Oñati Socio-Legal Series (ISSN: 2079-5971)
Oñati International Institute for the Sociology of Law
Antigua Universidad s/n - Apdo. 28
20560 Oñati - Gipuzkoa (Spain)
opo@iisj.es / http://opo.iisj.net

The ethics of Artificial Intelligence: An analysis of ethical frameworks disciplining AI in justice and other contexts of application

OÑATI SOCIO-LEGAL SERIES VOLUME 12, ISSUE 3 (2022), 614–653: NORM, NORMAL AND DISRUPTION: THE ROLE OF LAW, KNOWLEDGE AND TECHNOLOGIES IN NORMALISING SOCIAL LIFE
DOI LINK: HTTPS://DOI.ORG/10.35295/OSLS.IISL/0000-0000-0000-1273
RECEIVED 9 APRIL 2021, ACCEPTED 10 FEBRUARY 2022, FIRST-ONLINE PUBLISHED 30 MARCH 2022, VERSION OF RECORD PUBLISHED 1 JUNE 2022

GIAMPIERO LUPO*

Abstract

The recent introduction of AI tools in the justice sector poses several ethical implications as risks for judges' independence and for procedural transparency, and discrimination biases. By developing ethical frameworks governing AI application, private and public agents have been increasingly dealing with risks pertaining to the use of AI. By inventorying and analyzing a set of ethical documents through content analysis, this study highlights the ethical implications involved in the application of AI. Moreover, by investigating the CEPEJ Charter (European Commission for the Effectiveness of Justice of the Council of Europe), the unique ethical document focusing on AI in justice, we were able to clarify potential differences between justice and other contexts of AI application with respect to risks prospected and the protection of ethical principles. The analysis confirms that the discipline of AI is a complex subject that

The author wishes to thank Matteo Cenci (research assistant) that contributed to the data gathering activities at the basis of the paper's argumentations. The author wishes to thank also Dr. Francesco Contini, Prof. Patricia Branco and Prof. Richard Mohr that organized the *Technologies of Normalization* seminars (September – October 2020) that are at the basis of this special issue: a preliminary version of this paper has been presented in one of the seminars. The research activities described in this paper are part of the ACT (Autonomy through Cyberjustice Technologies - www.ajcact.org) project, a research partnership led by the University of Montreal's Cyberjustice Laboratory (www.cyberjustice.ca), and funded by a Partnership Grant issued by the Social Sciences and Humanities Research Council of Canada (SSHRC). For this, the author wishes to thank Prof. Karim Benyekhlef, Director of the Cyberjustice Laboratory, and all the other ACT researchers for support.

* Giampiero Lupo received his PhD in Political Science-Comparative and European Politics in 2010 at the University of Siena. He has been a researcher at the University of Bologna working on deliberative democracy, quality of democracy, justice systems, e-justice. He currently works as a researcher at the IGSG-CNR (Research Institute on Legal Informatics and Justice Systems - National Research Council of Italy) participating to a set of international projects as *Building Interoperability for European Civil Proceedings Online*, *e-Codex*, *Towards Cyberjustice* and publishing in peer-reviewed books and articles the results of his research. His main scientific interests are Artificial Intelligence and justice, e-justice, quality of democracy, quality of justice systems, deliberative democracy. Email address: giampiero.lupo@bo.igsg.cnr.it

involves very different aspects and therefore needs a broad focus on all contexts of application.

Key words

e-justice; AI ethics; AI in justice systems; law and technology; legal informatics

Resumen

La reciente introducción de herramientas de IA en el sector de la justicia plantea varias implicaciones éticas, como riesgos para la independencia de los jueces y para la transparencia procesal, así como sesgos de discriminación. Mediante el desarrollo de marcos éticos que rigen la aplicación de la IA, los agentes privados y públicos se han enfrentado cada vez más a los riesgos relacionados con el uso de la IA. Al hacer inventario y análisis de un conjunto de documentos éticos mediante un análisis de contenido, este estudio pone de manifiesto las implicaciones éticas que conlleva la aplicación de la IA. Además, al investigar la Carta de la CEPEJ (Comisión Europea para la Eficacia de la Justicia del Consejo de Europa), el único documento ético centrado en la IA en la justicia, pudimos arrojar luz sobre las posibles diferencias entre la justicia y otros contextos de aplicación de la IA con respecto a los riesgos previstos y la protección de los principios éticos. El análisis confirma que la disciplina de la IA es un tema complejo que implica aspectos muy diferentes y, por lo tanto, necesita un enfoque amplio en todos los contextos de aplicación.

Palabras clave

e-justicia; ética de la IA; IA en los sistemas judiciales; derecho y tecnología; informática jurídica

Table of contents

1. Introduction

In recent years, there has been a growing diffusion of Artificial Intelligence (AI) in several social and working contexts. AI involves various technologies characterized by a machine mimicking "cognitive" functions associated with human mind, including "learning," "problem solving," and "natural language processing." (Russell and Norvig 2016). AI is applied in several fields, such as autonomous vehicles (drones and self-driving cars) or medical diagnosis. In the justice sector, there is a growing utilization of AI algorithms for applications supporting justice professionals' work. The application of AI for supporting justice professionals is not a true innovation considering that the first experiments with AI and law began in the 1980s (Rissland *et al.* 2005). At that time, expert systems or Case-Based Reasoning (CBR) was experimented as the first typologies of AI tools for legal professionals providing assistance in the process of legal problem solving. These systems provided intelligent research tools for case law and norms related to the case at hand (Susskind 1987). The majority of these pilots have been developed for supporting lawyers. For instance, HYPO is a system for modeling cases' argumentation in the field of US Trade Secrets Law (Skeem and Eno Louden 2007, Simshaw 2018, Lupo 2019).

More recently, AI tools for justice are increasingly diffusing – both in the free market of ICT for lawyers and in several justice systems – and are beginning to have a more relevant role between justice professionals' everyday activities. For instance, Ross, an expert system developed by Ross Intelligence that incorporates the IBM's Watson technology, utilizes AI to automate activities, such as legal searches that usually involve lawyers and law firms (https://rossintelligence.com/).[1] In addition, justice administration is beginning to introduce AI. As an example, the California and Wisconsin Department of Corrections and Rehabilitation (CDCR) introduced the COMPAS system, a research-based risk and needs assessment tool for criminal justice practitioners. CDCR judges utilize the system for the placement, supervision, and case management of offenders in community and secure settings by elaborating data gathered through a questionnaire used to determine overall risk potentials and criminogenic needs profile (Brennan *et al.* 2009, Liu *et al*. 2019).

The introduction of complex technologies capable of imitating and replacing human abilities can influence the "normality" (Mol 1998) of contests in which they are applied. In the case of AI, this may have practical implications on the use of data, the protection of privacy, the responsibility and accountability of systems, their reliability, as well as compliance with fundamental human rights' principles and the rule of law. With regard to AI in justice, concerns may appear similar to the ones posed by non AI e-Justice technologies (Contini and Fabri 2003, Velicogna 2007, 2018, Contini and Lanzara 2008, Reiling 2016): for instance, AI and non-AI technologies share the issues related to the required changes in work routines and procedures and the acceptance of such changes by justice professionals (Huijboom and Van den Broek 2011, Lupo 2019, Contini 2020).

[1] The system incorporates IBM's Watson technology and allows users to ask natural language questions as well as search for and provide legal information ranging from specific citations to full legal briefs (Nunez 2017).

However, the implications of the changes introduced by the use of AI in the justice sector have the potential of being much deeper and less controllable. The use of personal data by AI refers to different sizes and types of data compared to non-AI systems. Open and Big Data are the fuel of AI technologies. The term "Open data" refers to data organized in a database that are freely downloadable and re-employable without having to pay an operating license (Huijboom and Van den Broek 2011). The term "Big data" refers to a big set of data that can be subject to a computer process (open data or data employable with a not-for-free operating license, electronic messages, connection traces, and GPS signals) (Davenport *et al.* 2012). Three elements (the three V rules) distinguish Big Data from regular datasets: a large "volume" of data, large "variety," and high "velocity" (Hoffman and Podgurski 2013). The mentioned characteristics and the major availability of open and Big Data imply a greater exposure of AI to the risks of malicious use of data, of their counterfeiting, and of third-party unwanted access to confidential and personal information. Consequently, the support and monitoring of AI compliance with the rules relating to the use of data – in regard to the respect of privacy, data protection of sensitive data, and the use and abuse of data by third parties – is more difficult.

In addition, the use of AI in justice poses significant challenges to justice systems' fundamental values, particularly when systems are utilized for supporting judicial decision-making. The use of AI for supporting the decisions of judges (see the mentioned COMPAS' case) may result in an undesired and undue influence on judges that can undermine their independence and impartiality (CEPEJ 2018, Contini 2020, Santosuosso and Poletti 2020). AI systems utilized by judges, public prosecutors, police, and lawyers pursue very sensitive tasks that may have important consequences on the future outcome of a procedure (and for citizens accessing to the justice system). Therefore, the issues of systems' liability and safety and the question of determining responsibility for AI failure are fundamental (Lin *et al.* 2011). In addition, due to the complexity of AI systems and the mystery surrounding how algorithms work – often protected by trade secret laws – (Liu *et al.* 2019, Hammoud 2020) the determination of responsibility in case of failures (is it the developer's or judge's fault?) can be complex. Consequently, this may invalidate citizens' right to a redress when affected by a decision biased by a failed AI system (Lin *et al.* 2011, Contini 2020).[2]

These examples acknowledge that the development of AI technologies and their diffusion into the most disparate contexts of human life and work may result in an abrupt change for Canguilheim's "expected and loved order" (Angelides 2012). These changes may create concerns, uncertainties, or rejections. More specifically, the introduction of these technologies within justice systems may be perceived as an attack on the "normality" of the judicial administration context based on the compliance with procedural laws and fundamental values that refer to the generic concept of the rule of law (Donaldson *et al.* 2000, Rigano 2019, Završnik 2020, Santosuosso and Poletti 2020).

To mitigate these changes, several actors, from public to private, are resorting to normativity to curb and govern AI technological innovation and its associated risks. A

[2] The interference of AI systems for justice professionals on fundamental values may also affect lawyers. All lawyers (in any justice system) must comply with the rules of professional ethics. By developing AI for lawyers, third parties may run the risk of affecting lawyers' compliance to deontology, for instance, with systems that fail in protecting clients' sensitive personal data (Lupo 2019).

very proliferation of normative frameworks, guidelines, and collection of ethical principles disciplining the application of AI in several contests have been registered (Van Dijk and Casiraghi 2020). These frameworks set principles or guidelines that should limit AI harm to fundamental rights and values. The website "algorithmwatch" (https://algorithmwatch.org/) identified and listed 83 ethical documents drafted by different types of actors and in different languages.

The proliferation of ethical framework documents in the legal and policy discourse, as well as the growing importance of ethical expertise, ethical committees, and ethical advisory groups and boards, has been called as the "ethification" phenomenon (Contini 2020, Casiraghi and Van Dijk and Casiraghi 2020). This phenomenon seems a consequence of the necessity of protecting normality from the possible harms that AI technologies may cause. Moreover, the drafting of ethical documents instead of laws, meets the need of keeping pace with AI's rapid evolution: ethical documents' drafting is a more flexible practice to cope with unpredictable effects of emerging technologies, differently from the law that is more rigid, time consuming and may lag behind technological development (Van Dijk and Casiraghi 2020). The counterpart to flexibility is that ethical documents are policies or soft-law tools (Tallacchini 2009, 2015, Floridi 2018) and they are unbinding. However, in a regulatory context that has not yet entirely addressed the issue of the AI use's implications, AI ethical documents are significantly important because they may anticipate the "proto-constitutional discourse" (Gill *et al.* 2015) that leads to the crystallization of comprehensive and binding laws.

This relationship between ethics (in general and AI ethics) and law is controversial. Ethics can represent a means for transcending existing legal frameworks, for providing an opportunity to ignore them avoiding the law – the so-called ethics washing (Wagner 2018, Lohr *et al.* 2019, Daly *et al.* 2021) – and for ensuring that AI will not be regulated by law (Rességuier and Rodrigues 2020). In this sense, ethics is a regulatory tool favorable to those who have no interest in having their behavior regulated given that "ethics has no teeth" (Rességuier and Rodrigues 2020) that is it lacks of enforcement methods. However, ethics may represent a form of attention to reality as it evolves, thus providing a substantial contribution to law-making (Daly *et al.* 2019, Rességuier and Rodrigues 2020). Indeed, ethical documents are usually drafted by actors who have practical experience of the application context to be regulated. This may be the case also for AI ethical documents whose drafters may have a greater possibility of grasping the multifaceted implications and risks of AI use.

Given the importance of ethics in the regulation of AI, this paper will focus on the analysis of AI ethical documents that have been drafted in the recent years with the objective of clarifying which ethical principles and risk factors the documents mainly converge. Moreover, given that the major focus of the research is on the implications of AI in justice, a deeper qualitative analysis focused on the CEPEJ[3] *European Ethical Charter on the use of AI in judicial systems and their environment* (CEPEJ 2018). By investigating the unique ethical framework document that focuses on AI in justice systems, we could clarify potential differences between justice and other contexts of application with respect to risks prospected and protection of ethical principles. The analysis also confirms on the one hand, that the discipline of AI is a complex subject that involves

[3] European Commission for the Effectiveness of Justice of the Council of Europe (CEPEJ).

very different aspects and therefore needs a broad focus on all contexts of application; on the other hand, that due to the rapid diffusion of AI, it will soon be necessary to activate law-making processes to draft more binding norms.

In order to pursue the mentioned aims, a content analysis of a sample of AI ethical framework documents and a quantitative analysis of data gathered were performed. In particular, 108 documents (for documents' gathering and selection, please see Section 2) have been manually coded on the basis of their reference to ethical principles or issues related to the application of AI. The research activities (whose preliminary results are presented here) are part of an international project on AI and justice coordinated by the University of Montreal in which IGSG-CNR[4] institute is a partner: ACT – Accessibility Through Cyberjustice https://www.ajcact.org/).

In this paper, we will first introduce the methodology of the study (Section 2), consisting of the use of content analysis and quantitative analysis techniques. Second, the sample of ethical documents gathered will be described (Section 3). Section 4 presents the results of the content analysis that clarifies convergences toward principles between documents and identifies notable patterns. Section 5 is dedicated to the analysis of the unique ethical document drafted for disciplining AI in justice systems that is the CEPEJ Charter. The final section presents the conclusions of the analysis.

2. Methodology

The methodology that guided the analysis of the AI ethical documents is based on the use of content analyses' techniques. Content analysis is a research technique frequently utilized in the social sciences to make replicable and valid inferences through the interpretation and coding of textual material (Erlingsson and Brysiewicz 2017). This technique consists of the evaluation of texts, or other symbolic constructs (as documents, oral communication, and graphics), through procedures of analytical decomposition and classification to obtain quantitative statistical data (Rositi 1988). Although the method has been used frequently in the social sciences, only recently has it become more prevalent among organizational scholars (Mayring 2004). In this regard, it represents an important bridge between purely quantitative and purely qualitative research methods. To transform a large amount of text into a highly organized and concise summary of key results, content analysis provide for the assignment of codes to extracts of a test, as periods or paragraphs, on the basis of given rules and at various levels of abstraction (Tipaldo 2007).

The methodology of this analysis consisted of two phases: a) an inventory of ethical framework documents disciplining AI in several contexts of application (including justice) and b) the content analysis of selected documents consisting of the data gathering through coding and the relative quantitative analysis of the data gathered.

We performed a document search and inventory through a purposive sampling method to ensure a heterogeneous sample with respect to stakeholders, content, geography, and date (Etikan *et al.* 2016). This involved two activities: first, the gathering and selection of ethical documents listed in the website https://algorithmwatch.org/ and their selection and second, the gathering and selection of further documents through web research and

[4] Legal Informatics and Justice Systems—National Research Council of Italy.

by consultation of other ethical documents' reference lists (for a complete list of documents, see Table A.1 in the appendix). The documents have been selected with the aim of balancing four fundamental objectives: a) the inclusion in the study of the largest sample as far as our knowledge of language allows, b) the inclusion of influential documents (possibly related to the application of AI in justice), c) the maximum variation among documents, and d) the selection of only normative documents (declarations about how AI should be designed and deployed).

All of the selected documents focus on rules, principles, and values related to the use of AI. However, to allow for the construction of a large sample, and to include in the study the maximum possible information on the implications of AI application, we also included documents that refer to AI's closely equivalent terms as intelligent systems or robots (see for instance IEEE's *Ethically Aligned Design 2019* or UNESCO's *Report of COMEST on Robotics Ethics 2017*; see Table A.1 in the appendix).

In addition, we identified and selected the pages in the document in which guidelines were located (for instance various documents also annexed dedicated studies to the guidelines): content analysis only focused on the parts of the documents listing principles and guidelines. During this initial stage of the research project, we selected up to 108 AI ethics documents. Before starting the coding activities typical of content analysis, documents were classified on the basis of variables useful for the analysis as type of drafting body (public/private/no profit), date of issue, and country of drafting (see Section 3 describing the sample).

The content analysis that followed the inventory involved the hand-coding of guidelines' sentences (each sentence delimited by a full-stop) on the basis of their reference to ethical principles or issues related to the application of AI. Two researchers handled the coding of documents: a junior research assistant that coded documents in a first stage and a senior researcher that controlled and amended coding. Usually, nouns clearly referring to principles (for instance, "privacy") are directly found in the sentence and easily detected. However, occasionally, sentences required the paraphrases of their content to detect a principle. The content analysis of the 108 documents facilitated the retrieval of 70 principles differently distributed in the sample (for principles' definitions and coding, see Table A.2 in the appendix). The software utilized (Atlas ti 8) allowed us to extract quantitative data related to the distribution of principles in the 108 documents useful to perform a statistical analysis and to highlight interesting phenomena regarding the characteristics of documents and the principles mentioned.

The analysis design retraces the trend of literature that analyzed the role of ethics in regulating AI (Jobin *et al.* 2019, Hagendorff 2020, Schiff *et al.* 2020), with the objective of providing a deeper analysis of ethical documents. In addition, the main focus on the use of AI in justice with the investigation of the CEPEJ *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment* enriches the analysis and positions it in the context of judicial studies. The CEPEJ document has been selected for its uniqueness: at the time of writing this paper, CEPEJ's is the only ethical document focusing on AI in justice. Clearly this focus has the limitation of excluding non-Eurocentric approaches to AI in justice; this limitation might be overcome in further research on future regulative frameworks on AI in justice drafted outside the EU.

Other limitations regard the content analysis of ethical documents. In one sense, the content analysis technique is inherently reductive, particularly when dealing with complex text, and a potential high level of interpretation characterizes it. The potential high level of interpretation has been limited by involving two researchers in the hand coding of documents, with a senior researcher controlling the results of coding activities. Moreover, the reductive quality of the content analysis is compensated by its capacity of allowing a comparison of a large number of documents in line with the objectives of the research. Another set of limitations regard the sample of documents investigated. Despite our attempt to gather the largest and most representative sample – as far as the researchers' knowledge of language allows – it is plausible that an unknown number of documents have been excluded. Moreover, it is foreseeable that in the future, other ethical documents will be drafted. The analysis of the excluded documents – with the involvement of researchers that can read in the languages excluded in the research and the inclusion of documents drafted following this paper's publishing – may be the objective of future research.

3. The sample

This section describes the sample of ethical documents selected for the content analysis. The documents are described on the basis of selected variables useful for the purposes of the analysis as the country of the organization drafting the document or the document's target audience.

Starting from the organizations drafting the ethical documents selected, the sample equally includes (as drafters) private, public, and no profit bodies (respectively 33.3%, 36%, 11%, and 33.3% of the sample). This data acknowledges that the proliferation of ethical guidelines does not regard only public bodies, but also involves private organizations (as well as no profit organizations). In particular, the bodies drafting the selected documentation primarily include associations, public bodies, private companies, and research organizations (see Table 1).

TABLE 1

creator type	Freq.	Percent	Cum.
Lobby	3	2.78	2.78
association	15	13.89	16.67
business	23	21.30	37.96
conference	4	3.70	41.67
experts	6	5.56	47.22
government	33	30.56	77.78
journal	1	0.93	78.70
research	23	21.30	100.00
Total	108	100.00	

Table 1. Types of bodies drafting the ethical documents selected.

With reference to the countries issuing the documents, the inventorying and selection of documents for the analysis acknowledges the overproduction of ethical documents in Europe and North America (see Fig. 1), thus confirming results of precedent studies (Jobin *et al.* 2019). However, the sample cannot be considered completely representative,

given that the inventory and selection are influenced by our knowledge of language. This limit did not prevent us from analyzing documents – for instance those drafted in Asia (in English language) – however, it only allowed for the focus on those documents drafted in English, French, Spanish, and German.

FIGURE 1

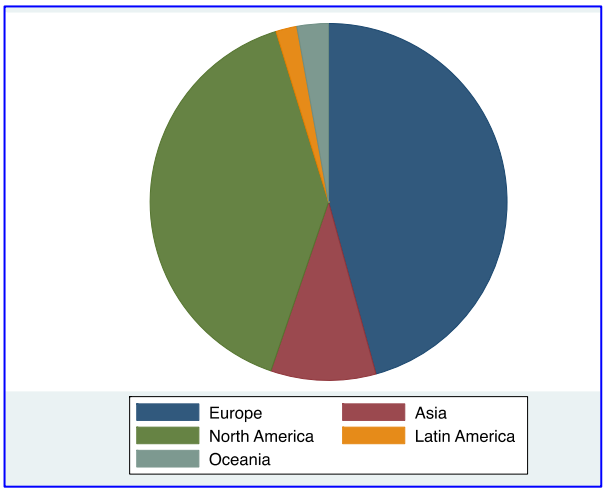

Figure 1. Country of bodies drafting the ethical documents selected.

In addition, the ethical documents selected have been characterized on the basis of the methodology utilized to draft the document. Documents' drafting methodologies include: 1) collegiate decision, 2) conference (documents drafted following a public conference), 3) deliberative processes (inclusive methods of citizens' participation), 4) through the involvement and constitution of expert groups, and 5) through an internal draft (therefore, involving only personnel part of the organization drafting the document). The analysis of data acknowledges that the most utilized methodology of drafting involves the constitution of an expert group (44.4%), followed by the "internal draft" (32.41%) and collegiate decisions (12.06%). The data on modality of drafting confirm that the subject covered by the ethical documents, namely, the discipline of the use of AI, is extremely technical and requires the involvement of experts in the field to understand the various implications at stake. By distinguishing between the types of organizations drafting the document, it is clear that while public bodies and no profit organizations preferred drafting documents through the involvement of expert groups (58.3% and 53.8%), private companies favored the internal draft, thus involving only companies' personnel (see Table 2). This acknowledges the intrinsic limitations of ethical documents drafted by private companies that may self-regulate AI for avoiding existing legal frameworks through ethics (the mentioned ethics washing; (Wagner 2018, Lohr *et al.* 2019, Daly *et al.* 2021) or anticipating potential future limitations of the use of this technology due to the approval of compulsory legislation.

TABLE 2

	Collegiate decision	Conference	Deliberative process	Expert group	Internal draft	Total
No profit	28.57	42.86	0.00	43.75	21.05	33.33
Private	7.14	0.00	0.00	12.50	68.42	30.56
Public	64.29	57.14	100.00	43.75	10.53	36.11

Table 2. Cross tabulation with data on type of organization and modality of drafting.

The ethical documents have been classified according to the involvement (or not) of multidisciplinary teams to draft the document. The data showed that the drafting of the majority of the documents in the sample (54.6%) involved multidisciplinary teams. This result integrates the data on the diffused use of expert groups for documents' drafting, thus acknowledging that the topic "AI implications and disciplining" is technical, but it also involves different scientific areas, from ICT to social sciences (Floridi 2018). Moreover, this data is consistent with the results of the content analysis (see Section 4) that confirms that one of the most widespread principles found in the documents analyzed supports the use of multidisciplinary teams for designing and implementing AI applications. As we will see later when discussing the results of the content analysis, interdisciplinary AI development teams are necessary given the variety of skills required to develop AI as computer science and mathematics, as well as neurosciences and psychology. In addition, teams must include experts in the specific AI's field of application (such as health or finance) and experts with a background in social sciences or policy to evaluate the broader societal impact of AI used to assist in making critical decisions (Marchant 2011). Furthermore, the CEPEJ Charter on the use of AI in justice quotes multidisciplinarity as an important precondition for responsible AI implementation. In addition, it indicates the multidisciplinarity of teams developing AI as a means to estimate and mitigate the potential risks of AI application in the social contexts where the risks of perpetuating discrimination are high such as the justice context.

The analysis also clarified the target audience of the documents selected (see Fig. 2). The target audience is quite variable and principally includes the general public (20%), private companies (19.5%), policy makers (29.8%), and developers (14.9%).[5] Justice professionals as a target audience are clearly underrepresented (only 2 in 108 documents, both drafted by EU institutions): this reflects the fact that currently, aside from several preliminary experiments, AI is not consistently diffused within justice systems (Santosuosso and Poletti 2020, Spajosević *et al*. 2020). This result is confirmed by the analysis of the variable describing the type of AI application the ethical document refers to. In only one case, the CEPEJ Ethical Charter, the ethical guideline refers to AI for justice. Indeed, the majority of the ethical documents refer to AI technology applied in generic context of application (58.3%), followed by business applications (15.7%) and data science applications (6.5%; see Fig. 3).

[5] The variable target audience is not mutually exclusive considering that many documents are addressed to more than one audience.

FIGURE 2

Figure 2. Target audience of the ethical documents investigated (histogram).

TABLE 3

Field of AI application	Freq.	Percent	Cum.
Business applications	17	15.74	15.74
Computer ICT	1	0.93	16.67
Consumers	2	1.85	18.52
Data Science	8	7.41	25.93
Finance	1	0.93	26.85
Generic	63	58.33	85.19
Health	2	1.85	87.04
Justice	1	0.93	87.96
Online services	1	0.93	88.89
Public administration	2	1.85	90.74
Research	4	3.70	94.44
Robotics	4	3.70	98.15
Talk Bots	1	0.93	99.07
Telco	1	0.93	100.00
Total	108	100.00	

Figure 3. Field of AI application of the ethical documents investigated.

As anticipated, the framework documents investigated represent forms of soft law. Therefore, they are not compulsory by definition, although they may anticipate a legislative discourse that can trigger the usual legislative decision-making. In our sample, ethical frameworks are in the majority of cases non-compulsory (102 over 108). The few exceptions to the "non-compulsory" rule are those ethical documents that are mandatory for institutions and individuals that are members or part of the organization that drafts the document. For instance, this is the case for the "Advisory statement on human ethics in artificial intelligence and big data research" (2017) of the National

Research Council of Canada (NRC)[6] that commits research institutions monitored by the NRC to support research on AI aligned to ethical principles. Another example regards documents drafted by private organizations that indicate specific and compulsory rules to employees involved in AI-based projects as the "Verivox" "Selbstverpflichtung zur Stärkung des Verbraucherschutzes auf digitalen Vergleichs- und Verbraucherplattformen" (2019).[7]

Connected to compulsoriness is the topic of assessment of compliance to the ethical principles included in the documents. Despite being non-compulsory in the majority of cases, the 44.5% of documents provide for an assessment mechanism. Where present, assessment mechanisms mainly involve regular compliance monitoring (93%).[8]

The previous argumentation allowed us to frame the documents inventoried on the basis of fundamental characteristics. The analysis confirms that the selection's objective of ensuring a maximum variation among documents has been reached.

4. The content analysis of AI ethical documents

This section describes the results of the content analysis in a quantitative fashion. It focuses on the distribution in the documents of "codes" related to the principles on the application of AI.

On the one hand, the analysis of data gathered through content analysis confirms that ethical documents converge on a set of principles and issues related to AI application. On the other hand, the analysis also acknowledges that principles that are less common among the documents are worth considering, because they may express implications on the use of AI that concern specific circumscribed contexts, but which can also have considerable effects in other fields of application.

[6] See Table A.1 in the appendix for the list of ethical documents including authors and date of issuing.

[7] Verivox is a German energy company: the document aims to protect consumers from unethical use of algorithms and AI utilized for supporting company's activities.

[8] Other methods of assessment registered: ethic commission, seal of approval, self-assessment, and ethical certification (all around 2%).

FIGURE 4

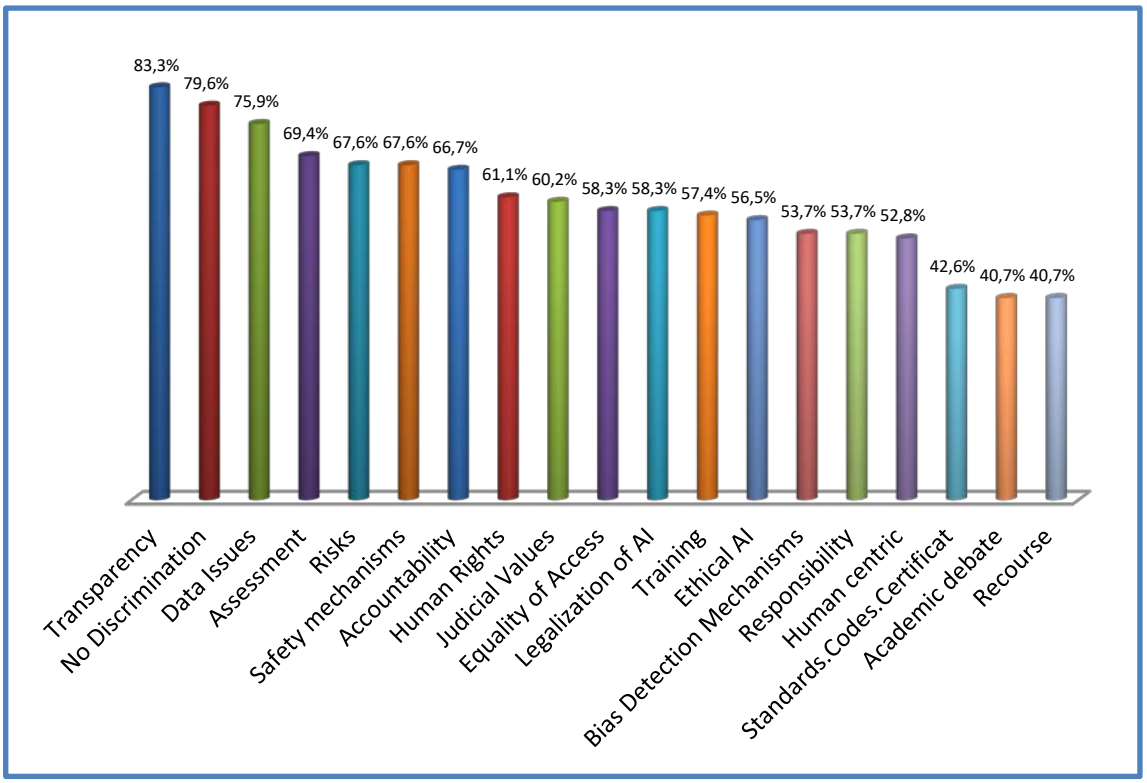

Figure 4. Histogram: percentage of documents in which a principle is present; first 20 scores (the complete distribution list is shown in Table A.2 in the appendix).

Fig. 4 highlights the histogram indicating the percentage of documents analyzed in which a code indicating a principle has been retrieved (see Table A.2 in the appendix for the definitions of each code).

As the histogram indicates, codes such as "transparency," "no discrimination," "data issues," "assessment," "risk of harm," "safety mechanisms," "accountability," "human rights," and "judicial values" are mentioned in 60% (or more) of the documents investigated. This result confirms the importance of the principles and implications of AI use already acknowledged in previous studies (Jobin *et al.* 2019, Hagendorff 2020, Schiff *et al.* 2020).

The analysis confirmed transparency (83.3%) as one of the most discussed principles within the AI ethics debate (Floridi 2018, Jobin *et al.* 2019). The code refers to the transparency of technology's functioning and procedures as well as the transparency of the AI organizations developing and using AI technology. With reference to AI technology, the accessibility of information on technological functioning is difficult given the complexity of the subject and the skills necessary to uncover the AI "black box" (Castelvecchi 2016). This principle is also important for AI digitalizing justice procedures, given that transparency of judicial procedures, rights, and norms is a fundamental rule of law value (Wallace 2003, Lupo 2019).

The code "No discrimination" (found in almost 80% of documents) refers to the risks that AI systems' outcomes are discriminatory, thus perpetrating discrimination against people, or groups of people, based on gender, race, culture, religion, age, or ethnicity. The code includes different types of discriminations mentioned in the investigated

documents as discrimination based on gender or against vulnerable groups. The non-discrimination principle is a fundamental value in liberal democratic countries that is often highlighted by international organizations such as the Council of Europe and UN (Cappelletti 1979, Sandefur 2009, Sherman 2013, ENCJ 2013). With the implementation of AI technology based on the use of large amounts of data, such as machine learning – utilized for supporting critical decision-making affecting citizens – the risks of perpetuating discriminations are severe. In particular, AI decision-making may reflect discrimination bias affecting the dataset utilized that may subsequently incorporate intentional or unintentional systemic human biases. These risks may considerably affect AI utilized in the justice systems, as the COMPAS case acknowledged (described in the introduction section). The ProPublica study (Washington 2018) found that COMPAS AI predictive analytics for judicial risk assessment has been trained with biased data that reproduce past discriminatory judges' decisions.

The code "data issues" (score 75.9%) indicates sentences referring to principles and risks related to the use of data, that is considered as the fuel of most of the AI systems (Lupo 2019). This is a "macro" code referring to issues as citizens' awareness of personal data storing, or the anonymization of personal data included in big datasets. In section 4.1, I will go into the details of the "data issues" macro-code unpacking it in the different normative principles mentioned by documents when regulating the use of data.

The code "assessment" refers to the mechanisms of impact assessment that are considered necessary to ensure that AI functioning is safe, responsible, and compliant with ethical principles. In addition, assessment is considerably diffused between the selected documents (69.4%). This indicates that the impact assessment tool is essential to reduce risks associated with improper use of AI or malfunctions, and this is also valid in extremely sensitive contexts as the judiciary.

The code "risks" refers to the sentences included in the documents describing possible harms that AI systems may cause. The high ranking of the "Risk of harm" code acknowledges the attitude of organizations and citizens that consider the application of AI a concern. This confirms results of a previous study (Jobin *et al.* 2019). This result is consistent with the scarce dissemination of the "Beneficial AI" code (37% of documents; see Table A.2 in the appendix) that refers to sentences indicating the potential positive outcomes resulting from the introduction of AI as an improvement of human productivity or well-being. The concern for the risks related to AI application is coherently associated with the diffused support for the use of safety mechanisms (the code "safety mechanisms" is present in 67.6% of the investigated documents). Safety mechanisms quoted in documents include cyber security tools, encryption systems protecting from data misuse, or repetitive production tests before technology goes live. The principle "accountability" is also fairly diffused between documents (66.7%) and refers to mechanisms of auditability and external control on system functioning and on the digitalized procedure. Accountability is also fundamental for justice systems, therefore indicating methods by which courts' and judges' activities are checked with respect to the rule of law values and efficiency (Mohr and Contini 2011). These control mechanisms are also desirable with regard to technologies that digitize judicial procedures, especially those based on AI, considering the enormous consequences that a biased system can cause to citizens accessing justice (Lupo 2019, Contini 2020).

The majority of the documents (61.1%) emphasize the adherence of AI on generic human rights as protection of human dignity, freedom of association, and freedom of information. This highlights the widespread concern that the introduction of AI technologies may harm fundamental human rights. The majority of documents when referring to human rights specifically quote fundamental international documents as the Universal Declaration of Human Rights or the European Convention on Human Rights (ECHR) quoted by the Ethics Guidelines for Trustworthy AI (*Ibidem* note 5).

Only a single ethical document refers to AI applied in justice systems, that is, the CEPEJ Ethical Charter (see Section 5). Despite this, the "judicial values" code is generally diffused within the sample (60.2%). This code is associated with sentences in documents referring to principles related to the generic concept of rule of law as due process, equal access to justice, fair conditions, and fair trial. Several ethical documents, even if focusing on other fields of application – such as The Toronto Declaration on AI (*Ibidem* note 5) – briefly refer to possible use of AI technologies in justice and therefore propose guidelines to prevent and limit potential harms. In addition, documents such as the Statement on Artificial Intelligence, Robotics and "Autonomous" Systems (*Ibidem* note 5), when focusing on potential harms originating from the use of AI, support an effective and equal access to the judicial system to obtain redress of an AI-based decision or compensation for harm caused by an AI system.

4.1. The issue of data use in AI

The issues related to the use of data are considerably important when referring to AI, given their dependency on the utilization of a large amount of data. This topic is also worth analyzing with reference to AI utilized in the justice system for two main reasons. First, AI technologies for justice rely on the availability of a large amount of data as the primary "fuel" for their functioning (Lupo 2019). For instance, predictive analytics tools as COMPAS obtain and utilize sensible data on arrested to assess the probability of recidivism (Brennan *et al.* 2009). Second, justice systems are growingly making available data in the form of freely downloadable databases, providing access to two types of case law data: public case law data and private structured data deriving from the courts' Case Management Systems. These forms of data usage in the justice system raise concerns with respect to privacy, data protection of sensitive data, and the use and abuse of data by third parties (Huijboom and Van den Broek 2011, Lupo 2019).

To analyze the topic of AI data use, the macro-code "data issues," indicating the quotes in documents referring to different types of problems and principles related to the use of data, have been unpacked in the different codes composing it (for the definitions of each code related to data issues, see Table A.2 in the appendix).

FIGURE 5

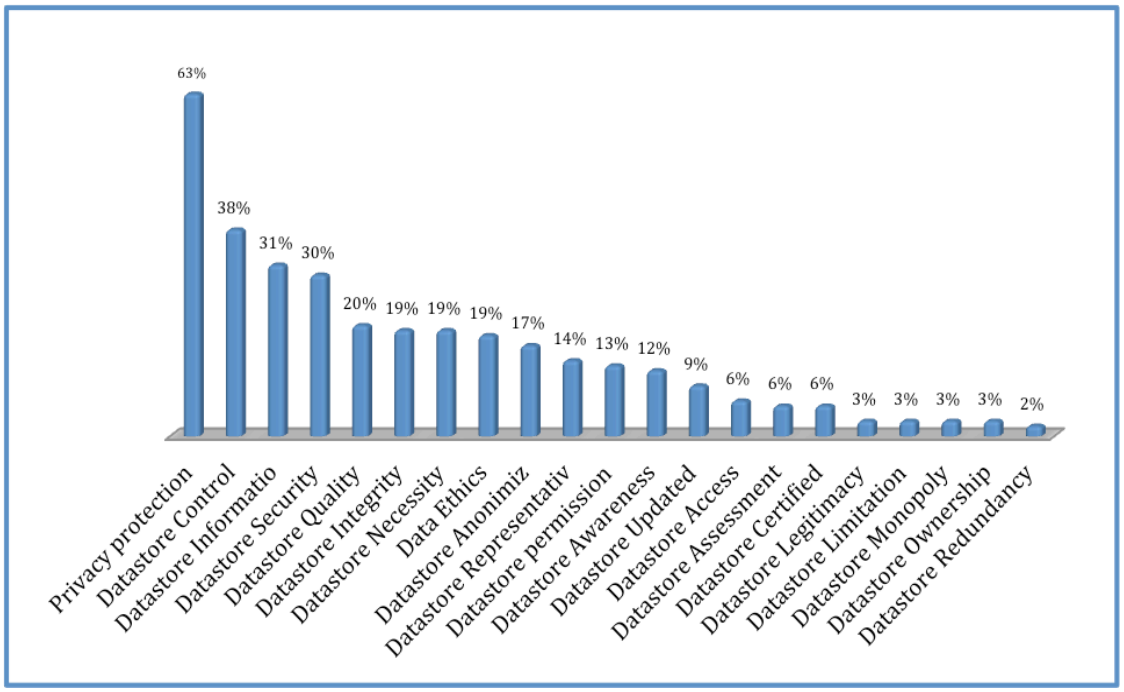

Figure 5. Histogram: percentage of documents in which a "data issues" code is present.

The histogram in Fig. 5 shows the distribution of the various codes belonging to "data issues" in the documents investigated expressed in percentages. The most diffused code is the one referring to the protection of privacy (63%). This can be considered an easily shared generic principle relating to the protection of personal data, which, precisely because of its generic value, is widespread in the majority of the ethical documents. Following privacy protection, we find the codes "data store control," "data store information," and "data store security" all diffused in more than 30% of documents. "Data store control" refers to the citizens' control over their personal data and over the storing of personal data in a dataset. "Data store information" refers to the transparency of the information provided to subjects of data storing regarding different aspects as methods of data anonymization or how the duration in which data are stored within a database. "Data store security," mentioned in 30% of ethical documents, indicates the protection mechanisms that ensure that stored data cannot be utilized in a malicious fashion, or that data are not stolen by unauthorized third parties.

In contrast to AI, the regulation of personal data sharing and usage is well established. Various regulatory instruments have arisen in different contexts such as the EU GDPR (General Data Protection Regulation; *ibidem* note 2), the California Consumer Privacy Act (CCPA 2018), or the Brazilian General Personal Data Protection Law (LGPD).[9] The GDPR, in particular, represents a ground-breaking normative instrument that unifies the EU regulation on data protection, thus simplifying the regulatory environment for EU and international businesses. The primary aim of the regulations is to provide

[9] It is worth mentioning that California Consumer Privacy Act (CCPA 2018) as well as the Brazilian General Personal Data Protection Law (LGPD) have been widely influenced by the EU GDPR (Erickson 2018, Wilkinson 2018, Barrett 2019, Thomas 2020).

individuals with control over their personal data. In addition, it addresses the transfer of personal data outside the EU and EEA (European Economic Area) areas.

The majority of the principles related to data protection quoted in the framework documents (see Fig. 5) can be retrieved in the GDPR regulation as the anonymization of data or the subject's consent on data storing processes. Simultaneously, the GDPR is mentioned in 15% of the ethical documents. Between the documents quoting GDPR, 60% are European (EU-based countries plus United Kingdom), and 40% are non-European, thus demonstrating that GDPR has also impacted countries outside the EU.

The analysis revealed that with reference to data use, ethical documents do not anticipate regular normativity (as may seem the case of principles regulating AI in the framework documents). Instead, they reproduce concepts already discussed in the previous regulatory processes by adapting them to AI, a context of application analogous to the one of personal data protection, as it is characterized by a technology based on the use of large datasets.

4.2. The internal coherence of ethical documents

By comparing distributions of principles through Spearman correlations,[10] it is possible to investigate the internal coherence of ethical documents, with reference to the normative receipts they support, and to the means indicated to reach the desired results. In this case, data utilized do not refer to the number of documents in which a coded principle is present but on the number of codes that are present in each document. Here codes refer to continuous variables that vary between each observation (that is, each ethical document). In Table 3, the Spearman correlations of a number of the analyzed codes are included on the basis of literature indications regarding potential conflicts (or relations worth considering) between AI ethical principles.

The significant correlation between safety and impact assessment (r_s 0.26) acknowledges that ethical documents coherently indicate impact assessment as a measure fundamental to overcome potential harms coming from the use of AI. Data also indicate that safe systems need to provide for accountability measures, as the positive correlation between safety and accountability acknowledges (r_s 0.32).[11] As mentioned, accountability is considerably important with regard to justice systems (and ICT applied in justice; Liu *et al. 2019)*, thus indicating methods for checking and monitoring courts' and judges' activities with respect to the rule of law values and efficiency (Mohr and Contini 2011).

Ensuring the safety of AI use also encompasses the regulation of AI tools with binding norms: the data show a significant correlation between the principles "AI legalization" (indicating documents' support for AI regulation) and "Safety mechanisms." In

[10] The high risk of not respecting the normal distribution assumption for the considered variables makes the Spearman correlation method preferable to Pearson's. The Spearman's rank correlation is a nonparametric test that measures the strength and direction of association between two variables that are measured on an ordinal or continuous scale. The Spearman correlation coefficient is a useful test when Pearson's correlation cannot be run due to violations of normality, a non-linear relationship or when ordinal variables are being used (Croux and Dehon 2010).

[11] The correlation is also significant if we only consider documents drafted by private organizations, thus indicating that private companies support the enablement of mechanisms of external accountability as opposed to preferring more "internal" security measures.

addition, the support of judicial values (a principle present in 60.2% of documents) is positively associated with "AI legalization." These results confirm the importance of initiating a legislative path that facilitates the establishing of compulsory rules for the introduction and use of safe and ethical AI in different contexts. The European Commission's 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) seems a response (at least for the EU member states) to the need to regulate the use of AI in the various application sectors expressed in the ethical documents. The EC proposal addresses the risks generated by specific uses of AI through a set of rules affecting developers and users. The legal framework for AI proposes an approach based on four different levels of risk on the basis of types of AI technology: unacceptable risk, high risk, limited risk, and minimal risk.

In addition, the data indicate an apparently incoherent positive and significant correlation between protection of privacy and open data (r_s 0.30). The result may be clarified by looking at the significant and positive relationship between open data and the anonymization of data (r_s 0.25). This may indicate that open data and privacy protection may be coherently supported if mechanisms of personal data anonymization are in place. This aspect also considerably affects the justice context, considering the amount of relevant data that the justice system can produce, useful data for statistical or system management purposes that cannot be disseminated without effective anonymization and pseudonymization tools.

It is useful here to consider the principle of non-discrimination that was found to be one of the most prevalent among the documents investigated (79.6%). This is an essential principle for AI in justice that the COMPAS case (Brennan *et al.* 2009), one of the most resounding cases of AI in justice misuse, has acknowledged. The analysis confirmed positive and significant correlations between the code "non-discrimination" and principles as "AI legalization" (r_s 0.30), accountability (r_s 0.37), and "multidisciplinarity of developers" (r_s 0.33). The latter data are particularly interesting, thus indicating that the inclusion of different types of skills and educational backgrounds, including social sciences experts, may be a consistent strategy to support non-discriminatory AI when applied to sensitive decision-making as in justice.

TABLE 3

Correlations between Items	R_s
Safety mechanisms – impact assessment	0.26*
Safety mechanisms – accountability	0.32**
Safety mechanisms – accountability (only private)	0.38*
AI legalization – safety mechanisms	0.33**
AI legalization – judicial values	0.32**
Privacy – open data	0.30**
Open data – anonymization of personal data	0.25*
Accountability – trust	0.28*
No discrimination – AI legalization	0.30**
No discrimination – accountability	0.37**
No discrimination – multidisciplinarity of developers	0.33**

Table 3. Correlations between items. Note: ** significant at p≤0.001; * significant at p≤0.005.

5. The analysis of CEPEJ Charter on the use of AI in justice

In December 2018, CEPEJ adopted the European Ethical Charter on the use of AI in judicial systems and their environment. The Charter is the first – and currently the only – example of a framework document that defines ethical principles relating to the use of AI in judicial systems. The document is addressed to policy makers, legislators, and justice professionals that must encounter the development of AI in national judicial systems. In the charter, the CEPEJ supports the idea that the application of AI in the field of justice can be an opportunity to improve the efficiency and quality of justice. However, it also necessitates that AI must be developed responsibly and in agreement with fundamental rights guaranteed in the European Convention on Human Rights (ECHR) and the Council of Europe Convention on the Protection of Personal Data.[12] Being only a collection of ethical principles, and not a normative text, the charter is not compulsory. However, when the charter refers to fundamental rights enshrined for example by the ECHR convention, it mentions norms with legal value and therefore mandatory. In addition, despite its non-compulsory nature, the document provides for a form of assessment based on a self-evaluation scale annexed to the Charter available for any actor planning to develop AI in justice. Moreover, the Charter – in addition to the list of guidelines on AI in justice – includes a report that investigates the opportunities and issues related to the application of AI for processing judicial decisions and data. Our analysis will primarily focus on the guidelines, thus excluding the

[12] Council of Europe. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. Treaty No.108.

annexed report to allow a comparison with the AI framework documents investigated in the previous sections.

The CEPEJ Charter identifies five macro-principles to be complied with when developing AI in justice: 1) respect of fundamental rights; 2) non-discrimination; 3) quality and security; 4) transparency, impartiality, and fairness; and 5) "under user control" principle. The argumentation within the Charter's text describing the five macro-principles is broad and allows for the inclusion of several other principles relating to the application of AI in judicial systems. Table 4 lists the principles mentioned in the CEPEJ Charter with the number of quotes in the document for each principle; aside, it included the distribution of Charter's principles in the sample of documents investigated.

This analysis highlights that the most quoted principles in the Charter are also largely diffused in the sample of investigated documents. For instance, this is the case for transparency, non-discrimination, and human rights. In the Charter, transparency refers to the diffusion to the users and to the public affected by a legal decision supported by AI, of clear and familiar language communication on the AI service offered, on the technology that have been developed, and on the risks of error. However, the Charter also refers to "technical transparency" (for example, open source code and documentation), which is occasionally restricted by the protection of trade secrets. In both cases, the Charter's authors acknowledge the importance of striking a balance "between the intellectual property of certain processing methods and the need for transparency" (CEPEJ 2018, p. 11). Non-discrimination is also largely quoted in the Charter, confirming a trend observed in the other ethical guidelines investigated. The Charter supports the use of corrective measures when an AI system for justice uses sensitive data in the development as well as deployment phases to avoid the reproduction or aggravation of discriminations already existing in the dataset utilized. The types of discrimination quoted and to be avoided following the Charter text are "racial or ethnic origin, socio-economic background, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health-related data, or data concerning sexual life or sexual orientation" (CEPEJ 2018, p. 9).

The many quotes of the "human rights" principle in the Charter confirm its objective stated in the introduction to support the responsible development of AI "with due regard for the fundamental rights of individuals as set forth in the ECHR and the Convention on the Protection of Personal Data" (CEPEJ 2018, p. 6). In addition, it is not surprising the number of citations that refer to judicial values, given that charter's focus is on AI developed in judicial systems. In particular, the charter quotes several principles related to a generic value of the rule of law, such as the importance of ensuring the equality of arms and the respect for the adversarial process or the development of AI that does not hinder judges' independence and impartiality. Furthermore, the report annexed to the charter details the risks for judicial values that may result from the introduction of AI. For instance, the use of machine learning tools supporting judicial decision-making and the consequential undue influence on judges' impartial and independent decisions.

The quotes indicating the multidisciplinarity of developers as an important precondition for responsible AI implementation are also of interest. The Charter's authors indicate the

multidisciplinarity of scientific analysis for AI development as a means to estimate the potential risks of AI application in social contexts when the risks of perpetuating discrimination are high. In addition, the focus on multidisciplinarity is coherent with the methodologies utilized to draft the Charter that consisted in the involvement of experts belonging to different scientific backgrounds: law, human rights, judicial systems, philosophy and epistemology, and computer science (Zlătescu and Zlătescu 2019).

TABLE 4

Principle	*No. of quotes in CEPEJ Charter*	*Distribution of principles in the documents analyzed*
Transparency	12	Transparency 71,3%
Judicial Value	8	Judicial Value 59,3%
No Discrimination	8	No Discrimination 78,7%
Human Rights	6	HumanRights 60,2%
Multidisciplinarity	5	Multidisciplinarity 29,6%
Safety Mechanisms	3	Safety Mechanisms 66,7%
Datastore Certified	3	DatastoreCertified 4,6%
Datastore Quality	3	DatastoreQuality 19,4%
Recourse	3	Recourse 39,8%
AI Legalization	3	AI Legalization 57,4%
Accountability	3	Accountability 65,7%
Human Centric	3	Human Centric 51,9%
EthicByDesign	3	EthicByDesign 14,8%
Academic Debate	2	Academic Debate 39,8%
Datastore Integrity	2	DatastoreIntegrity 18,5%
Equality Access	2	EqualityAccess 57,4%
Data Privacy	2	DataPrivacy 62,0%
Standard.Code.Guid.	2	StandardCodeGuideline 41,7%
Risk	1	Risk 66,7%
Datastore Security	1	DatastoreSecurity 28,7%
Insurance	1	Insurance 3,7%
Intellectual Property	1	IntellectualProperty 6,5%
Monitoring AI	1	MonitoringAI 33,3%
Open Source	1	OpenSource 6,5%
Traceability	1	Traceability 12,0%
Training	1	Training 56,5%
User Feedback	1	UserFeedback 9,3%
Awareness	1	MacroAwareness 39,8%
DataSensitive Cons	1	DataSensitiveCons 0,9%

Table 4. Number of principles' quotes in CEPEJ Charter and their distribution in the framework documents' sample.

The analytical comparison between the Charter's quotes and the distribution of principles in the sample of ethical documents resumed in Table 4, highlights the

similarities between AI in justice and AI applied in other contexts in terms of risks for fundamental rights and ethics. Apart from several intrinsic peculiarities that regard the influence of AI technology on judicial procedures, AI in justice may imply similar risks to the ones predicted with AI applied in other contexts. Consequentially, if we want to regulate AI in justice, we cannot fail to consider the experience of other contexts of application and the efforts of previously drafted ethical documents that indicate principles that may also have important implications for justice.

In this regard, looking at the principles found in the documents, but not included in the charter, it is evident that a number of these principles may be relevant to AI in justice. Several of the missing principles are listed in Table 5 with their relative distribution in the sample of ethical documents.

TABLE 5

Assessment	69.4%	Human Machine Harmony	5.6%
Consent	21.3%	Interoperability	12.0%
Consistency of Output	12.0%	No Profiling	7.4%
Datastore Anonymiz	16.7%	Replicability	2.8%
Datastore Control	38.0%	Stakeholders Involvement	38.0%
Determine Responsibility	53.7%	Surveillance Avoid	13.0%
Ecology	18.5%	WorkForce Challenge	26.9%
Free of NoTec	17.6%	Board Ethical Advisory	16.7%

Table 5. Several missing principles in CEPEJ Charter and their distribution in the sample.

The absence of the "assessment" principle here is immediately evident. "Assessment" indicates the establishment of mechanisms of evaluation of AI development's compliance with ethical principles and norms. This principle, which is retrieved in almost 70% of documents, is important also for AI in justice. Therefore, it may indicate the assessment of AI compliance not only to generic ethical principles but also to rule of law norms. It is fair to say that the evaluation of AI is not entirely excluded from the document, since a self-evaluation form that any developers or policy makers can use accompanies the Charter. However, it is surprising that no reference is made to assessment in the discussion of the five macro-principles that constitute the charter. A discussion on "Consent" (present in 21.3% of the ethical documents investigated) is also not included in the Charter. This principle indicates the necessity to obtain the consent of the subject affected by a decision supported by an AI technology. This principle is particularly important with reference to the use of AI in justice, given that decisions that can deeply affect the life of a party in a lawsuit (imagine for instance a judge's decision on detention) may be supported by a deceptive AI system (see the mentioned COMPAS case (Washington 2018). The issue of determining responsibility in cases of AI failure is connected to consent. The Charter lacks a discussion on responsibility (the principle "determine responsibility" is present in 53.7% of the ethical documents investigated). This discussion should include which actors are to be held liable in the event of AI failure: who uses the systems (such as the judge or other legal practitioners) or the developers. The technological failures of AI can be attributed to developers, as well as

to those who request and pay for the development of a system that is for instance, in the case of AI in justice, the Ministries of Justice. In contrast, the failures due to misuse could be attributable to the users, but this is not necessarily the case: for example, misuse can be caused by a lack of knowledge of complex AI systems due to the absence of adequate training. As acknowledged in previous literature (Contini and Lanzara 2008), a scarce diffusion of technological literacy between justice professionals may hinder the digitalization of justice procedures, and training may be useful to overcome this issue.[13] In the case of AI, and of issues and biases related to AI use in justice, training can be a useful strategy to open the AI black box, diffusing information on potential risks to users and reducing opportunities for failures.

The Charter focuses on different parts of the text on the implications of the use of large datasets that is typical of AI systems, also quoting the GDPR. Despite this, two principles related to the use of data mentioned in the sample of framework documents are excluded in the Charter – the anonymization of personal data and the user control on data stored. In addition, for AI in justice, the users' control on the storing of personal data or of data relative to procedures should be regulated, especially if data feeds AI or machine learning tools. Moreover, it is important that the storing and gathering of sensitive data by the judicial administration are protected from undesired identification of citizens.

The risks related to profiling have already affected the judicial context: the practice of judges' profiling by the statistical modeling of the decisions of individual judges is well established in countries such as the US with products such as Lex Machina and Context by Lexis Nexis and Gavelytics (Gavaghan 2017). This type of practice may entail the risks of so-called "forum shopping": the possibility for lawyers to bring cases to a certain jurisdiction on the basis of an assessment of the probabilities to win the case calculated through judicial profiling technologies. To restrict this practice, France issued a law that made it illegal to engage in "judicial analytics" and any practice to evaluate, analyze, compare, or predict the behavior of individual judges (Abiteboul and G'Sell 2019, Morison and Harkens 2020).[14] This argumentation confirms that future attempts to discipline AI in justice must clearly define the borders of digital profiling, both of citizens that access to justice and of judicial professionals.

Furthermore, the principle related to stakeholders' involvement during AI development, diffused in almost 40% of the ethical documents investigated, is not included in the Charter. The ICT design principles (Hanseth and Lyytinen 2016) and the literature on e-justice (Bailey *et al.* 2013, Lupo and Bailey 2014) support the idea of users' and stakeholders' involvement for the development of e-justice. Best practices emphasize the advantages of a staged, iterative process that incorporates inclusion and feedback from key stakeholders (Fersini *et al.* 2010). This may be of considerable importance also for AI in justice thus bringing two advantages: first, the inclusion of stakeholders allows

[13] Lanzara (2016) stressed that successful ICT systems have to achieve the right balance between system's maximum level of feasible simplicity and its maximum level of manageable complexity. As noted by Lanzara, systems that are simplified to a point that undermines their functionalities, value, and usefulness are highly unlikely to attract users and may, in fact, drive users to offline procedures (Lanzara 2016). However, systems cannot be so complex as to be beyond the technological capacity of most users. Therefore, training is the most effective means to raise the bar of manageable complexity by improving the technological capacity of users.

[14] LOI no. 2019-222 du 23 mars 2019 de programmation 2018–2022 et de réforme pour la justice (1).

developers to take advantage of users' knowledge and suggestions; second, it expands prospects for stakeholder acceptance of technological change, and it increases ownership in and championing of the success of the project (Contini and Lanzara 2008).

The Charter also lacks a discussion on the possible challenges posed by AI introduction and the consequential automation of justice professionals' jobs. The 26.9% of documents deal with the possible implications related to the automation of work processes due to AI and the replacement of the workforce with consequent reduction of available jobs. This aspect may also regard justice systems' professionals as for instance lawyers. The majority of AI systems for lawyers are applied for routine tasks that regard the processing of large amount of data and documents (as due diligence operations; (Roodman 2012). These tasks include the analysis of documentation usually performed by a trainee in law firms. By pursuing these assignments, trainees used to acquire the necessary skills to perform the profession. However, the introduction of AI may reduce these important training foundations in the law field.

Finally, the principles introduced in the Charter do not include the constitution of an advisory board that assesses the risks of AI introduction, as suggested in 16.7% of documents within the sample. Given the many risks related to AI introduction in the judiciary, the creation and involvement of such an actor could be fundamental to safeguard rule of law and access to justice. In addition, the operation of ethical assessment could be delegated to single individuals in each organization introducing AI in their routine operations as indicated by GDPR for what regards privacy and the protection of personal data: the regulation provides for a Data Protection Officer which has the function of supporting data management in line with the GDPR regulation.

The argumentation in this section, which highlighted a number of gaps of the Ethical Charter, does not represent a criticism for the work of the CEPEJ, which had the merit of drafting the first ethical guideline for AI in justice through the involvement of experts of different scientific fields. In addition to this, the ethical guideline has been enriched with a report annexed to the guidelines drafted thanks to experts' contributions that provides an in-depth analysis on the use of AI in judicial systems (in particular on AI applications processing judicial decisions and data). [15]

On the basis of this, focusing on the CEPEJ charter, a very comprehensive and multidisciplinary study on the use of AI in justice, allowed us to produce an overview of the risks of AI in the justice domain. The discussion that arose highlighted that disciplining AI in justice requires a broader spectrum of analysis that does not only involve the judicial sphere but also other fields of application if and when the "proto-constitutional discourse" (Gill *et al.* 2015) constituted by CEPEJ ethical framework will be the basis for comprehensive and binding norms.

[15] The analysis annexed to CEPEJ guidelines focuses in depth on the following topics: 1. Stating of the use of AI in the judicial systems of Council of Europe (CoE) member States; 2. Overview of open data policies relating to judicial decisions in the judicial systems of CoE member states; 3. Operating characteristics of AI applied to judicial decisions 4. AI legal reasoning; 5. AI judicial prediction' potentialities and limitations; 6 Modality of AI application in civil, commercial and administrative justice; 7. Specific issues related to AI in criminal justice; 8. AI in justice and protection of personal data; 9. Public debate on AI in justice and cyberethics; 10. Uses of AI in justice "to be encouraged", to be used "with precaution", to be applied "following additional scientific studies", to be applied "with extreme reservations".

6. Conclusion

The development of AI technology, and its introduction into the operational processes of different contexts, promises to revolutionize people's lives and their relationship with machines. This change is "revolutionary" not so much because human beings are developing new and innovative technologies, but because such technologies can act as intelligent agents that receive perceptions from the external environment and perform actions autonomously (Russell and Norvig 2002, Santosuosso and Poletti 2020). Consequently, this can result in an abrupt change in the "expected and loved order" described by Canguilheim (Angelides 2012) that may bring anxiety, concerns, or rejection. The creation of autonomous machines can really modify the "normal" as we have previously interpreted it, and moreover, due to the normative power of technology (Lanzara 2009, 2016), it can create a normativity that may conflict with the actual institutional and constitutional setting. The phenomenon of the massive production of ethical guidelines, collections of principles, and framework documents – the so-called "ethification" phenomena – (Van Dijk and Casiraghi 2020) represents society's reaction to the attack on "normality" that autonomous technologies may bring. This normative production, characterized by increased flexibility and rapidity in comparison to traditional law-making (useful to cope with the rapid AI technological evolution), is a first reaction to technological change. In addition, ethification may anticipate the "proto-constitutional discourse" (Gill *et al.* 2015) that eventuates in the issuing of binding norms.

Norms may have the capacity of curbing technological change and limiting its borders and its implications for the pre-established normal order as Foucault stated (Foucault 1977). However, the new order brought by AI technological innovation is characterized by unpredictable elements and new scientific knowledge that require consideration (Foucault 1977). Consequentially, norms must integrate scientific knowledge above all based on the empirical experience of AI application that may put in evidence the main risks for the context in which the technology is introduced. This operation can be more complicated for those application areas where the introduction of AI is still in its infancy.

This is the case of AI in justice, a field of application that is seeing the first experiments of this technology in judicial institutions, with few pilot cases becoming effective applications (Santosuosso and Poletti 2020, Spajosević *et al.* 2020) and the investments of few ICT companies for the development of applications in support of lawyers. From this perspective, the CEPEJ's attempt to discipline AI is undertaken in a context rife with uncertainty, because the empirical experience we have on issues related to AI in justice is limited. As noted through the analysis, the guidelines indicated by the Charter do not cover all aspects related to the application of AI. Consequently, they should be integrated. By comparing the Charter with the other framework documents analyzed, we identified other principles that could be considered to integrate the Charter, thus revealing that by looking at other application contexts in which AI is already utilized, it is possible to fill the gap of empirical experience that characterizes AI in the judiciary.

Consequentially, it is desirable that when the debate on AI in justice will be channeled in the legislative decision-making, the focus will not be limited to this single application's context, but it will benefit from the experience of all the areas of application, to grasp the multifaceted aspects of AI implications. A greater diffusion of these systems in the judiciary will also pave the way for more comprehensive

regulations. As highlighted by other forms of complex technology, experience, trial and error, as well as disasters and incidents – see, for instance, the case of modern aviation regulation – (Downer 2010), are the foundation of normative production. Only with a major diffusion of AI in the justice and other fundamental sectors may we acquire the fundamental experience useful to grasp its main implications for fundamental rights and values and to adequately regulate AI development and deployment.

References

Abiteboul, S., and G'Sell, F., 2019. *Les algorithmes pourraient-ils remplacer les juges?* [online]. Paris: Dalloz. Available from: https://hal.archives-ouvertes.fr/hal-02304016 [Accessed 15 February 2022].

Angelides, S., 2012. Disorder as "Pseudo-Idea". *Atlantis: Critical Studies in Gender, Culture & Social Justice*, 35(2), 10–20.

Bailey, J., Burkell, J., and Reynolds, G., 2013. Access to Justice for All: Towards an Expansive Vision of Justice and Technology. *Windsor Yearbook of Access to Justice* [online], 31(2), 181. Available from: https://doi.org/10.22329/wyaj.v31i2.4419 [Accessed 15 February 2022].

Barrett, C., 2019. Are the EU GDPR and the California CCPA becoming the de facto global standards for data privacy and protection? *Scitech Lawyer*, 15(3), 24–29.

Brennan, T., Dieterich, W., and Ehret, B., 2009. Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, 36(1), 21–40.

Cappelletti, M., 1979. Accesso alla giustizia: conclusione di un progetto internazionale di ricerca giuridico-sociologica. *Il Foro Italiano*, vol. 102, 53/54-59/60.

Castelvecchi, D., 2016. Can we open the black box of AI? *Nature News* [online], 538(7623), 20. Available from: https://doi.org/10.1038/538020a [Accessed 15 February 2022].

Contini, F., 2020. Artificial intelligence and the transformation of humans, law and technology interactions in judicial proceedings. *Law, Technology and Humans* [online], 2(1), 4. Available from: https://doi.org/10.5204/lthj.v2i1.1478 [Accessed 15 February 2022].

Contini, F., and Fabri, M., 2003. Judicial electronic data interchange in Europe. *Judicial electronic data interchange in Europe: Applications, policies and trends*, 1–26.

Contini, F., and Lanzara, G., eds., 2008. *ICT and innovation in the public sector: European studies in the making of e-government*. Cham: Springer.

Croux, C., and Dehon, C., 2010. Influence functions of the Spearman and Kendall correlation measures. *Statistical methods & applications*, 19(4), 497–515.

Daly, A., Devitt, S.K., and Mann, M., 2021. AI Ethics Needs Good Data. *In:* P. Verdegem, ed., *AI for Everyone? Critical Perspectives* [online]. London: University of Westminster Press. Available from: https://doi.org/10.16997/book55.g [Accessed 15 February 2022].

Daly, A., *et al*., 2019. Artificial intelligence, governance and ethics: Global perspectives. *The Chinese University of Hong Kong Faculty of Law Research Paper* [online], (2019–15). Available from: https://doi.org/10.2139/ssrn.3414805 [Accessed 15 February 2022].

Davenport, T.H., Barth, P., and Bean, R., 2012. How "big data" is different. *MIT Sloan Management Review* [online], 30 July. Available from: https://sloanreview.mit.edu/article/how-big-data-is-different/ [Accessed 15 February 2022].

Donaldson, M.S., Corrigan, J.M., and Kohn, L.T., eds., 2000. *To Err is Human: Building a Safer Health System.* Washington, DC: National Academies Press.

Downer, J., 2010. Trust and technology: the social foundations of aviation regulation. *The British Journal of Sociology,* 61(1), 83–106.

Erickson, A., 2018. Comparative Analysis of the EU's GDPR and Brazil's LGPD: Enforcement Challenges with the LGPD. *Brooklyn Journal of International Law* [online], 44(2), 859. Available from: https://brooklynworks.brooklaw.edu/bjil/vol44/iss2/9/ [Accessed 15 February 2022].

Erlingsson, C., and Brysiewicz, P., 2017. A hands-on guide to doing content analysis. *African Journal of Emergency Medicine* [online], 7(3), 93–99. Available from: https://doi.org/10.1016/j.afjem.2017.08.001 [Accessed 15 February 2022].

Etikan, I., Musa, S.A., and Alkassim, R.S., 2016. Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics* [online], 5(1), 1–4. Available from: https://doi.org/10.11648/j.ajtas.20160501.11 [Accessed 15 February 2022].

European Commission for the Effectiveness of Justice of the Council of Europe (CEPEJ), 2018. *European Ethical Charter on the use of artificial intelligence in judicial systems and their environment* [online]. Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3–4 December 2018). Strasbourg: European Commission for the Efficiency of Justice, Council of Europe. Available from: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c [Accessed 15 February 2022].

European Network of Councils for the Judiciary (ENCJ), 2013. *ENCJ Working Group: Judicial Ethics Report 2009–2010* [online]. Report of the European Network of Councils for the Judiciary. Available from: https://www.encj.eu/images/stories/pdf/ethics/judicialethicsdeontologiefinal.pdf [Accessed 15 February 2022].

Fersini, E., *et al*., 2010. *Semantics and machine learning: A new generation of court management systems* [online]. International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management. Valencia, 25–28 October.

Floridi, L., 2018. Soft ethics and the governance of the digital. *Philosophy & Technology* [online], 31(1), 1–8. Available from: https://doi.org/10.1007/s13347-018-0303-9 [Accessed 15 February 2022].

Foucault, M., 1977. Historia de la medicalización. *Educación médica y salud* [online], 11(1), 3–25. Available from: https://iris.paho.org/bitstream/handle/10665.2/3182/Educacion%20medica%20y%20salud%20%2811%29%2C%201.pdf?sequence=1&isAllowed=y [Accessed 15 February 2022].

Gavaghan, C., 2017. Lex Machina: Techno-Regulatory Mechanisms and Rules by Design. *Otago Law Review* [online], 15(1), 123. Available from: http://hdl.handle.net/10523/9199 [Accessed 15 February 2022].

Gill, L., Redeker, D., and Gasser, U., 2015. Towards digital constitutionalism? Mapping attempts to craft an internet bill of rights. Mapping Attempts to Craft an Internet Bill of Rights (November 9, 2015). *Berkman Center Research Publication* [online], no. 2015-15. Available from: https://doi.org/10.2139/ssrn.2687120 [Accessed 15 February 2022].

Hagendorff, T., 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* [online], 30(1), 99–120. Available from: https://doi.org/10.1007/s11023-020-09517-8 [Accessed 15 February 2022].

Hammoud, H., 2020. *Trade Secrets and Artificial Intelligence: Opportunities & Challenges* [online]. Available from: https://doi.org/10.2139/ssrn.3759349 [Accessed 15 February 2022].

Hanseth, O., and Lyytinen, K., 2016. Design theory for dynamic complexity in information infrastructures: the case of building internet. *In:* L.P. Willcocks, C.Sauer and M.C. Lacity, eds., *Enacting Research Methods in Information Systems.* Cham: Palgrave Macmillan, 104–142.

Hoffman, S., and Podgurski, A., 2013. Big bad data: Law, public health, and biomedical databases. *Journal of Law, Medicine & Ethics*, 41(S1), 56–60.

Huijboom, N., and Van den Broek, T., 2011. Open data: an international comparison of strategies. *European journal of ePractice*, 12(1), 4–16.

Jobin, A., Ienca, M., and Vayena, E., 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.

Lanzara, G.F., 2009. Building digital institutions: ICT and the rise of assemblages in government. *In:* F. Contini and G.F. Lanzara, eds., *ICT and Innovation in the Public Sector. Technology, Work and Globalization*. London: Palgrave Macmillan, 9–48.

Lanzara, G.F., 2016. *Shifting practices: Reflections on technology, practice, and innovation*. Cambridge, MA: MIT Press.

Lin, P., Abney, K., and Bekey, G., 2011. Robot ethics: Mapping the issues for a mechanized world. *Artificial Intelligence*, 175(5–6), 942–949.

Liu, H.W., Lin, C.F., and Chen, Y.J., 2019. Beyond *State v Loomis*: artificial intelligence, government algorithmization and accountability. *International Journal of Law and Information Technology*, 27(2), 122–141.

Lohr, J.D., Maxwell, W.J, and Watts, P., 2019. Legal Practitioners' Approach to Regulating AI Risks. *In:* K. Yeung and M. Lodge, eds., *Algorithmic Regulation*. Oxford University Press, 224–247.

Lupo, G., 2019. Regulating (Artificial) Intelligence in Justice: How Normative Frameworks Protect Citizens from the Risks Related to AI Use in the Judiciary. *European Quarterly of Political Attitudes and Mentalities* [online], 8(2), 75–96. Available from: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-62463-8 [Accessed 15 February 2022].

Lupo, G., and Bailey, J., 2014. Designing and implementing e-Justice Systems: Some lessons learned from EU and Canadian Examples. *Laws* [online], 3(2), 353–387. Available from: https://doi.org/10.3390/laws3020353 [Accessed 15 February 2022].

Marchant, G.E., 2011. The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight. *In:* G. Marchant, B. Allenby and J. Herkert, eds., *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight. The International Library of Ethics, Law and Technology* (vol 7). Dordrecht: Springer, 35–44.

Mayring, P., 2004. Qualitative content analysis. *In:* U. Flick, E.v. Kardorff and I. Steinke, eds., *A companion to qualitative research.* London/Thousand Oaks: Sage, 159–176.

Mohr, R., and Contini, F., 2011. Reassembling the Legal: "The Wonders of Modern Science" in Court-Related Proceedings. *Griffith Law Review*, 20(4), 994–1019.

Mol, A., 1998. Lived reality and the multiplicity of norms: a critical tribute to George Canguilhem. *Economy and Society*, 27 (2–3), 274–284.

Morison, J., and Harkens, A., 2020. Algorithmic justice: dispute resolution and the robot judge? *In:* M.F. Moscati, M. Palmer and M. Roberts, eds., *Comparative Dispute Resolution.* Cheltenham: Edward Elgar, 339–352.

Nunez, C., 2017. Artificial intelligence and legal ethics: Whether AI lawyers can make ethical decisions. *Tulane Journal of Technology & Intellectual Property* [online], vol. 20, 189. Available from: https://journals.tulane.edu/TIP/article/view/2682 [Accessed 15 February 2022].

Reiling, D., 2016. *Technology for justice: How information technology can support judicial reform.* Amsterdam University Press.

Rességuier, A., and Rodrigues, R., 2020. AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society* [online], 7(2). Available from: https://doi.org/10.1177/2053951720942541 [Accessed 15 February 2022].

Rigano, C., 2019. Using artificial intelligence to address criminal justice needs. *National Institute of Justice Journal* [online], 280, 1–10. Available from: https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs [Accessed 15 February 2022].

Rissland, E.L., Ashley, K.D., and Branting, L.K., 2005. Case-based reasoning and law. *The Knowledge Engineering Review*, 20(3), 293–298.

Roodman, D., 2012. *Due diligence: An impertinent inquiry into microfinance.* Washington, DC: CGD Books.

Rositi, F., 1988. Analisi del contenuto. *In:* M. Livolsi and F. Rositi, eds., *La 644icercar sull'industria culturale.* Rome: NIS, 59–94.

Russell, S.J., and Norvig, P., 2002. *Artificial intelligence: A modern approach*. 2nd ed. Hoboken: Prentice Hall.

Russell, S.J., and Norvig, P., 2016. *Artificial intelligence: A modern approach*. 3rd ed. Malaysia/London: Pearson Education.

Sandefur, R.L., 2009. Access to justice: Classical approaches and new directions. *In:* R.L. Sandefur, ed., *Access to Justice.* Bingley: Emerald, ix–xvii.

Santosuosso, A., and Poletti, D., 2020. *Intelligenza artificiale e diritto: Perché le tecnologie di IA sono una grande opportunità per il diritto.* Milan: Mondadori Università.

Schiff, D., *et al*., 2020. What's Next for AI Ethics, Policy, and Governance? A Global Overview. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* New York, 7–9 February. New York: Association for Computing Machinery.

Sherman, J., 2013. *Court Information Management Policy Framework to Accommodate the Digital Environment. Discussion Paper* (U14-23/2013E-PDF) [online]. Ottawa: Canadian Judicial Council. Available from: https://cjc-ccm.ca/cmslib/general/AJC/Policy%20Framework%20to%20Accommodate%20the%20Digital%20Environment%202013-03.pdf [Accessed 15 February 2022].

Simshaw, D., 2018. Ethical issues in robo-lawyering: The need for guidance on developing and using artificial intelligence in the practice of law. *Hastings Law Journal* [online], vol. 70, 173. Available from: https://www.hastingslawjournal.org/wp-content/uploads/70.1-Simshaw.pdf [Accessed 15 February 2022].

Skeem, J., and Eno Louden, J., 2007. *Assessment of evidence on the quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)* [online]. Report prepared for the California Department of Corrections and Rehabilitation by the Center for Public Policy Research. Davis: University of California. Available from: https://webfiles.uci.edu/skeem/Downloads_files/CDCR%20Skeem%26%20EnoLouden%20COMPASeval%20SECONDREVISION%20final%20Dec%2028%2007.pdf [Accessed 15 February 2022].

Spajosević, D., *et al*., 2020. *Study on the use of innovative technologies in the justice field* [online]. Final report. Prepared by the Directorate-General for Justice and Consumers. Brussels: European Commission. Available from: https://op.europa.eu/en/publication-detail/-/publication/4fb8e194-f634-11ea-991b-01aa75ed71a1/language-en [Accessed 15 February 2022].

Susskind, R.E., 1987. Expert systems in law: out of the research laboratory and into the marketplace. *Proceedings of the 1st international conference on Artificial intelligence and law.* Boston, 27–29 May. New York: Association for Computing Machinery.

Tallacchini, M., 2009. Governing by values. EU ethics: soft tool, hard effects. *Minerva*, 47(3), 281–306.

Tallacchini, M., 2015. To Bind or Not to Bind? European Ethics as Dolft law. *In:* S. Hilgartner, C. Miller and R. Hagendijk, eds., *Science and Democracy: Making Knowledge and Making Power in the Biosciences and Beyond*. Abingdon: Routledge, 156–175.

Thomas, I., 2020. Getting ready for the California Consumer Privacy Act: Building on General Data Protection Regulation preparedness. *Applied Marketing Analytics,* 5(3), 210–222.

Tipaldo, G., 2007. L'analisi del contenuto nella ricerca sociale: Spunti per una riflessione multidisciplinare. *Quaderni di Ricerca del Dipartimento di Scienze sociali dell'Università di Torino*, 9.

Van Dijk, N., and Casiraghi, S. 2020. "The "ethification" of privacy and data protection law in the European Union. The Case of Artificial Intelligence. *Brussels Privacy Hub Working Paper* [online], 6(22). Available from: https://brusselsprivacyhub.eu/publications/BPH-Working-Paper-VOL6-N22.pdf [Accessed 15 February 2022].

Velicogna, M., 2007. Justice systems and ICT-What can be learned from Europe. *Utrecht Law Review* [online], 3(1), 129–147. Available from: https://doi.org/10.18352/ulr.41 [Accessed 15 February 2022].

Velicogna, M., 2018. e-Justice in Europe: From national experiences to EU cross-border service provision. *In:* L. Alcaide Muñoz and M. Rodríguez Bolívar, eds., *International E-Government Development*. Cham: Palgrave Macmillan, 39–72.

Wagner, B., 2018. Ethics as an escape from regulation: From ethics-washing to ethics-shopping. Being profiling. *Cogitas ergo sum*, 1–7.

Wallace, A., 2003. *Overview of public access and privacy issues.* Paper delivered at Queensland University of Technology Conference. 6 November.

Washington, A.L., 2018. How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate. *Colorado Technology Law Journal*, 17(1), 131–161.

Wilkinson, S., 2018. Brazil's new General Data Protection Law. *Journal of Data Protection & Privacy*, 2(2), 107–115.

Završnik, A., 2020. Criminal justice, artificial intelligence systems, and human rights. *ERA Forum* [online], 20. Available from: https://doi.org/10.1007/s12027-020-00602-0 [Accessed 15 February 2022].

Zlătescu, I.M., and Zlătescu, P.E., 2019. Implementation of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment. *Law Review: Judicial Doctrine & Case-Law* [online], 10, 237–242. Available from: http://internationallawreview.eu/fisiere/pdf/23_Zlatescu__Supliment_Law_Review_SRDE.pdf [Accessed 15 February 2022].

Appendix

Tab. A.1. Ethical documents, Authors, Date of issuing

Author	Ethical Document	Date
Association for computing machine	*ACM code for Ethics*	2018
Accenture	*Accenture - Universal principles of data ethics*	2016
Personal Data Protection Commission Singapore	*A Proposed Model of Artificial Intelligence Governance Framework*	2019
National Research Council Canada	*Advisory Statement on Human Ethics in Artificial Intelligence and Big Data Research (2017)*	2019
AI Now	*AI NOW Ethical document on AI*	2018
Information Technology Industry Council	*AI policies and principles*	2017
Atomium – EISMD (AI4People)	*AI4People's Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*	2018
ADEL	*Algorithm Data Ethics Label*	2018
Smart Dubai Office	*Artificial Intelligence Ethics and Principles*	2019
Future of Life Institute	*Asilomar AI Principles*	2017
Institute for Business Ethics	*Business Ethics and Artificial Intelligence*	2018
Leadership Conference on Civil and Human Rights	*Civil Rights Principles for the Era of Big Data*	2014
DataEthics.eu	*Data Ethics on AI*	2018
Critical Engineering Working Group	*Critical Engineering Working Manifesto*	2011
Ekspertgruppen om DATAETIK (Danish Expert Group on Data Ethics)	*Data for the Benefit of the People*	2018
Government of Canada	*Directive on Automated Decision-Making*	2019
International Conference of Data Protection and Privacy Commissioners ICDPPC	*Declaration on Ethics and Data Protection in Artifical Intelligence*	2018
Center for Democracy & technology (CDT)	*Digital Decisions*	2017
CIGREF & Syntec Numérique	*Digital Ethics*	2018
IEEE	*Ethically Aligned Design – 1st Edition*	2019
EC- High Level Expert Group on AI	*Ethics guidelines for trustworthy AI*	2019
European Commission for the Efficiency of Justice (CEPEJ)	*European ethical charter on the use of Artificial Intelligence in judicial systems and their environment*	2018
IBM	*Everyday Ethics for Artificial Intelligence*	2019

Datenschutzkonferenz	*Hambacher Erklärung zur Künstlichen Intelligenz – Sieben datenschutzrechtliche Anforderungen (Conference of the Independent Federal and State Data Protection Supervisory Authorities Germany: Hambach Declaration on Artificial Intelligence – Seven data protection obligations)*	2019
The Holberton-Turing Oath	*Holberton Turing Oath*	2018
IBM	*IBM's Principles for Trust and Transparency*	2018
Bundesverband KI	*KI Gütesiegel (AI Seal of Quality)*	2019
Google	*Objectives for AI Applications*	2018
Open AI	*Open AI Charter*	2018
Microsoft	*Our Approach to AI*	/
Google | People + AI Research (PAIR)	*People + AI Guidebook*	/
FAT/ML (Fairness, accountability and transparency in machine learning)	*Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*	/
OECD Legal Instruments	*Recommendation of the Council on Artificial Intelligence*	2019
Google AI	*Responsible AI Practices*	/
Treasury Board of Canada Secretariat	*Responsible AI in the Government of Canada*	2019
Microsoft	*Responsible bots: 10 guidelines for developers of conversational AI*	2018
Treasury Board of Canada Secretariat	*Responsible use of artificial intelligence (AI)*	2019
SAP	*SAP's guiding principles for Artificial Intelligence*	2018
DataforGood	*Serment d'Hippocrate pour Data Scientist (Hippocratic Oath for Data Scientist)*	2018
Association for Computing Machinery	*Statement on Algorithmic Transparency and Accountability*	2017
Partnership On AI	*Tenets Partnership on AI*	
Microsoft	*The Future Computed – Artificial intelligence and its role in society*	2018
The Good Technology Collective	*The Good Technology Standard (GTS:2019-Draft-1)*	2018
Japanese Society for AI	*The Japanese Society for Artificial Intelligence Ethical Guidelines*	2017
Amnesty International & Access Now	*The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems*	2018
UNI Global Union	*TOP 10 PRINCIPLES FOR ETHICAL ARTIFICIAL INTELLIGENCE*	2018
CIGI Gentre for International Governance Innovation	*Toward a G20 Framework for Artificial Intelligence in the Workplace (CIGI Paper No. 178)*	2018
The Public Voice	*Universal Guidelines for Artificial Intelligence*	2018

Artificial Intelligence Industry Alliance	*Joint Pledge on Artificial Intelligence Industry Self-Discipline*	2017
National New Generation Artificial Intelligence Governance Expert Committee	*Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence*	2019
EPSRC	*Principles of robotics*	2010
European Group on Ethics in Science and New Technologies (EC)	*Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems (PDF)*	2018
IA LATAM	*Declaración de Ética para desarrollo y uso de la Inteligencia Artificial*	2017
Verivox	*Verivox/Pro7 Selbstverpflichtung (Self-commitment)*	2019
Kakao corporation	*Kakao Algorithm Ethics*	/
Monetary authority of Singapore	*Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector (PDF)*	2018
SAGE	*The Ethics of Code: Developing AI for Business with Five Core Principles*	2017
The Alan Turing Institute	*Understanding artificial intelligence ethics and safety*	2019
UK Government	*A guide to using Artificial Intelligence in the public sector*	2019
UNESCO	*Report of COMEST on Robotics Ethics*	2017
University of Notre Dame	*A Code of Ethics for the Human Robot Interaction (PDF)*	2014
B debate	*Barcelona Declaration for AI*	2019
Chinese government	*BEIJING AI PRINCIPLES*	2019
Royal Australian and New Zealand College of Radiologists	*Ethical Principles for AI in Medicine*	2019
PLOS Computational Biology	*Ten simple rules for responsible big data research*	2017
The ethics centre	*The-Ethics-Centre PRINCIPLES-FOR-GOOD-TECHNOLOGY*	2018
KI + VERWALTUNG	*9 Theses On opportunities and risks, democratic legitimacy and constitutional control in the algorithmization of administration*	2018
ind.ie	*Ethical Design Manifesto*	2017
Access Now	*Human rights in the Age of AI*	2018
New York Times	*How to Regulate Artificial Intelligence*	2017
Council of Europe	*Unboxing artificial intelligence: 10 steps to protect human rights*	2019
Council of Europe	*Declaration on the manipulative capabilities of algorithmic processes*	2019
Council of Europe	*Recommendation about Technological convergence, artificial intelligence and human rights*	2017

Council of Europe - T- PD	*Guidelines on the data protection implications of artificial intelligence*	2019
European commission	*AI FOR EUROPE*	2018
The Conference toward AI Network Society	*Draft AI R&D Guidelines*	2017
Data&Society	*Governing Artificial Intelligence: Upholding human rights and dignity*	2019
Institute for the Future, Omidyar Network	*Ethical OS framework*	2018
Global Network Initiative	*GNI-Principles-on-Freedom-of-Expression-and-Privacy*	2017
O'Reilly	*Data Ethics Checklist*	2018
World economic forum	*4 Steps to developing Responsible AI*	2019
Comité Consultatif National d'Ethique (CCNE) and Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistene (CERNA)	*Digital technology and healthcare, which ethical issues for which regulations?*	2018
French parliament	*Statement for a meaningful AI*	2018
G20 report	*Human centred principles*	2019
Federal government	*Nationale Strategie für Künstliche Intelligenz*	2018
Task Force on Artificial Intelligence of the Agency for Digital Italy	*AI at the service of Citizens*	2018
British Embassy in Mexico through the Prosperity Fund	*Towards an AI strategy in Mexico*	2018
New zeland human rights commission	*Privacy, Data and Technology: Human Rights Challenges in the Digital Age*	2018
UK government	*Data Ethics framework*	2018
Executive Office of the President of the US, National Science and Technology Council Committee on Technology	*Preparing for the future of AI*	2016
integrate.ai	*Responsible AI in Consumer Enterprise*	2018
New America	*joint Pledge on Artificial Intelligence Industry Self-Discipline (Draft for Comment)*	2019
Telia company	*Guiding principles on trusted AI ethics*	2019
DrivenData	*Deon- An ethics Checklist for data scientist*	/
Fast.ai	*AI ethics resources*	2018
H5	*A 'principled' artificial intelligence could improve justice*	2017
IDEO	*AI Ethical compass*	2018
Center for Democracy & technology (CDT)	*Digital Decision*	2017
Verbraucherzentrale Bundesverband e.V. Associazione consumatori	*Algorithmenbasierte Entscheidungsprozesse (Algorithm-based decision-making processes)*	2017
Bertelsmann Stiftung & iRights.lab	*AlgoRules*	2019

Ethkikkommission des Bundesministeriums für Verkehr und digitale Infrastruktur BMVI (Federal Ministry of Transport and Digital Infrastructure Ethics committee)	*Automatisiertes und Vernetztes Fahren (Automated and interconnected driving)*	2017
Institute for Digital Ethics (IDE) at the Stuttgart Media University	*10 ethische Leitlinien für die Digitalisierung von Unternehmen (10 ethical guidelines for the digitisation of enterprises)*	2017
Bitkom	*Empfehlungen für den verantwortlichen Einsatz von KI und automatisierten Entscheidungen - Corporate Digital Responsibility and Decision Making*	2018
Gesellschaft für Informatik	*Ethische Leitlinien (Ethical Guidelines)*	2018
Deutsche Telekom	*KI Richtlinien Deutsche Telekom (AI Guidelines)*	2018
Université de Montréal	*Montreal Declaration for Responsible AI*	2018
Telefonica	*Principos (Principles)*	2018
DIRECTORATE-GENERAL FOR INTERNAL POLICIES	*EUROPEAN CIVIL LAW RULES IN ROBOTICS EP*	2016

Tab. A.2. Principles, codes, definitions and their distribution in ethical documents

Principle's Code	% in Docs	Definitions
Transparency	83,3%	Transparency of information on AI functioning and procedures.
No Discriminat	79,6%	Avoid discrimination of any nature as gender, race and wealth.
Data Issues	75,9%	Macro-code including all issues related to data use.
Assessment	69,4%	Inclusion of mechanisms of evaluation of compliance to ethical guidelines.
Risks	67,6%	Risk of harm resulting from the use of AI.
Safety mechanisms	67,6%	Means to ensure safety of AI use.
Accountability	66,7%	Measures of accountability, auditability and external control on AI functioning.
HumanRights	61,1%	Reference to human rights and international and EU norms on fundamental rights.
Judicial Values	60,2%	References in the texts on judicial values referring to generic rule of law.
Equality of Access	58,3%	Inclusiveness and equality of access to AI technologies.
Legalization of AI	58,3%	Sentences exhorting binding regulation of AI use.
Training	57,4%	Support for training and education of AI users.
Ethical AI	56,5%	AI respectful of AI principles.
Bias Detection Mechanisms	53,7%	Mechanisms of assessment and detection of systems' biases.
Determine Responsibility	53,7%	Determine allocation of juridical responsibility in case of AI failure.
Human centric	52,8%	Human control over AI functioning.

Standards.Codes. Certificat	42,6%	Support for creation and adherence to AI standards, codes, certifications.
Academicdebate	40,7%	Support for academic research and debate on AI and its implications.
Recourse	40,7%	Ensure redress opportunities against biased decisions based on the use of AI.
AwarenessofAI	40,7%	Citizens' awareness of the use of AI supporting decisions affecting them.
Trust	39,8%	Support trust in AI technologies.
StakeholdersInvolvement	38,0%	Involvement of stakeholders for AI development.
AI Beneficial	37,0%	Sentences indicating potential positive outcomes caused by AI use as economic and well being improvement.
MonitoringAI	34,3%	Monitoring of AI safe and responsible functioning.
Multidisciplinarity	30,6%	Support for multidisciplinary teams developing AI.
WorkFChallenge	26,9%	Sentences indicating AI challenges to work force and employment.
Misuse	21,3%	Sentences warning against AI misuse.
SocialContext	20,4%	Data for AI representative of social context.
Robustness	19,4%	Robustness of AI functioning and results.
Ecology	18,5%	Sustain AI minimizing ecological impact.
GovernanceFramework	18,5%	Governance framework for AI diffusion in society.
Blackbox issue	17,6%	Issue of obscure functioning of AI and algorithms also due to intellectual property protection.
FreeofNoTec	17,6%	Support right of citizens not to utilize technology or AI tools.
Sustainability	17,6%	Sustainability of AI use.
TensionsbtwPrinciples	17,6%	Sentences indicating potential tensions between principles.
BestPractice	16,7%	Support for research and application of best practices for AI development and use.
BoardEthicalAdvisoryDIC	16,7%	Support for creation of an ethical advisory board for AI use.
JustifiableAiDIC	14,8%	Use of AI only if justifiable and necessary.
Metrics	13,0%	Develop metrics for AI evaluation.
SurveillanceAvoid	13,0%	Avoid use of AI for surveillance.
Traceability	13,0%	Traceability of AI use and results.
ConsistenofOutput	12,0%	AI output consistent with data input.
Interoperability	12,0%	AI interoperability with other systems already in place.
ErrorCorrecion andMitigation	11,1%	Mechanisms for AI errors' correction and mitigation.
OpenData	11,1%	Favour the use and diffusion of open data.
WarFare	11,1%	Sentences warning against the use of AI for warfare.
Efficient	10,2%	Efficiency of AI technology.
UserFeedback	10,2%	Take into account of user feedbacks when developing AI.
DemocracSupport	9,3%	Support for democratic values when developing AI.

GovernmentControlofAI	8,3%	Government control over the use and development of AI.
IntellectualProperty	7,4%	Protection of intellectual property related to AI development.
OpenSource	7,4%	Sentences sustaining use of open source for AI development.
ProfilingNO	7,4%	Avoid use of AI for citizens' profiling.
Solidarity	6,5%	Sustain AI pro-socially and supporting interpersonal solidarity.
HumanMachineHarmony	5,6%	Support for human-machine harmony.
HumanMimicking	5,6%	Sentences warning of the use of AI for human and human emotions' mimicking that may be utilized for people manipulation.
Predictability	5,6%	Predictability of AI results.
Insurance	4,6%	Provide for insurance mechanisms for AI damages.
Reimboursement	4,6%	Citizens' reimbursements in case of harms caused by AI.
UserAssistance	4,6%	Provide for user assistance for AI systems' utilizers.
ConflictofInter	3,7%	Minimize conflicts of interest among developers and stakeholders.
ConsumerProtec	3,7%	Protection of consumers.
InternatNatCooperation	3,7%	International cooperation for developing rules for responsible and ethical AI.
Resilience	3,7%	Sentences supporting AI resilience against cybersecurity attacks.
NoCoercion	2,8%	No coercion for the use of AI.
Legitimacy	2,8%	Legitimacy of AI use.
Replicability	2,8%	Replicability of results obtained with AI use.
Responsiveness	2,8%	AI responsive to user needs.
Adaptability	0,9%	Adaptability of AI systems.
Authenticat	0,9%	Systems of authentication for AI use.

Data issues macro code		
Privacy Protection	63%	Generic reference to protection of privacy.
Datastore Control	38%	Subjects' control over the storing of their data.
Datastore Information	31%	Diffusion of information on data storing to subjects.
Datastore Security	30%	Security of data storing methods.
Datastore Quality	20%	Quality of data stored.
Datastore Integrity	19%	Integrity of data stored.
Datastore Necessity	19%	Always store only data that are necessary.
Data Ethics	19%	Data storing supporting ethical values.
Datastore Anonymiz	17%	Store data protected with anonymization and pseudonymization mechanisms.
Datastore Representativ	14%	Representativeness of data stored.
Datastore permission	13%	Consent on the storing of personal data.
Datastore Awareness	12%	Subject's awareness of personal data stored.

Datastore Updated	9%	Data stored updated.
Datastore Access	6%	Accessibility of personal data stored by subjects.
Datastore Assessment	6%	Evaluation of data store compliance with ethics and norms.
Datastore Certified	6%	Respect data store certification.
Datastore Legitimacy	3%	Legitimacy of data stored.
Datastore Limitation	3%	Limitation to personal data stored.
Datastore Monopoly	3%	Avoid companies' monopoly on the storing of personal data.
Datastore Ownership	3%	Respect ownership of subjects' of data storing.
Datastore redundancy	2%	Avoid redundancy of data storing.