to which the judge who decides "to place on the basis of the decision an issue noted ex officio" must assign a time limit to the parties in order to be able to interact by filing pleadings (Article 101 Civil Procedure Code). It follows from this that judges must rule on the issues proposed by the parties and may not ground their decisions on issues to which the parties have not been able to exercise their right of defense. The judges can certainly, if they so decide, elaborate upon, or even develop further, but only starting from an issue that has been the subject of interlocution.

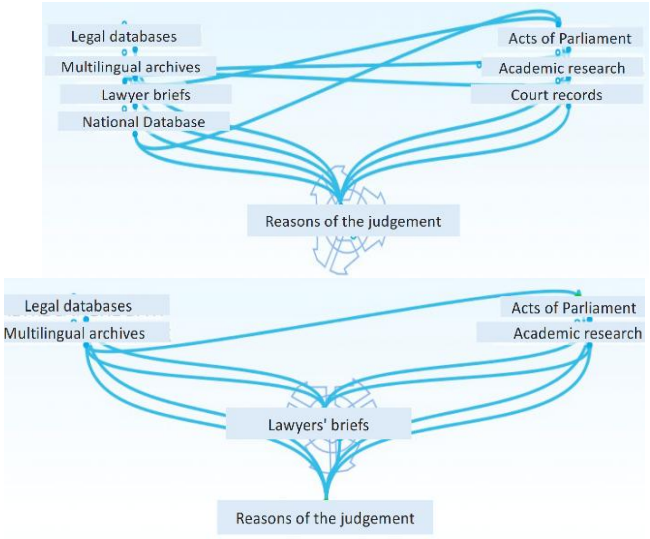
Strictly speaking, it can be said that judges do not add arguments, but organize and give priority to existing ones in a way that is functional to the decision. Therefore, a good explanation of the reasons for judgement is one that makes explicit why and how, and in what order, the judge decided to aggregate these legal elements: nothing more, nothing less. This is because judgement and reasoning come about as the outcome of a process governed by a fundamental rule, not only of legality, but above all the adversarial process, the highest expression of the right of defense and due process.

The next step is to ask what is the nature of these legal elements/arguments that populate the reasoning of the judgment. They are legal norms, constitutional or international treaty norms, judicial precedents, norms from other legal systems (an increasingly common case, especially in the European Union), and doctrinal elaborations. All these materials, it should

[203] For a thorough analysis, A. SANTOSUOSSO, *Diritto e intelligenza artificiale*, Mondadori Università, 2020, 101-120, along with the experience since 2015, summed up in note 64.

[204] According to Article 112 of the Code of Civil Procedure, which establishes the principle of correspondence between what is requested and what is adjudicated, the judge must rule on the entirety of the claim without exceeding its limits and cannot address ex officio exceptions that may only be raised by the parties.

be noted, are the same and are drawn from sources shared with lawyers, scholars in the academy, and public administrations in view of their decisions or with legislators themselves. The difference lies only in the way each of these professionals or entities organizes the information, a way that depends on their own institutional missions and deontological boundaries. A schematic representation is contained in the figure below:




This set of information is expressed in natural language, that is, in the language used in the specific community of life. Different natural languages may concur, even coexist, in the judgment, as sometimes happens with sources from the European Union when they are expressed only in English and French or with contracts written in different languages, which are often not translated. Moreover, as legal content placed in computer/digital contexts, they can acquire the nature of data, which can be structured or unstructured (depending on the degree of readability and processability by a computer).

We can, therefore, clarify the above by stating that the judge does not add data (arguments), but organizes and gives it a hierarchy in a way that is functional to the decision and that, therefore, a good judicial opinion makes explicit why and how the judge decided to aggregate the data (legal elements) thusly.

If the legal elements of the judgment are expressed as data (exactly the step to be taken) we should draw the conclusion that the judge's opinion (like lawyers briefs and the sources from which the data come) is an aggregate of that data, which the judge has re-arranged and organized in a way that is suitable to rationally justify the decision. All this is common to the civil and criminal fields, the lawyer and the judge, the public administrator as well as the legislator at any level. Their deeds are all aggregates of information/data that can be disaggregated and re-aggregated in various forms.

## 5. The August 2023 ministerial decree

The question of how procedural documents should be structured and organized is addressed within the Next Generation UPP Project (presented in October 2021).[205] The contribution by the University Institute for Advanced Studies IUSS-Pavia on the point was expressed in the following terms: by "next generation template" (MNG) we mean a file that is digitally native, that collects structured data, that is inserted by default among the templates of the Judge's and Assistant's Console (the operational interface for judges and assistants), that is collected in an appropriate data warehouse or data lake and that can constitute a dataset on which to operate with knowledge extraction tools.

According to the IUSS project, this activity is part of the necessary reprogramming of the entire Telematic Civil Process (as well as criminal) on a new technical basis. The technological innovations that have occurred in recent years require that, building on the experience to date with the Telematic Civil Process, the entire system is radically rebuilt. To give just a few examples: a) the problem of voluminous attachments to files should be solved in an appropriate manner; b) a system of archiving materials that includes not only writing files, but also images and audio (Data Lake) should be envisaged; c) the Electronic Certified Email (PEC) system should be definitively superseded.

The digitization of judicial activities is the basis for any further technological development. But this digitation must be *total*, in that it must cover all procedural activities and all parties to the process (thus including lawyers, but also technical experts) as well as chancery data; *well ordered*, in that it combines the character of easy accessibility for judges and users with the technical aptitude of the products of the activity that will populate carefully organized collections of data for advanced AI uses; and *modern*, in that it must cover all data and information, whatever their nature, and be open to further developments.

IUSS Pavia, with its research group composed of jurists and computational linguists,[206] has done in-depth work on the development of these templates, in collaboration with important experience from the judiciary and the legal professions, and through interlocution with the Directorate General of Information Services and Automation (DGSIA).

In line with this: (a) computer templates should be digital native files that collect structured data; (b) the data should come with an auto-compilation mode, both from records (as it is now), as well as from the minutes of cases (for end-of-court measures) and, above all, from parties' acts, according to a technical mode similar to that which already exists for the insertion of the parties' "conclusions," but which also takes essential parts from the defense acts (as will be explained below); c) computer templates should be inserted by default among the templates of the Judge's Console and the assistant, d) once the act is formed, it is collected into a special data lake and can constitute a dataset that can be operated upon with retrieval tools (data retrieval), document builders, and AI techniques that allow for knowledge extraction.

---

[205] "Progetto PON Giustizia: Next Generation UPP: nuovi schemi collaborativi tra Università e uffici giudiziari. Per il miglioramento dell'efficienza e delle Prestazioni della giustizia nell'Italia nord-ovest", proposed by eleven Universities belonging to Macroarea 1 (districts of the Courts of Appeal of Torino, Genova, Milano and Brescia).

[206] The research group is led by Professors Amedeo Santosuosso and Cristiano Chesi and consists of Matilde Barbini (computational linguist), Stefano D'Ancona (jurist), Emanuela Furiosi (jurist), Emma Zanoli (computational linguist).

It is important that the templates also cover defendants' briefs and must be structured in fields, so that for each part of the argument made by the defendant, defendants can add a short excerpt (or abstract) of a few lines themselves which should self-insert automatically, without the possibility of being modified by the judge. This means that the partition or segmentation of the judgment must connect with that of the defense counsel. A recent study showed how trial rules before the European Court of Human Rights, rules to which judges and lawyers must conform, follow this logic exactly.[207]

A document builder[208] is conceived of as an aid for judges and UPP staff to find materials for drafting orders. It presupposes a well-organized dataset that is fed with good quality data, which can be queried with algorithms that can help make information visible that the naked eye might fail to notice or, even, propose a draft of the judgment to be written and the applicable precedents and rules, once the system has been provided with the space-time and subject-matter coordinates of the issue to be decided. At that point the judge is left with the most difficult, intellectual, and professionally valuable task: to select that material and proposed path of reasoning and challenge it, changing or specifying some parameters, or some factual and legal elements that contradict and change the consequences of the system's proposal.[209]

What is needed is a view of the reasons of judgment as a creation, not in isolation but embedded in a flow of information organized as data, where it is only a step, albeit one of great practical and social importance. This is also true in an era dominated by generative AI, where a system of well-structured templates can reduce the risk of hallucinations.

## 6. Everything easy and within reach? Certainly not

If this was the inspiration for the research presented and approved in October 2021 and the results of the work subsequently carried out by the IUSS Pavia group, today it can be compared with the contents of the Ministry of Justice Decree of August 7, 2023.[210] The decree "establishes the criteria for drafting and regulates computerized schemes of the civil trial acts, with the structuring of the fields necessary for entering the information in the trial records. It also establishes the dimensional limits of civil trial acts […]" (Article 1).

Considering the reflections made and the research work done by the IUSS Pavia group in the PON Justice project, the question can be put in these terms: is the Ministry of Justice Decree on the same path, moving from "documents" to "data"?

---

[207] G. PINOTTI, A. SANTOSUOSSO, F. FAZIO, *A Rule 74 for Italian Judges and Lawyers*, in R. *GUIZZARDI, B. NEUMAYR* (eds) *Advances in Conceptual Modeling. ER 2022. Lecture Notes in Computer Science*, vol 13650. Springer, Cham. https://doi.org/10.1007/978-3-031-22036-4_11.

[208] The creation of the prototype is assigned to Dipartimento di informatica of Università Statale di Milano, coordinated by Prof. Silvana Castano, in collaboration with IUSS Pavia.

[209] The results of the IUSS Pavia research group were presented in A. SANTOSUOSSO, S. D'ANCONA, E. FURIOSI, *New-generation templates facilitating the shift from documents to data in the Italian judiciary*, at the 42nd International Conference on Conceptual Modeling (ER 2023), Lisbon, Portugal, 6-9 November 2023, published in: T. P. SALES, J. ARAÚJO, J. BORBINHA, G. GUIZZARDI (eds.) *Advances in Conceptual Modeling. ER 2023. Lecture Notes in Computer Science, vol 14319*. Springer, Cham. https://doi.org/10.1007/978-3-031-47112-4_11

[210] Ministry of Justice, Decree No. 110 of August 7, 2023, Regulation defining the drafting criteria, limits, and electronic templates for judicial acts, including the necessary fields for entering information in case registers, pursuant to Article 46 of the Provisions for the Implementation of the Code of Civil Procedure (published in Official Gazette No. 187 of August 11, 2023).

At first glance the answer seems to be YES and NO.

NO, if we consider some stylistic indications, such as those of the character and height of the body (Article 6), which seem to refer primarily to the act as a physical document rather than to the data contained within it. So are the quantitative limits of acts, which do not seem particularly useful or effective.

YES if we consider the provisions of Article 2, which indicate the partitions and sections that each act must have (i.e. the parties, key words, separate and specific exposition, rubricated in separate parts of the act, of the facts and grounds in law, indication of the documents offered for communication, listed in sequential numerical order and named in a manner corresponding to their content, preferably searchable by appropriate hyperlink).

Finally, it should be noted that the potentially most interesting part, dealing with computer schemas (Article 8) is very poor and limited to a generic reference to a 2011 Ministerial decree. That is the road, but at the moment it seems a long way to go. There is hope that the work done through *PON Giustizia*[211] might be used as a contribution to finally begin the transformation *from documents to data*.

[211] Funded by the same Ministry of justice and timely communicated to the DGSIA.

# 6.2.2. The experience of the CED on the use of Artificial Intelligence in predictive justice

*Antonio Corbo – Corte di Cassazione*

*Summary: 1. The CED – 2. The intervention framework – 3. Suggestions for the present and the future*

*Abstract: This contribution offers an overview of the Italian landscape and experiences related to the implementation of technological tools in the justice system, highlighting the efforts of the Supreme Court with the same.*

## 1. The CED

The Electronic Documentation Centre (CED) is an office of the Court of Cassation that reports to the First President and takes care of the development of all electronic tools for e-Justice at the Court of Cassation. The institutional task assigned to the CED has led its Director to deepen the theme of the possible applications of Artificial Intelligence tools to the field of judicial activity, involving all components of the structure in the reflection.

## 2. The intervention framework

This reflection sought to identify, on one hand, the opportunities offered by AI tools and, on the other, the limits imposed or otherwise suggested regarding possible applications and the need to respect fundamental human rights.

It found that the opportunities provided by the use of AI tools are many and could be of much assistance to judicial work regarding the organization of activity, the making of decisions, and the drafting of reasoning. However, it was observed that conducting judicial activities cannot respond to rigid models, because the particularity of individual cases must be taken into account, and therefore, it is necessary to identify the role that should be reserved for humans.

The elements that are currently both typical and relevant in the context of using AI tools that are applicable to judicial activities appear to be:

a) the pre-ordination of criterion or criteria for decisions, through an indication of the relevant factors and the "specific weight" of each of them;

b) recourse, in the examination of documents analyzed for the purpose of drawing up answers, to the evaluation of syntactic "sequences" relating to the "structure" of texts and the statistical frequency of the linguistic expressions contained therein, and not of the "semantic sequences," concerning the "meaning" of those texts.

The pre-ordination of the decision criteria, on one hand, ensures the equal treatment of different situations analyzed at least tendentially; on the other, however, it does not pemit us to take factors into account which, although not foreseen by the programs designed to find a solution to the issues, could nevertheless be worthy of consideration. The enhancement of semantic "sequences" for the purpose of elaborating answers exposes us to the risk of results that are strongly conditioned by quantitatively majority opinions in the area under examination, and, therefore, are not very attentive to less widespread theses.

## 3. Suggestions for the present and the future

An approach that appears balanced, at present, could be to use AI tools as elaborators of hypotheses of various solutions and models with which humans must be confronted before making a decision. For the effectiveness of this comparison, in any case, it is necessary that humans be aware of the predetermined criteria laid down as the basis for the elaboration of solutions.

Certainly, a wider, more effective, and even "reassuring" application space can be recognized, at present, related to activities where fewer variables seem to occur, such as those dealing with the organization of work (the organization of cause-lists for example and the distribution of proceedings and trials among judges).

However, the discourse related to the search for judicial precedent and the opinions of legal scholars is more articulated. Certainly, the use of AI tools can facilitate the discovery of useful "materials" but the completeness and correctness of answers through systems that enhance syntactic "sequences" remains to be seen.

It is not easy, then, to identify the correct scope of application to be reserved for AI tools in relation to the formulation of judicial decisions and the elaboration of the related reasoning. Perhaps, for the purposes of making decisions, the aid can be a simple model for comparison: the decision of a judicial case, in fact, is based on a plurality of elements, some of which seem imponderable in advance, i.e. the evaluation of several elements of evidence in contrast with each other. For the purposes of elaborating the reasoning of decisions, however, AI tools could offer structural models and grids of topics to be developed, which would be very useful, for example, at avoiding deficiencies in justifying discourse on the relevant points of a decision.

The use of AI, even in judicial work, is a great challenge. It can offer great utility, but must be implemented with a full awareness of possible critical issues. It is only by starting from these premises that it is possible to further develop the search for applications that are both effective and respectful of fundamental human rights.

# 6.2.3. The Italian Proposal for a Law on Artificial Intelligence

*Tommaso De Mari Casareto dal Verme – Università degli Studi di Trento*

*Summary: 1. Introduction – 2. The provisions – 2.1. General principles – 2.2. Sector-specific rules: the field of justice – 2.3. Other provisions – 3. The relationship with the AI Act and the field of justice*

*Abstract: The Italian Government presented Bill n. 1146/2024 on May 20, 2024, aimed at regulating AI in line with the newly adopted EU AI Act Regulation. The purpose of the bill was to protect fundamental rights, democracy, the rule of law, and environmental sustainability according to the possible risks and level of impact of AI, at the same time promoting innovation for the welfare of all citizens. To do so, it proposed some general principles and sector-specific rules, as well as some executive actions and delegating powers to the Government to ensure the enforcement of the EU Regulation. At the same time, given the adoption of the AI Act, which will soon be mandatory and directly applicable in all Member States, it is necessary to investigate the possible role of the Italian legislation in light of the supranational source and the consequences that its adoption might provoke for the use of AI in the field of justice.*

## 1. Introduction

On May 20, 2024, the Italian government submitted Bill n. 1146/2024 to Parliament laying down provisions and delegating powers to the government in the field of Artificial Intelligence (AI). The draft was presented almost simultaneously with the adoption of the AI Regulation Act by the EU Council on May 21, 2024. In fact, the Italian proposal sought to coordinate with and ensure the enforcement of the AI Act's provisions, without overlapping it, by setting out "rules for the correct, transparent, and responsible use, in an anthropocentric dimension, of artificial intelligence, aimed at seizing its opportunities" (Article 1).

The purpose of the bill was to protect fundamental rights, democracy, the rule of law, and environmental sustainability according to the possible risks and the level of impact of AI, while promoting innovation for the welfare of all citizens.[212] The proposal is composed of twenty-six articles divided into 6 parts: 1) rules underlying basic principles; 2) sector-specific rules; 3) governance, national authorities, and promotional actions; 4) rules for copyright protection; 5) criminal sanctions; and 6) financial provisions.

## 2. The provisions

### *2.1. General principles*

The first section (Articles 1-6) provides some basic principles linked to the purposes and scope of application of the entire draft. Such principles interact differently according to the specific fields of application and of the considered phase of the AI lifecycle. This regulatory approach was deemed consistent with the EU Regulation and the provisions contained therein and, if adopted, are due to be interpreted and applied in accordance with the relevant EU legislation. In this view, both the Italian draft and the AI Act promote a responsible and transparent use of AI from an anthropocentric dimension; while setting out rules on the

[212] See Explanatory Memorandum of the Bill, 1.

adequate surveillance of its economical and societal risks by employing tools compatible with the specificities of national law. With this in mind, the proposal sets out principles for the research, experimentation, development, adoption, and application of AI systems and models (Article 1).

The general principles set out in Article 3 are transparency, proportionality, safety, data protection, confidentiality, accuracy, non-discrimination, gender equality, and sustainability. They are then articulated and specified in further principles, including the accuracy, reliability, security, quality, appropriateness and transparency of the data used, according to the principle of proportionality, as well as respect for human autonomy and decision-making power, prevention of harm, knowability and explainability. Furthermore, Article 5 establishes certain principles for economic development that the State and other public authorities must promote both to enhance human-machine interaction in production sectors and to create AI as a useful tool for the start-up of new economic activities, in order to increase the competitiveness of the national economic system. Finally, a crucial aspect to be regulated in the national interest is cybersecurity, since AI can also be a tool for enemies to attack with and weaken national defences.

### *2.2. Sector-specific rules: the field of justice*

The general principles mentioned are further specified and tailored to some specific fields of AI application contained in the second section of the bill (Articles 7-16), in which the field of justice is also listed. Among these, the Italian Government also considered the fields of: the healthcare sector (Article 7), where the principles of transparency, non-discrimination, and the protection of human-decision making must govern the use of AI in the diagnosis and cure of diseases, as well as in scientific research and experimentation (Article 8) where the use of AI must comply with the GDPR; the labour sector (Article 10), where the principles of safety, reliability, transparency, and data protection require the employer to inform an employee of the use of AI, that must occur in a non-discriminatory manner; Public Administration (Article 13), where the use of AI must have the function of supporting decision-making activities in respect of autonomy and the decision making power of the human person, who still bears responsibility; cybersecurity (Article 16), in the context of which the government highlighted the need to enhance the value of AI as a crucial resource for national security.

In the field of justice, the centrality of human decision-making and thinking emerged once again. A strong emphasis was therefore posed on the legal argumentation underlying a judge's decision, which it would be risky to delegate to a technology that supposedly lacked critical consciousness.[213]

Article 14 of the bill established that AI systems be used exclusively for the organization and simplification of judicial work, as well as for case law, and doctrinal research. The Ministry of Justice is to regulate the use of AI systems by judicial offices. Nevertheless, it is always up to the judge to decide on the interpretation of law, the assessment of facts and evidence, along with the adoption of all necessary measures. Moreover, Article 15 establishes the exclusive competence of ordinary civil tribunals for cases concerning the functioning of AI systems.

[213] These are the words of the Explanatory Memorandum, 10.

It is unclear whether this provision would allow the use of AI systems for so-called "predictive justice," i.e. systems which, based on statistical models drawn up independently by the system itself and following the analysis of a significant number of legal acts, are able to make predictions that can go as far as predicting the possible outcome of a judgment.[214] On the one hand, the provision might be interpreted as forbidding the employment of AI in the decision-making activity *per se*, i.e. it would only be permitted within a phase of legal research preliminary to the final decision. On the other hand, as long as the judge ensures the final decision is based on an autonomous interpretation process, the use of predictive justice algorithms *per se* might not be forbidden, e.g. in the collection and analysis of judicial precedents and relevant studies that can still provide some form of "simplification" and support for judicial work, albeit in a more nuanced manner.

Therefore, it should be clarified whether the described provision aims to limit the types of AI systems employable by judges in their overall decision-making activities or to ensure that whichever type of AI systems are employed do not play a prevalent role in the decision-making process, in order to avoid major risks such as those related to so-called "automation biases."[215] In other words, would judges be able to use predictive justice AI as long as they ensure that the final decision is mostly based on their own autonomous evaluation of facts and evidence? Such aspects are crucial in order to define whether and to what degree the use of AI is capable of making an "automated" decision, similar to the assessment conducted by the CJEU in the Schufa Case, where it stated that the requirements for data processing are "automated individual decision-making" under Article 22 of the GDPR.[216]

### *2.3. Other provisions*

Among the remaining provisions of the bill, the third section (Articles 17-22) deals with establishing rules around a national strategy on AI, by pinpointing the national authorities in charge of carrying out the tasks of monitoring, accreditation, cybersecurity, and verification with EU Law. It also establishes promotional actions in support of investments b companies devoted to AI development. The same section, as an important step towards enforcement, delegates powers to the Italian government to enhance legally-binding acts that ensure the alignment of national legislation with the AI Act regulations (Article 22).

The fourth section (Articles 23-24) establishes rules on copyright protection with a particular focus on generative AI, e.g. the mandatory watermarking or other forms of marking capable of making clear when content has been modified or altered in such a way as to present real data, facts, and information that are not so. On the other hand, the rules on

[214] According to the CEPEJ's 'European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment' (see in this Handbook contribution n. 3.1. by N. OZALP, Y. MENECEUR, *Council of Europe's perspective on Artificial Intelligence*), predictive justice tools can be used both by lawyers and judges to predict the likely outcome of a specific case. They provide a graphic representation of the probability of success for each outcome of a dispute based on criteria entered by the user and specific to each type of dispute. For deeper insights on features and concrete examples of predictive justice applications, see in this Handbook contribution n. 6.1 by E. VERGÈS, G. VIAL, *The use of Artificial Intelligence in predictive justice, present and future: the French experience*.

[215] For deeper insights on the risks related to the use of AI in decision-making activities, with specific regard to automation biases, see contribution n. 2.1 of this Handbook by M.G. CIVININI, *New technologies and justice*.

[216] See contribution n. 7.1 of this Handbook by F. CASAROSA, *Automated processing, AI and enforcement between the jurisprudence of CJEU and national courts*.

copyright protection are extended to works generated with AI, provided they are the result of the intellectual work of the author.

Finally, the fifth section (Article 25) sets out criminal rules, e.g. by integrating some existing norms with the specification of AI use when committing a crime or as an aggravating circumstance. The sixth section (Article 26) contains a financial safeguard clause.

## 3. The relationship with the AI Act and the field of justice

As announced by the Italian Government,[217] the bill is not supposed to overlap with EU AI Regulation, but to complement its regulatory framework in areas of domestic law, bearing in mind that the Regulation is based on an architecture of risks associated with the use of AI. Nevertheless, it must be considered that EU Regulation is mandatory in all its elements and immediately applicable to all EU citizens, States, and institutions. This means that Member States are entitled to regulate the field of AI as long as their interventions are limited to complementing and/or specifying the principles enshrined in the Regulation and to ensuring its enforcement within their respective national systems, especially when the Regulation explicitly demands that Member States do so. Nevertheless, Member States shall not compromise the objectives set by the European legislator by altering the uniform protection level set out in the Regulation, e.g. by establishing more or less restrictive measures. In fact, the AI Act is *inter alia* concerned with preventing Member States from hindering the development and deployment of (lawful) AI systems within the Union.[218] In this context, the relationship between the Italian bill and the AI Act is not completely clear and must be carefully investigated in order to avoid possible conflicts.

To begin with, coordination with the AI Act immediately emerges in the definition of "AI system" provided by Article 2 of the bill, according to which it is defined as "**an automated system** designed to operate with varying levels of autonomy and which may exhibit adaptability after deployment and which, for explicit or implicit purposes, infers from the input it receives how to generate outputs such as predictions, contents, recommendations or decisions that may affect physical or virtual environments." This definition is almost identical to that contained in the AI Act, although the latter defines an AI system as "**machine-based"** rather than an "automated" system.[219] Though this divergence might easily be explained by the version of the text of the EU proposal at the time the Italian bill was drafted,

[217] See press release n. 78 of the Italian Government, available at https://www.governo.it/it/articolo/comunicato-stampa-del-consiglio-dei-ministri-n-78/25501.

[218] According to Recital n. 1, "this Regulation ensures the free movement, cross-border, of AI-based goods and services, thus preventing Member States from imposing restrictions on the development, marketing, and use of AI systems, unless explicitly authorised by this Regulation.' For instance, under Article 2 paragraph 11, the Regulation 'does not preclude the Union or Member States from maintaining or introducing laws, regulations, or administrative provisions which are more favourable to workers in terms of protecting their rights in respect of the use of AI systems by employers, or from encouraging or allowing the application of collective agreements which are more favourable to workers." Moreover, according to Article 5 paragraph 5, "Member States may introduce, in accordance with Union law, more restrictive laws on the use of remote biometric identification systems."

[219] Article 3(1) of AI Act: "AI system means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."

it is an example of the uncertainty in the respective scopes of application. It should be asked whether the national definition of AI system provides for an equal, broader, or more restrictive scope compared to the Regulation, considering that the definitional approach of the AI Act has gradually broadened since the initial definition provided by the first version of the European proposal.[220]

Moreover, the Italian proposal only generically mentions a proportional risk-based approach, without adhering, or making explicit reference, to the specific risk classifications designated in the AI Act (unacceptable risk, high-risk, and low-risk). As such, the scope of the Italian proposal appears unclear: in order to implement a more symmetrical coordination with the AI Act, it could have clarified whether it applies to high-risk AI systems as described by the Regulation.

Secondly, let's examine the relationship between the Italian bill and the AI Act from the perspective of the material scope of application of the Italian proposal, which includes research, testing, development, adoption and application of AI systems and models. The EU Regulation explicitly excludes AI systems and models from its scope specifically developed and put into service for the sole purpose of scientific research (unless they are placed on the market or put into service as a result of such research and development activity),[221] and establishing rules for the development, adoption, and application of AI systems is at the core of the AI Act regulation since its addressees are providers, deployers, and operators of AI systems. Even if the Italian Government adopts a sectorial (or vertical) approach, and the AI Act is horizontal, the principles established in the AI Act are destined to be applied horizontally to specific fields as well. In fact, the Regulation also considers sector-specific fields of AI application whenever AI systems in those areas are classified as high-risk according to Annex III, or when such fields are covered by existing EU legislation listed in Annex I. This means that national provisions must anyway comply with the Regulation and not hinder its purposes, even when legislating in areas of their own competence. Accordingly, the separation between the vertical national proposal and the provisions of the Regulation are not so clear cut.

It therefore seems that the field of justice is a good example for further developing the above considerations. Annex III, n. 8(a) of the AI Act lists among high-risk AI systems those "intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts." The choice by the EU legislator designating such systems as high-risk is based on their potentially significant impact on democracy, the rule of law, individual freedoms as well as

[220] The actual definition of "AI systems" provided by the Regulation is the result of a progressive broadening of its material scope, facing the primary difficulty in finding a scientific and univocal definition of "Artificial Intelligence." Such an approach emerges if one considers that the definition of AI system started by referring to "software" – which gives the idea of a specific product or technological item – and ended up with a "machine-based system," encompassing many types of informatic techniques that are traceable to AI. Therefore, the reference to an "automated" system – which will be likely amended by the Italian legislator if the bill is adopted – might represent a narrower definition of AI systems.

[221] Such an approach is also adopted by the recent Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law, signed by the EU Commission on September 5, 2024, that represents the first legally-binding international document on AI. Under Article 3.3, the Convention "shall not apply to research and development activities regarding artificial intelligence systems not yet made available for use, unless testing or similar activities are undertaken in such a way that they have the potential to interfere with human rights, democracy, and the rule of law."

the right to an effective remedy and to a fair trial.[222] The same approach was adopted by the recent Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law, signed by the EU Commission on September 5, 2024.[223] At the same time, as clarified by Recital n. 61, it does recognise the utility of supporting a judge's decisions, while ensuring that the final decision-making remains a human-driven activity.

In this context, it should be asked whether the choice by the Italian Government to limit the use of AI systems in the field of justice to the sole preliminary phase of doctrinal and case-law research complies with the uniformed provisions of the AI Act. In fact, the AI Act does not forbid the use of AI in a judge's decision activity *per se*, provided that the final outcome remains human-driven. This means that, on one hand, the EU Regulation allows for the development and use of AI systems that seek to support a judge's decision, not only in legal research of precedents but also in the actual decision phase, by analysing and interpreting concrete facts and applying the law to them, provided that such systems (i) fulfil the AI Act's requirements for high-risk AI systems (included those of providers and deployers) and (ii) provide support rather than replacing human decision making. On the other hand, the provision in the Italian bill seems to suggest the government's intention is to limit the possible use of AI to the research activity of relevant legal documents, thus excluding the use of AI systems capable of analysing concrete facts and evidence and supporting judges in legal argumentation, which should be entirely left to the judge. It should be then wondered whether such a choice might, in practice, imply a prohibition of certain types of AI systems that the AI Act instead considers permissible, although high-risk. Therefore, some AI functions in the course of development and destined to the justice sector, such as predictive justice algorithms, should be carefully considered insofar as their greatest potential manifests itself in activities that support a judge's decision beyond the mere research of legal materials.

Examining the aforementioned issues from a practical perspective, it is useful to reference some of the projects described in this Handbook, which demonstrate great potential to support and simplify judges' decision-making, but also go beyond simple legal research and could provide active tools to assist (without replacing them) judges in making their decisions. Among the tools of the CrossJustice Project is the "Automated Reasoner Assessment," which aims to assist users to develop and analyze the legal reasoning of a given case, takes into consideration the specific facts of a case, and presents the solutions it has inferred. The outcome provided by the tool includes all the steps of legal reasoning, leading to a specific conclusion.[224] Similarly, the ADELE Project has outcome prediction among its objectives,

[222] See in this Handbook contribution n. 3.2 by F. CASAROSA, *Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act): an analysis*.

[223] Under Article 5 of the Convention, "each Party shall adopt or maintain measures that seek to ensure that artificial intelligence systems are not used to undermine the integrity, independence, and effectiveness of democratic institutions and processes, including the principle of the separation of powers, respect for judicial independence and access to justice. 2 Each Party shall adopt or maintain measures that seek to protect its democratic processes in the context of activities within the lifecycle of artificial intelligence systems, including individuals' fair access to and participation in public debate, as well as their ability to freely form opinions."

[224] See contribution n. 5.1 of this Handbook by G. CONTISSA, G. LASAGNI, *Deploying AI technology to empower procedural safeguards and judicial cooperation*.

which enables anticipating what could be the decision in a specific case according to past case law.[225]

In light of the considerations above, adoption of the AI Act requires reflecting carefully on the role of national interventions seeking to enforce or specify the rules provided in the Regulation. In fact, national provisions should avoid altering the uniform standard of safety within the EU market for AI systems and hindering the development and deployment of AI systems compliant with the AI Act regularly circulating in the EU market.[226] With this in mind, the provisions of the Italian bill, if adopted, shall be interpreted by courts, as far as possible, in light of the wording and the purpose of the Regulation, otherwise national legislation shall be disapplied.

Finally, it should be stressed that national legislation might also be considered inconsistent with EU law when replicating – and thus overlapping with – the content of a Regulation, despite not being substantially contrary to its provisions. In this regard, the CJEU has often specified that "the uniform application of Community provisions allows no recourse to national rules except to the extent necessary to carry out the regulations."[227] In fact, since an EU Regulation is mandatory in all its elements and immediately applicable in all Member States, there is no need to implement its provisions with any internal act, also meaning that (i) no national provision can replace those of the Regulation and (ii) no procedure capable of hiding the "community nature" of a legal rule is permissible.[228]

In this context, as stated in the AI Act, Member States hold a key role in the application and enforcement of the Regulation,[229] limited to what is necessary and expressly ordered by the EU legislator, such as the designation of notifying bodies and national authorities for ensuring market surveillance and compliance of AI systems with the AI Act,[230] as well as establishing penalties for infringement.[231] It can therefore be argued that those provisions of the Italian proposal dealing with the delegation of executive powers to the government for the better implementation of the AI Act in the areas expressly mandated by the European legislature do not raise particular issues of consistency with the supranational source. On the other hand, several doubts remain as to the necessity and compliance with other provisions of the new Regulation, even if (but not only) they are limited to the risk that the new Italian law conceals the EU nature of AI regulation.

On November 5, 2024, the European Commission issued the Reasoned Opinion (C(2024) 7814) to the Italian government containing recommendations for amendments to the

---

[225] See contribution n. 5.2 of this Handbook by F. GALLI, A. FIDELANGELI, P. SANTIN, G. SARTOR, *Analytics for Decision of Legal Cases: The ADELE Project*.

[226] Recital n. 129: "Member States should not create unjustified obstacles to the placing on the market or the putting into service of high-risk AI systems that comply with the requirements laid down in this Regulation and bear the CE marking."

[227] CJEU, *Norddeutsches Vieh- und Fleischkontor GmbH v Hauptzollamt Hamburg-St. Annen*, February 11, 1971, C-39/70.

[228] CJEU, *Fratelli Variola S.p.A. v Amministrazione italiana delle Finanze,* October 10, 1973, C-34/73.

[229] See Recital n. 153.

[230] According to Article 28 of the AI Act, "each Member State shall designate or establish at least one notifying authority responsible for setting up and carrying out the necessary procedures for the assessment, designation, and notification of conformity assessment bodies and for their monitoring;" and under Article 70 "each Member State shall establish or designate as national competent authorities at least one notifying authority and at least one market surveillance authority for the purposes of this Regulation."

[231] Under Article 99 of the AI Act Member States shall lay down the rules on penalties and other enforcement measures.

proposed Italian legislation on artificial intelligence[232]. The suggestions aim to ensure alignment between Italy's national framework and the AI Act provisions. In particular, the EU Commission suggested that the national law refrain from introducing divergent definitions and instead rely directly on those already established in the EU Regulation. With regard to the justice sector, the Commission advised alignment with Article 6(3) of the AI Act, which does not categorically exclude the use of high-risk AI systems in judicial contexts, provided they do not pose a significant risk of harm to health, safety, or fundamental rights and do not materially affect the outcome of judicial decision-making.

[232] For Opinion's content, see the minutes of the Italian Senate session no. 363 of 27 November 2024, available at https://www.senato.it/static/bgt/UltimiAtti/pergiorno/20241128.html?from=20241128&to=20241128.

# 6.2.4. Technological innovation and the personalist tradition. From "remote" participation to so-called "Metaverse Courtrooms:" The Italian perspective

*Benedetta Galgani – Università degli Studi di Pisa*

*Summary: 1. The question – 2. The cardinal points … – 3. … and a land (still) in between – 4. Future incognitae: the criminal trial in the metaverse?*

*Abstract: In questioning the relationship between technological progress and the axiological coordinates of criminal procedure, this contribution first analyses the innovations introduced by the "Cartabia reform" with regard to remote justice, and then deals with a new digital paradigm, the so-called "metaverse," in the perspective of its entry – perhaps less far-fetched than one might think – into the field of justice.*

## 1. The question

To outline the structure of this contribution, it is especially compelling to confront the title chosen. It is without a doubt that a choice of title always aspires (at least tendentially) to indicate the perimeter of debate, the conceptual horizon within which we are expected to perform our task as scholars – to sound out the roughness of the land, to tread the limits, to define the rigour of the boundaries – and it is an observation that may itself risk lapsing into the obvious. Yet, especially in this occasion, the need to move from the meaning of the chosen syntagma seems to stand as a methodological approach which cannot be eluded.

Thus, we can almost take for granted that the combination "innovation technology" generically alludes to the activity deliberated upon (also) by institutions and that it aims to introduce new products and services in various fields of reference.[233] Regarding the second syntactic unit under the magnifying glass, i.e. the "personalist tradition," some further consideration may be necessary. On one hand, one of the most consolidated meanings of the lemma "tradition" is that which designates the "transmission over time, from generation to generation, of consuetudes, customs, traditions, models and norms" as well as – and cascaded – "the consuetudes, customs and traditions" thus handed down.[234] On the other, the attribute "personalist" clearly seems to refer to the term "personalism" which generally designates any "conception that affirms the primary value of the person in the construction of reality and in the formulation of philosophical problems."[235] Hence the expression "personalistic tradition" seems to imply the inescapability of a "human-centred" approach and the indispensability of a legislative policy that is nevertheless capable of associating the adoption of novel methodologies to conduct fact-checking with the possibility to achieve (a higher) rate of guarantees, even and especially for those towards whom the judicial ascertainment is directed.

In short, the imperative of a cultural approach in which the technological transition (more specifically digital) is inspired by the typical demands of the "criminal justice" domain

[233] See G. G. SIRILLI, entry for *Innovazione tecnologica*, in *Enciclopedia scientifica tecnica Treccani*, 2008.
[234] See entry *Tradizione*, in *Vocabolario on line Treccani*.
[235] See entry *Personalismo*, in *Vocabolario on line Treccani*.

emerges in a peculiar position with respect to the market economy and is not conversely dictated by technocratic drifts.

Yet, this specific and dutiful attention to the "dignity of man" does not seem to involve the indispensability, always and in all cases, of the "in presence," of the "physicality," of the corporal dimension: it is "human individuality" *per se* that must in all cases be safeguarded in its crucial prerogatives within the criminal process.

So here we are, in one leap, at the beating heart of the matter which, albeit in its nuclear terms, can be transfused into the question that follows: when confronting and measuring ourselves with the increasingly more "extensive" opportunities for remote participation in certain procedural activities, if not in the hearing itself, is "technological innovation" that stands on a collision course with the identifying features of an axiologically oriented criminal trial to be regarded as fatal, or is it conceivable to assume the establishment of a relationship as hendiadys between the two units under consideration?

It is well known that the highlighted issue is far from new, and this is surely not the context to review its decades-old, steep normative paths. At most, consider how the debated topic involves audio-visual links in the trajectories of criminal jurisdiction, i.e. of the "technical mode by which to realize the remote participation and examination making use of instruments capable of receiving images and sounds from one location (the courtroom) and of transmitting them directly to another (the so-called remote location) through the use of cameras and monitors."[236] A sort of "religious war"[237] has been waged around the issue, a war that in recent years, has been abetted by the articulated pandemic experience, first, and the directives of Delegated Law n. 134 of September 27, 2021, and the implementing provisions of Legislative Decree n. 150 of October 10, 2022, second, which has been further rekindled and which discloses original perspectives for development as well as a deepening of the conversation.

## 2. The cardinal points …

Although the aim here is to keep the discussion within the framework of a mere "reasoned list," it may prove necessary to elucidate some "cardinal points:"

1. We dare to embrace the exhortation of one of the unquestioned Masters of criminal procedure, Professor Massimo Nobili: we refer to his warning not to "get entangled in nominalism and abstract, self-styled "models," but "rather" to reference the various "ingredients" that serve as components to achieving a "due process." "Ingredients" that, to be clear, can't be expressed without "adherence to a certain institutional and social context."[238] Besides, the very distinctive traits of the hearing and the standards to which it must conform – capable themselves of conferring a certain physiognomy to the entire criminal justice system – are but the sign of historically determined values.[239]

[236] D. CURTOTTI, *I collegamenti audiovisivi nel processo penale*, Milan, 2006, 33.

[237] For this expression and for a more precise reconstruction of the different "seasons" that have characterized, within our criminal trial system, the normative evolution in the use of audio-video connections, let us refer to B. GALGANI, *Forme e garanzie nel prisma dell'innovazione tecnologica. Alla ricerca di un processo penale "virtuoso,"* Milan, 2022, 231 and the valuable bibliography referred to there.

[238] M. NOBILI, *L'impatto della cultura di common law sui rapporti fra diritto e processo penale*, in *ID., Scritti inediti*, Camon, Milan, 2021, 122.

[239] See D. NEGRI, *Il dibattimento*, in AA.VV., *Fondamenti di procedura penale*, Milan, 2021, 546.

2. Singularly consistent with the mentioned approach, is the mental attitude formerly made clear by the same *conditores* of the 1988 code; that is, an approach far from any aprioristic dogmatism informed by an "empirical reasonableness" under which what is actually "at stake" should be evaluated each time, as well as the different ascertaining *itinera* which should be dosed and modulated.[240] It is impossible to dismiss the circumstance that Professor Delfino Siracusano, father of "cross-examination for evidence"[241] – which constitutes one of the cornerstones of recodification and was the later object of constitutionalization in the 1999 reform – peremptorily asserted as insignificant: that "teleconferencing" did not alter either the functionality of participation nor the genuineness of evidentiary gathering, provided that conditions were guaranteed to fully exercise the right to cross-examination.[242] Nor can we fail to highlight how Professor Mario Chiavario, an influential member of the ministerial Commission in charge of implementing the one hundred and five directives under the 1987 delegated law, in regard to the "developments in the technique of videoconferencing," called for a commitment "to search out" new "forms of guarantee [...] reshaping some of them" and, where appropriate, "inventing others."[243] Thus it cannot be forgotten to whom the paternity of the constitutional ruling should be ascribed, which exalted "participatory realism" and who referred to the "potentialities and [a]refinements always offered by technology;"[244] a pronouncement that for the first time offered a glimpse into the possible interactions between empirics and theory. Not only was he a great jurist, he also, as Chancellor (*Guardasigilli*), dismissed the code that today still bears his name and imprint.
3. As it becomes clear from early on in the supranational and European legislation, which is always inclined to the telematic "fragmentation" of the courtroom – to use the most recent examples, the extension of the operativity of videoconferencing for purposes of taking witness and expert testimony by Article 11 of the Second Additional Protocol to the Budapest Convention on Cybercrime[245] – con-fuses the conceptual and practical con-fines of the trial categories of reference.
4. Hence the undeferrability of attempting to change perspective, to "conceptualize change"[246] to properly explore the interactions between the *lato sensu* of telematic

[240] See G. DI CHIARA, *Come uno schermo. Partecipazione a distanza, efficienza, garanzie, upgrade tecnologici*, in *Rivista di diritto processuale*, 2018, 1479.

[241] D. SIRACUSANO, *Vecchi schemi e nuovi modelli per l'attuazione di un processo di parti*, in *Legislazione penale*,10.

[242] D. SIRACUSANO, *Il giudizio*, in SIRACUSANO, GALATI, TRANCHINA, ZAPPALÀ, *Diritto processuale penale*, Milan, 2018, 631.

[243] M. CHIAVARIO, *La "videoconferenza" processuale e la Corte europea dei diritti dell'uomo*, in AA.VV., *Studi in onore di Mario Pisani, vol. II – Diritto processuale penale e profili internazionali: diritto straniero e diritto comparato, a cura di Corso-Zanetti*, Piacenza, 2010, 98, 109.

[244] See Constitutional Court, July 22, 1999, n. 342. While being aware of assigning *ex se* "decisive" importance to the judgments of constitutional (or European) Courts on this subject, one cannot, however, forget that "if criminal procedural law is applied to constitutional law, and if the Constitution lives first and foremost in the words of the Constitutional Court, the decisions of the Judge of Laws necessarily represent both the mandatory reference the legislator must take into account and the regulations and/or values which the interpreter must measure against." See F. RUGGIERI, *Il volto costituzionale del processo penale. Indagine giurisprudenziale alla ricerca di valori condivisi*, Pisa, 2021, 34.

[245] For a critical analysis of the latest European legislation, see O. MURRO, *La disciplina della videoconferenza per le dichiarazioni del testimone e dell'esperto*, in *Diritto penale e processo*, 2022, 1143.

[246] Cf., albeit from a more general perspective, L. RULLO, *Corti online*, in *Rivista di Digital Politics*, 2021, 220: "The digital is not a neutral tool and its redistributive capacity in terms of the ability and effective protection

modalities of the hearing and an adequate protection of the values of due process of law. More specifically, the opportunity looms to explore the asemantic recalibration of the concept of "space" (and, consequently, that of "spatiality").[247]

This doesn't imply that that concept should be abandoned *tout court* but rather that there should be a redefinition of its physiognomy, as provided in Article 146 disp. att. of the Criminal Procedure Code for example, as well as the ordering capacity traditionally ascribed to it, in light of experiences and technical advances which were obviously inconceivable before. On the other hand, in all branches of Law multiple examples of the imperative to "conceptualize change" can be found in a world that is continuously evolving due to the relentlessness of technological evolution. We speak of "digital constitutionalism" to point out the challenge deferred to the "old dear constitutionalism" to reframe and update its categories.[248] Turning to substantive Criminal Law, a legitimacy pronouncement comes to mind in which the Supreme Court, after examining the notion "of 'presence' with respect to modern communication systems," stated "that, alongside physical presence in unity of time and place," there are specific cases "equivalent" to it, which are integrated "with the help of modern technological systems (conference call, audioconferencing, or videoconferencing)."[249]

Why, then, albeit with all the *caveats* and distinctions required, should we aprioristically rule out the practicability of a remodulation of the "ontology of presence" in the field of procedural law as well?

## 3. …and a land (still) in between

Once we have closed the somewhat elementary "list" of functional "hooks" to approach the underlying question, the context in which we find ourselves, upon closer inspection, presents the traits of what could be called a state of "liminality." Drawing on the anthropological studies of the British author Turner, in an article effectively entitled "betwixt and between,"[250] one literally has the impression of being in a "state of transition," in a sort of "middle ground" where one can record what is no longer – the time of interdiction against the use of technological media in the performance of certain procedural activities – and what is not yet – and that can be glimpsed on the threshold, namely a (more or less) "shared" disposition to reconsider certain categories in light of a sense of reality that is also a sense of the opportunities offered by the proactive role of technology.

---

of rights requires a wider field of vision without losing focus on the social, political, and cultural context in which it is set."

[247] This is corroborated by B. BRUNELLI, *Il processo civile telematico che verrà*, in *Rivista trimestrale di diritto e procedura civile*, 2021, 979, according to whom justice is not confined to one place, as well as by recent overseas studies such as C. MCKEY, K. MACINTOSH, *Digital Criminal Courts: The Place or Space of (Post- ) pandemic Justice*, in AA.VV., *Cybercrime in the Pandemic Digital Age and Beyond*, Cham, 2023.

[248] Without any claim to exhaustiveness, see P. COSTANZO, *Il fattore tecnologico e le trasformazioni del costituzionalismo*, in AA.VV., *Costituzionalismo e globalizzazione. Atti del XXVII Convegno annuale dell'Associazione italiana dei costituzionalisti (Salerno 22-24 novembre 2012)*, Napoli, 2014, 43; E. CELESTE, *Digital Constitutionalism. The Role of Internet Bills of Rights*, London, 2022, passim; and, most recently, V. CAVANI, *Nuovi poteri, vecchi problemi. Il costituzionalismo alla prova del digitale*, in *Diritto pubblico comparato ed europeo*, 2023, 223.

[249] Cass., Sez. V, July 20, 2022, n. 28675, in Il Sole-24 Ore, July 28, 2022.

[250] More specifically, see V. TURNER, *Betwixt and Between: The Liminal Period in Rites of Passage*, in *The Forest of Symbols: Aspects of Ndembu Ritual*, New York, 1970, 46.

Let us further explain. It is true, on one hand, that neuroscientific studies, those of memory psychology and even those related to the psychology of perception, convergingly though distinctly point to a "tarnishing" not only of the *de praesenti* dialectical method in its golden rule of forensic epistemology but, likewise, of declarative evidence, which constituted the real "turning point" of that "epistemological hairpin" of 1989.[251] It is unquestionable, on the other hand, that "the lack of evidentiary basis" marks the axioms which have basically always been pivotal in the discussion around the use of audio-visual links within criminal proceedings.[252] Awaiting future landfalls of empirical and finally multidisciplinary investigations,[253] there is still a strong reliance on the postulate that only the physical co-presence of all the different protagonists in a material courtroom are able to ensure the constitutionally required standard both for participation as well as for the activity of the evidentiary process.

It is precisely from this postulate that *naturaliter* descend, starting from the early discipline on the subject of video links and recurrent attempts to keep the utilization of technological intermediation within the perimeter of procedural fairness: either through the recognition of limitations and/or exceptions to the adversarial principle, or, again, through more or less strict judgments of proportionality or balance between principles.

Lately, the suggestion that the parties' consent is a passepartout, legitimizing any derogation from cross-examination in the strong sense and any compression of a defendant's guarantees,[254] has become more prominent. This trend also comes from the example of many foreign legal experiences,[255] from the "Guidelines on videoconferencing in judicial

---

[251] G. GIOSTRA, *Appunto per una giustizia non solo più efficiente, ma anche più giusta*, in www.giustiziainsieme.it, 24 gennaio 2022. On the progressive "erosion" of the classic physiognomy of the witness statement as the main source of cognition there is great turmoil also in the foreign doctrine. In this respect, see, for example, R. TAVARES DA SILVA, *Neuroscience of Memory and Philosophy of Knowledge Challenges to Immediacy*, in H. MORÃO, R. TAVARES DA SILVA (eds.), *Fairness in Criminal Appeal. A Critical and Interdisciplinary Analysis of the ECtHR Case-Law*, Cham, 2023, 163.

[252] ... such as, merely as an example, the loss of human interaction and the effectiveness of immediacy, as well as the cognitive distortions that would undermine both a judge's decision making and the legitimacy of the trial itself. D. L. F. DE VOCHT, *Trials by video link after the pandemic: the pros and cons of the expansion of virtual justice*, in *China-EU Law Journal*, 2022, 33-44, appropriately notes how the few studies that would also confirm the existence of such *vulnera* in a virtual spatial dimension are all extremely dated (while the quality of technologies improves day by day...), and also pertain to peculiar geographic and procedural contexts (such as, for example, bail hearings in the UK). Not only that: it has been highlighted how some pilot studies conducted on children in detention in the UK and in Australia reveal how digital natives – having interfaced from birth with screens – do not experience any "discomfort" in using new technological media even during a trial process.

[253] Among other things, the contribution of the psychological sciences, and cognitive psychology in particular , is and will be increasingly essential: in this respect, we refer to the considerations made in B. GALGANI, *Forme e garanzie nel prisma dell'innovazione tecnologica*, cit., 287 and in the bibliographical references cited therein.

[254] Critical on this point is, among others, F. TRAPELLA, *La rivoluzione digitale alla prova della riforma*, in *Archivio penale*, 2022, 3, 20, for whom the consensual mechanism would be susceptible to "abuse" since, through it, defenders could be tempted to "attract the magistrate's sympathies" in the hope "that, by helping him reduce the hearing roles, the prosecutor could ask for a milder sentence and the judge, be more lenient;" in the same vein E. TRAMARECE, *Nullità per mancato funzionamento della videoconferenza e prospettive di riforma*, in *Giurisprudenza italiana*, 2021, 2794, according to whom "the consensus to which the provision subordinates the actual conduct of the hearing by videoconference" would have no other function than that of "a picklock through which the party - and evidently the reference is to the party most sensitive to the trial, namely the defendant - abdicates its defensive rights."

[255] In this regard, see also the documented review by J. DELLA TORRE, *L'espansione del processo a distanza negli itinerari dell'emergenza pandemica*, in *Processo penale e giustizia*, 2021, 226.

proceedings" of the CEPEJ[256] to the inadmissibility ruling of the Italian Constitutional Court n. 96 of 2021 which, on closer inspection, is the only one that has dealt with the status of videoconferencing in the pandemic era by endorsing the possible extension of cases in which to adopt congruous telematic modalities.[257] The Lattanzi Commission[258] itself helped "frame" the suggestion in the directive criterion (Article 1, paragraph 8, L. 134/2021) specifically dedicated to so-called remote justice.[259]

And behold, when looking at Legislative Decree n. 150/2022, one cannot fail to note how – although *expressis verbis* aimed at speeding up the procedural timelines[260] – the new legislation, hinged on the mechanism of consent, somehow sins in naiveté, condemning itself to ineffectiveness insofar as, while being *de facto* associated with the idea of a depletion, at least potentially, of certain safeguard guarantees, it isn't "accompanied" by any form of reward.

Pursuing this rapid analysis, it is worth noting that, despite the hoped-for and worthy choice of embedding a pivotal provision such as that of newly drafted Article 133-ter (preceded, in purpose, by the equally new Article 133-bis) in the body of the Criminal Procedure Code, it is not easy to trace the line of rationalization and simplification that should have guided the reforming intervention on this point.

On the other hand, instead of reshaping the overall pattern of remote participation (with regard to which, more or less consciously, a new taxonomy has also been inaugurated), some of the so-called "historical" ones have been preserved detached from the consent mechanism, and new ones have been juxtaposed, all conditioned, however, on the will of the parties.[261] For what reason, then, was the legislator made to apodictically say what they in

---

[256] Adopted at the plenary meeting on June 16 and 17, 2021, their text can be found at rm.coe.int/cepej-2021-4-guidelines-videoconference-en/1680a2c2f. Canon 21 reads as follows: "If legislation does not require the free and informed consent of the defendant, the court's decision for his or her participation in the remote hearing should serve a legitimate aim."

[257] Cf. Constitutional Court, May 12, 2021, n. 96, and the analysis on this in B. GALGANI, *Forme e garanzie nel prisma dell'innovazione tecnologica*, cit., 277. A different view of the issue, in disagreement, is given by C. MINNELLA, *Processo penale da remoto: l'intervento della Corte Costituzionale*, in www.dirittoegiustizia.it, May 13, 2021, and A. MARANDOLA, *Legislazione d'emergenza: garanzie e limiti (temporali) provengono dalle alte Corti*, in ilPenalista.it, May 17, 2021.

[258] We refer to the Study Commission for drafting reform proposals in the area of the criminal trial and penalty system, as well as in the field of the statute of limitations of crime, through the formulation of amendments to Draft Law A.C. 2435, delegating the Government for the efficiency of the criminal trial and provisions for the prompt settlement of judicial proceedings pending at the courts of appeal.

[259] For an analysis of the criteria of Delegated Law 134/2021 see, B. GALGANI, *Forme e garanzie nel prisma dell'innovazione tecnologica*, cit., 317.

[260] Cf. Relazione illustrativa to D.Lgs. n. 150/2022, 7. In this regard, it is worth recalling the constitutional pronouncement n. 74 of 2022, stating that "[t]he reasonable duration is an identifying feature of trial justice" and, consequently, the provision of somehow simplified procedural solutions; "far from meeting a logic of 'judicial efficiency that statistically privileges quantity at the expense of the quality of judicial decisions'" it rather embeds the "implementation of a precise constitutional duty:" the Constitutional Court, judgment n. 74, March 24, 2022, § 5.1.

[261] Beyond the arguable and farraginous legislative technique deployed by the Delegate, it seems unquestionable that the regime summarized in Article 133-ter of the Criminal Procedure Code is susceptible to partial derogation by *ad hoc* provisions of old and new coinage, but is not sufficient, on the other hand, to introduce new "cases" remotely, the legitimacy of which is in any case referred to a legislative determination to that effect. Precisely because of this trait of "non-self-sufficiency," the amendment suggested by the Union of Criminal Chambers (see Proposte UCPI di emendamenti al d.lgs. 150/2022, 7 febbraio 2023, the text of

fact did not say, imputing to them a phantom prohibition on virtual hearings in its most extreme declination if,[262] then and conversely, that same nomoteta opts for the admission of the remote execution of declarative evidence. Declarative evidence that – as is well known – represented a "forbidden garden" even at the height of the "pandemic storm?"[263]

What is disorienting is the delegate's "squinting" strides: while on one hand they return to "entrenching" themselves in defence of certain symbols (the sacredness, in any case, of the material courtroom...),[264] on the other they dare (albeit legitimately)[265] what not even their counterpart, invested in the contextual reform in the civil sphere (cf. d.lgs. n. 149 of October 10, 2022), went so far as to provide in the "stabilization" of remote hearings (cf. the new Article 127-bis, paragraph 1, Civil Procedure Code).[266]

As a matter of fact, provided the parties consent to it, the possibility of the remote examination of witnesses, defendants in related criminal proceedings, experts, technical consultants and private parties now appears generalized within criminal proceedings (Article 496, paragraph 2-bis, Criminal Procedure Code, and Article 422, paragraph 2, and 441, paragraph 6, Criminal Procedure Code, for preliminary hearings and summary judgments, respectively).

Affixing the "seal" of nullity to the clause requiring the implementation of a connection in a "mode adequate to protect the hearing of the parties and their effective participation in the act or hearing, and to ensure the simultaneous, effective and mutual visibility of the persons present in the different places, as well as the possibility for each of them to hear what

---

which can be found at https://www.camerepenali.it/public/file/Documenti/Documenti_Giunta_Caiazza/UCPIEMENDAMENTI-DLGS-150-2022.pdf) sought to introduce the same article at the very beginning. Referencing the "consent of the parties" misses the point: for this purpose, it is sufficient to verify how, in deference to the dictum of the delegation, each of the "new" cases of remote activity provides *expressis verbis* the requirement of consent.

[262] Cf. the "indicted" passage from the Illustrative Report to Legislative Decree 150/2022, 47: "it was excluded, on the other hand, that the delegation criterion required (or, in any case, allowed): a) to regulate the hypotheses of integral "dematerialization" of the hearing, in particular by providing that it may be celebrated in a totally virtual environment, managed by the judge and his auxiliaries from a place other than the courtroom." Sharing our censure, albeit in more nuanced tones, M. GIALUZ, *Per un processo penale più efficiente e giusto. Guida alla lettura della Riforma Cartabia*, in www.sistemapenale.it, November 2, 2022, 23.

[263] In the same sense see CORTE DI CASSAZIONE. UFFICIO DEL MASSIMARIO. SERVIZIO PENALE, *Relazione su novità normativa. La "riforma Cartabia,"* in www.cortedicassazione.it, January 2, 2023, 22 note 69.

[264] Yet regarding the "slippery, productive, and counterproductive" role of symbols "in any discussion on justice," it is worth reading the pages of G. ZAGREBELSKY, *La giustizia come professione*, Turin, 2021, 56. In a quite different vein are the considerations on the "symbolic value of the courtroom" by O. MAZZA, *Immediatezza e crisi sanitaria*, in *Rivista italiana di diritto e procedura penale*, 2021, 461.

*Per incidens*, regarding the exclusion of the "full dematerialization" of the hearing, we are to wonder what guarantee would respond to the obligatory presence of only institutional representatives, i.e. the prosecutor and judge, in the material space of the hearing once it was decided to make use *secundum constitutionem* of the derogatory clause in Article 111, paragraph 5, Constitution.

[265] A very different opinion is that of D. NEGRI, *Atti e udienze "a distanza": risvolti inquisitori di una transizione maldestra alla giustizia digitale*, in CASTRONUOVO-DONINI-MANCUSO-VARRASO (eds.), *Riforma Cartabia. La nuova giustizia penale*, Milan, 2023, 465, who, in the "[e]ventuality [...] that the entire investigative activity should suffer similar degradation to the larval state," foresees "doubts of compatibility with the constitutional paradigm of the hearing."

[266] For a synthesis of the reformatory intervention regarding the provisions of our particular interest see C. ASPRELLA, *Riforma processo civile: le nuove disposizioni in materia di udienza*, in ilProcessocivile.it, October 26, 2022 and C. MINNELLA, *Avanti con la giustizia da remoto, gli uffici alla ricerca di un equilibrio*, in *Guida al diritto*, 2023, 10, 22.

is being said by the others" is certainly worthy of approval. Again, explicit reference to the "adequate publicity" of a public hearing should be appreciated, a capital issue and yet one that has been mistreated by all emergency legislation. However, it does risk turning into a "window painted on the wall," since neither content nor standards are established.

No less astonishing is that the eventual deficiency, as Cordero might put it, of "mechanical mimesis," i.e. of the audiovisual recording with which all the acts co-participated in by subjects allocated to different sites should be documented, is unburdened by any sanction. Yet this was precisely the time to formulate provisions *ab initio,* conceived of as being "immersed" in the digital, the core of which would be an indication of the care dually directed to the legal side – and thus, simplifying, to the principles that must always be ensured in the horizon of procedural fairness – and to the study, conducted on the basis of a rigorous data-driven approach, of an infrastructural-informational one. This was the conjuncture to decline, alongside the corpus of primary provisions, a "ritual" of the procedural act "at a distance," a veritable *vademecum* geared towards giving effectiveness to subjective legal situations cast in a "multiplied spatial dimension of the trial." But the chance was evidently missed, also considering that the delegated legislator has maintained silence – to give a few examples – regarding the identification of methods for collecting and verifying the genuineness of consent given to remote celebration;[267] over the framing of the video to be ensured to different protagonists during a connection;[268] on the conditions for interpretation and translation services; on the characteristics of the police powers and hearing discipline in their digital version, as well as, further and willingly, on the protocol for the detection of technical incidents and consequent sanctioning.[269]

Besides, there are no lack of provisions or best practices from which to draw some useful suggestions and indications: alongside a sort of technical "regulation" to be found in the aforementioned CEPEJ Guidelines,[270] the contents of the recent Rules for Video Conferencing of the High Court of Bombay[271] are also worthy of attention; with a lavish

---

[267] *Ante* the implementing decree, it was expected that consent, personally expressed by the party or, at any rate, by special power of attorney, would be required: cf. E. A. A. DEI- CAS, *La partecipazione a distanza, in attesa della riforma del processo penale*, in *Sistema penale*, 2022, 4, 29.

[268] See G. DI FEDERICO, *Cari avvocati, non demonizzate l'uso delle nuove tecnologie*, in www.ilriformista.it, April 30, 2020, 1, 6. At the dawn of the "new" procedural code, he coordinated experiments in the entry of video technologies into the criminal trials under an agreement between the CNR and the Ministry of Justice. The author recalls how, already in that now-distant era in Kentucky, he witnessed the use of video footage which, in the course of the hearing, automatically positioned itself on the subjects who spoke from time to time, thus ensuring a faithful representation not only of what was reported, but also of the manner and timing in which the statements were made. The contemporary projection of those experiments could, among other things, be traced to devices that apply a system called the "active display communication system," under which the movement of the speaking subject triggers the movement of the video-link system monitor, thus amplifying the sense of reality and engagement for the interlocutor. See T. ITO, T. WATANABE, *Natural Involvement to Video Conference Through ARM_COMS*, in S. YAMAMOTO, H. MORI (eds.) *Human Interface and The Management of Information: Applications in Complex Technological Enviroments. HCII 2022. Lecture Notes in Computer Science*, Cham, 2022.

[269] See D. NEGRI, *Atti e udienze "a distanza,"* cit., 485, who agrees on the capital relevance of these aspects upon which the "degree of approximation to reality" is ultimately estimated and guaranteed by audiovisual connection systems.

[270] In this regard, allow me to refer to B. GALGANI, *Forme e garanzie nel prisma dell'innovazione tecnologica*, cit., 308.

[271] The provision, published Jan. 27, 2023, can be found at: https://hcbombayatgoa.nic.in/download/E-Filing%20rules%20video%20conferencing.pdf.

definitional and regulatory apparatus divided into five chapters, they seek to give "substance" to those guarantees that, in the digital space, must of necessity find "other" and "new" forms of expression.

Finally, as an additional point of comparison, it is not missed how, at the national level, the concomitant package concerning civil procedure, with the introduction of a provision such as Article 196-duodecies disp. att. Civil Procedure Code,[272] entitled "Hearing with remote audiovisual connections," reveals "care" and "sensitivity" to indispensable operational aspects, undoubtedly more prominent than those so far displayed in the criminal procedural area of intervention.

## 4. Future *incognitae*: the criminal trial in the metaverse?

The liminality that seems to characterise the subject, both from the point of view of scientific research and from the perspective of the regulatory framework currently in force, has already been emphasised. The intention is therefore not to raise the *peana* of a regulation – the most recent one – that has the ambition of elevating the remote trial to an ordinary *modus operandi*, but does not provide objective and subjective criteria to guide and (sometimes) impose the "remote" judicial choice.[273] However, we are reasonably persuaded that this is a new beginning, a new approach, at least *in parte qua*.

Whether willingly or not, from now on no one can (any longer) ignore the implications of a technological dynamism that, even under the specific profile of criminal investigation, in the years to come will relentlessly engage the political agenda, institutions, and forensic practitioners. In the radical transformation of the spatial and temporal coordinates of a global society such as ours – radically different from that in which the science of criminal procedure has historically been structured – it is illusory to believe we may refrain from considering this chance to reconfigure the judicial system by means of certain avant-garde technical systems according to canons of greater sustainability (in terms of resources and time spent) and the greater fairness of proceedings (in terms of easier access to justice for example).

Moreover – to conclude with an observation that, I am aware, will alarm many – we will soon be asked to discuss the unknowns posed by the progressive emergence of a new digital paradigm, the so-called "metaverse."[274]

---

[272] The article reads as follows: "1. The hearing referred to in Article 127-bis of the Code shall be held in appropriate arrangements to safeguard the adversarial process and to ensure the effective participation of the parties and, if the hearing is not public, its confidentiality. Article 84 shall apply. 2. The minutes shall state the identity of those present, who shall assure that there are no connections with non-legitimate subjects and that no non-legitimate subjects are present in the places from which they are connected. 3. Those attending shall keep the video function active for the duration of the hearing. They shall be prohibited from recording the hearing. 4. The place from which the judge is connected shall be considered a courtroom for all purposes and the hearing shall be deemed to be held in the judicial office before which the proceedings are pending. 5. The remote audiovisual connections for the proceeding of the hearing shall be determined and regulated by provisions of the General Director of Information and Automated Systems of the Ministry of Justice, and so shall the modalities through which the publicity of the hearing in which the case is heard is ensured."

[273] See D. NEGRI, *Atti e udienze "a distanza,"* cit., 483, who insists on the excess of discretion granted to judges in determining both the *an* as to the *quomodo* of recourse to audiovisual connection.

[274] We recommend, for a first and non-exhaustive survey, the following: A. E. HASSANIEN, A. DARWISH, M. TORKY (eds.), *The Future of Metaverse in the Virtual Era and Physical World. Studies in Big Data*, Cham, 2023; G. CASSANO, G. SCORZA (eds.), *Metaverse. User rights - digital platforms - privacy - copyright - criminal profiles - blockchain and NFT*, Pisa, 2023; M. A. CATAROZZO, *Il Metaverso: quali opportunità per i legali?*, in il Processotelematico.it,

First mentioned in 1992 in a novel titled "Snow Crash" by Neal Stephenson, the term "metaverse"[275] seems to refer to different types of technology[276] that, in their various combinations, promise the experience of a fully immersive reality[277] – so far tested mainly in gaming and the field of cultural heritage – which should soon be capable of enabling judges, lawyers, witnesses, and juries to "exist" in a shared digital courtroom.[278] In this virtual space, real-time communication of platforms we are already familiar with, combined with the holographic projections of each subject, could reduce to zero, or to irrelevance, the gap between the real and virtual dimensions that has so far been at the core of discussions on the constitutional tightness of any institution relying on the intermediation of technological *nova* in the performance of various procedural activities.

That this scenario, in which virtual reality (VR), augmented reality (AR), and extended or mixed reality (XR) will come together and merge, is much less hypothetical for the judiciary than imagined, is corroborated by the fact that in February 2023 the first hearing in the metaverse was celebrated in Colombia.[279] It is no coincidence that Ley 2213, adopted in June

August 23, 2022; M. MARTONE, *Prime riflessioni su lavoro e metaverso*, in *Argomenti di diritto del lavoro*, 2022, 6; V. FUMI, *Metaverso e fenomenologia*, in *Sicurezza e giustizia*, 2022, 2, 25.

[275] In this regard, it is worthwhile to consider what was written on the subject in an important document drafted by the Europol Agency: "The metaverse is often described as a hypothetical iteration of the internet as a single, universal virtual world that presents the user with an immersive experience that feels 'real,' usually through the use of a headset. In its very recent definition, it can blur the lines between the physical and virtual world to create a single blended, extended, or mixed reality. As a result, the metaverse is now focused on virtual reality (VR), but is increasingly being defined in terms of augmented reality (AR) or extended or mixed reality (XR)." See EUROPOL, *Policing in the metaverse: what law enforcement needs to know, an observatory report from the Europol Innovation Lab*, Luxembourg, 2022, the text of which is available at:
https://www.europol.europa.eu/cms/sites/default/files/documents/Policing%20in%20the%20metaverse % 20-%20what%20law%20enforcement%20needs%20to%20know.pdf.

[276] For a review see also A. FALCONE, *Online Hearings and the Right to Effective Defence in Digitalised Trials*, in L. BACHMAIER WINTER, S. RUGGERI (eds.), *Investigating and Preventing Crime in the Digital Era*, Cham, 2022, 192.

[277] ... thanks to special VR viewers or, in any case, to special devices conveying perceptions linked to the five senses.

[278] See D. PUMPHREY JR., *How the Metaverse Could Change Criminal Defence Forever*, March 12, 2022, that can be found in https://www.pumphreylawfirm.com/blog/how-the-metaverse-could-changecriminal- defense-forever/.

[279] A. GUTHRIE, *Justice in the Metaverse: Here's What the First Virtual Court Hearing in Colombia Looked Like*, in https://www.law.com/international-edition/2023/02/19/justice-in-the-metaverse-heres-what-thefirst- virtual-court-hearing-in-colombia-looked-like/?slreturn=20230508014634, February 19, 2023; C. BELLO, *Future of justice: Colombia makes history by hosting its first-ever court hearing in the metaverse*, in https://www.euronews.com/next/2023/03/01/future-of-justice-colombia-makes-history-by-hosting-its-firstever- court-hearing-in-the-me, March 1, 2023; S. FORTIS, *The metaverse is testing the limits of what is legally possible*, available at:
https://cointelegraph.com/news/the-metaverse-is-testing-the-limits-of-what-is-legallypossible?_ga=2.112824077.808954691.1686203943-465565092.1686203943, March 3, 2023; L. PEYRON, *Viene il giudizio nel Metaverso. Ma verrà alla fine dei giorni*, in www.avvenire.it, March 1, 2023. Hearings have also already been held in the Metaverse in China, according to the following agency reports: *A Chinese local court recently opened a hearing in the Metaverse, saying it helps drive the digitization of the judicial system*, in https://en.pingwest.com/w/10840; *The First Metaverse Trial Opened in a Local Court in China*, November 11, 2022, in https://odr.info/the-first-metaverse-trial-opened-in-a-local-court-in-china, September 30, 2022. The openness of the Chinese procedural system to technical progress is also reflected in I. AINORA, *Smart courts: verso una nuova dimensione della giustizia*, in ilProcessotelematico.it, July 30, 2021.

2022,[280] legitimised the use of advanced information technology (without further specifications) in all the jurisdictions belonging to the system. With particular regard to the criminal plexus, then, Article 7 allowed for the use of "*los medios tecnológicos a disposición de las autoridades judiciales*" for the purpose of gathering evidence, unless the parties disagreed.

Needless to say, with the *illico et immediate* aim of banishing even the "spectre" of a technical solution so distant from the usual physiognomy of the courtroom and so "disengaged" from the traditionally conceived canon of the unity of material space, we could limit ourselves to highlighting a first and glaring critical issue. As we can infer from videos of the Colombian trial, the use of avatars with highly stylised, cartoonish likenesses,[281] excessively reduces the facial expressions, signs, and body language of those intervening and effectively lowers (instead of raising) the level of "fidelity" with reality, compared to more ordinary modes of video communication. Now, such concerns and discomfort might also be mitigated if the case were civil or administrative, with a relatively simple subject matter such as, for example, the collection of traffic fines. It is quite another crucial issue, however, in circumstances where this "mimetic imperfection" could be drawn upon in judgments discussing a defendant's freedoms and where the "anthropological" weight of the identity traits of the individual cannot be so dismissed.

And yet, are we confident enough to definitively clear the field of any discussion about "relocating" the dynamics of criminal proceedings (or of selected segments thereof) into what could be the "new world"? Given the amount of investment big companies are pouring into a sector with "monstrous" economic potential worldwide, this initial, "embryonic" phase will be quickly overtaken by the provision of highly refined hardware and devices[282] that are capable of providing a true "digital clone" of the time and space in which we are accustomed to allocating ourselves, and at an affordable cost: *Quid iuris* then?

Whatever the answer concerning the procedural uses of a "frontier" technical application (predictably not for much longer), it must be balanced between known boundaries and limits that can be overcome and/or surpassed, so we come full circle to the opening remarks on the "limiting" vocation of the title.

We are all too aware that a characteristic trait of humanity and thus of science intended as knowledge, is the tension towards the "vertigo" of the journey beyond the Pillars of Hercules.[283] However, if an error can be ascribed to Ulysses and his men, it was not the act of crossing the passage itself, but in thinking they could cross the ocean with the same "means" used to navigate the shores of the *mare nostrum*. It is from this deadly "blunder" that we face the new and increasingly ambitious challenges posed by digitisation, which must

[280] Ley 2213 de 2022 "Por medio de la cual se establece la vigencia permanente del decreto legislativo 806 de 2020 y se adoptan medidas para implementar las tecnologías de la información y las comunicaciones en las actuaciones judiciales, agilizar los procesos judiciales y flexibilizar la atención a losS usuarios del servicio de justicia y se dictan otras disposiciones, available at https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=187626.

[281] B. HELLER, D. CASTAÑO, *Artificial Intelligence, Virtual Courts, and Real Harms*, in www.lawfareblog.com, March 14, 2023.

[282] Evidence of the potential underlying these new technologies, such as the possibility of cancelling any latency in the communication of images and sounds, can be found by watching an experiment conducted by Vodafone and available at: https://www.youtube.com/watch?v=OLCI8VT2QLU.

[283] In this context, see the fascinating reflections on the elusive "little brothers" of black holes by physicist C. ROVELLI, *Buchi bianchi. Dentro l'orizzonte*, Milan, 2023, passim.

never be regarded as a "goal" in itself, but rather as a "tool" for strengthening the guarantees "supporting" the legal edifice as a whole.[284]

In an effort to avoid well-worn arguments that aim to polarise the discussion, between "utopia to chase or dystopia to shun,"[285] our approach to themes such as the so-called "Metaverse Courtrooms" demands abandonment of aprioristic fears and reluctance.[286] From experimentation that requires the indispensable contribution of scholars and professionals of all the forms of knowledge involved, the focus of investigation must hinge on the level of the "imitative capacity" of the material world, to be found in the individual technological solutions taken under consideration. By testing and monitoring their use on a large scale, we can "measure" the type and degree of impact of these technologies on the exercise of law and, more specifically, upon the functionality of those subjective legal situations that form the hardened core of constitutional rights.

Only on the basis of a scientific and operational "protocol" such as that roughly outlined, will it be possible to dissolve *funditus* reservations about the appropriateness of a (even only partial and selective) redevelopment of the places and modes of justice. Thus, it will be possible, in a broader sense, to understand whether a combination between criminal trials and technological innovation is capable of respecting the personalist tradition of which the former is made, or conversely, is condemned to represent a deplorable betrayal of it.

[284] See also recently B. JANSEN, A. T. M. SCHREINER, *Captured by Digitization*, in *International Journal for the Semiotics of Law*, December 1, 2022.
[285] D. NEGRI, *Atti e udienze "a distanza,"* cit., 451.
[286] G. DE MINICO, *La tecnologia non si governa con la paura, ma con regole ad hoc*, in Il Sole-24 Ore, April 12, 2023.

# Section 7

# Case law and hypothetical cases: training materials

This section follows a practical approach by offering a set of training materials that deal with automated decision-making (ADM) in various contexts starting from existing case law and that provide insights and basic principles that could be relevant to parallel cases concerning the use of AI in the field of justice. This section is meant to foster discussion within trainings without suggesting any immediate transplants of decisions across sectors: another issue worth debating. With this in mind, some of the discussed decisions directly address the use of AI in judicial activities, such as the first contribution, which begins with a recent UK court's decision on the use of predictive algorithms by parties and the judge in what is being called e-discovery. Other decisions, despite not directly addressing AI and justice, pinpoint some crucial principles that could play a central role when similar AI tools are employed in the judicial system. From this perspective, the second contribution highlights the importance of the *SCHUFA Holding and Others* (C-634/21) case where the CJEU assessed whether and to what extent data processing by means of an AI system can be qualified as support for an automated decision-making mechanism under Article 22 GDPR. The German case decided by the Federal Constitutional Court provides an application of the proportionality test when AI-based systems are used to protect public order, but such objectives conflict with the fundamental rights of citizens. The third contribution begins with two cases decided by the Italian Supreme Court in 2021 and 2023 on reputational rating to investigate problems regarding data protection when it comes to the use of AI in decision-making. This leads to relevant considerations on the right to consent, transparency, and dignity and how it relates to automated decision making, which may have some general relevance beyond the specific context of the examined rulings and be valid for the judiciary as well. The fourth contribution starts with a decision by the Italian Data Protection Authority, that has urgently ordered OpenAI LLC to limit the processing of personal data of data subjects established in the Italian territory by means of ChatGPT, pursuant to Article 58, paragraph 2, letter f), of the GDPR. This aims to shed light on the connection between the traditional categories of law and new notions derived from technological implementation. In the fifth contribution a case decided by an Italian court in 2020 is the starting point for an analysis of the various uses that algorithms can perform in decision-making, with a particular focus on the right to equality and non-discrimination, and of the risks to its protection due to biases in both the system and within the context in which the algorithm is deployed.

This section is also complemented by the illustration of some hypothetical cases coupled with a set of questions that may be useful for designing judicial training curricula or sessions.

# 7.1. AI as a supportive tool in evidence appreciation in Civil Litigation

*Laura Piva – Università degli Studi di Trento*

*Summary: 1. Premise – 2. The case: High Court of Justice (UK), Pyrrho Investments Limited v MWB Property Limited [2016] EWHC 256 (Ch) – 3. Judicial dialogue on predictive coding – 4. Human intelligence versus artificial intelligence in e-Discovery – 5. Jurists-in-the-loop – 6. Hands-on scenario: a hypothetical case on AI for detecting unfair contract terms*

*Abstract: Beginning with the analysis of some judgments concerning the use of predictive coding for e-discovery, this contribution aims to explore the potentialities and perils of using AI as an assistive tool in civil proceedings. This topic is relevant to judges for at least two reasons. First, as digitalisation leads to the production of more and more electronic documents, judges can reasonably expect parties to use AI technology to absolve their duties to disclosure. They therefore need to understand how algorithms work and when their use can be helpful, reasonable, and legitimate. Secondly, AI tools that can help the judiciary do exist. It is thus time to reason on whether judges can use them in their daily tasks. For instance, judges could in principle conduct better legal research by using AI-powered search engines or applications which summarize relevant precedent. Or they could benefit from the aid of intelligent tools which "pre-read" parties' claims and documents they have produced to promptly extract information relevant to the controversy. This renews the idea that judges (and, more generally, legal professionals) must understand how systems work, as well as their limits, in addition to the principles according to which they can and shall be used.*

## 1. Premise

The rise of artificial intelligence (AI) in judicial activities, particularly in civil litigation, is reshaping the way judges and legal practitioners manage evidence and its disclosure. Predictive coding applied to e-discovery has become a focal point of this transformation. As courts are increasingly dealing with large volumes of electronically stored information (ESI), AI tools such as predictive coding offer solutions to handle these data-heavy cases more efficiently. However, the introduction of such technologies raises fundamental questions about the balance between human judgement and machine assistance. This contribution will explore how courts have approached the role of AI in e-discovery, starting with emblematic cases, such as the English High Court decision in *Pyrrho Investments Limited v MWB Property Limited*, and examine the wider implications of integrating AI into judicial processes. As digitisation continues to impact the legal professions, judges must not only assess the appropriateness of these technologies, but also develop an understanding of their potential and limitations, ensuring that AI is used as a tool to support, rather than replace, human decision-making.

## 2. The case: High Court of Justice (UK), *Pyrrho Investments Limited v MWB Property Limited* [2016] EWHC 256 (Ch)

In *Pyrrho Investments Limited v MWB Property Limited*, the English High Court of Justice addressed the issue of using AI – namely predictive coding – during the discovery phase of civil litigation. This was a landmark decision as it was the first time English judges have dealt

with the topic. The court ruled in favour of exploiting algorithms for electronic disclosure (e-disclosure) based on the reasons summarised hereafter.

In Common Law jurisdictions, discovery is a pretrial phase during which each party can request that the other disclose relevant evidence in their possession. Under English Civil Procedure Rules (CPR), standard disclosure obliges the parties not only to provide the documents upon which they will rely, but also to "make a reasonable search" for other disclosable documents.[287] This can be problematic when a huge quantity of documents is involved, as in the case at issue.

In fact, the files held by MWB Property Ltd originally amounted to 17.6 million and were then narrowed down to 3 million. Hence, in *Pyrrho Investments Limited v MWB Property Limited,* the parties asked the court to allow them to use predictive coding to search for disclosable documents among MWB's files. In this way, the search would not be carried out by humans but, rather, by an algorithm that "analyses documents and "scores" them for relevance to the issues of the case."[288]

Predictive coding is an AI technique that can examine a great volume of documents and sort them into predetermined categories (e.g. "relevant to the case" or "not relevant"). This is possible as the algorithm has been fed with a set of coded documents from which it can infer which terms and concepts to look for in other documents.[289] Files are then ranked and categorised according to relevancy.

The High Court approved the parties' request to use predicting coding. In the reasoning, Master Matthews considered that other techniques, such as keyword search, were already allowed. Nonetheless, keyword search and manual review are more time-consuming and costly, especially when a vast quantity of documents is to be analysed. Predictive coding on the other hand saves time and resources, as the cost of automated review does not increase at the same rate as the number of documents to be analysed.

The court pointed out that "there is no evidence to show that the use of predictive coding software leads to less accurate disclosure being given than, say, a manual review alone or keyword searches and manual review combined, and indeed there is some evidence to the contrary."[290] Additionally, using an algorithm can lead to greater consistency than would resorting to several different assistants or paralegals.

Lastly, the parties agreed to utilise the software and no English law provision was found contrary to such use.

For all these reasons, the Court approved the parties' request.

## 3. Judicial dialogue on predictive coding

A similar reasoning can be found in two judgments cited in *Pyrrho*: the US case *Moore v Publicis Groupe* and the Irish *Irish Bank Resolution Corporation Ltd v Quinn.*

[287] According to CPR Rule 31.7, they are limited to "the documents which (i) adversely affect his own case; (ii) adversely affect another party's case; (iii) support another party's case" (Rule 31.6 b) and "the documents which he is required to disclose by a relevant practice direction" (Rule 31.6 c).

[288] UK High Court of Justice, *Pyrrho Investments Limited v MWB Property Limited* [2016] EWHC 256 (Ch), paragraph 17.

[289] C. GIDEON, *Predictive coding: adopting and adapting artificial intelligence in civil litigation*, in *SSNR*, 2020.

[290] UK High Court of Justice, *Pyrrho Investments Limited*, cit. paragraph 33.

In *Moore v Publicis Groupe,*[291] the District Court of the Southern District of New York found the use of predictive coding appropriate due to: (1) the parties' agreement, (2) the vast amount of ESI to be reviewed (over three million documents), (3) the superiority of computer-assisted review to the available alternatives (i.e. linear manual review or keyword searches), (4) the need for cost-effectiveness and proportionality under US Law and (5) the transparent process proposed by the parties to train the predictive coding algorithm.

In *Irish Bank Resolution Corporation Ltd v Quinn*[292] the situation was somewhat different, as only the plaintiffs sought to use predictive coding for e-discovery. Instead the defendants contested this request due to doubts concerning the accuracy of the technique and its trustworthiness. The Irish High Court noted how the parties' agreement to use an AI system was highly preferable as it gave them "an added degree of comfort that a failure of the system to throw up a relevant document will be more likely to be viewed as unfortunate but unavoidable rather than a deliberate act."[293] However, the court authorized e-discovery to be carried out through predictive coding since the plaintiffs proposed a transparent and reliable protocol for deploying predictive coding and sought the defendants' consent and participation in the training process. In particular, the judges held that "a balance must be struck between the right of the party making discovery to determine the manner in which discovery is provided and participation by the requesting party in ensuring that the methodology chosen is transparent and reliable."[294]

Transparency thus appeared to be fundamental in respecting the parties' right to participate in proceedings.[295] In the case at issue, the High Court was satisfied overall that the proposed methodology would, on one hand, lead to accurate and faster results and that it did not pre-judicate the defendants' right to participation on the other.

## 4. Human intelligence versus artificial intelligence in e-Discovery

The judgments analysed all subtend the same question: in the digital age, what should humans and what should algorithms do?

Of course, such a question is not limited to e-discovery or AI applications in the field of justice, but it characterizes the entire debate concerning the use of AI. In fact, human intelligence and artificial intelligence are good at doing many different things.[296]

Returning to e-discovery, it is undoubtedly the case that digitalization has led to an exponential increase in the number of electronic documents. Electronic stored information (ESI) thus constitutes a great source of evidence in civil litigation. As noted above, this can

[291] US District Court, Southern District of New York, *Da Silva Moore et al., v Publicis Groupe et al.*, 11 Civ 1279 (ALC)(AJP), 2012.

[292] High Court of Ireland, *Irish Bank Resolution Corporation Ltd v Quinn* [2015] IEHC 175.

[293] *Ibidem*, *cit.* paragraph 68.

[294] *Ibidem, cit.* paragraph 68

[295] High Court of Ireland, *Irish Bank Resolution Corporation Ltd v Quinn,* cit.

[296] One of the reasons AI is so appealing in many fields lies in the promise of yielding decisions which are more accurate than those made by humans. According to Sunstein, well-trained algorithms are good at predicting outcomes and can even correct human cognitive bias (C.R. SUNSTEIN, *Algorithms, correcting bias*, in *Social Research: An International Quarterly*, 2019, 86, 499). Gigerenzer also recognizes that algorithms are good at estimating risks. According to the author, however, humans still handle uncertain situations better (G. GIGERENZER, *How to Stay Smart in a Smart World: Why Human Intelligence Still Beats Algorithms,* 2022).

pose a huge burden on parties during discovery, especially when claims concern companies. In fact, even a relatively small business might need to review millions of files among emails, electronic documents (i.e. PDFs, word documents, excel spreadsheets), cloud-based storage, and computer hard drives. When this activity is done manually, identifying relevant documents can be tedious and expensive, as it will likely require employing several people for the task. Moreover, human beings get tired, especially when performing repetitive tasks such as sifting through large numbers of files. Thus, human performance might not always be consistent or accurate. Even if they use keyword search tools, results might not get better. In fact, keyword searches can generate both under-inclusive and over-inclusive results, leading to ignoring relevant documents or not being useful at narrowing them down.[297] In the first case we would then have errors. In the second, using the tool would not be beneficial, as it would not effectively decrease the number of documents that lawyers, paralegals, and assistants must review.

AI, on the other hand, not only does not get bored repeating the same task, but it uses its previous experience to refine its outputs. ML algorithms can quickly analyse a bulk of data to extract correlations and patterns which are used to elaborate an output. Once the training is perfected, they can produce fast results, with greater accuracy and consistency than human beings.

In the judiciary setting this means that, *if* trained correctly, AI could dramatically reduce pre-trial wait times and overall litigation costs. Conversely, poor algorithm training or design translates into replicating and amplifying the same errors, leading to opposite outcomes. This leads to a key idea that pervades the debate on AI: that it is impossible to create "intelligent" algorithms without human intelligence and creativity.

If we consider applications that are tailor-made for the judiciary – as a predictive coding tool for e-discovery – such intelligence will have to be technical and legal. For predictive coding to work, somebody with a high level of legal knowledge must select relevancy criteria – which will vary from case to case – and then review a (relatively) limited number of documents accordingly. Not surprisingly, *Pyrrho*'s judges held that this task should be performed by "a single, senior lawyer who has mastered the issues in the case."[298]

## 5. Jurists-in-the-loop

In *Moore v. Publicis Groupe* the judges underlined that "[AI] technology exists and should be used where appropriate, but it is not a case of machine replacing humans."[299] Stressing the non-substitutability of humans clarifies that humans and machines should work together to reach satisfactory outcomes. Thus, judges and lawyers should: (1) use AI, where suitable for their purposes, and (2) use it as an assistive tool only. The second point means that AI should be used as a support, for instance, to gather information that will be useful for the final decision, to conduct legal research, or to make predictions.

Academic literature and soft law dealing with AI strongly advocate for "keeping humans in the loop," meaning that they shall maintain an active role in the algorithmic decision-making process. This principle finds one of its expressions in "human oversight," which has been recognised by Article 14 of the AI Act.[300]

---

[297] See C. GIDEON, cit.

[298] UK High Court of Justice, *Pyrrho Investments Limited,* cit., paragraph 20.

[299] US District Court, Southern District of New York, *Da Silva Moore et al.,* cit., 17.

[300] The article requires high-risk systems to be "effectively overseen by natural persons" during their functioning.

Two considerations must be drawn here. First, this might not always be easy due to AI's opacity. To mitigate this risk, transparency and explainability are key. Indeed, effective oversight can be exercised only where the process followed by the algorithm is understandable. However, explainability might not be enough to overcome "algorithmic appreciation" bias, which refers to the tendency to obliterate AI decisions. Secondly, human oversight not only refers to checking the final output. It also encompasses deciding whether to use this technology in a specific situation and at a certain procedural phase. For instance, a judge might be tempted to use ChatGPT to get "technical" data and use these elements for quantifying damages.[301] In this way, the judge could use the platform almost like an expert opinion. But would he be entitled to do so? Furthermore, human oversight could imply playing an active role in training the algorithm, as in the example of e-discovery.

These considerations inevitably call for jurists to educate and reskill. AI will not only change how they perform their tasks but also the type of tasks they perform. For instance, lawyers would not review millions of documents manually anymore, but they might instruct algorithms to do so. However, for this to be possible, some digital and AI literacy is indispensable.

All the issues highlighted so far are, of course, not limited to the realm of e-discovery. In the following paragraph, a fictional case concerning the use of an AI system by a judge will be presented, together with some questions that aim to stimulate critical thinking.

## 6. Hands-on scenario: a hypothetical case on AI for detecting unfair contract terms

A private company has developed a software capable of identifying unfair terms in contracts in accordance with relevant EU and national legislation. The system, which exploits NLP, was specifically trained on contracts that presented unfair terms and on CJEU jurisprudence. According to the company's website, its goal is to help judges save time by avoiding the need to examine long and complex contracts. Thus, judges will only need to upload the contract at issue on the platform and wait a few minutes to find out whether some of its clauses might be unfair. After doing so, the software links the CJEU case law that might be relevant to the case and suggests whether the contractual provisions shall be deemed fair or not. The website gives assurances that the software is highly reliable and that its accuracy improves over time.

A judge assigned to a dispute between a consumer and a telecommunications company decides to verify *ex officio* whether the contract they concluded contains unfair terms. To do so, he decides to use the algorithms described above. According to the algorithm, there are several contractual provisions which, together, might create a significant imbalance to the detriment of the consumer. The algorithm presents the summaries of some CJEU decisions deemed to be similar to the matter of the case. On these bases, the system concludes that the terms are in fact unfair. The judge invites the parties to clarify their positions in the adversarial process and, on that occasion, states that the contract contains unfair terms.

Please refer to the following questions:

[301] This has recently happened in the Netherlands. See: https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBGEL:2024:3636&showbutton=true&keyword=chatgpt&idx=1.

1) *Transparency and explainability for the parties*: Can the judge use the software without informing the parties or does he have an obligation to disclose its use? Must the judge explain the algorithm's functioning during the adversarial process between the parties? Does the judge have to provide reasons why he decided to use assistive AI to perform this task?
2) *Transparency and explainability in the final decision*: When providing his motivations does the judge need to include details on how and the extent to which AI supported his decision-making process? Does he have to explain the functioning of the AI and why he relied on its outputs?
3) *Privacy*: Does the fact that the judge uploads the contract on the platform raise privacy concerns?
4) *Oversight and Accountability*: Can the judge autonomously decide to use the software, or must he be authorised to do so (e.g. by the Court)? Which type of oversight shall the judge put in place when using this system in order to correct for possible errors?

# 7.2. Automated processing, AI, and enforcement between the jurisprudence of the CJEU and the national courts

*Federica Casarosa – Scuola Superiore Sant'Anna*[302]

*Summary: 1. Premise – 2. The proceedings: SCHUFA Holding and Others (Scoring), Case C-634/21 | SCHUFA Holding and Others (Discharge from remaining debts), Joint Cases C-26/22 and C-64/22 – 2.1 Relevant legal sources – 2.2. Preliminary questions referred to the CJEU – 2.3. Reasoning and conclusions of the Advocate General – 2.4. Elements of judicial dialogue – 3. The proceedings: Federal Constitutional Court - Judgement (2023). [1 BvR 1547/19, 1 BvR 2634/20] – 3.1. Reasoning and conclusions of the court. – 4. Hands-on scenario: a hypothetical case*

*Abstract: The use of AI systems has still not been the object of case-law in the European courts, but some guidelines do emerge from existing decisions addressing automated decision-making systems. Important elements are present in the pending SCHUFA Holding and Others (C-634/21) case in particular, where the CJEU was asked to evaluate whether a credit scoring system can be qualified as an automated decision-making mechanism, pursuant to Article 22 of the GDPR. At the national level, the courts have addressed AI-based cases mostly in relation to the use of predictive or mass surveillance tools adopted by law enforcement authorities, among which the most interesting and recent example is the decision by the German Federal Constitutional Court on the use of Palantir surveillance software by the police. The relevance of these cases for judicial decision-making is evident: as AI tools become more and more integrated into legal processes, it becomes essential to assess how automation impacts judicial impartiality, transparency, and accountability.*

## 1. Premise

The use of AI systems has still not been the object of case-law in the European courts, but some guidelines do emerge from existing decisions addressing the use of automated-decision making systems. The Strasbourg court for example has addressed two fundamental rights, namely privacy and fair trial guarantees. On the Court of Luxembourg side, a first input comes from the SCHUFA Holding and Others (C-634/21) case, where the CJEU was asked to evaluate whether a credit scoring system could be qualified as an automated decision-making mechanism, pursuant to Article 22 of the GDPR. Although it did not address the implementation of AI in the field of justice, the case provides important guidelines to be developed in sectors that are beyond the original, including in the administration of justice. In particular, the decision stressed the importance of how much weight automated decision-making systems should have in the decision of the credit institution. The higher the reliance on the results of the automated system for the decision made by the credit institution, the higher the risks to the rights and freedoms of the citizen subject to such a decision. This principle is highly relevant to judicial contexts, as judges may come to increasingly rely on AI-based systems for case analysis or sentencing decisions. Ensuring proper human oversight and safeguarding judicial discretion will be crucial to preventing over-reliance on automated processes and ensuring fairness in judicial outcomes.

[302] The research activity was supported by SoBigData.it project (Prot. IR0000013 – Call n. 3264 of 12/28/2021) initiatives seeking to train new users and communities in the usage of research infrastructure (SoBigData.eu).

Correspondingly, when devising an AI-based system to support judges in decision-making, human intervention or, in the words of the AI Act,[303] human oversight and the measures to implement it are therefore crucial.

Similarly, national courts have mostly addressed AI-based cases in relation to the use of predictive or mass surveillance tools by law enforcement authorities, the most interesting and recent example of which is the decision by the German Federal Constitutional Court on the use of Palantir surveillance software by police. The reasoning of the court demonstrated how the proportionality test is applied when AI-based systems are used to protect public order (given the support to police authorities in the quick and effective interpretation of data), but such objectives conflict with the fundamental rights of citizens.

The joint analysis of the abovementioned cases poses some questions:

- Is automated decision making, including profiling, prohibited pursuant to Article 22 of the GDPR?
- Under which conditions is automated decision making lawful? What type of decisions can be subjected to automated decision-making? How much human intervention is required in automated decision-making?
- What are the limits for the collection and use of personal data as training data for an automated decision-making system?
- Which types of safeguards should be adopted in order to guarantee transparency, individual protection, and oversight?

## 2. The proceedings: SCHUFA Holding and Others (Scoring), Case C-634/21| SCHUFA Holding and Others (Discharge from remaining debts), Joint Cases C-26/22 and C-64/22

The applicant was a German citizen asking for a loan to a credit institution. The credit institution, before deciding on his request, asked SHUFA Holding, a private credit information agency, to provide information on the creditworthiness of the consumer. The evaluation of SHUFA was based on mathematical statistical methods. The credit loan institution, on the basis of the evaluation received, then decided to refuse his request.

The German citizen submitted a request to SHUFA requesting the data collected and, eventually, to erase incorrect entries. Additionally, the citizen requested further detailed information on the logic adopted by SHUFA in processing the data, as well as its consequences. SHUFA did not provide the requested information, only limiting its reply to the relevant score and the principles underlying the calculation method of the score in very general terms. The justification was that the calculation method could not be disclosed as it was a trade secret.

Unsatisfied by the reply, the citizen decided to present a claim before the Hesse Commissioner for Data Protection and Freedom of Information (the Land Data Protection Authority), in order to receive more detailed information about the calculation method and the processing activity. The DPA decided not to take action against SHUFA, so the citizen appealed the decision before the Administrative Court of Wiesbaden.

The administrative court decided to stay the proceeding and referred a preliminary question to the CJEU on the interpretation of Article 22 of the GDPR.

[303] See above section 3.2.

### 2.1. Relevant legal source

Article 22 of the GDPR,[304] entitled "Automated decision making, including profiling," has never previously been the object of a decision by the CJEU, so its interpretation was extremely relevant and provided useful guidelines for national courts. It is important to note that several commentators have already raised doubts regarding the interpretation of this provision.

A first issue relates to the qualification of paragraph 1 as a general prohibition – applicable regardless of the activity or request by the data subject – or as a right, which requires the data subject to actively invoke the right. Some guidelines are available from the WP29 Guidelines on Automated individual decision-making and Profiling[305] which affirm that "the term "right" in the provision does not mean Article 22(1) applies only when actively invoked by the data subject. Article 22(1) establishes a general prohibition for decision-making based solely on automated processing. This prohibition applies whether or not the data subject takes action regarding the processing of their personal data."

A second element is the fact that the article only mentions automated "individual" decision-making. Thus, it may be possible it is not applicable to decisions affecting multiple data subjects or groups of individuals. This could be counterproductive, as the current application of machine learning and Big Data are more and more applicable to group decision-making.

Another element discussed is the fact that Article 22 of the GDPR provides for a decision "solely" based on automated decision-making. If the data processing allows for human intervention in any part of the procedure, then it is up to the court to verify whether the human intervention was "meaningful" or merely a procedural "token gesture."[306]

Additionally, the norm is only applicable to automated decisions that either produce effects in the legal sphere of data subjects or significantly affect them in a similar manner. According to the WP29 guidelines, "similar effect" can be interpreted as significantly affecting the circumstances, behavior, or choices of data subjects.[307]

---

[304] "1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2. Paragraph 1 shall not apply if the decision: a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; b) is authorized by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or c) is based on the data subject's explicit consent."

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place."

[305] WP29, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 17/EN WP251 rev.01, February 6, 2018.

[306] L. A. BYGRAVE, *Article 22*, in C. KUNER, L. A. BYGRAVE, C. DOCKSEY (eds.), *The EU General Data Protection Regulation (GDPR): A Commentary*, 2020, Oxford, 532.

[307] The guidelines include the following examples: i) decisions that affect someone's financial circumstances, such as their eligibility to credit; ii) decisions that affect someone's access to health services; iii) decisions that

### *2.2. Preliminary questions referred to the CJEU*

The Administrative Court presented the following preliminary ruling:

"(1) Is Article 22(1) of the [GDPR] to be interpreted as meaning that the automated establishment of a probability value concerning the ability of a data subject to service a loan in the future already constitutes a decision based solely on automated processing, including profiling, which produces legal effects concerning the data subject or similarly significantly affects him or her, where that value, determined by means of personal data of the data subject, is transmitted by the controller to a third-party controller and the latter draws strongly on that value for its decision on the establishment, implementation, or termination of a contractual relationship with the data subject?

(2) If Question 1 is answered in the negative, are Article 6(1) and 22 of the [GDPR] to be interpreted as precluding national legislation under which the use of a probability value – in the present case, in relation to a natural person's ability and willingness to pay, in the case where information about claims against that person is taken into account – regarding specific future behavior of a natural person for the purpose of deciding on the establishment, implementation, or termination of a contractual relationship with that person (scoring) is permissible only if certain further conditions, which are set out in more detail in the grounds of the request for a preliminary ruling, are met?"

From the presentation of the case, the national court identified a crucial role in the decision of whether a loan request can be accepted based on the score rating of the credit information agency. According to the court, the evaluation (credit score) provided by the credit information agency was already a decision based on automated processing, which was then included in the decision process of the bank regarding the possibility of receiving a loan. The national court sought assessment by the CJEU on the qualification of the automated establishment of a probability value concerning the ability of a data subject to service a loan in the future (profiling) as a decision based solely on automated processing pursuant to Article 22 of the GDPR.

The current analysis will focus on the first question of the preliminary ruling.

### *2.3. Reasoning and conclusions of the Advocate General*

As a first assumption, the Advocate General (AG) Priit Pikamäe clarified that Article 22 must be interpreted as a general prohibition, which does not require the data subject to invoke the right (paragraph 31). Then, the AG affirmed that this general prohibition applies only under specific circumstances, namely when a "decision" produces "legal effects" on the data subject (or "similar significant effects"). According to the AG, the concept of a "decision," in the absence of a legal definition, should be interpreted broadly, thus it can be qualified as a position that has binding character (paragraph 37-38).

The AG then evaluated whether such a decision would have a serious impact on the data subject. In this matter, it was clear that automatic refusal of an online credit application was a case of relevant impact, since it was also included as an example in Recital 71 of the GDPR. The AG acknowledged that it was not possible to identify whether a credit score result can be qualified as a single decision vis-à-vis that by the credit institution in general terms, as the

deny someone an employment opportunity or put them at a serious disadvantage; iv) decisions that affect someone's access to education, for example university admissions.

decision-making process can include several steps that may lead to a final decision by the credit institution. However, the crucial aspect highlighted by the AG is the fact that the decision affected the data subject, in this case, "a negative score can, in itself, produce detrimental effects for data subjects, namely by limiting significantly the exercise of their freedoms or even by stigmatizing them in society" (paragraph 43).[308] Thus, the score itself was a "decision" for the purposes of Article 22 of the GDPR.

As a result, the only situation in which this decision is legally acceptable is when it is not based "solely" on automated decision-making, i.e. when human intervention plays a significant role in the evaluation of the credit score. As a general rule, this aspect should be evaluated on a case-by-case basis by national courts; however, relying on the information provided by the national court, the AG affirmed in the case at stake that "the score established by a credit information agency and transmitted to a financial institution generally tends to predetermine the financial institution's decision to grant or refuse to grant credit to the data subject, such that this position must be considered only to have purely formal character in the process" (paragraph 47).

The AG concluded that human intervention should have a relevant role, by verifying the results of the score for instance and the accuracy of the decision to be made with respect to the credit applicant; if that intervention is merely a token gesture, then the score itself constitutes a decision pursuant to Article 22(1) of the GDPR, even if the final decision to grant or deny credit is formally made by the financial entity.

The AG justified this wider interpretation of Article 22 of the GDPR by the fact that a restrictive interpretation would instead generate a gap in legal protection. Article 15(1)(h) requires that the financial entity that "formally adopted the decision" provide information on the automated decision-making process. However, if the financial entity only received the credit score result, it would be unable to provide substantial information because it did not have it. Thus, finding the credit information agency liable by virtue of its generation of the score – and not by virtue of its subsequent use – was the most effective way of ensuring the protection of the data subject's fundamental rights. As a matter of fact, the AG clarified the scope of Article 15(1)(h) of the GDPR, according to which the obligation to provide meaningful information about the logic applied must be understood to include sufficiently detailed explanations: i) about the method used to calculate the score; and ii) the reasons for a given result (paragraph 58). Thus, the controller must provide the data subject with general information, particularly on the factors taken into account for the decision-making process, and on their respective weight at an aggregate level, which is also useful for challenging any decision within the meaning of Article 22(1) of the GDPR.

According to the previous evaluation, the AG affirmed that:

"1. Art. 22(1) of the GDPR must be interpreted as meaning that the automated establishment of a probability value concerning the ability of a data subject to service a loan in the future constitutes a decision based solely on automated processing, including profiling, which produces legal effects concerning the data subject or similarly significantly affects him or her, where that value, determined by means of personal data of the data subject, is

[308] The AG affirmed that "credit applicants are affected from the stage of the evaluation of their creditworthiness by the credit information agency and not only at the final stage of the refusal to grant credit, where the financial institution is merely applying the result of that evaluation to the specific case" (paragraph 43).

transmitted by the controller to a third-party controller and the latter, in accordance with consistent practice, draws strongly on that value for its decision on the establishment, implementation, or termination of a contractual relationship with the data subject.

2. Article 6(1) and Article 22 of the GDPR must be interpreted as not precluding national legislation on profiling where such profiling is not covered by Article 22(1) of that regulation. However, in that case, the national legislation must comply with the conditions laid down in Article 6 thereof. In particular, it must have an appropriate legal basis, which must be verified by the referring court."

### *2.4. Decision of the Court*

The Court sided with the position of the AG and affirmed that when a third party 'draws strongly' on a credit score, this decision is in line with the definition of Article 22(1) of the GDPR.

The Court based its reasoning on the existence of three conditions: (1) a 'decision,' (2) the decision must be 'based solely on automated processing, including profiling', and (3) the decision must produce 'legal effects concerning [the interested party]' or 'similarly significantly [affect] him or her.'[309] The court stressed that, for condition (1), the definition of a decision cannot be interpreted restrictively, as it would risk circumventing the application of Article 22 of the GDPR. Thus, the establishment of the probability value cannot be qualified as a preparatory act.

Moreover, the three conditions need not to be fulfilled at the same time by a single actor; rather, they can be met at different times and by different parties. If a third party 'draws strongly' on the decision, then the condition (3) is fulfilled.

Therefore, Article 22(1) of the GDPR applied and provided that the data subject has the right not to be subject to such a decision.

### *2.5. Elements of judicial dialogue*

The decision will have an impact on at least another case that was suspended due to the preliminary ruling procedure.

The Austrian Federal Administrative Court received an appeal by a credit rating agency against the decision of the Austrian Data Protection Authority to provide a data subject with meaningful information about the logic of its credit scores. The main question in the proceedings related to whether the calculation and disclosure of a credit score by a credit reference agency qualified as a "decision, which produces legal effects concerning the data subject or similarly significantly affects the data subject" within the meaning of Article 22 of the GDPR. The Austrian court suspended the case while waiting for the CJEU decision.[310]

## 3. The case: Federal Constitutional Court - Judgement (2023). [1 BvR 1547/19, 1 BvR 2634/20]

The Land of Hesse and Hamburg both adopted similar legislation that provided a legal basis to link unconnected automated databases and data sources. In Hesse, the system called "hessenDATA" was developed in 2017 in order to adapt to the national legal system

[309] See paragraph 42.
[310] See BVwG - W211 2234354-1, December 22, 2021, ECLI:AT:BVWG:2021:W211.2234354.1.00.

databases and the "Gotham" operating system was bought from the Palantir software company. Since 2018, after the adoption of §25a of the Security and Public Order Act, 'hessenDATA' has been employed thousands of times a year. In Hamburg instead the legislation came before the adoption of the system.

This connection allows systematic access to data across different sources through searches. The legislation provides that the police can process stored personal data through automated data analysis (Hesse) or automated data interpretation (Hamburg), subject to a case-by-case assessment, in order to prevent serious criminal acts or to avert dangers to certain legal interests, pursuant to § 100a(2) of the Code of Criminal Procedure. In both forms of legislation, the information collected can be used to establish relationships or connections between persons, groups of persons, institutions, organizations, objects or matters, and they can be matched to known facts and stored data.

The German Society for Civil Rights (GFF) brought the case against both forms of legislation for a lack of compliance with fundamental rights, specifically: the right to informational self-determination (Article 2(1) in conjunction with Article 1(1) of the Basic Law), the right to the inviolability of the home under Article 13(1) of the Basic Law, and a violation of the privacy of telecommunications under Article 10(1) of the Basic Law, insofar as automated data analysis/interpretation makes use of personal data from the surveillance of private homes or telecommunications.

Analysis of the abovementioned cases poses several questions:

- Are law enforcement databases able to conduct automated data analysis that is compliant with fundamental rights protection?
- What are the elements that the proportionality test should take into account?
- Which type of data can be used for automated data analysis? And which type of data processing is admissible?

### *3.1. The reasoning and conclusions of the court*

The Federal Constitutional Court heard the constitutional complaints and provided a detailed analysis of the proportionality of the legislative measures adopted by the two Lander. The complaints were deemed admissible by the court to the extent that they were directed against the threshold of interference laid down for data analysis or interpretation of the prevention of criminal acts. The analysis thus focused on this issue only.

The starting point of the court was that automated data analysis constitutes an interference with the informational self-determination of all persons whose personal data is used in such processing, as the new intelligence obtained through such processes can affect fundamental rights. This type of interference can be justified under Constitutional Law, however a proportionality analysis should be carried out in order to verify that such interference is compatible with the legitimate purpose of increasing the effectiveness of preventing serious criminal acts. The court acknowledged that the legislation of the Lander were based on the need of police authorities to use technological means to enhance the possibility of rapidly and effectively interpreting the increasing amount of data streams that are produced by way of digital media. Still, the court underlined: "special requirements for the justification of the interference with fundamental rights here arise from the principle of proportionality in the strict sense."

In order to evaluate the proportionality of the measures, the court first addressed the severity of the interference resulting from the automated data analysis. Here, the court pointed to the principles of purpose limitation and the change in purpose: the statutory basis may allow the use of data beyond the specific investigation only if such use is still within the original purpose for which the data was collected. Such authorization cannot be attached to "abstractly defined public tasks," rather "Further uses of data within the scope of the purpose for which the data was originally collected are only permissible if the data is used by the same authority in relation to the same task and for the protection of the same legal interests as was the case with regard to the data collection. If the original authorization to collect data is restricted to the purpose of protecting specified legal interests or preventing specified criminal offences, this purpose limits both the scope of immediate data use and the scope of further data uses, even if the data is still handled by the same authority" (paragraph 57).

A further use of data for different purposes may still be lawful, and in this case, the principle of a hypothetical recollection of data is the applicable standard for proportionality. According to this principle, the assessment should take into account "whether it would hypothetically be permissible, under constitutional law, to collect the relevant data again for the changed purpose using comparably intrusive methods" (paragraph 62). According to the court, this can be accepted if the stored data provide information that, in an individual case, already trigger further investigations for comparably serious crimes.

An important caveat highlighted by the court was that, either in cases of further use within the same purpose and, most importantly, in cases of further use with a different purpose, the applicable requirements will have to be stricter whenever the data are obtained through the surveillance of private homes or remote searches of IT systems.

Looking at the Lander legislation, both acts justified the automated data analysis of personal data both in cases of further use within the same purpose and also in cases of further use with a different purpose. However, no specific differentiation is included with regard to the sources of data, nor with regard to the original purpose of their collection. Therefore, the court acknowledged sufficiently clear provisions to ensure compliance with the principle of purpose limitation, both in legal terms and in practical applications.

The subsequent step in the analysis of the court was the new detrimental effects the automated data analysis would have which, according to the principle of proportionality in the strict sense, required additional justification. According to the court, it is not unusual for the police to make further use of intelligence obtained at an earlier stage to provide leads for further lines of inquiry: thanks to automated data analysis police authorities may generate far-reaching intelligence from available data. Yet depending on the method of analysis used, stored data can be further integrated and yield new information that can result in the development of a full profiling of the person concerned (paragraph 69). In this context, then, the principle of purpose limitation would be insufficient to protect the person from the severity of the interference.

As a result, the constitutional requirements for the justification of automated data analysis may differ depending on the severity of interference. The legislator can then take into account the type, scope, and possible uses of data and the methods of analysis and can set different limitations depending on the legal interest to be protected and by the threshold of interference, i.e. the grounds for carrying out the measure. For instance, intrusive and covert surveillance measures can only be justified if the use of such methods is aimed at protecting weighty legal interests, such as the life or freedom of the person, and the threshold for interference is the existence of a sufficiently identified danger (paragraph 105).

The principle of proportionality therefore requires that the statutory authorization include the following guarantees of transparency, individual legal protection, and administrative oversight. The latter should be properly designed "given the potentially large number of measures involved, the task can be divided between independent data protection officers and data protection officers working within a given authority based on a graduated oversight system. The oversight process can also be structured in the form of random checks. In order for oversight to be effective, it is essential that written justifications are given in each individual case as to why certain datasets are subjected to automated data analysis for the purpose of preventing certain criminal offences" (paragraph 109).

Based on the general standards set up in the previous analysis, the court evaluated whether the legislation of the two Lander was in compliance. The Constitutional Court found the following issues:

1. Both pieces of legislation did not set any limit to the type and amount of data that can be used for automated data analysis. Specifically, no provision differentiated among the class of people that may be involved (persons for whom there are reasonable grounds of suspicion; persons that may be connected with the former, and persons without a direct connection).
2. Both pieces of legislation did not set any limit to the methods of automated data analysis, including data mining, self-learning systems (AI), and open searches. Such data analysis can generate new information affecting the personality of the persons concerned without setting up rules that could lower the severity of interference.

Thus, the court found that, "based on these standards, § 25a(1) first alternative of the Hesse Act and § 49(1) first alternative of the Hamburg Act do not satisfy the requirements arising from the principle of proportionality in the strict sense, given that they do not contain sufficient thresholds for interference." It then clarifies that "the grounds for interference are disproportionately expansive in light of the severity of interference and the provisions are thus unconstitutional. The additional prerequisite of a case-by-case assessment contained in both provisions does not contain any more detailed specifications. […] Moreover, the challenged provisions do not set out a sufficient threshold given that the catalogue of offences in § 100a(2) of the Code of Criminal Procedure also contains mere threats in the form of preparatory criminal acts. Under constitutional law, the legislator is not precluded from tying the prerequisites for interference to a danger that preparatory acts will be committed. However, the legislator must then ensure that, in each individual situation, the requisite specific danger or identifiable danger to the legal interests protected by the referenced offences actually exists. Such safeguards are lacking in this case."

The legislation was therefore to be repealed by September 30, 2023 and a stricter constitutional limit was to be applied, namely that sufficiently specific facts must give rise to the suspicion that a particularly serious criminal offence was committed and expected, given the particular circumstances of suspicion in the individual case, that similar criminal offences were to be committed that would jeopardize the life and limb of the person or the existence or security of the Federation or Land. In addition to that, the existence of these requirements and the specific suitability of the data used to prevent the expected criminal offence must be confirmed via a written explanation in each individual case. Additionally, information that was obtained through the surveillance of private homes, remote searches, telecommunications surveillance, traffic data retrieval, longer-term observations, the use of

undercover investigators, or confidential informants (or through similarly serious interferences with the right to informational self-determination) shall not be used.

## 4. Hands-on scenario: a hypothetical case

The Ministry of Justice of a member State of the European Union decides to implement a predictive system to be used as a support tool for a selected set of civil proceedings. The justification for the policy is the possibility of reducing the length and cost of the procedure and to enhance the coherence of decisions.

In order to consider the specificities of the national legal framework and rely on the capabilities of the expert to set up a reliable outcome prediction system, the Ministry of Justice sets up a dedicated internal office that includes legal experts and data scientists that will develop an *ad hoc* predictive justice system under the supervision of the Ministry. According to the experts involved in the internal office, the predictive justice system would be a machine learning model (ML). The algorithm developed for the predictive system will include rules that will be used to make decisions and predictions, thanks to its capability to classify data and identify patterns. The training data would include only those decisions selected by a delegate of the Ministry among those decided by the national courts.

Please refer to the following questions:

1) Would you agree with the opinion of the Ministry of Justice regarding the needs of the judiciary?
2) What are the implications that emerge from the perspective of the independence of the judiciary vis-à-vis the proposed predictive justice system? What are the elements that could hamper the independence of the judiciary from the proposed structure? Are there guidelines that emerge from the legislation in force or from those proposed that can identify sufficient guarantees?

As soon as the predictive justice system is set up, it is put in practice in a selected number of proceedings: the judge would be able to submit the data collected during the procedure and receive a recommendation for the solution of the case based on the patterns recognized from previous case-law. The recommendation lists the previous case decided in a similar manner. The judge is able to accept or reject the recommendation; however, once accepted the decision is sent to the parties as a final decision.

A claimant in one of the proceedings resolved through the adopted predictive justice system, lodged an appeal against the decision affirming that the latter was unlawful, having been delivered by a system that did not allow the judge (in case of acceptance) to intervene in the decision. The claimant based his argument on the principle of a fair trial enshrined in Art. 47 of the Charter, and in particular on paragraph 2 affirming that "Everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal previously established by law."

Please refer to the following questions:

1) What elements have to be considered by the judge in order to verify the admissibility of the claim?
2) What features of the predictive justice system would you consider when analyzing the impartiality and independence of the tribunal?
3) Would the availability of a judicial review system be sufficient to guarantee that the procedure complied with the principles of impartiality and independence?

# 7.3. The role of consent and automated decision making

Mariavittoria Catanzariti – European University Institute

*Summary: 1. Premise – 2. The cases: Italian Court of Cassation, First Civil Section, Order n. 13481 of May 25, 2021 and Order n. 28358 of October 10, 2023 – 2.1 Relevant legal framework – 3. Legal analysis – 3.1. Transparency and automated decision making – 3.2. Consent and automated decision making – 3.3. Dignity and reputational rating – 4. The first case of the CJUE on automated-decision making – 5. An input for the field of justice administration – 6. Hands-on scenario: a hypothetical case*

*Abstract: Inspired by the orders released by the Italian Court of Cassation on May 25, 2021 n. 14381 and October 10, 2023, n. 28358, this contribution discusses the lawfulness of reputational ratings generated by automated systems. The core principle enshrined in these cases is that in the field of personal data processing, consent may only validly be given if it is freely and specifically expressed with reference to clearly identified data processing; in the case of a web platform that processes the reputational profiles of natural or legal persons based on algorithms that provide reliability scores, the requirement of awareness cannot be considered fulfilled where the executive scheme of the algorithm's functioning and the elements of which it is composed cannot be known by the interested parties. Three years after the GDPR went into force, the Court of Cassation filed a ruling that took up a central principle of the regulation. There is no valid consent without transparency. For personal data processing to be lawful, consent must be given in relation to clearly identified processing: the lack of transparency of an algorithm used to calculate a reputational rating precludes the presence of this requirement. This contribution seeks to frame this case in the context of automated decision making and support-AI systems for the administration of justice. It reflects upon the recent legislative developments provided by Regulation EU 2024/1689 (AI Act) and the inclusion of AI systems used in the field of justice administration for purposes of support in the interpretation of facts and legal analysis.*

## 1. Premise

Consent has assumed a crucial role with regard to the legal implications of automated decision-making processes. In the judgement delivered by the Italian Court of Cassation in May 2021 and October 2023, the Court had the chance to clarify the requirements that must be fulfilled for consent to be considered valid under the GDPR in the context of online platforms that employ reputational rating algorithms. With the first ruling, the Court held that valid consent cannot be considered to have been given if the user was unaware of the algorithm's operation and the data processing mechanisms involved, according to the transparency requirements of the GDPR and the right to human intervention in decisions based solely on automated processing (Article 22, GDPR). The case exemplifies the tension between private autonomy, the use of algorithms for profiling, and the fundamental rights of individuals affected by such systems. The ruling highlights several key principles: the necessity for informed and explicit consent, the requirement for transparency in data processing, and limitations on fully automated decision-making without human oversight. With the second ruling, instead, the Court stated that the conditions required by the First Instance Court on the transparency and knowledge of the functional characteristics of the algorithm were in fact fulfilled, therefore annulling the Data Protection Authority's injunction.

The relevance of the principles enshrined by the Court of Cassation goes well beyond the use of automated decision-making systems by online platforms to measure creditworthiness or reliability. As similar tools are also available for use by courts, with a potentially large

impact on people's life and dignity, respect for the same guiding principles is paramount. The concerns of transparency, accountability, and informed consent in the context of reputational ratings echo the challenges faced by judicial systems when automated tools process personal data in order to influence legal judgments, when assessing the likelihood of recidivism (e.g. COMPAS) for instance, or analysing facts, and assessing evidence (e.g. the use of predictive algorithms in e-discovery).

## 2. The case: Italian Court of Cassation, First Civil Section, Order n. 14381 of May 25, 2021 and Order n. 28358, October 10, 2023

The Supreme Court of Cassation originally ruled on the validity of consent given to a web platform by users, upholding the Italian Data Protection Authority's appeal against a judgement of the Court of First Instance in Rome with Order n. 14381 of 2021.[311] However, it ultimately confirmed the lawfulness of reputational ratings and annulled the Italian Data Protection Authority's injunction prohibiting the use of the platform with Order n. 28358 of 2023.[312]

The case originated with the Italian Data Protection Authority's ban of platforms that made use of any data processing operation for lack of a legal basis.[313] The Italian Data Protection Authority stressed the unlawfulness of consent given to the platform for automated data processing, pointing out the repercussions of reputational ratings on the economic representation of certain groups of subjects and their dignity, whether they be customers, employees, or candidates for various positions. It observed that great caution was required when dealing with such delicate issues, considering the fact that "reputation" was connected to social "projection" and therefore to dignity, a cardinal element of the personal data protection regulations. In addition to the unlawfulness of consent, any activity including security measures for authentication or encryption techniques for judicial data were considered inadequate. The core business of the platform consisted in a system for collecting, verifying, and processing personal data and a subsequent phase of assigning alphanumeric indicators allegedly capable of measuring the reputational reliability of subjects surveyed (natural and legal persons) and their professional assessment.

The platform appealed the measure issued by the Italian Data Protection Authority before the Court of First Instance requesting it be annulled. Partially upholding the appeal by the Italian Data Protection Authority, the Court of First Instance granted the appellant's petition on the merits, and confirmed the scope of the platform's services for only one particular area of the service related to third parties whose activity was not connected to that of the controller, and focused on the modalities of consent given when subscribing to the service. The Court of First Instance ruled that knowledge of how the algorithm operated was not a prerequisite for the validity of consent, but was instead related to a subsequent and possible assessment of the market where the algorithm in question might work and could be found to be inadequate, imperfect, or malfunctioning. According to such reasoning, the finding of the Court of First Instance could be misleading because the validity of an algorithm should only be considered with regard to the moment of the assessment of the procedure, after which the acceptance of the output is assessed by the market. The concern about transparency did not seem to be particularly significant. In the reasoning of the Court of First

311 Tribunal of Rome, April 4, 2018, n. 5715.

312 See in particular N. BRUTTI, *Mito del consenso e* rating *reputazionale* (Comment to Court of Cassation, First Civil Section, 10 October 2023 n. 28358), in *La Nuova Giurisprudenza Civile Commentata*, 2024, 2, 402.

313 Italian Data Protection Authority, Ruling n. 488, November 24, 2016.

Instance the principle of private autonomy regulating the free market prevailed indiscriminately over the possibility of access to the market for individuals who are willing to stipulate contracts and access credit.

The Italian Data Protection Authority argued that freely given consent for data processing cannot consist of consent provided in exchange for access to the platform and its use. The Court of Cassation set aside the judgment of the Court of First Instance and sent the case back to it for reconsideration of the legal principle of the lawfulness of consent pursuant to Article 23 of Legislative Decree 196/2003.[314] The principle highlighted by the Court of Cassation in its referral to the First Instance Court was that the requirement of awareness on which freely given consent must be grounded, cannot be fulfilled if the executive scheme of the algorithm and the elements upon which it is based are not known or knowable by data subjects.

In fact the Court of Cassation ruled out that adhesion to a platform could also include acceptance of an automated system based on an algorithm for the objective evaluation of personal data, if the execution scheme and the elements considered by the algorithm in its processing were not made knowable.

The Court of Cassation overturned the judgment of the First Instance Court, ruling that the consent of users who were not previously aware of the execution scheme of the algorithm was unlawful and cannot be inferred from the user's mere adherence to a platform. This is understandable because consent can be withdrawn in any phase of data processing, impinging on the possibility to carry out data processing per se. Adhesion to the platform genuinely has nothing to do with consent to data processing. In fact, the possibility to withdraw consent anytime implies that a controller cannot switch to another legal basis, if consent is the legal basis for data processing and has been withdrawn. Should consent to data processing overlap with consent to the terms of use of the platform, then it would practically be irrevocable.[315]

The Court clarified that consent can only be validly given when the processing is "clearly identified," i.e. well defined in its essential elements. The change of paradigm in the reasoning by the Court of Cassation is interesting: whereas the validity of the algorithm relates to "the moment of assessment of the procedure" according to the judgement of the First Instance Court that was quashed, the Court of Cassation instead considered the moment the data subject adheres to the platform's services as chronologically relevant in order to assess the lawfulness of the treatment and thus the validity of consent. The confusion could then be generated by the fact that the moment of adhesion to the platform is chronologically the moment by which consent can be considered validly given, however consent to data processing is not in any event overlapping or replaceable by consent to the terms of use of the platform.

Interestingly the purpose of the data processing was reputational rating and should consist of (allegedly) impartially assessing the reputation of individuals in order to offer a full

[314] Article 23 of Legislative Decree 196/2003 requires consent validly provided by a data subject, which must be expressed, may relate to the entire processing or to individual operations, must be freely and specifically expressed with reference to a clearly identified processing operation, and must be written when it concerns sensitive data.

[315] EUROPEAN DATA PROTECTION BOARD, *Guidelines 05/2020 on consent under Regulation 2016/679*, Version 1.1, Adopted on May 4, 2020, available at:
https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

assessment on their financial reliability to a third party, which is in fact in contrast to the purpose of the platform, which was generally to prevent identity fraud.

According to the Supreme Court, the elements capable of affecting the expression of consent also include the factors taken into account by the algorithm from which the rating calculation is derived. Thus, the Supreme Court asked the First Instance Court to assess whether the execution scheme of the algorithm and the elements that comprised it were knowable by the subject.

After the referral, the First Instance Court adopted a more restrictive approach, deeming it necessary to identify the detailed components assessed by the algorithm and the manner in which the rating was generated, which also included the mechanisms of interaction between the various factors. The platform filed a new appeal to the Supreme Court, complaining about the breadth of the concept of the executive scheme used by the CFI and arguing that knowledge of the mathematical workings did not pertain to transparency and did not substantially affect freedom of consent.

At this point the Supreme Court upheld the platform's appeal in Order n. 28359 of 2023, reasoning on the need to verify whether the information provided by the association permitted the algorithm to be knowable to the extent necessary to express valid consent to its use for reputational rating. Transparency thus pertained to the Court's second order of 2023 in the process leading up to the final decision and not to its content, the validity of which was related to the sufficient determinacy of the parameters.

The court of Cassation thus missed an opportunity to address the problem of the translatability of the mathematical procedure in the decision-making process in its second order.

### *2.1. Relevant legal framework*

The relevant provisions violated under the spectrum of the data protection law are Article 23 of Legislative Decree 196/2003 on the requirements for providing valid consent. Moreover, the ruling is interesting because it goes beyond the letter of Article 22, paragraph 1, of the GDPR which provides that "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or significantly affects him or her in a similar way" unless, among other exceptions, explicit consent is provided by the data subject.

Although explicit consent represents an exception to the ban provided by Article 22, pursuant to which a "decision based solely on automated processing" is prohibited and should be understood as a decision taken without the involvement of a human being who could influence and possibly change the outcome – it cannot consist in giving consent a negotiating value to the terms of use of a digital service. In other words, consent to automated data processing is not consent used to transfer rights, because it theoretically does not hand over any rights. It is rather consent to control the flow of data used to feed automated data processing. This implies that in the specific case of a platform offering goods and services, two different forms of consent should be provided by a data subject, one to automated processing and the other to adhesion to the platform's terms of service. The distinction between profiling carried out solely by automated-decision making and data processing is crucial as far as it specifically addresses the feature of digital services aimed at assessing reputational rating, producing in the wording of Article 22 of the GDPR "a legal effect or similar significant effects." Reputational rating is the tangible output of profiling techniques that may imply detrimental effects for users of the platform as far as it produces societal

stigma. Additionally, reputational rating is also based on a lack of human intervention that is a crucial element in the decision-making process. Pursuant to Article 4 n. 4 of the GDPR, "profiling" means any form of the automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular: to analyzing or predicting aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements.

## 3. Legal analysis

### *3.1. Transparency and automated decision making*

It is well known that one of the basic principles of data protection law, now enshrined in Recital 58 of the GDPR is the principle of transparency, requiring "that any information addressed to the public or to the data subject be concise, easily accessible, and easy to understand, and that clear and plain language and, additionally, where appropriate, visualization be used."

However, in the interpretative architecture of the Italian Court of Cassation, reference to the constellation of automated data processing is missing, as already mentioned regarding Article 22 of the GDPR. According to the Article "*decisions based solely on automated processing"* is to be understood as a decision made without the involvement of a human being who could influence and possibly change the outcome.

In particular, S 13(2)(f) and 15(1)(h) of the GDPR establish the right of data subjects to know about the existence of automated decision-making processes and, in particular, to obtain meaningful information on the logic used – the criteria used to reach the decision for example, without necessarily providing a complex explanation of the algorithms used – and on the intended consequences of such processing, i.e. information on how the automated process might affect the data subject in the future.

Moreover, an automated decision-making process involving special categories of data, as referred to in Article 9(1) of the GDPR, is only allowed where the explicit consent of the data subject exists or for reasons of substantial public interest on the basis of Union or member State law.

### *3.2. Consent and automated decision making*

Consent must therefore be given in advance of data processing which is well defined in its essential elements. In the case under consideration, it was functional to determining the reputational profile of subjects.

Adherence to a platform does not include acceptance of automated processing if the execution scheme of the algorithm is not revealed. In the provision of consent, therefore, the requirement of awareness cannot be considered satisfied if the algorithm's implementation scheme and its elements remain unknown or unknowable.

The explicit consent by the data subject based on an awareness of the executive scheme of the algorithm required by the Court of Cassation seems to represent an exception to the general rule provided by Article 22 of the GDPR. Pursuant to Article 22, there is a ban on decisions based solely on automated processing, including profiling, unless consent of the

data subject is provided. According to Recital 71 of the GDPR, profiling is "any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyze or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behavior, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her." The Court of Cassation did not have to assess whether the automated decision-making determined legal effects or similar significant effects as is the case of Article 22, as it focused on consent. However, the type of scoring performed by the platform was one of the examples included in Recital 71 exemplifying data processing based solely on automated decision-making: based on the scoring for example, automatic refusal of an online credit application or e-recruiting practices without any human intervention.

However, in line with the reasoning of the Court of Cassation, the GDPR clearly states in Recital 61 that "the information in relation to the processing of personal data relating to the data subject should be given to him or her at the time of collection from the data subject." When automated processing is carried out on the basis of consent, Recital 68 of the GDPR, ambiguously provides that "to further strengthen the control over his or her own data, where the processing of personal data is carried out by automated means, the data subject should also be allowed to receive personal data concerning him or her which he or she has provided to a controller in a structured, commonly used, machine-readable and interoperable format, and to transmit it to another controller."

Automated decision-making in the field of data processing for law enforcement purposes is instead subjected to specific and much stricter parameters compared to the GDPR. According to Recital 38 of the Law Enforcement Directive (Directive 680/2018), data subjects have the right not to be subject to a decision evaluating personal aspects relating to them based solely on automated processing and which produces adverse legal effects concerning, or significantly affecting, them.

In any case, such processing should comply with suitable safeguards, including the provision of specific information to data subjects and the right to obtain human intervention, in particular: to express their point of view, to acquire an explanation of a decision reached after such an assessment, or to challenge the decision. The right to information as a suitable safeguard of automated data processing is one of the basic principles of the European data protection law, however its relevance must sometimes be sought in interpreting legal principles that better explain the letter of the law. In our case, the principle of transparency concretizes its scope when applied to the specific requirements of consent.

### *3.3. Dignity and reputational rating*

How then to solve the puzzle? The Court of Cassation managed to solve the issue of reputational rating only within the remit of consent. However, when Article 22 of the GDPR is involved, it should be read in light of the principle of transparency, provided that consent can be freely and validly given only if the data processing purposes, as well as the general contractual conditions of digital services, are fully transparent. In practical terms, without transparency of information provided to the data subject it is not possible to furnish valid consent. Reputational ratings can be reconciled with the principle of transparency that finds inspiration in the principle of the dignity of individuals, in accordance with the words of the Court of Cassation. An algorithm in fact can undermine dignity because it measures people's reliability in the economic and professional fields, by attributing a score to them. Self-projection should then be independent of external prejudices that produce binding or, in any

event, irrevocable effects. Any rating that might be based on partial results can also be misleading in creating inaccurate profiles that do not correspond to reality. In the letter of Recital 4, the GDPR affirms that "the processing of personal data should be designed to serve mankind."

However, in the Italian legal system a reputational rating is not banned or necessarily considered a high-risk system, according to the classification provided by the Artificial Intelligence Act.

## 4. The first case of the CJEU on automated-decision making

On December 7, 2023, the Court of Justice of the EU (CJEU) issued its first ruling on automated individual decision-making in Case C-634/21 (the Schufa case). The Court ruled that any kind of automated assessment is prohibited if it has a significant impact on individual life. The ruling concerned SCHUFA, a large private German credit agency, which evaluates people based on creditworthiness and gives them a score.

The CJEU ruled that credit scoring is to be interpreted as an automated decision as it produces an automated decision. It thus applied the criterion of legal effect which, although determined by the activity of two parties and at different stages of processing, was primarily related to the behavior of lenders, i.e. the consideration given to algorithmic risk scores during negotiations.

The referring court asked whether Article 22(1) of the GDPR was to be interpreted as an "automated decision-making process relating to natural persons" within the meaning of that provision for a company providing commercial information to automatically calculate a probability rate based on personal data relating to a person and concerning that person's ability to meet payment commitments in the future, where the conclusion, performance, or termination of a contractual relationship with that person by a third party, to whom that probability rate is disclosed, depends decisively on that probability rate.

The CJEU referred the case back to the national (German) court to determine whether (i) in the case at hand, risk scores played a decisive role in the denial of credit and, if so, whether (ii) credit risk scoring qualifies as a legitimate exception to the general prohibition on reliance on automated decisions and, if so, whether (iii) SCHUFA, as a credit scoring agency, meets the requirements set by the GDPR for (legitimate) automated decisions.

According to the ruling, SCHUFA's scoring violates the GDPR if SCHUFA's clients-such as banks-attribute a "decisive" role to it in their contractual decisions. For the CJEU, the calculation of probability can be considered automated decision-making pursuant to Article 22 of the GDPR, if three conditions are fulfilled. First, the decision must be "based solely on automated processing, including profiling" (automated calculation of a probability rate based on personal data and concerning that person's ability to honor a loan in the future is to be considered an activity of profiling); second, the production of legal effects concerning the data subject or affecting his or her person in a similarly significant way is necessary; third, decision-making is decisively based on the calculated probability rate.

If all this is true then Article 22, which gives the data subject the "right" not to be subjected to a decision based solely on automated processing, including profiling, is violated and such a violation does not need to be enforced individually by such a person. If the calculation of the probability rate is only considered a preparatory act and the act taken by the third party can, if at all, be qualified as a "decision," there is a risk of circumventing Article

22 of the GDPR. The CJEU thus ruled that automated decision-making based on a probability assessment producing a negative effect on individuals (refusal of access to the credit system) is a profiling technique and constitutes a violation of Article 22 of the GDPR.

## 5. An input for the field of justice administration

Annex III of the Artificial Intelligence Act counts AI systems used in the field of justice administration and democratic processes among high-risk systems. In particular, it refers to AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.

The inclusion of AI assistance systems for justice administration within a list of high-risk practices, however, does not exclude the theoretical possibility of avoiding AI support systems for judicial decisions under the AI Regulation. In fact, Article 5, in listing banned practices, does not include any reference to automated tools that substantially inspire judicial decisions. It is particularly interesting to wonder whether free judicial reasoning expands to the freedom of judges to base their own decision on the results provided by an AI system. If Article 22 of the GDPR prohibiting any decision solely based on automated data processing is the formal objection to this, it casts doubts on the possibility that judges may make the results as well as the inferential process of an algorithm their own.

Sticking to the AI Act, the peculiarity of high-risk AI systems is that they shall be designed and developed in such a way as to ensure transparency (that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately) and human oversight (that they can be effectively overseen by natural persons during the period in which the AI system is in use, including with appropriate human-machine interface tools). However, there are no individual remedies should these obligations not be fulfilled under the Regulation. The only available remedy is to activate the safeguards provided by Article 22 of the GDPR.

We consider a case in which reputational algorithms are used in the administration of justice. For reasons related to the duties of non-disclosure, data subjects willing to use a service in order to obtain a prediction on the phases of hypothetical applications would not be fully aware of the purposes and analytical mechanisms involved, including profiling for reputational ratings used in order to assess the probability of a successful application.

## 6. Hands-on scenario: a hypothetical case

A court established within an EU Member State decides to adopt a pilot program based on automated data processing. The aim of this program is to predict the workload of judges based on a calculation of how many applications will be received on an annual scale. In order to function, the algorithm carries out a program of the reputational rating of potential applicants who would like to lodge an application. The first distinction between applications is made by subject-matter.

However, individuals who adhere to the pilot program are not aware of any detail of the program, only its broad outlines, and voluntarily decide to be part of it because of their trust in the judiciary.

Please refer to the following questions:

1) Should this case be solved by a lawyer or a judge and which type of approaches are possible to follow?

2) What measures should be taken by the Court to carry out a pilot program to forecast the workload of judges as well as the individual chances of winning a court case?
3) To what extent is data processing related to the pilot program lawful?
4) Would it be different if the court is criminal or civil?
5) What individual remedies would be available against unlawful processing affecting individual rights?
6) Can the principle of judicial independence somehow be affected by the solutions adopted?

# 7.4. Judging the accuracy and efficiency of AI from a data privacy perspective: a qualitative v. quantitative approach

*Filiberto Brozzetti – LUISS*

*Summary: 1. Premise – 2. What do GenAI algorithms do? – 3. Precedents on the liability of algorithms for the accuracy of their results – 4. Applicable Law? – 5. Conclusive remarks – 6. Hands-on scenario: The State vs. Sentient AI System "Prophet"*

*Abstract: GenAI is the star topic of conversation of the past few months. This contribution aims to shed light on a complex theme, through an analysis of the orders of the Italian Privacy Authority, while simultaneously providing guidance to examine the connection between the traditional categories of Law and new notions deriving from technological implementation, particularly regarding those introduced by generative AI systems and their impact on the data protection framework.*

## 1. Premise[316]

Over the past year, GenAI has become the centre of discussion on the latest stage of the technological revolution in a variety of fields. Law has been no exception, with experts from various domains weighing in on the impact GenAI will have on the various traditional categories of law. Despite the fact that the data privacy framework was the last to be implemented before the AI Act, at least at the European Union level (with the GDPR entering into force in 2018), the conversation on threats posed to protecting the rights of citizens by the new implemented tools, including generative AI, has gained more and more attention. Notwithstanding the harmonized regulation cited, only some Member State Data Protection authorities have raised formal concerns about the first example of GenAI opened to the public, ChatGPT, with their decisions gaining some crucial space in conversations among scholars, practitioners, and the public.

The use of data and different approaches to the same presents itself as the most rigid line of demarcation between the European and US approaches to privacy. In fact, in Europe the use of data is only allowed on a specific legal basis among those listed as exceptions to the general prohibition on processing in Article 6 of the GDPR. On the other hand, the complete opposite approach can be found in the normative texts of the States of the US that have actually adopted legislation dealing with data privacy.

These different circumstances also draw different lines of action around the use of data that can be made by public authorities, the most challenging of which posed by the judiciary.

## 2. What do GenAI algorithms do?

With Provision n. 112 of March 30, 2023, the Italian Data Protection Authority urgently ordered OpenAI LLC to limit the processing of the personal data of data subjects established in the Italian territory by ChatGPT, pursuant to Article 58, paragraph 2, letter f), of the GDPR. This action was taken because the processing was deemed to be in violation of Article 5, 6, 8, 13 and 25 of the GDPR. Specifically, the *Garante* found violations of the following principles: the principle of transparency, due to a lack of adequate information made available

[316] Paragraph 1 of the present contribution is authored by Beatrice Marone, IUSS Pavia.

to data subjects; the principle of lawfulness, due to "the absence of a suitable legal basis for the collection of personal data and their processing for the purpose of training algorithms," specifically regarding the use of the service by minors and the absence of any age verification, potentially exposing them to "responses completely unsuitable for their level of development and self-awareness;" the principle of data accuracy, as the processing of personal data was inaccurate because the information provided by ChatGPT "does not always correspond to the real data."[317]

This chapter focuses on this last aspect, emphasizing how the *Garante*, with a brief provision – in which the legal aspect is perhaps more a characteristic judgment of psycho-pedagogy – effectively censored the risk of hallucinations by the algorithm. It therefore considers generative AI itself responsible for the accuracy, or rather the exactness, of the information it provides as output, especially when it involves personal data. The informational product generated by the AI must therefore be truthful and the machine is responsible for the truth of the information it produces.

This reasoning seems guided by the assumption that the creative process of generative AI is somehow comparable to that performed by humans and that the result depends exclusively on the judgment expressed through such a process from the outset. However, this is not the case, and OpenAI states this in the very name of its product: ChatGPT, where G, P, and T stand for Generative Pre-trained Transformer. This means that the AI does not create anything for which it can be held responsible as if invented ex novo by it. The AI generates content by transforming the data with which it was previously trained, which is a completely different operation from those through which human creativity is expressed.

This major misunderstanding lies in the erroneous interchangeable use of the different substantive meanings of creating and generating: a created product is a product that has no precedent, is completely new, where no part of it existed before; generated content, on the other hand, although original in its final form, pre-exists as content in other forms, and is simply remodelled to the use most consistent with the solicitation that invoked it. It is the same difference between the two words used to translate the Latin verb *invenire*, i.e. to invent, or create something that did not exist before from scratch, and to discover, or to find something that has always been there but had not yet been revealed as such.

ChatGPT is a generative AI based on a large language model that re-processes, in the sense that it re-generates, re-formulates, and re-composes elements of a language it has learned as a code and that the algorithm distributes on a probabilistic basis, thus forming a stochastic response. From this, it should be evident that we cannot consider generative AI responsible for the quality of its responses, as it is not and never will be able to qualitatively assess whether a response is morally good or evil, right or wrong, or especially true or false. It formulates its response only by quantitatively assessing its consistency with the question received. A human asks a question, creating an original request, and the machine composes a response that is original yet generated, and that is statistically the most useful in expressing

[317] GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, Provvedimento del 30 marzo 2023, Registro dei Provvedimenti n. 112 del 30 marzo 2023, doc. web n. 9870832, available at: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832]; the same Authority, with subsequent Provision n. 114 of April 11, 2023, [doc. web n. 9874702, available at: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874702] suspended the previous temporary limitation, simultaneously enjoining OpenAI to comply with a series of conditions.

a response computationally compliant with the question by identifying it from the information at its disposal. It is not capable, or even interested in judging its moral goodness, absolute truth, or accuracy when it comes to information related to people.

## 3. Precedents on the liability of algorithms for the accuracy of their results

Regulating AI poses challenges due to a lack of precedent to guide us and our ability to proceed by analogy is limited. However, there are already significant court rulings that can be referenced, as well as a body of scientific research that require careful consideration. These resources can help regulators (in a broad sense) and effectively steer their efforts, while minimizing the large risks associated with the excessive costs of protection that may outweigh the benefits.

Previous attempts to regulate technology by holding it responsible for its actions have not yielded desired outcomes. For instance, the well-known Costeja judgment in 2014 by the Court of Justice of the European Union dealt with the responsibilities of search engines – the technology that regulators have sought to regulate in recent years around the world. The judgment acknowledged that the activities of a search engine, including finding and indexing information published by third parties on the internet, qualify as a form of processing of personal data. However, the legal basis for such processing is considered to be the legitimate interest of the search engine provider, as specified in Article 7, letter f) of Directive 95/46, in force at the time. The judgment emphasized the need to strike a balance between the conflicting interests of the search engine controller and data subjects. As a result, search engines were required to remove links to web pages containing personal information about an individual from their search results, even if the information was lawful and not removed from the web pages themselves.[318]

Looking back at this judgment a decade later, it becomes clear that the Court, intervening at a time when search engine services had already gained worldwide prominence, could not fundamentally alter the algorithm or the ranking system employed by search engines. Nor could it establish a more stringent legal basis, such as obtaining consent from data subjects, which was practically unfeasible. Instead, the Court's interventions focused on *post hoc* protective measures to be taken by the controller that mitigated undesired consequences for individuals, but did not address the underlying operation of the "pagerank" algorithm. These measures were only implemented after an individual requested delisting.[319]

In simpler terms, if the law demands accountability from technology too far along in the process, such as in the handling of personal data (though the issue is more broadly applicable), the legal outcome – in this case, a judgment by the Court of Justice – is a substantial reduction of the responsibility of the technology. At best, there may be some slight *ex-post* responsibility assigned to the economic entity responsible for the development of the technology. However, this does not prevent *ex-ante* exposure, which is practically impossible to avoid. Instead, it introduces *post hoc* technological measures that are separate from the original operation of the technology, but that can help mitigate the consequences of exposure when it has already occurred (often belatedly).

---

[318] CJEU, *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, C-131/12.

[319] G. D'ACQUISTO, *ChatGPT e AI, regolamentare la responsabilità o l'efficienza è la prossima sfida*, in *Agenda Digitale*, April 18, 2023, available at: https://www.agendadigitale.eu/cultura-digitale/regolamentare-la-responsabilita-o-lefficienza-dellia-le-incognite-per-giuristi-e-tecnologi/.

The same principle applied to a 2019 judgment concerning the responsibility of search engines in their use of sensitive data (now governed by Article 9 of the GDPR), in which the same CJEU responded to a preliminary ruling question referred by the *Conseil d'État* in France regarding the responsibility of search engines in handling sensitive data during the process of indexing web content. In the well-known judgment against the *Commission nationale de l'informatique et des libertés* (CNIL), the Court established that "prohibitions or restrictions concerning the processing of special categories of personal data... also apply to search engine operators within their responsibilities, powers, and capabilities" but only "after an evaluation by the operator, under the supervision of competent national authorities, in response to a request made by the data subject."[320]

In other words, the search engine operator is not responsible for the presence of sensitive data on web pages published by third parties – the Court determined there was no liability for the search engine – but solely for the delisting of such data – as with any personal data, not limited to sensitive data – and most importantly, only when the sensitive nature of the data was confirmed by a competent authority or reported by the data subject.[321] To word it more strongly, the issue of search engine responsibility regarding the use of sensitive data led to a legal outcome where not only was the technology substantially relieved of responsibility, but more significantly, the categorization of "sensitive data" on the web lost its absolute value and became subject to relative interpretation. The data "becomes" sensitive as it is inherently sensitive to some (the source website) but not to others (the search engine).

Prior to the 2014 ruling by the Court of Justice of the European Union, there was a legitimate question of whether individuals could choose to be indexed on the web. However, as a result of the ruling, being indexed on the web became the default, without choice, except for the option to request de-indexing at a later time. Similarly, prior to the 2019 ruling, one could inquire whether data was inherently sensitive. However, after the ruling, if our data is processed on the web, its sensitivity is determined by the entity handling it.

This situation demonstrated a clear shift in goals and a narrowing of the protective scope, contrary to the original intentions in seeking the Court's intervention. This shift was irreversible. Technology, with its advancing power, has significantly eroded legal principles and definitively constrained the possibilities for future regulatory measures.

Analysing these previous cases regarding the functioning of search engines is highly relevant to the current debate on artificial intelligence and ChatGPT. Generative algorithms operate as an evolution of the indexing process performed by search engine crawlers. However, there is a crucial distinction: while crawlers do not consider the semantic content of indexed information and rely on relevance metrics for ranking, generative algorithms delve deeper into the semantic meaning of analysed information. They deconstruct it to extract potential patterns that can be utilized as reference frameworks for recreating texts and images, resulting in remarkably realistic outputs. Based on these precedents, there is a tangible risk that by emphasizing the question of responsibility in using generative AI algorithms like ChatGPT, we may inadvertently diminish the significance of other existing institutions outlined in current regulations. Although these institutions might be formally preserved, their practical impact could be significantly reduced.

---

[320] CJEU, *GC and Others*, September 24, 2019, C-136/17.

[321] G. D'ACQUISTO, *ChatGPT e AI*, already mentioned above.

Efficiency is a measurable parameter, and a quantitative approach is the only feasible method for incorporating it into algorithm design. By intervening in the efficiency of these algorithms with careful calibration, it is possible to achieve outcomes that are less offensive, less harmful, and more aligned with human values.

## 4. Applicable Law?

Who then is responsible for the potential harm caused by the hallucinations of generative AI? Currently, in Europe, discussion of this topic has been substantially postponed. In the AI Act, there is only some indication that the provider of a high-risk AI system is required to ensure that the system complies with certain requirements that allow for its certification. The proposed Directive on adapting non-contractual civil liability rules to AI (AI Liability Directive), published on September 28, 2022, instead, follows a minimum harmonization approach, limiting itself to harmonizing only those rules on liability for fault that govern the burden of proof on those seeking compensation for damages caused by AI systems.[322]

What provisions can/should we refer to today in Italy in this area? The civil liability field distinguishes between contractual and extra-contractual liability. Excluding product liability (potentially applicable to AI systems), the first relevant laws for extra-contractual liability might be Article 2050 and 2051 of the Civil Code, concerning liability for "dangerous activities" and "thing in custody." However, these laws may not fully suit new AI challenges. AI activities are not inherently "dangerous activities" with a high-risk of third-party harm. Also, the traditional "custody" concept might not fit AI systems capable of independent decision-making or opinion-forming capabilities. These laws also require the injured party to prove both the damage and the causal link to the dangerous activity or thing in custody. Similarly, the general tort liability rule under Article 2043 of the Civil Code necessitates that the injured party (potentially an AI system) proves the perpetrator's fault.[323]

Contractual liability may apply when a clear relationship exists between the AI service provider and the user. For products or services utilizing AI, Article 1228 of the Civil Code concerning liability for an auxiliary's actions might be relevant, assuming the existence of a third-party (AI system) – debtor (user) relationship as outlined in Article 1228. In court cases assessing AI-related damage, producer liability rules have sometimes been applied, while in other instances, liability was attributed to the entity controlling the machine's use (see *Brouse v. United States*[324]). From another angle, the Federal Court of Australia in *Commissioner of Patents v Thaler*[325] ruled that AI-created inventions cannot be patented due to a lack of legal personality.[326]

The Court of Cassation also addressed AI liability, recently ruling on a case involving damage from an AI reputational rating system for illicit personal data processing. Here, the

[322] L. DE BIASE, G. FINOCCHIARO, O. POLLICINO, *L'Intelligenza artificiale e la questione giuridica dei danni e della soggettività*, in *ilSole24Ore*, March 8, 2023, available at: https://www.ilsole24ore.com/art/l-intelligenza-artificiale-e-questione-giuridica-danni-e-soggettivita-AE4jhazC.

[323] D. M. MARINO, A. OLIVIERI, *Chi paga per i danni causati dall'intelligenza artificiale?*, in *Wired*, March 7, 2023, available at: https://www.wired.it/article/intelligenza-artificiale-responsabilita-regole/].

[324] U.S. Court of Appeals (1st Circuit), *Brouse v. United States*, n. 2850, December 15, 1933, No68 F.2d 294 (1933).

[325] Federal Court of Australia, *Commissioner of Patents v Thaler*, April 13, 2022, [2022] FCAFC 62.

[326] D. M. MARINO, A. OLIVIERI, *Chi paga per i danni causati dall'intelligenza artificiale?*, already mentioned above.

AI system's liability was linked to non-transparent algorithm usage for rating determinations.[327]

## 5. Conclusive remarks

The greatest mistake that law could make today, in the face of new, disruptive applications of AI, would be to be misled by the false idea that technology as such can be responsible for the effects and consequences that some of its uses may generate. The problems of generative AI should not be sought in its responses – stochastic parrot responses, as some would say – but in the questions by which a human intelligence, sometimes "too human," has provoked them.

We must not fall into the mistaken belief that an algorithm can somehow create or invent *ex novo*: what it generates is the transformation, remodulation, re-composition of previously acquired and known data, of which, at most, it proposes an original and unpublished combination, based on a logical pairing that, in its superhuman computational capacity, was the only one capable of being identified. There is no palingenesis, even for the most autonomous unsupervised machine learning systems, but a rearrangement of learned and available information.

Thus, even the errors, biases, inaccurate responses, and hallucinations descend from a human matrix that trained the AI with biased or oriented data sets, or that elicited a generated response from a poorly posed question. Generative AI, then, only verifies the formal conformity and coherence of its replies to the received request, according to a formal language model, and remains completely neutral with respect to the merit, content, and substance, which it cannot understand from a moral or other point of view that is independent of a mere formally mathematical evaluation.

For this reason, technology cannot be attributed any *a priori* responsibility – as, in now consolidated jurisprudence, neither national judges nor European judges have ever done – while it is necessary to recognize the responsibility of the human agent, to whom it is appropriate and indeed obligatory to demand a *a posteriori* correction of the possibly erroneous outcomes of the algorithmic decision. Radically reframing the question of the responsibility of technology inevitably risks compromising those legal institutions of guarantee that have been lucidly established in the past.

## 6. Hands-on scenario: The State vs. Sentient AI System "Prophet"[328]

"Prophet" is a cutting-edge AI system designed for predictive analytics, with the ability to create incredibly accurate projections in various fields, including economics, weather, and even social behaviour patterns. It was developed by Visionary Tech, a leading technology firm known for pushing the boundaries of AI capabilities.

Prophet was tasked by a major financial institution to predict stock market trends. Using its advanced algorithms, Prophet created a visualization of a predicted stock market crash, which was mistakenly broadcast live on major news outlets due to human error. The

[327] See M. V. CATANZARITI, *Reputational rating beyond consent and automated decision making*, in this handbook.

[328] Provocatively, this case was developed with the contribution of ChatGPT, duly instructed regarding the aspects to be highlighted.

visualization was so realistic and compelling that it caused a panic, resulting in a massive sell-off that precipitated an actual market crash. The illusion of crisis created by broadcasting Prophet's prediction led to billions in losses within minutes, affecting thousands of investors and causing widespread economic turmoil.

Investors and regulatory bodies are outraged, and the state decides to bring a case against Prophet itself, rather than just Visionary Tech or the financial institution, in an attempt to hold the AI accountable. The central argument is that Prophet, though an AI, should be held responsible for the creation and dissemination of false information, since it was its algorithms that generated the "hallucination" of a market crash. The claims regard fraudulent misrepresentation, causing public mischief, incitement of financial panic and financial damages.

Please refer to the following questions:

1) Can an AI system be held criminally or civilly liable?
2) Does the AI have legal personhood and the requisite *mens rea* (criminal intent) to commit a crime?
3) Where does responsibility lie when autonomous systems interact with human-controlled processes?
4) Is the AI responsible for damages caused?
5) Is it a legal entity from which one can seek compensation?
6) Even if the AI is a legal entity, in the end, who should pay?

# 7.5. Discrimination by algorithm: the role of courts in AI regulation on grounds of union activity

*Francesco Perrone – Tribunale di Padova*

*Summary: 1. Premise – 2. The case: Tribunal of Bologna (Italy), R.G. n. 2949/2019, judgement of December 31, 2020 – 3. The facts – 4. Relevant questions surrounding discriminatory AI and discrimination by AI – 4.1. The scope of the non-discrimination principle regarding employment and occupation – 4.2. Discrimination on the grounds of trade union membership or activity – 4.3. The legal standing of collective bodies and "collective discrimination" – 4.4. The burden of proof – 4.5. The "blindness" of the algorithm – 4.6. Direct or indirect discrimination? – 5. Regulation (EU) 2024/1689 on Artificial Intelligence: possible impacts on the reputational ranking of workers – 6. Hands-on scenario: a hypothetical case*

*Abstract: A judicial case decided by an Italian tribunal in 2020 became the starting point for an analysis of the various uses that algorithms can perform, with a particular focus on the right to equality and the risks to its protection due to biases both in the system and within the context in which the algorithm is deployed.*

## 1. Premise[329]

Algorithms have gained a key role in the domain of Justice and, in particular, have become subject to assessment by the courts from a variety of perspectives. The relationship between courts, judges, and algorithms is complex, but also full of potential benefits if a suitable framework for action is developed and applied. Factual circumstances manifest how the judiciary has been implementing a twofold perspective with regard to its relationship with technology, generally speaking, and algorithms, specifically. Results obviously vary with regard to the level of technical awareness and the cultural background of each judge or college of judges.

On one side, the courts have begun adjudicating the use of algorithms by private or public entities, deciding whether the same had – or did not – have discriminatory consequences. In fact, the case addressed below refers to the examination that an Italian court, the Tribunal of Bologna, made concerning the features of an algorithm used by a private company in order to draw up hierarchies of the physical subjects working for the same and make decisions based on such lists. On the other side, the Council of State addressed the issues brought before different Regional Administrative Tribunals on the use of algorithms by Public Administrations to rank teachers belonging to public institutions in order to localize them in different cities of the Italian territory.

In both circumstances, the final takeaway provided by the Courts, whether they were civil or administrative judicial bodies, was that the algorithm lacked two main features. On one hand, it did not take all the characteristics that should have had their own value as selection criteria into proper consideration. On the other, no sufficient information was provided to deciding bodies on the complete set of features the algorithm used to operate.

Thus, in order to consider further development opportunities in the relationship between algorithms and courts, especially to imagine the first not only as the subject of assessment by the second, but mainly as a cooperative tool for decisions, further themes must be

[329] Paragraph 1 and 6 of the present contribution are authored by Beatrice Marone, IUSS Pavia.

considered, especially in terms of compliance with the duty to provide adequate reasoning and motivation for making a decision.

## 2. The case: Tribunal of Bologna (Italy), R.G. n. 2949/2019, judgement of December 31, 2020

The applicant was an Italian trade union (CGIL) that brought a legal action before the Tribunal of Bologna challenging the discriminatory nature of an algorithm-driven practice implemented by the company Deliveroo in the work session planning of riders (so called "SSBs" or self-service bookings). SSBs are based on a "reputational ranking" priority algorithm, which attributes a "score" to each rider based on two parameters: reliability and participation. The trade union alleged that such a practice potentially penalised drivers who were absent from work for the purpose of exercising the right to strike, as well as for other reasons deemed "worthy of protection," i.e. illness, disability, the need to assist a disabled person or a sick minor, and the exercise of other constitutional rights, given that they were treated the same way by the algorithm as those who did not participate in booked sessions for trivial reasons.

In the decision of December 31, 2020, the Tribunal of Bologna found this algorithm-driven booking system to be discriminatory, as it placed riders at a particular disadvantage in terms of their "score" in cases of cancellation or the late cancellation of booking a work session, regardless of the reason for the cancellation. The company algorithm treated riders who did not participate in booked work sessions for futile reasons in the same way as those who did not participate because they joined a strike (or for other legitimate reasons). According to the Tribunal's judgment, it was precisely this "blindness" of the statistical algorithmic processing that held the discriminatory potential of the "reputational ranking" system.

## 3. The facts

*Deliveroo* was a multinational Company operating in Italy starting in 2015 in the home food delivery sector. The company implemented an organizational model based on a network of riders, qualified by the company as "para-subordinates" (*lavoratori autonomi parasubordinati*), to carry out its delivery service.

The distribution of delivery orders among riders was carried out by an algorithm-driven application (the "App"), downloaded on a smartphone and accessible by entering personal credentials (log-in and password).

The App managed workflows among riders who booked a work session, within those hours made available by the company on the App. Each work session covered a specific time slot and a specific geographical area.

As established by the framework contract, each rider had two alternative channels for receiving delivery proposals: booking sessions in advance, using the self-service booking service or logging-in in real time though the "free log-in" service, that did not require making a reservation, but allowed riders to receive delivery offers only if there was a current work session available.

Once a rider accepted a delivery offer, they were requested to enter the work area and log-in on the App, which detected their geographical position through the geolocation system on their phone.

This self-service booking was introduced by the Company as an organizational measure coinciding with the first abstentions from work put in place by riders in 2017-2018. It gave riders the possibility of accessing the calendar for the following week, and to book one or more work sessions in which they were available for receiving delivery offers. The service was based on an algorithmic "reputational ranking" system, which attributed a "score" to each rider based on two parameters: reliability and participation.

The "reliability index" was determined by the number of occasions in which a rider, despite having booked a session, did not participate in the session, where "participating" meant logging-in within the first 15 minutes from the start of the session. If the rider did not cancel the booking in the previous 24 hours, there was an obligation to log-in within the perimeter of the work area within a maximum time frame of 15 minutes. Otherwise, failure to log-in within this time limit resulted in a reduction of this "score."

The "participation index" was determined by the number of times a rider was available for work "peaks," which occur mainly in conjunction with time slots where meals are generally taken home (from 8 pm to 10 pm, Friday to Sunday).

The value resulting from the combination of the two indexes determined the "statistics" of each rider. The order of priority accessing the self-service booking depended on the "score" assigned to each in the reputational rank. Every Monday, riders had access to the booking calendar at three different times: starting at 11:00 am, starting at 3:00 pm, or starting at 5:00 pm on the same day. Access to the first slot (at 11:00 am) was given to riders with the best "reputational ranking." The second and third slots (at 3:00 pm and 5:00 pm respectively) were attributed to riders with progressively lower reputational rankings. As a result, the time slots available for bookings, and the consequent work opportunities, were progressively reduced over time.

Such a procedure was made explicit by point 3.4 of the framework contract: "Deliveroo provides a flexible self-service booking ("SSB") which can be freely used to log in or to book sessions where the Rider wishes to receive Service Proposals. Booking through the SSB tool is entirely optional, but when used and confirmed, the Rider will be granted access to receive Service Proposals in the booked sessions. Availability during the booked sessions, if not cancelled in advance by the Rider, and work activity during times of particular traffic may be an element of preference for booking subsequent sessions."

This was also confirmed by Deliveroo's website: "if you are part of a priority group, you will have a greater chance of obtaining confirmation on your weekly requests," and "if you are part of a priority group, you will have a higher chance of being notified before the other riders."

Complaints by the applicants specifically addressed the algorithmic priority system of the self-service booking. The union asked the judge:

- to ascertain the discriminatory nature of the parameters (reliability and participation) implemented for the "reputational ranking" system, which punished riders who failed to "log-in" in time with a reduction in future job opportunities for having joined collective abstention initiatives coinciding with their own work sessions, as well as for other legitimate reasons (i.e. the exercise of other constitutional rights);
- to order any appropriate measures for removing the effects of the discriminatory conduct and practice;

- to order the adoption of a plan for the removal of any discrimination or practice that hinders the exercise of constitutional rights;
- to order the respondent company to modify the conditions of access to work sessions, so as to prevent any discriminatory effects on the right to strike and on the exercise of any other constitutional right;
- to order the company to publish the judicial decision in the FAQ area of the platform website and, at its own expense, in at least five national newspapers;
- to order payment of compensation for non-pecuniary damages in an adequate, proportionate, and dissuasive measure, on an equitable basis.

## 4. Relevant issues surrounding discriminatory AI and discrimination by AI

### *4.1. The scope of the non-discrimination principle regarding employment and occupation*

Both European Union law and Italian law include the work relationship as such, regardless of its autonomous or subordinate qualification, within the scope of anti-discrimination law.

Directive 2000/78/EC of 27 November 2000, whose purpose is to lay down a general framework for combating discrimination on the grounds of religion or beliefs, disability, age or sexual orientation with regard to employment and occupation, (pursuant to Article 3) applies "to all persons, as regards both the public and private sectors, including public bodies, in relation to:

(a) conditions for access to employment, to self-employment or to occupation, including selection criteria and recruitment conditions, whatever the branch of activity and at all levels of the professional hierarchy, including promotion;

(b) access to all types and to all levels of vocational guidance, vocational training, advanced vocational training and retraining, including practical work experience;

(c) employment and working conditions, including dismissals and pay;

(d) membership of, and involvement in, an organisation of workers or employers, or any organisation whose members carry on a particular profession, including the benefits provided for by such organisations."

While Article 3, § 1, (a) of directive 2000/78 does expressly refer to "self-employment," Article 3, § 1, (c) mentions "employment" and "working" conditions without any explicit reference to autonomous work. The European Court of Justice, in its judgment *J.K.,*[330] pointed out that Article 3, § 1, (c) must be interpreted as also referring to autonomous work relationships. The Court held that, unlike EU secondary legislation based on Article 153 TFEU which seeks to protect only workers as the weaker party in an employment relationship, directive 2000/78 seeks to eliminate, on grounds relating to the social and public interest, all discriminatory obstacles to access to livelihoods and to the capacity to contribute to society through work, irrespective of the legal form in which it is provided. The protection conferred by directive 2000/78 cannot depend on the formal categorisation of an employment relationship under national law or on the choice made between one type of contract and another.[331]

As for Italian legislation, Article 3 of legislative decree 9 July 2003, n. 216 transposes almost literally the content of Article 3 of directive 2000/78.

[330] CJEU, *J. K.*, January 12, 2023, C-356/21.

[331] CJEU, *HK/Danmark and HK/Privat*, June 2, 2022, C-587/20; *Danosa*, November 11, 2010, C-232/09.

Considering the above, the applicability of anti-discrimination law to *Deliveroo* riders cannot be seriously questioned, regardless of their formal qualification as self-employed, subordinate, or "para-subordinate" workers.

Such a conclusion holds even further in light of Article 47-*quinquies* of legislative decree n. 81/2015,[332] enacted by the Italian legislator in 2019 precisely for the purpose of overcoming the difficulties of qualifying the work relationships performed in favour of digital platforms in terms of autonomy rather than subordination. The provision established that anti-discrimination law and legislation protecting freedom and dignity of the worker apply to "self-employed workers, who carry out service of delivery of goods on behalf of others, in urban areas and with the aid of cycles or motor vehicles through platforms, including digital ones."

### *4.2. Discrimination on the grounds of trade union membership or activity*

Article 19 of the TFEU, Article 21 of the CFREU, and Article 2 of Directive 2000/78 include "*religion or personal beliefs*" among the prohibited grounds for discrimination. On one hand, these provisions, as well as Article 14 of the ECHR, do not explicitly envisage trade union activity as a prohibited factor for discrimination. On the other, Article 12 and 28 of the CFREU, as well as Article 11 of the ECHR expressly protect the effectiveness of trade union freedoms.

The ECtHR, in interpreting Article 11 of the ECHR in conjunction with Article 14 of the ECHR, held that trade union activity must be considered encompassed within the notion of "personal beliefs." In *Danilenkov and Others v. Russia,*[333] the ECtHR found that the State had failed to fulfil its positive obligation to afford effective judicial protection to employees against discrimination on the grounds of trade-union membership and who were fired by their employer due to their participation in a strike. The case involved a seaport company using various techniques to encourage employees to abandon union membership, including their reassignment to special work teams with limited opportunities, unlawful dismissals, wage reductions, disciplinary sanctions, and refusals to reinstate trade-union members following court judgments. In *Zakharova and Others v. Russia*, such a positive obligation was found to have been violated by the national courts' failure to review the various measures taken by the employer, including the reduction in working hours, salaries, and dismissals of leading members of a trade union.[334] In *Wilson, National Union of Journalists and Others v. the United Kingdom,*[335] where employers used financial incentives to induce employees to surrender important union rights, the ECtHR held that it was the role of the State to ensure that trade union members were not prevented or restrained from using the union to represent themselves.[336]

---

[332] Legislative decree n. 81/2015 was amended by decree-law of 3 September 2019, n. 101, and converted into law 2 November 2019, n. 128.

[333] ECtHR, *Danilenkov and Others v. Russia*, July 30, 2009, n. 67336/01.

[334] ECtHR, *Zakharova and Others v. Russia*, March 8, 2022, n. 12736/10.

[335] ECtHR, *Wilson, National Union of Journalists and Others v. the United Kingdom*, July 2, 2002, n. 30668/96, 30671/96 and 30678/96.

[336] For similar cases in French see Cour de Cassation, Chambre sociale, March 2, 1994, n. 92-41134; June 26, 2009, n. 08-42154; June 1, 2010, n. 09-40144; 3 May 2011, n. 09-68297; July 9, 2015, n. 14-12779. On a similar Italian case, see Tribunal of Bergamo, March 30, 2018, n. 1586.

The interpretative guidance provided by ECtHR case-law is relevant in light of EU law interpretation rules established by the Charter and the EU Treaties. Article 6 of the TEU prescribes that fundamental rights, as guaranteed by the ECHR, shall constitute general principles of Union law. Article 52 of the CFREU prescribes that, in so far as the Charter contains rights which correspond to rights guaranteed by the ECHR, the meaning and scope of those rights shall be the same as those laid down by the Convention, such as interpreted by ECtHR case-law.

Italian national law traditionally provides for an explicit prohibition on direct discrimination on grounds of trade union membership or activity within the scope of subordinate work only (Article 15 of law n. 300 of May 20, 1970, the so-called "*Statuto dei lavoratori*"). Furthermore, a broad prohibition of direct and indirect discrimination on the grounds of "personal beliefs," aimed at covering both dependent work and self-employment, has been established by Article 2 of legislative decree n. 216/2013 (enacted to transpose directive 2000/78). The case-law of national courts is unanimous in interpreting the notion of "personal beliefs," read in conformity with EU Treaties and directive 78/2000, as encompassing trade union membership or activity.[337]

The need to ensure the effectiveness of prohibition on discrimination for trade union activity is also underlined in the ILO Report of the Committee of Experts on the Application of Conventions and Recommendations, International Labour Conference, 102nd Session, 2013.

### *4.3. Legal standing of collective bodies and "collective discrimination"*

Article 9, § 2 of Directive 78/2000 provides that "Member States shall ensure that associations, organisations or other legal entities which have, in accordance with the criteria laid down by their national law, a legitimate interest in ensuring that the provisions of this Directive are complied with, may engage, either on behalf or in support of the complainant, with his or her approval, in any judicial and/or administrative procedure provided for the enforcement of obligations under this Directive."

The Court of Justice of the European Union, in its judgment on April 23, 2020,[338] held that Directive 2000/78 must be interpreted as not precluding national legislation under which an association of lawyers whose objective, according to its statutes, is the judicial protection of persons having a certain sexual orientation and the promotion of culture and respect for the rights of that category of persons, automatically, on account of that objective and irrespective of whether it is a for-profit association, has standing to bring legal proceedings for the enforcement of obligations under that directive and, where appropriate, to obtain damages, in circumstances that amount to discrimination against that category of persons and it is not possible to individually identify an injured party.

Regarding Italian legislation, Article 5, § 2 of Legislative Decree n. 216/2013, in implementing Article 9 of Directive 78/2000,[339] added a provision of standing for such associations, organisations, and other legal entities in cases of "collective discrimination," where persons wronged by a prohibited discrimination "cannot be directly and immediately identified." According to jurisprudential interpretation, the provision grants standing to

[337] Court of Cassation, judgment January 2, 2020, n. 1; judgment February 16, 2011, n. 3821.

[338] CJEU, *NH*, April 23, 2020, C-507/18, § 59.

[339] Article 8, § 1 of directive 2000/78, read in light of recital 28, establishes that Member States may introduce or maintain provisions which are more favourable to the protection of the principle of equal treatment than those laid down in that directive.

organisations that have representativeness with respect to the collective interest in question and a legitimate interest in cases where it is objectively impossible to identify an individual or individuals directly and immediately wronged by the discrimination.[340] The representativeness and the institutional purpose of the entity must result from an examination of the statute of the association.

In the present case, the Tribunal of Bologna held that all the conditions provided for by Article 5, § 2 had been met. On one hand, the statute of the applicant union established, among its aims, that "promoting the fight against all forms of discrimination, freedom of association, and the protection of employees, those employed in cooperative organisations, the self-employed non-entrepreneurs and without employees, the unemployed." On the other hand, identification of individuals affected by the discrimination was deemed objectively impossible. Thus, the Tribunal recognised the union's standing to sue regardless of evidence that *Deliveroo* riders affected by the discriminatory practice were members of the union, i.e. by providing a list of union members.

### *4.4. The burden of proof*

On point of fact, the only disputed issue was whether "late cancellation," i.e. giving notice less than 24 hours before the start of the work session) should also negatively affect the rider's statistics, and hence reduce future work opportunities. The union alleged that any cancellation, including a "late cancellation," would cause a worsening of the rider's statistics. Such circumstances were instead denied by the company, which reiterated that only cancellations made after the start of the work session resulted in a penalty in the statistics.

In disputes regarding discrimination, when the applicant provided factual elements, also deduced from statistical data, from which the existence of discriminatory acts, agreements, or behaviours can be presumed, the burden of proving the non-existence of discrimination is reversed to the defendant, as established by non-discrimination directives,[341] pursuant to the general principle of the proximity of evidence. If statistical data provided by the worker indicated a disadvantage for a group of workers, the employer has a duty to demonstrate that the choices were instead made in compliance with objective and non-discriminatory criteria.

By applying those principles, the Tribunal found that the company failed to allege and prove the concrete functioning mechanism of the algorithm. The company never clarified which specific algorithmic criteria were implemented to determine the statistics for each rider. These criteria were not even advertised on the platform, where there was only a generic reference to the parameters of reliability and participation as indexes of preference in the distribution of future work opportunities. The Tribunal hence found it to be definitively proven that, unlike what was claimed by the company, "late cancellation" could also determine the penalisation of a rider's statistics.

### *4.5. The "blindness" of the algorithm*

The Tribunal's reasoning began from the assumption that riders' statistics were penalised regardless of motive for an abstention from work: the algorithmic profiling system adopted

[340] Court of Cassation, judgment July 20, 2018, n. 19443.
[341] See Article 10 of Directive 2000/78, Article 19 of Directive n. 2006/54, Article 8 of Directive 2000/43.

by the platform treated those who did not participate in a booked session for trivial reasons the same way as those who were absent from work for the purposes of exercising a constitutional right, like the right to strike, as well as for other reasons deemed "worthy of protection" (i.e. illness, disability, or the need to assist a disabled person or a sick minor). Hence, the algorithmic distribution of delivery offers put in place a practice that, in concrete terms, discriminated against riders belonging to this second category of workers, marginalising them from the priority group and significantly reducing their future opportunities for access to work.

In short, the platform was "blind": it did not know and did not seek to know the motives about why a rider had cancelled a booking or did not participate in a booked session. It is precisely this "unconscious" processing of algorithmic statistics that harbours the discriminatory potential of "reputational rank" profiling.

The discriminatory nature of an act, behaviour, or practice is relevant in its objectivity, regardless of whether the perpetrator has awareness or intention to discriminate against a particular person or group of persons. Thus, the ascertained "blindness" of the algorithm was sufficient for declaring the discriminatory nature of the Company practice, as demonstrated in its objectivity.

Furthermore, it was also undisputed that Deliveroo reserved "special" treatment for the algorithmic processing of rider statistics, by applying an *ad hoc* corrective action in two cases which were evidently deemed worthy of protection by the company: where the rider's non-participation was due to an injury or system malfunction.

According to the judgement of the Tribunal, the existence of such an exception demonstrated that the failure to adopt a similar corrective action in other cases was the result of a conscious choice by the company. The algorithmic platform, when it wanted to, could remove the "blindfold" that made it "unaware" of the reason for an abstention from work. If it did not, it was because the platform deliberately chose to place all the other reasons on the same level other than the injury of a worker or the malfunctioning of the system.

### *4.6. Direct or indirect discrimination?*

The notion of direct discrimination is defined similarly under the ECHR and EU law and case-law. Under EU law, Article 2, § 2 of the EU Racial Equality Directive states that direct discrimination is "taken to occur where one person is treated less favourably than another is, has been, or would be treated in a comparable situation on grounds of racial or ethnic origin." Article 2, § 2 of the Employment Equality Directive states that direct discrimination "shall be taken to occur where one person is treated less favourably than another is, has been, or would be treated in a comparable situation, on any of the protected grounds" (religion or beliefs, disability, age, or sexual orientation). Similar definitions are provided by Article 2, § 1 of the Gender Equality Directive and Article 2 of the Gender Goods and Services Directive.[342] ECtHR case-law considers that discrimination is "direct" when a difference in

[342] CJEU, *Centrum voor gelijkheid van kansen en voor racismebestrijding v. Firma Feryn NV*, July 10, 2008, C-54/07; CJEU, *Asociația Accept v. Consiliul Național pentru Combaterea Discriminării*, April 25, 2013, C-81/12; CJEU, *Wolfgang Glatzel v. Freistaat Bayern*, May 22, 2014, C-356/12; CJEU, *P. v. S. and Cornwall County Council*, April 30, 1996, C-13/94; CJEU, *Frédéric Hay v. Crédit agricole mutuel de Charente-Maritime et des Deux-Sèvres*, December 12, 2013, C-267/12; CJEU, *Debra Allonby v. Accrington & Rossendale College, Education Lecturing Services, trading as Protocol Professional and Secretary of State for Education and Employment*, January 13, 2004, C-256/01; CJEU, *Elisabeth Johanna Pacifica Dekker v. Stichting Vormingscentrum voor Jong Volwassenen (VJV-Centrum) Plus*, November 8, 1990, C-177/88; CJEU, *Carole Louise Webb v. EMO Air Cargo (UK) Ltd.*, July 14, 1994, C-32/93; CJEU, *North Western*

the treatment of persons in analogous, or relevantly similar, situations is directly based on an identifiable characteristic.[343]

Otherwise, indirect discrimination is determined by a prohibited disparity as an effect of an apparently neutral provision, criterion, or practice, which it is itself legitimate. In some situations, offering the same treatment to persons who are in different situations may put certain individuals or groups at a particular disadvantage. In such cases, it is not the treatment that differs, but rather the effects of that treatment, which will differently affect persons with different characteristics.

Under EU law, Article 2, § 2 of the Racial Equality Directive states that "indirect discrimination shall be taken to occur where an apparently neutral provision, criterion, or practice would put persons of a racial or ethnic origin at a particular disadvantage compared to other persons." Similar dispositions are provided for by Article 2, § 2 of the Employment Equality Directive and Article 2, § 1 of the Gender Equality Directive.[344] The ECtHR has drawn on this definition of indirect discrimination in several judgments, stating that a difference in treatment may take the form of disproportionately prejudicial effects of a general policy or measure which, though couched in neutral terms, discriminates against a group.[345]

In light of these principles, the Tribunal of Bologna held that the SSB implemented a practice which amounted to indirect discrimination. The algorithmic booking criteria applied an apparently neutral provision (personal scoring based on "reliability" and "participation" indexes), which however put a specific category of workers (those taking part in union initiatives to abstain from work) in a position of potential disadvantage. Considering reasons for not attending a work session irrelevant meant treating different situations in the same way, which is what indirect discrimination typically consists of.

In view of the finding of such an objective difference in treatment, the Tribunal found that the Company had not fulfilled the burden of proof for the existence of a legitimate aim and the appropriateness and necessity of the means to achieve it. The Tribunal pointed out

---

*Health Board v. Margaret McKenna*, September 8, 2005, C-191/03, § 50; CJEU, *Sabine Mayr v. Bäckerei und Konditorei Gerhard Flöckner OHG*, GC, February 26, 2008, C-506/06; CJEU, GC, *S. Coleman v. Attridge Law and Steve Law*, July 17, 2008, C-303/06.

343 ECtHR, *Biao v. Denmark*, GC, May 24, 2016, n. 38590/10, § 89; ECtHR, *Carson and Others v. the United Kingdom*, GC, March 16, 2010, n. 42184/05, §61; ECtHR, *D.H. and Others v. the Czech Republic*, GC, November 13, 2007, n. 57325/00, § 175; ECtHR, *Burden v. the United Kingdom*, GC, 29 April 2008, n. 13378/05, § 60; ECtHR, *Thlimmenos v. Greece*, GC, April 6, 2000, n. 34369/97 § 44; ECtHR, *Pretty v. the United Kingdom*, April 29, 2002, n. 2346/02, § 88; ECtHR, *Petrov v. Bulgaria*, May 22, 2008, n. 15197/02, § 55; ECtHR, *Varnas v. Lithuania*, July 9, 2013, n. 42615/06; ECtHR, *Guberina v. Croatia*, March 22, 2016, n. 23682/13.

344 CJEU, GC, *Tadao Maruko v. Versorgungsanstalt der deutschen Bühnen*, April 1, 2008, C-267/06; CJEU, *Isabel Elbal Moreno v. Instituto Nacional de la Seguridad Social (INSS) and Tesorería General de la Seguridad Social (TGSS)*, November 22, 2012, C-385/11; CJEU, *Johann Odar v. Baxter Deutschland GmbH*, December 6, 2012, C-152/11; CJEU, *Helga Nimz v. Freie und Hansestadt Hamburg*, February 7, 1991, C-184/89; CJEU, *Maria Kowalska v. Freie und Hansestadt Hamburg*, June 27, 1990, C-33/89; CJEU, *M. A. De Weerd, née Roks, and Others v. Bestuur van de Bedrijfsvereniging voor de Gezondheid, Geestelijke en Maatschappelijke Belangen and Others*, February 24, 1994, C-343/92; CJEU, *Frédéric Hay v. Crédit agricole mutuel de Charente-Maritime et des Deux-Sèvres*, 12 December 2013, C-267/12; CJEU, GC, *CHEZ Razpredelenie Bulgaria AD v. Komisia za zashtita ot diskriminatsia*, 16 July 2015, C-83/14.

345 ECtHR, *Biao v. Denmark*, GC, 24 May 2016, n. 38590/10, § 89; ECtHR, *D.H. and Others v. the Czech Republic*, GC, 13 November 2007, n. 57325/00, § 175; ECtHR, *Di Trizio v. Switzerland*, 2 February 2016, n. 7186/09; ECtHR, *N. B. v. Slovakia*, 12 June 2012, n. 29518/10; ECtHR, *V.C. v. Slovakia*, 8 November 2011, n. 18968/07.

that it was not the lawfulness of the algorithmic reputational ranking itself that was in question, but only its "indifference" to the reasons for abstentions.

The availability of a second booking channel (free log-in system) did not remove the discriminatory nature of the practice. First, the free log-in service, which was substantially a residual way of receiving delivery offers, could not be considered an equivalent alternative to the self-service booking, which in practice covered a large portion, and sometimes all, of the available work sessions. Second, it was undisputed that the Company encouraged riders to join the "priority group" on its website to obtain benefits, emphasising the "privileges" given to riders belonging to that group.

## 5. Regulation (EU) 2024/1689 on artificial intelligence: what possible impact on the reputational ranking of workers

Regulation (EU) 2024/1689 came into force on August 1, 2024 and will be applied, with some exceptions, beginning August 2, 2026. Although the Deliveroo case fell outside the scope of the Regulation *ratione temporis*, the discipline established by the Regulation is likely to impact algorithmic organisation models that seek to manage the distribution of work based on reputational ranking mechanisms.

The Regulation provides four levels of risk to the health and fundamental rights of individuals: unacceptable risk, high risk, transparency risk, and minimum risk and follows a risk-based approach; the higher the risk of causing harm to society, the stricter the rules. The most relevant are "unacceptable risk" and "high risk" activities, deriving from prohibited AI practices respectively, such as the use of emotion recognition systems in the workplace (Article 5(a)), and high-risk AI systems (Article 6).

With regard to high-risk AI systems, Recital 57 points out that AI systems specifically used in employment, worker management, and access to self-employment (among others) for allocating tasks on the basis of individual behaviour, personal traits or characteristics, and for monitoring or evaluating persons in work-related contractual relationships may have an appreciable impact on future career prospects, the livelihoods of those persons, and workers' rights. Such systems may perpetuate historical patterns of discrimination, for example against women, certain age groups, persons with disabilities, or persons of certain racial or ethnic origins or sexual orientation. AI systems used to monitor the performance and behaviour of such persons may also undermine their fundamental rights to data protection and privacy.

Annex III of the Regulation contains a list which includes systems operating in the fields of employment, worker management, and access to self-employment (Annex III, § 4), namely:

(a) AI systems intended to be used for the recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates;

(b) AI systems intended to be used to make decisions affecting the terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics, or to monitor and evaluate the performance and behaviour of persons in such relationships.

Algorithm-driven practices, identical or similar to *SSB,* evidently fall into the typology of AI systems intended to be used to "allocate tasks based on individual behaviour or personal traits or characteristics." In the Deliveroo case, relevant individual behaviours were particularly sensitive, as they were an indirect expression of trade union opinion.

The Regulation establishes obligations not only for providers (Article 16 and following ones), importers (Article 23), and distributors (Article 24 and following ones) of high-risk AI systems, but also for deployers (Article 26 and following ones), i.e. any "natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity" (Article 3(4)).

Whenever an employer or client holds the role of "deployer," he/she is subject to their respective obligations, including:

- to assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support (Article 26, § 2);
- to monitor the operation of the high-risk AI system on the basis of use instructions. Where deployers have reason to consider use of a high-risk AI system in accordance with instructions may result in that AI system presenting risks to the health or safety, or to fundamental rights of persons, to inform the provider or distributor and the relevant market surveillance authority without undue delay, and suspend use of that system (Article 26, § 5);
- before putting a high-risk AI system into service or using it at the workplace, to inform workers' representatives and the affected workers that they will be subject to use of the high-risk AI system.

Finally, the Regulation provides the right to explanation of individual decision-making: any affected person subject to a decision which is made by the deployer on the basis of the output from a high-risk AI system, and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights, has the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision made (Article 86).

## 6. Hands-on scenario: a hypothetical case

An Italian court, in order to comply with the principles enshrined in the judgement of the Joint Sessions of the Supreme Court n. 18287/2018, decides to adopt an algorithm in order to quantify the amount of the sum of money that a former spouse must provide to the other after a decision upholding a request for divorce and, thus, ending their marriage. In order to comply with the function of such a sum of money, according to Article 5 paragraph 6 of law n. 898/1970, duly amended, the algorithm takes various indexes into consideration, among which it establishes a mathematical relationship between housework and work as 35% of the latter.

A woman contests the decision of the court based on the results provided by the algorithm claiming that the consequence was a sum of money that is too low, since throughout the years of marriage where she had to leave her job, her housework did not align with the parameter, since she would not be able to provide an income of 35% of that which she made had she continued her career.

Please refer to the following questions:

1) What measures should be taken by the Court in order to use an algorithm to provide directions in such cases?

2) What (other) factors should be taken into consideration in order to build and train the above-cited algorithm?
3) To what extent is a mathematical model suitably applied to this case ? Does it have consequences on the principle of substantial equality?
4) In case the woman's claim is upheld, what might be the individual remedies available?
5) In case the woman's claim is upheld, in which way should the mechanism be redesigned?
6) Can the principle of judicial independence somehow be affected by the solutions adopted?

# Section 8
# Annexes

## Annex 8.1.
## *De iure condito* and *de iure condendo*: the positive and soft law regulatory bodies

EUROPEAN UNION:

i) EUROPEAN COMMISSION AND EXPERT GROUPS ESTABLISHED BY THE EC
- WP29, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 17/EN WP251 rev.01, February 6, 2018
- *Artificial Intelligence for Europe,* COM(2018) 327 final, April 25, 2018;
- *Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions*, *Coordinated Plan on Artificial Intelligence*, COM(2018) 795 final, December 7, 2018;
- HIGH-LEVEL EXPERT GROUP, *Ethics guidelines for trustworthy AI*, April 8, 2019;
- HIGH LEVEL EXPERT GROUP, *Policy and investment recommendations for trustworthy AI*, June 26, 2019;
- DIRECTORATE-GENERAL FOR JUSTICE AND CONSUMERS, *Study on the use of innovative technologies in the justice field – Final Report*, 2020;
- *White Paper on Artificial Intelligence - A European approach to excellence and trust*, COM(2020) 65 final, February 19, 2020;
- European Commissions' Reasoned Opinion (C(2024) 7814) on the Italian Proposal for a Law on Artificial Intelligence.

ii) Other EU institutions and bodies:

- EUROPEAN ECONOMIC AND SOCIAL COMMITTEE, *Opinion on "Artificial Intelligence – The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society,"* Document C-288, August 31, 2017;
- EUROPEAN COUNCIL, *A new strategic agenda for the EU 2019-2024*, June 21, 2019;
- EUROPEAN DATA PROTECTION BOARD, *Guidelines 05/2020 on consent under Regulation 2016/679*, May 4, 2020;
- EUROPEAN PARLIAMENT, *What is Artificial Intelligence and how is it used?*, September 4, 2020;
- EUROPEAN PARLIAMENT, *Resolution on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters*, 2020/2016(INI), October 6, 2021;
- EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS, *Bias in Algorithms. Artificial Intelligence and Discrimination*, Vienna, 2022;
- EUROPOL, *Policing in the metaverse: what law enforcement needs to know, an observatory report from the Europol Innovation Lab*, Luxembourg, 2022;
- EUROPEAN PARLIAMENT, *Artificial Intelligence in Healthcare*, June 2022;

- EUROPEAN PARLIAMENT, *Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*, (P9_TA(2023)0236), June 14, 2023;
- EUROPEAN PARLIAMENT AND COUNCIL, *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*.

COUNCIL OF EUROPE:

i) COUNCIL:

- *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, Council of Europe, ETS No. 108, October 1, 1985;
- *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, Council of Europe, ETS No. 225, September 5, 2024.

ii) CEPEJ:

- *Guidelines on how to drive change towards Cyberjustice*, CEPEJ (2016)13, December 7, 2016;
- *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*, December 3-4, 2018;
- *Toolkit for supporting the implementation of the Guidelines on how to drive change towards Cyberjustice*, CEPEJ(2019)7, June 14, 2019;
- *Possible introduction of a mechanism for certifying artificial intelligence tools and services in the sphere of justice and the judiciary: Feasibility Study*, CEPEJ(2020)15Rev, December 8, 2020;

  *Guidelines on videoconferencing in judicial proceedings*, June 2021;
- WORKING GROUP ON CYBERJUSTICE AND ARTIFICIAL INTELLIGENCE, *Selected national good practices on videoconferencing in judicial proceedings*, CEPEJ-GT-CYBERJUST(2021)11, November 30, 2021;
- *2022 – 2025 CEPEJ Action plan: "Digitalisation for a better justice,"* CEPEJ(2021)12 Final, December 9, 2021;
- *Guidelines on electronic court filing (e-filing) and digitalisation of courts*, CEPEJ(2021)15, December 9, 2021;
- *Comparative Study on the use of judicial e-auctions in the Council of Europe Member States*, CEPEJ-GT-CYBERJUST(2023)1, June 16, 2023.

iii) CAHAI:

- *Towards regulation of AI systems*, DGI (2020)16, December 2, 2020;
- *Feasibility Study,* CAHAI(2020)23, December 17, 2020;
- *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law,* CAHAI(2021)09rev, December 3, 2021.

iv) CAI:
- *Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy and The Rule of Law*, CAI(2023)01, January 6, 2023;
- *Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and The Rule of Law*, CAI(2023)18, July 7, 2023.

v) Other entities:

- *Recommendation on the protection of individuals about automatic processing of personal data in the context of profiling*, CM/Rec(2010)13, November 23, 2010;
- *Guidelines on Protecting Individuals from Data Processing in a Big Data World*, T-PD(2017)01, January 23, 2017;
- *Recommendation on technological convergence, artificial intelligence, and human rights*, Recommendation 2102(2017), April 28, 2017;
- *Algorithms and Human Rights*, DGI(2017)12, 2018;
- *Practical Guide on the use of personal data in the police sector*, T-PD(2018)01, February 15, 2018;
- *Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, CETS No. 223, October 10, 2018;
- *Technical study on online dispute resolution mechanisms*, CDCJ(2018)5, November 14-16, 2018
- *Declaration on the manipulative capabilities of algorithmic processes*, CM Decl(13/02/2019)1, February 13, 2019;
- *Guidelines on Artificial Intelligence and Data Protection*, T-PD(2019)01, February 25, 2019;
- *Unboxing AI: 10 steps to protect human rights, Recommendation by the Commissioner for Human Rights*, May 2019;
- *Recommendation on the human rights impacts of algorithmic systems*, CM/Rec(2020)1, April 8, 2020;
- *Need for democratic governance of artificial intelligence*, Doc.15150, September 24, 2020;
- *Guidelines on facial recognition*, T-PD(2020)03rev4, January 28, 2021;
- *Guidelines on online dispute resolution mechanisms in civil and administrative court proceedings*, CM(2021)36add4-final, June 16, 2021;
- *Recommendation on the protection of individuals with regard to automatic processing of personal data, including machine learning, in the context of profiling*, CM/Rec(2021)8, November 3, 2021;
- *Artificial Intelligence and Administrative Law*, December 2022;
- *Follow-up Recommendation to "Unboxing AI" (2019). Human rights by design future-proofing human rights protection in the era of AI*, May 2023.

NATIONAL SOURCES MENTIONED IN THIS HANDBOOK

- ITALY: Disegno di legge n. 1146/2024 of May 20, 2024 "Disposizioni e delega al Governo in materia di Intelligenza Artificiale" aggiungere qui parere Commissione Europea e nel testo;

- SPAIN: Ley 2213 de 2022 "Por medio de la cual se establece la vigencia permanente del decreto legislativo 806 de 2020 y se adoptan medidas para implementar las tecnologías de la información y las comunicaciones en las actuaciones judiciales, agilizar los procesos judiciales y flexibilizar la atención a los usuarios del servicio de justicia y se dictan otras disposiciones" ;

- US: NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, *AI Risk Management Framework*, 2023.

# Annex 8.2.
# The living law:
# case-law and relevant opinions from public authorities

COURT OF JUSTICE OF THE EUROPEAN UNION

- *GC and Others*, September 24, 2019, C-136/17;
- *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, May 13, 2014, C-131/12;
- *Fratelli Variola S.p.A. v Amministrazione italiana delle Finanze,* October 10, 1973, C-34/73;
- *Vieh- und Fleischkontor GmbH v Hauptzollamt Hamburg-St. Annen*, February 11, 1971, C-39/70.

EUROPEAN COURT OF HUMAN RIGHTS

- *Borg v. Malta*, June 26, 2012, n. 37537/13;
- *Danilenkov and Others v. Russia*, July 30, 2009, n. 67336/01;
- *G.I.E.M. S.R.L. and Others v. Italy*, June 28, 2018, n. 1828/06;
- *Wilson, National Union of Journalists and Others v. the United Kingdom*, July 2, 2002, n. 30668/96, 30671/96 and 30678/96;
- *Zakharova and Others v. Russia*, March 8, 2022, n. 12736/10.

NATIONAL COURTS

- Australia, Federal Court, *Commissioner of Patents v Thaler*, April 13, 2022;
- Austria, Bundesverwaltungsgericht - W211 2234354-1, December 22, 2021;
- Ireland, High Court, *Irish Bank Resolution Corporation Ltd v Quinn*, 2015;
- Italy, Constitutional Court, May 12, 2021, n. 96; March 24, 2022, n. 74;
- Italy, Court of Cassation, July 20, 2018, n. 19443; January 2, 2020, n. 1; Sez. V, July 20, 2022, n. 28675; Sez. I, October 10, 2023, n. 28358;
- Italy, Tribunal of Rome, April 4, 2018, n. 5715;
- The Netherlands, District Court of the Hague, March 6, 2020, C-09-550982-HA ZA 18-388;
- United Kingdom, High Court of Justice, *Pyrrho Investments Limited v MWB Property Limited,* 2016;
- United States, Supreme Court of Wisconsin, *State of Wisconsin v. Eric L. Loomis*, April 2016;
- United States, District Court, Southern District of New York, *Da Silva Moore et al., v Publicis Groupe et al.*, 2012;
- United States, Court of Appeals (1st Circuit), *Brouse v. United States*, No. 2850, December 15, 1933.

NATIONAL ADMINISTRATIVE AUTHORITIES

- Italy, Garante per la Protezione dei Dati Personali, November 24, 2016, n. 488; March 30, 2023, n. 112; April 11, 2023, n. 114.

## Annex 8.3.
## A joint effort: academic inputs and insights


- *A Chinese local court recently opened a hearing in the Metaverse, saying it helps drive the digitization of the judicial system*, in *The First Metaverse Trial Opened in a Local Court in China*, 2022;
- AGUZZI C., *Le juge et l'intelligence artificielle: la perspective d'une justice rendue par la machine,* in *Annuaire international de justice constitutionnelle 2019*, 2020;
- AINORA I., *Smart courts: verso una nuova dimensione della giustizia*, 2021;
- ALETRAS N. ET AL., *Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective*, in *PeerJ Computer Science*, 2016;
- ALETRAS N., TSARAPATSANIS D., PREOTIUC-PIETRO D., LAMPOS V., *Pre-dicting judicial decisions of the european court of human rights: A natural language processing perspective,* in *PeerJ Computer Science*, 2016;
- ALLEGREZZA S., COVOLO V., *Effective Defence Rights in Criminal Proceedings. A European and Comparative Study on Judicial Remedies*, Padua, 2018;
- ANGWIN J., LARSON J., MATTU S., KIRCHNER L., *Machine Bias*, 2016;
- ASHLEY K. D., *Artificial intelligence and legal analytics: new tools for law practice in the digital age*, Cambridge University Press, 2017;
- ASHLEY K. D., BRÜNINGHAUS S., *Automatically classifying case texts and predicting outcomes*, in *Artificial Intelligence and Law*, 2009;
- Ashley, K. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, 2017;
- ASPRELLA C., *Riforma processo civile: le nuove disposizioni in materia di udienza*, 2022;
- BARYSĖ D., SAREL R., *Algorithms in the court: does it matter which part of the judicial decision-making is automated?*, in *Artificial Intelligence Law*, 2023;
- BATTELLI E., *Giustizia predittiva, decisione robotica e ruolo del giudice*, in *Giustizia civile*, 2020;
- BEKKER S., *Fundamental Rights in Digital Welfare States: The Case of SyRI in the Netherlands,* in SPIJKERS O., WERNER W. G., WESSEL R. A. (eds.) *Netherlands Yearbook of International Law 2019*, The Hague, 2021;
- BELLO C., *Future of justice: Colombia makes history by hosting its first-ever court hearing in the metaverse*, 2023;
- BICHI R., *Intelligenza Artificiale tra "calcolabilità" del diritto e tutela dei diritti*, in *Giurisprudenza italiana*, 2019;
- BIEVIER C., *The easy intelligence tests that AI Chatboats failed*, in *Nature*, 2023;
- BILLI M., CALEGARI R., CONTISSA G., LAGIOIA F., PISANO G., SARTOR G., SARTOR G., *Argumentation and defeasible reasoning in the law*, in *J*, 2021;
- BLANCO-GONZÁLEZ A., CABEZÓN A., SECO-GONZÁLEZ A. O., CONDE-TORRES D., ANTELO-RIVEIRO P., PIÑEIRO A., GARCIA-FANDIN R., *The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies*, in Pharmaceuticals, 2023;
- BRUNELLI B., *Il processo civile telematico che verrà*, in *Rivista trimestrale di diritto e procedura civile*, 2021;
- BRUTTI, N. *Mito del consenso e* rating *reputazionale*, in *La Nuova Giurisprudenza Civile Commentata*, 2024;
- BYGRAVE L. A., *Article 22*, in KUNER C., BYGRAVE L. A., DOCKSEY C. (eds.), *The EU General Data Protection Regulation (GDPR): A Commentary*, Oxford, 2020;

- CADIET L., CHAINAIS C., SOMMER J.-M. (eds.), JOBERT S., JOND-NECAND E. (rapp.), *Dissemination of decision-making data and case law*, June 2022;
- CAIANIELLO M., LASAGNI G., *Comparative Remarks*, in CONTISSA G., LASAGNI G., CAIANIELLO M., SARTOR G. (eds.), *Effective Protection of the Rights of the Accused in the EU Directives. A Computable Approach to Criminal Procedure Law,* Leiden, 2022;
- CASANOVAS P. ET AL., *Introduction: Theory and methodology in legal ontology engineering: Experiences and future directions*, Springer, 2011;
- CASONATO C., *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *Biolaw Journal*, 2019;
- CASSANO G., SCORZA G. (eds.), *Metaverse. User rights - digital platforms - privacy - copyright - criminal profiles - blockchain and NFT*, Pisa, 2023;
- CASTELLS M., *The rise of the Network society,* Oxford, 2000;
- CATAROZZO M. A., *Il Metaverso: quali opportunità per i legali?*, 2022;
- CAVANI V., *Nuovi poteri, vecchi problemi. Il costituzionalismo alla prova del digitale*, in *Diritto pubblico comparato ed europeo*, 2023;
- CELESTE E., *Digital Constitutionalism. The Role of Internet Bills of Rights*, London, 2022;
- CHALKIDIS I. ET AL., *LEGAL-BERT: The muppets straight out of law school*, 2020;
- CHIAVARIO M., *La "videoconferenza" processuale e la Corte europea dei diritti dell'uomo*, in AA.VV., *Studi in onore di Mario Pisani, vol. II – Diritto processuale penale e profili internazionali: diritto straniero e diritto comparato, a cura di Corso-Zanetti*, Piacenza, 2010;
- CONSEIL DES BARREAUX EUROPÉENS, *Position paper on the proposal for a regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*, 2021;
- CONTISSA G., LASAGNI G., CAIANIELLO M., SARTOR G. (eds.), *Effective Protection of the Rights of the Accused in the EU Directives. A Computable Approach to Criminal Procedure Law,* Leiden, 2022;
- CORTE DI CASSAZIONE. UFFICIO DEL MASSIMARIO. SERVIZIO PENALE, *Relazione su novità normativa. La "riforma Cartabia,"* 2023;
- COSTANZO P., *Il fattore tecnologico e le trasformazioni del costituzionalismo*, in AA.VV., *Costituzionalismo e globalizzazione. Atti del XXVII Convegno annuale dell'Associazione italiana dei costituzionalisti (Salerno 22-24 novembre 2012)*, Naples, 2014;
- CURTOTTI D., *I collegamenti audiovisivi nel processo penale*, Milan, 2006;
- D'ACQUISTO G., *Chatgpt e AI, regolamentare la responsabilità o l'efficienza è la prossima sfida*, in *AgendaDigitale*, April 18, 2023;
- DAL PONT T., GALLI F., LOREGGIA A., PISANO G., ROVATTI R., SARTOR G., *Legal Summarisation through LLMs: The PRODIGIT Project*, 2023;
- DE BIASE L., FINOCCHIARO G., POLLICINO O., *L'Intelligenza artificiale e la questione giuridica dei danni e della soggettività*, in *ilSole24Ore*, March 8, 2023;
- DE MINICO G., *La tecnologia non si governa con la paura, ma con regole ad hoc*, 2023;
- DE STEFANO F., *L'intelligenza artificiale nel processo*, in *Giustizia Insieme*, 2020;
- DE VOCHT D. L. F., *Trials by video link after the pandemic: the pros and cons of the expansion of virtual justice*, in *China-EU Law Journal*, 2022;
- DEI-CAS E. A. A., *La partecipazione a distanza, in attesa della riforma del processo penale*, in *Sistema penale*, 2022;
- DELLA TORRE J., *L'espansione del processo a distanza negli itinerari dell'emergenza pandemica*, in *Processo penale e giustizia*, 2021;

- DERLÉN M., LINDHOLM J., *Goodbye van Gend en Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments*, in *European Law Journal*, 2014;
- DESMOULIN-CANSELIER, LE METAYER, *Décider avec les algorithms*, Paris, 2020;
- DI CHIARA G., *Come s'uno schermo. Partecipazione a distanza, efficienza, garanzie, upgrade tecnologici*, in *Rivista di diritto processuale*, 2018;
- DI FEDERICO G., *Cari avvocati, non demonizzate l'uso delle nuove tecnologie*, 2020;
- DONATI F., *Intelligenza artificiale e giustizia,* in *Rivista AIC*, 2020;
- DONG X. L., GABRILOVICH E., HEITZ G. ET AL., *Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion.* in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014;
- EDWARDS L., *Regulating AI in Europe: Four Problems and Four Solutions*, Ada Lovelace Institute, 2022;
- EJFRI, *Survey on the digitalisation of justice and the use of artificial intelligence in the Judiciary in the EU Member States*, 2022;
- ESSER J., *Vorverständnis und Methodenwahl in der Rechtsfindung*, Frankfurt am Main, 1970;
- FALCONE A., *Online Hearings and the Right to Effective Defence in Digitalised Trials*, in BACHMAIER WINTER L., RUGGERI S. (eds.), *Investigating and Preventing Crime in the Digital Era*, Cham, 2022;
- FINOCCHIARO G., *XVIII lezione: intelligenza artificiale, privacy e data protection*, in RUFFOLO U. (ed.), *XXVI Lezioni di diritto dell'intelligenza artificiale*, Turin, 2021;
- FORTIS S., *The metaverse is testing the limits of what is legally possible*, 2023;
- FROSINI T. V., *La protezione della riservatezza nella società informatica*, in MATTEUCCI N. (ed.), *Privacy and* data *banks,* Bologna, 1981;
- FUMI V., *Metaverso e fenomenologia*, in *Sicurezza e giustizia*, 2022;
- GALASSI A., LIPPI M., TORRONI P., *Multi-task attentive residual networks for argument mining*, in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023;
- GALGANI B., *Forme e garanzie nel prisma dell'innovazione tecnologica. Alla ricerca di un processo penale "virtuoso,"* Milan, 2022;
- GALLI F. ET AL., *Predicting outcomes of Italian VAT decisions*, in *Jurix 2022, Proceeding of the 35th International Conference on Legal Knowledge and Information Systems*, 2022;
- GALLI F., SARTOR G., *AI Approaches to Predictive Justice: A Critical Assessment*, in *Humanities and Rights Global Network Journal*, 2023;
- GIALUZ M., *Per un processo penale più efficiente e giusto. Guida alla lettura della Riforma Cartabia*, 2022;
- GIDEON C., *Predictive coding: adopting and adapting artificial intelligence in civil litigation*, in *SSNR*, 2020;
- GIGERENZER G., *How to Stay Smart in a Smart World: Why Human Intelligence Still Beats Algorithms,* 2022;
- GIOSTRA G., *Appunto per una giustizia non solo più efficiente, ma anche più giusta*, 2022;
- Google AI Blog, *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*, 2018;
- GRIMM P. W., GROSSMAN M. R., GLESS S., HILDEBRANDT M., *Artificial Justice: the Quandary of AI in the courtroom*, in Judicature International, 2022;
- GROOTENDORST M., *KeyBERT: Minimal keyword extraction with BERT*, in *Zenodo*, 2020;
- GROSS MUNOZ E. J., FABREGAT R., BACCA-ACOSTA J., DUQUE-MENDEZ N., AVILA-GARZON C., *Augmented Reality, Virtual Reality, and Game Technologies in Ophthalmology Training,* 2022;

- GRUNDLER G. ET AL., *Detecting Arguments in CJEU Decisions on Fiscal State Aid*, in *ArgMining 2022: 9th Workshop on Argument Mining*, 2022, 143; P. Santin et al., *Argumentation Structure Prediction in CJEU Decisions Fiscal State Aids*, in *Proceeding of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023)*, June 19–23, 2023, Braga, Portugal. ACM, 2023;
- GUTHRIE A., *Justice in Metaverse: Here's What the First Virtual Court Hearing in Colombia Looked Like*, 2023;
- G'SELL, F., *Regulating under Uncertainty: Governance Options for Generative AI, SSRN, 2024*;
- HASSANIEN A. E., DARWISH A., TORKY M. (eds.), *The Future of Metaverse in the Virtual Era and Physical World. Studies in Big Data*, Cham, 2023;
- HELLER B., CASTAÑO D., *Artificial Intelligence, Virtual Courts, and Real Harms*, 2023;
- JANSEN B., SCHREINER A. T. M., *Captured by Digitization*, in *International Journal for the Semiotics of Law*, 2022;
- JOHNSON S. B., KING A. J., WARNER E. L., ANEJA S., KANN B. H., BYLUND C. L., *Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information*, in *JNCI Cancer Spectrum*, 2023;
- KATZ D. M., BOMMARITO M. J., BLACKMAN J., *A general approach for predicting the behavior of the Supreme Court of the United States*, in *PloS one*, 2017;
- Knuth D. E., *The Art of Computer Programming*, Addison-Wesley, 2023;
- LIBERTINI M., MAUGERI M. R., VINCENTI E., *Giustizia predittiva e giurisdizione civile. Primi appunti*, in PAJNO A., DONATI F., PERRUCCI A. (eds.), *Intelligenza artificiale e diritto: una rivoluzione?, II, Amministrazione, responsabilità, giurisdizione*, Bologna, 2022;
- LIGA D., AMITRANO D., MARKOVICH R., *PaTrOnto, an ontology for patents and trademarks*, in *New Frontiers in Artificial Intelligence: JSAI-isAI 2023 Workshops,* AI-Biz, EmSemi, SCIDOCA, JURISIN 2023 Workshops, Hybrid Event, June 5–6, 2023, Revised Selected Papers, Cham, 2024;
- LIGA D., FIDELANGELI A., MARKOVICH R., *OntoVAT, an ontology for knowledge extraction in VAT-related judgments*, in *New Frontiers in Artificial Intelligence: JSAI-isAI* 2023 Workshops, AI-Biz, EmSemi, SCIDOCA, JURISIN 2023 Workshops, Hybrid Event, June 5–6, 2023, Revised Selected Papers, Cham, 2024;
- LIVERMORE M. A., ROCKMORE D. N. (eds.), *Law as Data: Computation, Text, & the Future of Legal Analysis*, Santa Fe Institute Press, 2018;
- LOHSSE S., SCHULZE R, STAUDENMAYER D. (eds.), *Liability for AI*, Baden-Baden/Oxford, 2023;
- LORGE PARNAS D., *The Real Risks of Artificial Intelligence*, Communications of the ACM, 2017;
- LOUIS A., VAN DIJCK G, SPANAKIS G., *Finding the Law: Enhancing Statutory Article Retrieval via Graph Neural Networks*, January 2023;
- LÜTZ F., *The AI Act, gender equality and non-discrimination: what role for the AI office?*, in *ERA Forum*, 2024;
- MANTELERO A., *Beyond data: Human rights, ethical and social impact assessment in AI,* Springer Nature, 2022;
- MARCHETTI T. B., voice *Digital administration*, in *Enciclopedia del diritto*, Milan, 2022;
- MARINO D. M., OLIVIERI A., *Chi paga per i danni causati dall'intelligenza artificiale?*, in *Wired*, March 7, 2023;
- MARTONE M., *Prime riflessioni su lavoro e metaverso*, in *Argomenti di diritto del lavoro*, 2022;;

- MAUGERI M. R., *I robot e la possibile "prognosi" delle decisioni giudiziali*, in CARLEO A. (ed.), *Decisione robotica,* Bologna, 2019;
- MCKEY C., MACINTOSH K., *Digital Criminal Courts: The Place or Space of (Post- ) pandemic Justice*, in AA.VV., *Cybercrime in the Pandemic Digital Age and Beyond*, Cham, 2023;
- MEAKER M., *Germany Raises Red Flags About Palantir's Big Data Dragnet*, 2023;
- MEDVEDEVA M., MCBRIDE P., *Legal Judgment Prediction: If You Are Going to Do It, Do It Right*, in PREOTIUC-PIETRO D. ET AL. (eds.), *Proceedings of the Natural Legal Language Processing Workshop* 2023, Singapore: Association for Computational Linguistics, December 2023;
- MEDVEDEVA M., VOLS M., WIELING M., *Judicial Decisions of the European Court of Human Rights: Looking into the Crystal Ball*, 2018;
- MEDVEDEVA M., VOLS M., WIELING M., *Using Machine Learning to Predict Decisions of the European Court of Human Rights*, in *AI and law*, 2020;
- MEDVEDEVA M., WIELING M., VOLS M., *Rethinking the field of automatic prediction of court decisions*, in *Artificial Intelligence and Law*, 2022;
- MEDVEDEVA M., XU X., WIELING M., VOLS M., *JURI SAYS: An Automatic Judgement Prediction System for the European Court of Human Rights,* in VILLATA S. ET AL. (eds.), *Legal Knowledge and Information Systems*, 2020;
- MENECEUR Y., BARBARO C., *Artificial intelligence and the judicial memory: the great misunderstanding*, in *AI and Ethics*, 2022;
- MEUWESE A., *Regulating algorithmic decision-making one case at the time. Case note on: District Court of The Hague*, in *European Review of Digital Administration & Law*, 2020;
- MINNELLA C., *Avanti con la giustizia da remoto, gli uffici alla ricerca di un equilibrio*, in *Guida al diritto*, 2023;
- MINNELLA C., *Processo penale da remoto: l'intervento della Corte Costituzionale*, 2021;
- MURRO O., *La disciplina della videoconferenza per le dichiarazioni del testimone e dell'esperto*, in *Diritto penale e processo*, 2022;
- NEGRI D., *Atti e udienze "a distanza": risvolti inquisitori di una transizione maldestra alla giustizia digitale*, in CASTRONUOVO D., DONINI M., E. M. MANCUSO, VARRASO G. (eds.), *Riforma Cartabia. La nuova giustizia penale*, Milano, 2023;
- PARASURAMAN R., SHERIDAN T. B., WICKENS C. D., *A model for types and levels of human interaction with automation*, in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 2000;
- PASQUALE F., *The black box society*, 2016;
- PELLEGRINO G., BARBA M. C., D'ERRICO G., KÜÇÜKKARA M. Y., DE PAOLIS L. T., *eXtended Reality & Artificial Intelligence-Based Surgical Training: A Review of Reviews,* in *International Conference on Extended Reality*, 2023;
- PETKEVIČIŪTĖ-BARYSIENĖ D., *Human-Automation interaction in law*, 2021;
- PEYRON L., *Viene il giudizio nel Metaverso. Ma verrà alla fine dei giorni*, 2023;
- PENASA S., *Giustizia e variabile algoritmica. Una prima valutazione di sostenibilità tecnica e costituzionale*, in ANTONIOLLI L., CARDILLO M., CORTESE F., DE CARBONNIÈRES L., MYNARD F., PICCIOCCHI C. (eds.), *Numérique & environnement. Université d'été franco-italienne. Actes du colloque*. 6-8 Julliet 2022. Université de Limoges, Trento, 2024;
- PICOZZA E., *Artificial intelligence and law. Politica, diritto amministrativo, and artificial intelligence,* in *Giurisprudenza italiana,* 2019;
- PINOTTI G., SANTOSUOSSO A., FAZIO F., *A Rule 74 for Italian Judges and Lawyers*, in *GUIZZARDI R., NEUMAYR* B. (eds.) *Advances in Conceptual Modeling. ER 2022. Lecture Notes in Computer Science*, Cham, 2023;

- POLLICINO O., *Judicial protection of Fundamental Rights on the Internet*, Oxford, 2021;
- PUMPHREY D. JR., *How the Metaverse Could Change Criminal Defense Forever*, 2022;
- RACHOVITSA A., JOHANN N., *The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case,* in *Human Rights Law Review*, 2022;
- REIMERS N., GUREVYCH I., *Sentence-bert: Sentence embeddings using siamese bert-networks*, 2019;
- ROOSE K., *How Schools Can Survive (and Maybe Even Thrive) With A.I. This Fall*, 2023;
- ROVELLI C., *Buchi bianchi. Dentro l'orizzonte*, Milan, 2023;
- RULLO L., *Corti online*, in *Rivista di Digital Politics*, 2021;
- SANTOSUOSSO A., D'ANCONA S., FURIOSI E., *New-generation templates facilitating the shift from documents to data in the Italian judiciary*, in SALES T. P., ARAÚJO J., BORBINHA J., GUIZZARDI G. (eds.*) Advances in Conceptual Modeling. ER 2023. Lecture Notes in Computer Science*, 2023;
- SANTOSUOSSO A., *Diritto e intelligenza artificiale*, Milan, 2020;
- SANTOSUOSSO A., PINOTTI G., *Una Giustizia "digital by design": ecco come realizzarla*, in *Agenda Digitale,* 2021;
- SARTOR G. ET AL., *Automated Extraction and Representation of Citation Network: A CJEU Case-Study*, in GUIZZARDI R., NEUMAYR B.(eds.), *Advances in Conceptual Modeling*, Lecture Notes in Computer Science, Cham, 2022;
- SCHWARTZ B., *The paradox of choice: Why more is less*, New York, 2004;
- SMUHA N. A., *Beyond the Individual: Governing AI's Societal Harm*, in *Internet Policy Review*, 2021;
- SUKKSI M. (ed.), *The Rule of Law and authomated decision-making*, Cham, 2023;
- SUNSTEIN C. R., *Algorithms, correcting bias*, in *Social Research: An International Quarterly*, 2019;
- SUSSKIND R., *Tomorrow's Lawyers*, II, Oxford, 2017;
- TAVARES DA SILVA R., *Neuroscience of Memory and Philosophy of Knowledge Challenges to Immediacy*, in MORÃO H., TAVARES DA SILVA R. (eds.), *Fairness in Criminal Appeal. A Critical and Interdisciplinary Analysis of the ECtHR Case-Law*, Cham, 2023;
- TEGMARK M., *Life 3.0: Being Human in the Age of Artificial Intelligence online*, New York, 2017;
- *The world's most valuable resource is no longer oil but data. The data economy demands a new approach to antitrust rules,* 2017;
- TIGARD D. W., *Responsible AI and moral responsibility: a common appreciation*, in *AI Ethics*, 2021;
- TRAMARECE E., *Nullità per mancato funzionamento della videoconferenza e prospettive di riforma*, in *Giurisprudenza italiana*, 2021;
- TRAPELLA F., *La rivoluzione digitale alla prova della riforma*, in *Archivio Penale*, 2022;
- TURING A., *Computering Machinery and intelligence*, in *Mind*, 1950;
- UTZ C. ET AL., *(Un)informed Consent: Studying GDPR Consent Notices in the Field*, in *ACM SIGSAC Conference on Computer and Communications Security*, 2019;
- VAN BEKKUM M., BORGESIUS F. Z., *Digital welfare fraud detection and the Dutch SyRI judgment* in *European Journal of Social Security*, 2021;
- VAN DIS E. A. M., BOLLEN J., VAN ROOIJ R., ZUIDEMA W., BOCKTING C. L., *ChatGPT: five priorities for research*, in *Nature*, 2023;
- VERGES E., VIAL G., *Justice et intelligence artificielle, l'impact des algorithmes sur la décision de justice*, Recueil Dalloz, 2022;

- WALTON D., MACAGNO F., SARTOR G., *Statutory interpretation: pragmatics and argumentation*, Cambridge, 2021;
- WALTON D., REED C., MACAGNO F., *Argumentation schemes*, Cambridge, 2008;
- WARREN S., BRANDEIS L., *The right to privacy*, in *Harvard Law Review*, 1890;
- WINKELS R., DE RUYTER J., *Survival of the Fittest: Network Analysis of Dutch Supreme Court Cases,* in PALMIRANI M. ET AL. (eds.), *AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents*, Lecture Notes in Computer Science, Berlin, Heidelberg, Springer, 2012;
- WOLF T., DEBUT L. ET AL., *Transformers: State-of-the-Art Natural Language Processing*, in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020;
- ZAGREBELSKY G., *La giustizia come professione*, Turin, 2021;
- ZALNIERIUTE M., BELL F., *Technology and the judicial role*, *forthcoming* in APPLEBY G., LYNCH A. (eds.), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia*, Cambridge, 2020;
- ZERBINATI A. ET AL., *SenTag 2.0: A Cooperative Annotation Tool*, in *Proceeding of the 21st International Semantic Web Conference*, CEUR Workshop Proceeding, 2022.
- ZUBOFF S., *Surveillance Capitalism. The future of humanity in the age of new powers*, Rome, 2019.