Document for Consultation

Draft UNESCO Guidelines for the Use of AI Systems in Courts and Tribunals

Why the Guidelines?

In 2023, a UNESCO survey on the use of AI systems by judicial operators found that 93% of judicial operators surveyed were familiar with AI technologies, with 44% already using AI tools such as ChatGPT for work-related activities. Further, only 9% of judicial operators surveyed reported that their organizations had issued guidelines or provided AI-related training, thereby underlining the need for guidance on the use of AI systems in courts and tribunals.

As part of UNESCO's AI and the Rule of Law programme, under the Global Judges Initiative, UNESCO has developed the draft Guidelines for the Use of AI Systems in Courts and Tribunals. The Guidelines aim to offer comprehensive guidance to courts and tribunals to ensure that the deployment of AI technologies aligns with the fundamental principles of justice, human rights, and the rule of law.

The guidelines are open for public consultation in English until **5 September 2024.**

UNESCO encourages stakeholders, including judicial professionals, legal experts, and the public, to review and provide feedback on the draft guidelines.

Share your comments and feedback through the form here.

Process of development of the draft Guidelines

September – October 2023	Desk research
September 2023– April 2024	Undertaking survey on the use of AI systems by judicial operators
May 2024	Survey findings launched at the WSIS+20 Forum in Geneva, Switzerland
May – June 2024	1st round of Expert Consultative Workshop and written expert comments on first draft of guidelines
June 2024	2nd round of written expert comments and second draft of guidelines
July 2024	Third draft of guidelines
August –September 2024	Open consultations on the third draft of the guidelines
September – October 2024	Consolidation of comments and finalization of the guidelines
November 2024	Publication and launch of the guidelines

Contributors to the draft Guidelines

The first draft of the Guidelines was prepared by Juan David Gutiérrez, PhD, Associate Professor, University of Los Andes, Colombia.

The following experts contributed to the consultative workshops and the drafts of the Guidelines:

Dr. Catherine Regis (Professor, Faculty of Law, University of Montreal), Judge Jean Aloise Ndiaye (Supreme Court of Senegal), Ameen Jauhar (Team Lead, Centre for Applied Law & Tech, Vidhi Centre for Legal Policy), Karine Caunes (Global Program Director, EU Law, Center for AI and Digital Policy), Alfonso Peralta Gutiérrez (Judge of the First Instance and Criminal Investigation, Granada, Spain), Anthony Wong (President of IFIP, International Federation for Information Processing), Amanda Leal (Associate, AI Governance at The Future Society), Dr. Krishna Ravi Srinivas (Adjunct Professor at NALSAR University of Law), Anca Radu (Researcher, European University Institute, Department of Law), Dr. Dory Reiling (Retired Senior Judge, International Court and ICT expert), Carolina Villadiego (Legal Counsel for Latin America, International Commission of Jurists), Grace Mutung'u (Researcher and Advocate of the High Court of Kenya), Matthieu Quinou (Doctor of Private Law and Attorney at Law at the Paris Bar, France), Martin Okumu Masiga (Secretary

General, Africa Jurists and Judges Forum), Colin van Noordt (Researcher), Felipe Guzmán (Consultant on regulatory instruments for the digital transformation of the Judicial Branch, Superior Council of the Judiciary of Colombia), Agneris Sampieri (Mexican Attorney), Dr. Miriam Stankovich (Principal Digital Policy Specialist, Center for Digital Acceleration, United States), Kalliopi Terzidou (Doctoral Researcher, Faculty of Law, University of Luxembourg), Giulia Gentile (Assistant Professor, Essex Law School, United Kingdom).

Table of Contents

Introduction

The interest in adopting artificial intelligence (AI) tools by courts and tribunals is growing (*Box No. 1*).[1] More recently, judges, judicial support staff, prosecutors, and lawyers around the globe have started to use chatbots powered by Large Language Models (LLMs) to draft legal documents, judicial decisions, and elaborate arguments in court hearings (*Box No. 2*).[2]

Box No. 1 – Use cases of AI by courts and law firms

- The Brazilian Supreme Court implemented VICTOR, a system that processes thousands of appeals brought to the court and facilitates the identification of cases that meet the "general repercussion" prerequisite.

- The Supreme Court of India deployed SUVAS, a software used to translate thousands of documents in English into ten vernacular languages and vice versa.

- Natural Language Processing (NLP) tools are used to predict judicial decisions from the Court of Justice of the European Union.

- In South Africa and Zimbabwe, law firms adopted AI tools for contract review and management and casework research.

Sources: Aletras et al. (2016), Aneja and Mathew (2023), Kufakwababa (2021), Medvedeva et al. (2020), Ministro do Superior Tribunal de Justiça (2020), and UNESCO (2023).

Box No. 2 – Use cases of generative AI by judges and lawyers

- In Argentina, judges have used LLMs to generate summaries of their decisions in plain and accessible language.

Law firms, legal service companies, and universities have developed generative AI systems based on LLMs, either independently or with tech companies, to conduct legal research and litigation work, add context to a case, summarize legal texts, and draft emails and contracts. These tools are not limited to general-purpose commercial LLMs; justices and lawyers can also use LLMs designed exclusively to carry out legal activities and open-source LLMs.

- Cases of judges and lawyers who issued judicial decisions or submitted legal documents that included references to non-existent rulings due to the use of AI chatbots have been reported in the United States, South Africa, and Brazil.

Sources: Adams (2023), Ambrogi (2023), Benetts et al. (2023), Gutiérrez (2024), LexisNexis (2023), Weiser (2023), Weiser and Bromwich (2023), and Witten (2023).

However, formal guidance on adequately using these tools for individuals or organizations in the justice sector is limited. However, there are a few published official principles, rules, or guidelines on how AI tools can be used ethically and responsibly for the administration of justice. These includes guidance published by Australia (2024), Brazil (2020), Canada (2023, 2024), New Zealand (2023), and the United Kingdom (2023).[3]

The need for guidance is accentuated by new AI laws that include specific rules regarding using these tools in judicial contexts. For example, the European Union's AI Act classifies as "high risk" those AI systems that are "intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution".[4] This classification will trigger obligations for the deployers of these AI systems, such as implementing a risk management system and ensuring human oversight of the tools.[5]

Additionally, AI systems integrated into decision-making processes may have cascading effects on the whole judicial system. On the one hand, the adoption of AI systems may benefit court users. For example, using AI tools for court management in the judiciary could contribute to ensuring the right to trial within a reasonable time. These AI systems can support pre-trial activities (e.g., automating the courts' filing system), court hearings (e.g., automatic translation), and post-sentencing proceedings (e.g., anonymizing court decisions).[6] On the other hand, such impacts may be harmful, for example, when the deployment contributes to systematically discriminating vulnerable individuals or groups of individuals in situations in which such tools are defective or misused. Hence, new measures at the organizational and individual levels could be required to prevent and mitigate adverse effects and to enhance the opportunities to improve the administration of justice.

Furthermore, AI systems represent a plurality of technologies that can potentially transform and disrupt the practice of law and judicial adjudication.[7] The *UNESCO Guidelines for the Use of AI Systems in Courts and Tribunals* (from now on *Guidelines*) offer guidance on the measures that the judicial sector could consider to enhance its capabilities in the context of digital transformation.

The *Guidelines* aim to contribute by proposing principles and recommendations for deploying and using AI tools to support the administration of justice. The *Guidelines* are relevant for organizations of the judiciary, such as courts and tribunals, and for individuals who are part of these judicial organizations, including magistrates, justices, judicial officers, and judicial support staff (from now on, collectively referred to as the judiciary).

Although the *Guidelines* were specifically designed to orient the judiciary's adoption and use of AI systems, they are pertinent to the legal profession broadly as well. In this sense, most of the principles and recommendations proposed in the *Guidelines* could be adapted for prosecutors, arbitrators, lawyers, bar associations and law societies, civil

servants in legal administration, and researchers. Civil society organizations such as professional associations and universities have recently published similar guidance.[8g]

Box No. 3 – Key definitions

- AI system: Computational systems that can "process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control" (UNESCO 2022, 10).

- Algorithmic audit: Examination of algorithmic systems that takes place after deployment to assess compliance with regulations or more targeted issues such as bias, fairness, transparency, explainability, security, or performance (Ada Lovelace Institute and DataKind UK 2020).

- Algorithmic impact assessments: Examination of algorithmic systems that takes place before deployment to assess future societal impacts and identify potential risks, which may focus on specific areas such as human rights, data protection, privacy, non-discrimination, and environment protection (Ada Lovelace Institute and DataKind UK 2020, Reisman et al. 2018).

- Algorithmic impact evaluations: Examination of algorithmic systems that takes place after deployment to assess its effects on users or impacts on affected populations (Ada Lovelace Institute and DataKind UK 2020).

- Generative AI system: Computational systems "that communicate in natural language, able to give answers to relatively complex questions and can create content (provide a text, picture, or sound) following a formulated question or instructions (prompt)" (CEPEJ-GT-CYBERJUST 2024, 2).

- Individual members of the judiciary: Magistrates, judges, justices, judicial officers, and judicial support staff.

- Judicial operator: A legal professional that is involved in the administration of justice or that has an active legal role in a judicial process, such as judges, judicial support staff, prosecutors, and lawyers.

- Large Language Models (LLMs): Generative AI models "that process textual inputs, known as prompts, and generate text outputs based on them. Their inputs, as well as outputs, can be in different text formats such as natural language, tabulated text, or even program code" (BSI 2024, 7).

- Organizations of the judiciary: Bodies that govern the judiciary, courts, and tribunals.

The *Guidelines* provide orientation for adopting and using AI tools and include specific recommendations for generative AI systems. These tools can help improve the quality of the judiciary's work by facilitating court management, document drafting, exploring specific topics, automating tasks, and supporting decision-making processes. Moreover, one of the key features of LLMs is that they facilitate the interaction between computers and users through natural language.

Although AI tools can support the core objectives of the judicial sector, the adoption of defective instruments and the negligent use of AI systems by the judiciary may also undermine human rights, such as fair trial and due process, access to justice and effective remedy, privacy and data protection, equality before the law, and non-discrimination, as well as judicial values such as impartiality, independence, and accountability.[9] Moreover, the misuse of AI systems may undermine society's trust in the judicial system.

AI tools are not a substitute for qualified legal reasoning, human judgment, or tailored legal advice. For example, AI chatbots powered by LLMs generate text by stitching together sequences of linguistic forms detected in its training data "according to probabilistic information about how they combine, but without any reference to meaning".[10] These AI tools cannot substitute for qualified legal reasoning because they have no rationale or contextual understanding of a legal problem. Furthermore, the use of LLMs for legal purposes by non-lawyers requires caution since the systems' output may appear authoritative and coherent while presenting facts that may be inaccurate and responses that are not consistent with legal facts.[11]

In sum, the *Guidelines* propose principles and specific guidance for the judiciary to prevent adopting AI tools that do not comply with human rights, prevent the misuse of AI systems, and enhance the benefits these tools can deliver.

The *Guidelines* are based on four main types of sources: a) UNESCO's "Recommendation on the Ethics of Artificial Intelligence"[12]; b) UNESCO's toolkits and guidance on AI, including the "Global Toolkit on AI and the Rule of Law for the Judiciary"[13] and the "Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence"[14]; c) the results of UNESCO's survey on the use of AI by judicial operators[15]; and d) diverse secondary sources and specialized literature, cited in the endnotes, including principles and guidelines published by the bodies that govern the judiciary, courts, and civil society organizations of diverse jurisdictions, including bar associations and law societies.

Guidelines for the use of AI systems by Courts and Tribunals

These *Guidelines* aim to advance universal respect for justice and the rule of law by offering guidance on using AI systems, including generative AI, by the judiciary. The *Guidelines* propose principles and specific recommendations for organizations and individuals of the

judiciary who intend to use AI systems to perform various tasks in the judicial sector while ensuring the protection, promotion, and respect of human rights and fundamental freedoms.

SECTION 1: PRINCIPLES

Below are the guiding questions for this section. Please provide your answers or comments in the <u>survey form</u>.

1. **Are the principles for the development, procurement and deployment of AI systems to be followed by organizations of the justice sector below adequate? If no, please provide reasons.**

2. **Are the principles for the use of AI systems to be followed by judicial operators below adequate? If no, please provide reasons.**

3. **Should we consider adding or deleting some principles? If yes, what principles should be added or deleted?**

4. **Should there be specific principles for generative AI systems? If yes, please provide reasons.**

1. Principles

Organizations and individual members of the judiciary should follow thirteen principles when adopting and using any type of AI system, including generative AI tools:

1.1. **Protection of human rights**: Adopt AI systems grounded in respect for human rights that allow the judiciary to respect, protect, and promote human rights while administering justice.[16] While all human rights that may be potentially affected by AI systems should be considered along the life cycle of these tools, the following four points are particularly important in the context of judicial proceedings:

   a. **Fairness:** Adopt AI systems that aim to attain their goals through processes that safeguard fairness and ensure inclusive technology access.[17]

   b. **Non-discrimination:** Prevent biased applications of AI systems and outcomes that reproduce, reinforce, perpetuate, or aggravate discrimination.[18]

   c. **Procedural fairness:** Assess the implications of AI systems for procedural fairness throughout the AI system's life cycle and prevent deployments that breach rights to procedural fairness.

    d. Personal data protection: Adopt AI systems that protect personal data treated for the administration of justice and deploy tools that contribute to anonymizing judicial decisions.[19] The judiciary should avoid using AI tools in ways that generate risks of disclosing such data or enable unauthorized access by third parties.[20]

1.2. Proportionality: Adopt AI systems that aim to achieve legitimate and proportional ends in the context of their use.[21]

1.3. Safety: Adopt AI systems that avoid, address, prevent, and eliminate unwanted harm.[22]

1.4. Information security: Adopt AI systems that protect confidential information in line with international standards for access to information.[23]

1.5. Awareness and informed use: Understand the functionalities, types of uses, potential impacts, limitations, and risks of available AI systems to make informed decisions about their implementation and be aware of the intended purpose for using a specific AI system to carry out judicial activities.[24]

1.6. Transparent use: Inform in a proper and timely manner when and how AI systems are used **and** how these tools work, especially when decisions made with or based on such tools can affect the rights and freedoms of individuals or communities.[25]

1.7. Accountability and auditability: Ensure accountability by informing and explaining why **certain** AI tools were adopted by the judiciary and ensuring traceability of the AI system's processes and outcomes, mainly when such tools are used for decision-making processes.[26] Undertake administrative, legal, and human measures to ensure that the AI systems may be audited during and after their deployment.[27]

1.8. Explainability: Adopt AI systems that are transparent in terms of how the system operates, how it was developed, its training data, its limitations (including its margin of error), its capabilities, and the purpose of the systems.[28] Explainability refers to making intelligible and providing insight – to deployers and users – into the inputs, outcomes, and functioning of AI systems, making these elements understandable and traceable for humans.[29]

1.9. Accuracy and reliability: Adopt and use accurate AI systems, meaning AI systems that can provide useful and pertinent information and produce correct outputs and predictions,[30] and reliable AI systems, meaning AI systems that work "properly with a range of inputs and in a range of situations".[31]

1.10. **Human oversight:** Do not rely exclusively on AI systems to adopt decisions or automate entire processes that may negatively impact the rights and freedoms of individuals or communities, and maintain an appropriate level of human control and involvement concerning all AI systems.[32]

1.11. **Human-centric design:** The development, deployment, and use of AI systems should follow human-centric design principles to complement and augment the judiciary's capacities and respect human dignity and autonomy.[33]

1.12. **Responsibility**: Organizations that deploy AI systems and individuals that use AI systems must assume responsibility for the decisions and actions taken with the support of AI tools without prejudice to the provider's potential liability in case the AI system is defective.[34]

1.13. **Multi-stakeholder governance and collaboration**: The organizations that are part of the judiciary should consult diverse stakeholders throughout the AI system's life cycle, especially those that may be directly or indirectly affected by its deployment. Allow meaningful participation of marginalized groups and incorporate their feedback in the development and use of AI tools used to make judgments or with the potential to impact any other significant legal issue.[35]

SECTION 2: SPECIFIC GUIDANCE FOR ORGANIZATIONS THAT ARE PART OF THE JUDICIARY

Below are the guiding questions for this section. Please provide your answers or comments in the <u>survey form</u>.

1. **Are there any other rules or standards that organizations should follow with regard to the development, procurement, and deployment of AI systems?**

2. **If yes, what are they? Please provide reasons for your response.**

3. **Should there be specific rules or standards for generative AI systems? If so, what are your suggestions?**

4. **Do you have any other comments on the specific guidance below?**

2. Specific guidance for organizations that are part of the judiciary

These recommendations apply to the bodies that govern the judiciary, courts, and tribunals that intend to adopt and use AI systems.

2.1. On the adoption of AI tools:

2.1.1. **Conduct algorithmic impact assessments before AI systems are deployed, including human rights and ethical assessments.** AI systems that support decision-making processes should be examined through algorithmic impact assessments that identify implications for access to justice and individual rights and potential risks. Such reports may inform whether the systems should be deployed and recommend the appropriate risk prevention, mitigation, redressal, and monitoring measures.[36] The algorithmic assessments should be available for public consultation. Algorithmic impact assessment guidelines and instruments, such as UNESCO's "Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence", may be used to examine potential impacts on human rights such as fair trial and due process, access to justice and effective remedy, privacy, and data protection, equality before the law, and non-discrimination.[37] These assessments should be done in consultation with the relevant stakeholder groups. Finally, other emerging frameworks could be considered for conducting AI algorithmic impact assessments (including some focusing on human rights).[38]

2.1.2. **Data transparency, quality, and integrity**. Adopt AI systems that offer greater transparency of their training data and allow deployers and users to assess the quality and integrity of such data.[39]

2.1.3. **Necessity, proportionality, suitability, and alignment**. The necessity, proportionality, and appropriateness of using an AI system to perform tasks should be assessed and established from the outset and aligned with the objectives of the organization using the AI system (e.g., that protects human rights and ensures the judiciary's independence and autonomy).

2.1.4. **Consult with relevant stakeholders**. Before adopting AI systems that are used for decision-making or tools that support management processes and that can significantly affect fundamental rights, such as access to justice, the organization should consult with relevant stakeholders about the opportunities, potential risks, and impacts of introducing such tools.

2.1.5. Compliance with the law and compatibility with organizational standards. Adopt systems developed and offered in compliance with general principles of international human rights law and local laws, such as data protection and privacy, antidiscrimination, procedural fairness, consumer, competition, and intellectual property laws. Furthermore, when an external provider supplies the AI system, ensure the tool's terms of use are compatible with local laws and organizational standards of the organizations that govern the judiciary.

2.1.6. Access to key information about AI systems. Obtain information from the AI system's developers and providers about the tool's adequate uses, the requirements for maintaining its optimal operation, and the limits and risks associated with its adoption. This point may require developers and providers to disclose sufficient information about how the model was developed (including the data used to train it) and how it works. Different transparency instruments, such as model cards, may be considered for providing access to such information.[40] Furthermore, the organizations that are part of the judiciary should adopt explainable AI tools, which refer to systems with the "ability to explain their rationale for decisions, characterize the strengths and weaknesses of their decision-making process, and convey an understanding of how they will behave in the future."[41]

2.1.7. Ensure the viability of the algorithmic audits conducted while AI systems are deployed. In the acquisition process and during the implementation process, ensure the developer or provider of the AI system agrees to allow and collaborate with algorithmic audits carried out or commissioned by the organization to external parties (e.g., firms specialized in conducting algorithmic audits that are contracted by the organization of the judiciary). If the AI system was developed internally, ensure the development of the internal capacity to monitor and audit the tool (with external guidance if needed), for example, by establishing "mechanisms that facilitate the system's suitability, such as ensuring traceability and logging of the AI system's processes and outcomes."[42] Furthermore, "establish processes for third parties (e.g., suppliers, consumers, distributors/vendors) or workers to report potential vulnerabilities, risks or biases in the AI system."[43]

2.1.8. Proactively disclose key information about the AI systems used by the judiciary. Publish information about the adopted AI systems, how they operate, how they are used, and the implications when such tools are deployed to automate decisions or support decision-making processes. For example, the bodies that

govern the judiciary can publish an online repository with key information about the AI systems adopted to administer justice. The repository can include data on how the AI systems work, who uses the tools, how they are used, and the implications for judicial decision-making processes.[44] The information on the repository should be updated periodically, at least on a yearly basis, and it should be clear when the last time that the repository was updated.

2.1.9. Conduct algorithmic impact evaluations of AI systems that have been deployed. Evaluate the effects of the deployed AI systems on users, affected populations, and society. The latter should include evaluating the impact on human rights and identifying the system's main effects (intended and not intended) over diverse user groups and populations.[45]

2.2.   On internal procedures and standards:

2.2.1. Human intervention. Ensure that the system allows for human intervention. In other words, human control and supervision, or at least monitoring, should occur during all the implementation stages and usage of the AI system.

2.2.2. Risk management systems. Regarding the AI systems that are used in decision-making processes, establish a risk management system that enables the organization to identify, detect, monitor, classify, diagnose, control, and prevent risks as well as mitigate harm. The management system should assign clear roles, responsibilities, and procedures along the life cycle of the AI system. In case of defective AI systems or systems that appear to be producing negative effects, the deployment should be suspended while an investigation is conducted.

2.2.3. Cybersecurity-enhancing measures. Adopt technical, managerial, and human measures to prevent, control, and mitigate cybersecurity risks and incidents. When the deployment and use of the AI system entails accessing cloud services, ensure that the level of protection offered by the provider matches local legal standards as well as organizational security standards.

2.2.4. Data governance. Establish a robust data governance framework and infrastructures to protect personal data and promote responsible data-sharing practices for personal and non-personal data. The organization should also establish "data protocols governing data access. These protocols should outline who can access data and under which circumstances. Only duly qualified personnel with the competence and need to access individual's data should be allowed to do so"[46].

2.2.5. Publish impact evaluations **and performance reports**. Carry out continuous and regular risk assessments of AI systems and publish periodic reports on the impacts and performance (e.g. effectiveness and efficiency) of the adopted AI systems in meeting the organization's objectives.

2.2.6. Enhanced privacy protections. Given the sensitive nature of personal and legal data that is handled in judicial processes, it is essential to have stringent privacy protections in place. (i) Data minimization. To mitigate the risk of data breaches, AI systems that require minimal personal data to function, particularly in cases involving sensitive personal information should be used. Hence, the organizations of the judiciary should deploy AI systems that are based on privacy enhancing processes. (ii) Consent protocols. Pursuant with local data protection and privacy laws, draft and implement consent protocols for treating personal data with AI tools within the judiciary. This will ensure that parties are informed and can control how their data is used. (iii) Data anonymization. Anonymization techniques should be employed when AI systems process personal data, especially in creating databases used for legal analytics or precedent study (e.g., anonymizing the parties' personal data from court decisions before the rulings are published).[47] However, this should be balanced with the need for having open datasets as well as the right to freedom of expression and access to information so as not to impede the development of digital justice systems. How these rights are balanced in practice depends on how each jurisdiction regulates the right to access courts documents and anonymization rules derived from data protection and privacy laws.

2.3. On human talent:

2.3.1. Provide access to training. Offer training opportunities to the members of the judiciary who will use the AI tools and ensure that they develop core competencies of AI literacy, particularly to identify the suitability of AI tools that can be used to carry out different types of tasks, understand how to use these tools and their potential impacts and risks in relation to the international and regional human rights law and national fundamental rights law, and assess their outputs critically.[48]

2.3.2. Adopt organization-specific guidelines. Adopt context and organizational-specific guidelines on which AI systems can be used, under what conditions, and the accepted and prohibited uses for specific processes and tasks. Moreover, the guidelines should include protocols for incident reporting and clear information on

how each actor is accountable within the organization. The guidelines should ensure members of the judiciary acknowledge and take responsibility for any materials they produce using AI.

2.4. On the use of generative AI systems:

2.4.1. Awareness of functions and limitations. Ensure that the members of the judiciary who will use the generative AI tools are aware of the tool's adequate uses, limitations and the risks associated (e.g. biased, incorrect outputs) with their adoption for the drafting of legal documents and supporting participation in legal proceedings, as well as supporting court management activities (e.g. translation).

2.4.2. Content authenticity and integrity. Ensure the authenticity and integrity of content in the judiciary. To achieve this, several measures should be taken. (i) All AI-generated legal documents, evidence presentations, or judicial opinions should be clearly labeled as AI-assisted, so that judges, lawyers, and parties involved are aware of the nature of the content. (ii) Robust systems should be implemented to track the development and modifications of AI-generated legal content. This is important for evidentiary purposes, as it ensures that all materials used in court can be verified for authenticity. For instance, if an AI tool has been used to create or modify a legal document, there should be a clear record of when and how it was used. (iii) Certification protocols should be developed for AI tools used in the judiciary, verifying that they meet international and local ethical guidelines as well as each jurisdiction's legal standards for accuracy and reliability. An example of this could be a certification process for an AI tool that is used to analyze evidence, ensuring that it is accurate and reliable enough to be used in a court of law.

2.4.3. Usage restrictions. Specific guidelines should be issued to govern the use of generative AI in the judiciary to prevent misuse and protect the integrity of the legal process. Certain applications of AI should be banned or limited considering their impact on human rights. For example, when the terms of use of a generative AI tool indicate that the user's prompts will be used by the provider to train its models or that third parties can access these prompts, then the use of such tool should be prohibited or restricted to prevent that the judiciary loses control of who can access confidential information or personal data. Moreover, for instance, the use of AI in certain sensitive areas, such as the unilateral generation of binding legal decisions or the creation of fabricated evidentiary material, should be prohibited. Special consideration should

be given to intellectual property rights' compliance by ensuring that any AI-generated content respects the intellectual property of the original text.

SECTION 3: SPECIFIC GUIDANCE FOR INDIVIDUAL MEMBERS OF THE JUDICIARY

Below are the guiding questions for this section. Please provide your answers or comments in the <u>survey form</u>.

1. **Are there any other rules or standards that individual members of the judiciary should follow with regard to the use of AI systems?**
2. **If so, which ones? Please provide reasons for your response.**
3. **Should there be specific rules or standards for the use of generative AI systems by individuals?**
4. **If yes, what are they? Please provide reasons for your response.**
5. **Any other comment for the specific guidances for individual members?**

3. Specific guidance for individual members of the judiciary

These recommendations are applicable to individuals that are part of the judiciary, including magistrates, judges, justices, judicial officers, and judicial support staff.

3.1. Preparation for the use of AI tools:

3.1.1. AI awareness, AI literacy and capacity building. Individuals should be aware of the functionalities, strengths, and accuracy of AI systems and their limitations, biases, and risks in the context of legal activities. The latter also includes awareness of the liabilities arising from these tools' negligent use. The organizations of the judiciary should ensure that individuals have access to education programs and ongoing training to develop critical AI literacy skills. These skills should allow individuals to understand AI fundamentals, identify the functionalities and purposes of diverse AI systems (particularly those deployed in their organization), understand the limits and risks of these tools, learn how to use the AI systems ethically and responsibly, and critically assess the outputs and performance of the tools.[49]

3.1.2. Use AI tools that have been tested through algorithmic impact assessments. As explained in a previous section, the organizations that are part of the judiciary should test AI systems through algorithmic impact assessments, including human

rights and ethical assessments. The assessment of AI tools that are meant to support decision-making processes is needed to identify implications for access to justice and individual rights and potential risks. However, if your organization does not assess its AI systems, ensure that you use tested tools (by the provider or expert third parties) and select those that have been the subject of human rights impact assessments.[50]

3.2.   On the use of AI tools:

3.2.1. Avoid overreliance on AI systems to make substantive decisions. Do not rely entirely on AI systems to make decisions on the merits of a case nor to decide on procedural issues that may affect human rights; instead, use AI tools' outputs to complement the legal analysis undertaken with other methods and sources of information.

3.2.2. Aligned use. If you are allowed to use AI systems that have not been formally vetted by your organization, make sure that the tool reflects intended values and goals of the administration of justice. Ensure that using these AI tools does not jeopardize the organization's objectives and the human rights of any individual affected by the tool.

3.2.3. Adhere to the terms of use. Read and comply with the AI systems' terms of use (published by the external supplier of the tool or provided by your organization when the system was developed internally). These terms usually indicate the adequate uses of the tools, the uses that are not allowed or prohibited, and the risks that should be avoided by deployers and users. However, consult with your organization about the terms of use of AI systems, especially if you consider that some terms may unduly restrict users' rights, deny transparency or absolve developer of any liability.

3.2.4. Transparency. Provide meaningful information on when the AI tool is used and how its use could affect individuals who are part of a judicial proceeding or beneficiaries of the legal work. Inform what tools were used and their versions. Ensure that materials produced or published using AI are acknowledged and distinguished using quotation marks or citations.

3.2.5. Responsibility and accountability. Take responsibility for using AI tools in legal tasks and adhere to standards for accountability in the use of AI developed by your organization. This recommendation implies both an *ex-ante* duty to disclose the use of an AI system and an *ex-post* duty to provide further information when required to determine responsibility.

3.2.6. **Opportunity to review decisions and contestability**. Provide parties, interested parties, or clients with an adequate opportunity to challenge and contest decisions taken with or supported by AI systems (e.g., decisions that are informed by the outputs produced by AI **predictive** tools), as well as basic information of how the AI system works, how it was trained, what inputs were used to operate the system, and how the outputs produced by the AI tool informed the decision.

3.2.7. **Proactive reports for preventing harm**. Inform the organization when there is suspicion of malfunctioning or potential or likely negative impacts and stop using the AI system if you notice that it creates potential harms on human rights.

3.3. **On the use of generative AI systems:**

3.3.1. **Protect personal and confidential data**. Do not include personal data or confidential information in prompts when using external generative AI tools.[51] Be aware that any information that you input into a public AI chatbot (e.g., as part of the prompt) should be treated as being made available to everyone, among others, because the terms of service of most of the AI companies that provide free access to these AI systems indicate that the inputs will be used to train future models.[52]

3.3.2. **Main uses for LLMs.** LLMs may be used for different tasks including, but not limited to, drafting basic legal documents, writing speeches and presentations, summarizing, translating, making grammatical corrections, modifying the tone of a text (e.g., informal to formal), improving its readability, exploring specific topics, and carrying out administrative tasks (e.g., drafting emails).[53] However, all the previous tasks involve verifying the AI system's output and cross-checking with reliable sources. AI tools can assist in locating material that you are familiar with and can evaluate on your own. On the contrary, these tools are ineffective means of researching to find information you cannot assess independently.[54]

3.3.3. **Unreliability as search engines and for legal analysis.** Commercial general-purpose LLMs are not reliable sources of information or adequate means for conducting legal analysis or carrying out mathematical tasks. "Even with the best prompts, the output may be inaccurate, incomplete, misleading, or biased."[55]

3.3.4. **Awareness of LLMs' limitations and risks**. Be aware that the outputs generated by current LLMs may include incorrect, imprecise, or fictitious information about factual, legal (laws and case law), and technical issues. LLMs can produce answers

that lack any reference to the real world and nonsensical text.[56] Moreover, members of the judiciary should be conscious of potential bias in the text produced by LLMs: "Imbalances in the training data can also lead to biases in the model."[57] Additionally, certain LLMs may have been trained on data up to a certain point in time (e.g., data available up to the point when they were last trained). Therefore, information about recent legal cases, laws, and facts may not have been included in the training data; thus, the output produced by the LLM-powered chatbot may be outdated or inaccurate.

3.3.5. **Verify outputs before using them**. The convincing structure of the text produced by an LLM should not lead to excessive trust in the factuality and veracity of the output.[58] "For various reasons, LLMs offer no guarantees regarding the factuality, quality, and desired formatting (e.g., specific code format) of their outputs."[59] Therefore, avoid excessive trust in AI tools by always ensuring that the output generated by generative AI systems is verified and contrasted with reliable sources of information before its use in legal materials and documents.

3.3.6. **Transparent use.** Disclose the use of generative AI systems for drafting text – rulings, opinions, and other documents that may have legal consequences – or when it is explicitly used in court hearings. For that purpose, distinguish the text produced by the AI chatbot used in a decision by employing quotation marks and a citation system.[60]

3.3.7. **Integrity**. Do not attribute synthetic text as your own. Additionally, the members of the judiciary should prevent potential infringements of copyright and intellectual property rights associated with the use of content produced by generative AI systems.

3.3.8. **Responsibility.** Take responsibility for using the output produced by generative AI systems to draft rulings and judicial decisions, as well as to inform your participation in court hearings.

<center>***</center>

End Notes

_____

[1] For a database that overviews over a hundred AI tools used by judicial operators around the World, see CEPEJ, 'Resource Centre Cyberjustice and AI' (European Commission for the efficiency of justice (CEPEJ) 2024) <https://public.tableau.com/app/profile/cepej/viz/ResourceCentreCyberjusticeandAI/AITOOLSINITIATIVESREPORT?publish=yes>. For a repository of AI tools and projects developed by Brazil's CNJ, see CNJ, 'Repositório Nacional de Projetos de Software e Versionamento de Arquivos do – Git.jus' (Conselho Nacional de Justiça (CNJ)) <https://www.cnj.jus.br/sistemas/git-jus/> accessed 27 May 2024. For an overview of different tasks carried out with AI systems in the justice sector, see F Bell and others, 'AI Decision-Makingand the Courts: A Guide for Judges, Tribunal Members and Court Administrators' (Australasian Institute of Judicial Administration 2022); UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (The United Nations Educational, Scientific and Cultural Organization (UNESCO) 2023) CI/DIT/2023/AIRoL/01 <https://unesdoc.unesco.org/ark:/48223/pf0000387331>. For reports and papers on specific AI tools used by judicial operators, see Nikolaos Aletras and others, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 PeerJ Computer Science 1; Urvashi Aneja and Dona Mathew, 'Artificial Intelligence in India's Judicial System: A Case of Organised Irresponsibility?' (Digital Futures Lab 2023) <https://www.responsibletech.in/post/smart-automation-and-artificial-intelligence-in-indias-judicial-system-a-case-of-organised-irresponsibility>; Elsa Estevez, Sebastián Linares and Pablo Fillottrani, 'PROMETEA: Transformando La Administración de Justicia Con Herramientas de Inteligencia Artificial' (Banco Interamericano de Desarrollo 2020) <http://dx.doi.org/10.18235/0002378>; Collen Zvandasara Kufakwababa, 'Artificial Intelligence Tools in Legal Work Automation: The Use and Perception of Tools for Document Discovery and Privilege Classification Processes in Southern African Legal Firms' (Stellenbosch University 2021) <http://hdl.handle.net/10019.1/109893>; Goretty Carolina Martinez, 'La Inteligencia Artificial y Su Aplicación al Campo Del Derecho' [2012] Alegatos 827; Masha Medvedeva, Michel Vols and Martijn Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights' (2020) 28 Artificial Intelligence and Law 237; Ministro do Superior Tribunal de Justiça, _Artificial Intelligence: Technology Applied to Conflict Resolution in the Brazilian Judiciary_ (Fundação Getulio Vargas 2020) <https://bdjur.stj.jus.br/jspui/handle/2011/156490>; Marion Oswald and others, 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality' (2018) 27 Information & Communications Technology Law 223; Víctor Saavedra and Juan Carlos Upegui, 'PretorIA y La Automatización Del Procesamiento de Causas de Derechos

Humanos' (Derechos Digitales América Latina y Dejusticia 2021) <https://www.derechosdigitales.org/wp-content/uploads/CPC_informe_Colombia.pdf>.

[2] Nathan-Ross Adams, 'Parker v Forsyth NO: Lessons for Using AI for Legal Research' (*Michalsons*, 11 July 2023) <https://www.michalsons.com/blog/parker-v-forsyth-no-lessons-for-using-ai-for-legal-research/66884>; Robert J Ambrogi, 'Four Months After Launching Its "Homegrown" GenAI Tool, Law Firm Gunderson Dettmer Reports On Results So Far, New Features, And A Surprise on Cost' (*Lawsites*, 20 December 2023) <https://www.lawnext.com/2023/12/four-months-after-launching-its-homegrown-genai-tool-law-firm-gunderson-dettmer-reports-on-results-so-far-new-features-and-a-surprise-on-cost.html>; Francesca Bennetts and others, 'Responsible AI: Navigating the Risks and Embracing the Possibilities' (*Allen Overy*, 11 December 2023) <https://www.allenovery.com/en-gb/global/news-and-insights/publications/responsible-ai-navigating-the-risks-and-embracing-the-possibilities>; Juan David Gutiérrez, 'Judges and Magistrates in Peru and Mexico Have ChatGPT Fever' (*Tech Policy Press*, 19 April 2023) <https://techpolicy.press/judges-and-magistrates-in-peru-and-mexico-have-chatgpt-fever/> accessed 16 October 2023; Juan David Gutiérrez, 'ChatGPT in Colombian Courts: Why We Need to Have a Conversation about the Digital Literacy of the Judiciary' (*VerfBlog*, 23 February 2023) <https://verfassungsblog.de/colombian-chatgpt/>; Juan David Gutiérrez, 'AI Technologies in the Judiciary: Critical Appraisal of Large Language Models in Judicial Decision-Making' in Regine Paul, Emma Carmel and Jennifer Cobbe (eds), *Handbook on Public Policy and AI* (Edward Elgar Publishing 2014) <https://papers.ssrn.com/abstract=4667572>; LexisNexis, 'LexisNexis Launches Lexis+ AI, a Generative AI Solution with Linked Hallucination-Free Legal Citations' (*LexisNexis*, 25 October 2023) <https://www.lexisnexis.com/community/pressroom/b/news/posts/lexisnexis-launches-lexis-ai-a-generative-ai-solution-with-hallucination-free-linked-legal-citations>; Benjamin Weiser, 'Here's What Happens When Your Lawyer Uses ChatGPT' *The New York Times* (27 May 2023) <https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>; Benjamin Weiser and Jonah E Bromwich, 'Michael Cohen Used Fake Cases Cited by A.I. to Seek an End to Court Supervision' *The New York Times* (29 December 2023) <https://www.nytimes.com/2023/12/29/nyregion/michael-cohen-ai-fake-cases.html?smid=nytcore-ios-share&referringSource=articleShare>; Mark Witten, 'Applying Generative AI to Law: Opportunities and Risks | Queen's Law' (*Queen's University*, 23 October 2023) <https://law.queensu.ca/news/Applying-generative-AI-to-law>; Kieran Woboditsch-Velasco, 'The Quiet Revolution: Generative AI's Rise in Canadian Law Firms' (*MyOpenCourt*, 25 November 2023) <https://myopencourt.org/the-quiet-revolution-generative-ais-rise-in-canadian-law-firms/> accessed 5 January 2024.

[3] For guidelines issued by the bodies that govern the judiciary and courts, see: UK Courts and Tribunals Judiciary, 'Artificial Intelligence (AI) - Guidance for Judicial Office Holders' (2023) <https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf>; Conselho Nacional de Justiça, 'Dispõe Sobre a Ética, a Transparência e a Governança Na Produção e No Uso de Inteligência Artificial No Poder Judiciário e Dá Outras Providências' (Conselho Nacional de Justiça (CNJ) 2020) Resolução Nº 332 <https://atos.cnj.jus.br/atos/detalhar/3429>; Courts of New Zealand, 'Guidelines for Use of Generative Artificial Intelligence in Courts and Tribunals' (2023) <https://www.courtsofnz.govt.nz/going-to-court/practice-directions/practice-guidelines/all-benches/guidelines-for-use-of-generative-artificial-intelligence-in-courts-and-tribunals/>; Federal Court, 'Interim Principles and Guidelines on the Court's Use of Artificial Intelligence' (Federal Court 2023) <https://www.fct-cf.gc.ca/en/pages/law-and-practice/artificial-intelligence>; Federal Court, 'Notice to the Parties and the Profession: The Use of Artificial Intelligence in Court Proceedings' (Federal Court 2024) <https://www.fct-cf.gc.ca/Content/assets/pdf/base/FC-Updated-AI-Notice-EN.pdf>; Queensland Courts, 'The Use of Generative Artificial Intelligence (AI) Guidelines for Responsible Use by Non-Lawyers' (2024) <https://www.courts.qld.gov.au/going-to-court/using-generative-ai>; Supreme Court of Victoria, 'Guidelines for Litigants: Responsible Use of Artificial Intelligence in Litigation' (Supreme Court of Victoria 2024) <https://www.supremecourt.vic.gov.au/forms-fees-and-services/forms-templates-and-guidelines/guideline-responsible-use-of-ai-in-litigation>. It is worth noting that the Courts of New Zealand published three separate guidelines for judges, judicial officers, tribunal members and judicial support staff; lawyers; and, non-lawyers, respectively. Similarly, the Federal Court of Canada issued "Principles and Guidelines on the Court's Use of Artificial Intelligence" and a Notice to parties, self-represented litigants, and interveners on the use of AI in Court proceedings. For guidelines issued by other public bodies see: CEPEJ, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' (European Commission for the efficiency of justice (CEPEJ) 2019) Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018).; CEPEJ-GT-CYBERJUST, 'Use of Generative Artificial Intelligence (AI) by Judicial Professionals in a Workrelated Context' (European Commission for the efficiency of justice (CEPEJ) CEPEJ Working group on Cyberjustice and Artificial Intelligence (CEPEJ-GT-CYBERJUST) 2024) Information Note <https://rm.coe.int/cepej-gt-cyberjust-2023-5final-en-note-on-generative-ai/1680ae8e01>.

[4] Point 8(a) of Annex III of the European Union's AI Act.

[5] See sections 2 - 5 of the Chapter III of the European Union's AI Act.

[6] Kalliopi Terzidou, 'The Use of Artificial Intelligence in the Judiciary and Its Compliance with the Right to a Fair Trial' (2022) 31 Journal of Judicial Administration 154; Kalliopi Terzidou, 'Automated

Anonymization of Court Decisions: Facilitating the Publication of Court Decisions through Algorithmic Systems', *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law* (Association for Computing Machinery 2023) <https://doi.org/10.1145/3594536.3595151>.

[7] For a framework of different types and classifications of AI tools, see OECD, 'OECD Framework for the Classification of AI Systems' <https://www.oecd-ilibrary.org/content/paper/cb6d9eca-en>.

[8] For guidelines issued by universities, law societies, bar associations, and professional associations, see: Bell and others (n 1); James E Baker, Laurie Hobart and Mathew Mittelsteadt, 'AI for Judges: A Framework' (Center for Security and Emerging Technology 2021) <https://doi.org/10.51593/20190019>; Juan G Corvalán and Mariana Sánchez Caparrós, 'Guía de Directrices Para El Uso de ChatGPT e IA Generativa de Texto En La Justicia' (UBA IALAB 2023) <https://ialab.com.ar/wp-content/uploads/2023/11/Guia-de-directrices-usos-de-ChatGPT-e-IA-generativa-en-la-justicia.pdf>; Fédération des Barreaux d'Europe, 'European Lawyers in the Era of ChatGPT: Guidelines on How Lawyers Should Take Advantage of the Opportunities Offered by Large Language Models and Generative AI' (New Technologies Commission of the European Bars Association 2023) <https://www.fbe.org/nt-commission-guidelines-on-generative-ai/>; The State Bar of California, 'Recommendations from Committee on Professional Responsibility and Conduct on Regulation of Use of Generative AI by Licensees' (2023); The Bar Council, 'Considerations When Using ChatGPT and Generative Artificial Intelligence Software Based on Large Language Models' (2024) <https://www.barcouncilethics.co.uk/wp-content/uploads/2024/01/Considerations-when-using-ChatGPT-and-Generative-AI-Software-based-on-large-language-models-January-2024.pdf>; Queensland Law Society, 'Guidance Statement No 37 Artificial Intelligence in Legal Practice' (2024) 37 <https://www.qls.com.au/Guidance-Statements/No-37-Artificial-Intelligence-in-Legal-Practice>; New South Wales Bar Association, 'Issues Arising from the Use of AI Language Models (Including ChatGPT) in Legal Practice' (2023) <https://inbrief.nswbar.asn.au/posts/9e292ee2fc90581f795ff1df0105692d/attachment/NSW%20Bar%20Association%20GPT%20AI%20Language%20Models%20Guidelines.pdf>. The American Association for the Advancement of Science (AAAS) developed materials for judges (papers and podcasts) on AI in the Courts, including recommendations for the use of AI in litigation and legal research, among others. The materials are available here: https://www.aaas.org/ai2/projects/law/judicialpapers.

[9] Bell and others (n 1); Juan David Gutiérrez, 'Retos Éticos de La Inteligencia Artificial En El Proceso Judicial.' in ICDP (ed), *Derecho Procesal. #NuevasTendencias. XLI Congreso Colombiano de Derecho Procesal* (Instituto Colombiano de Derecho Procesal (ICDP) y Universidad Libre 2020) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4011179>; MSI-NET, 'Algorithms and Human

Rights - Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications' (Council of Europe 2018) <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html#:~:text=The%20expert%20study%20examines%20the,law%20principles%20and%20judiciary%20processes.>; OHCHR, 'Taxonomy of Human Rights Risks Connected to Generative AI - Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI' (United Nations Human Rights Office of the High Commissioner (OHCHR) 2023) <https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf>; Terzidou, 'The Use of Artificial Intelligence in the Judiciary and Its Compliance with the Right to a Fair Trial' (n 6); Giulia Gentile, 'AI in the Courtroom and Judicial Independence: An EU Perspective' (*EUIdeas*, 22 August 2022) <https://euideas.eui.eu/2022/08/22/ai-in-the-courtroom-and-judicial-independence-an-eu-perspective/> accessed 26 October 2023.

[10] Emily M Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2021) 616 <https://doi.org/10.1145/3442188.3445922>.

[11] Matthew Dahl and others, 'Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models' <https://doi.org/10.48550/arXiv.2401.01301>; Varun Magesh and others, 'Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools'.

[12] UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' (2022) < https://unesdoc.unesco.org/ark:/48223/pf0000381137>. For an assessment of generative AI tools through the lens of the UNESCO recommendation see: UNESCO, 'Foundation Models Such as ChatGPT through the Prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence' (United Nations Educational, Scientific and Cultural Organization (UNESCO) 2023) <https://unesdoc.unesco.org/ark:/48223/pf0000385629>.

[13] UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (n 1).

[14] UNESCO, 'Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence' (The United Nations Educational, Scientific and Cultural Organization (UNESCO) 2023) <https://unesdoc.unesco.org/ark:/48223/pf0000386276>.

[15] UNESCO, 'UNESCO Global Judges' Initiative: Survey on the Use of AI Systems by Judicial Operators' (The United Nations Educational, Scientific and Cultural Organization (UNESCO) 2024) <https://unesdoc.unesco.org/ark:/48223/pf0000389786>.

[16] For an explanation of what does "respect, protection and promotion of human rights and fundamental freedoms and human dignity" entail in the context of AI technologies, refer to the Preamble and sections III.1 and III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence. For ethical impact evaluation guidance, see UNESCO's "Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence". For an overview of how human rights that may be affected by generative AI, see OHCHR (n 9).

[17] For an explanation of the principle of "fairness and non-discrimination", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

[18] Ibid.

[19] For examples on how AI tools have been use for automating the anonymization of judicial decisions, see UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (n 1); Terzidou, 'Automated Anonymization of Court Decisions: Facilitating the Publication of Court Decisions through Algorithmic Systems' (n 6).

[20] For an explanation of the principle of "privacy and data protection", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

[21] For an explanation of the principle of "proportionality and no harm", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

[22] For an explanation of the principle of "safety and security", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

[23] For a risk analysis published for authorities that aim at incorporating generative AI models into their workflows, as well as possible countermeasures, see BSI, 'Generative AI Models. Opportunities and Risks for Industry and Authorities' (Federal Office for Information Security - Bundesamt für Sicherheit in der Informationstechnik (BSI) 2024) <https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4>. For more information about access to information laws, see UNESCO's webpage: https://www.unesco.org/en/access-information-laws.

[24] For an explanation of the principle "Awareness and literacy, refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

25 For an explanation of the principle of "transparency and explainability", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence. For an overview of diverse algorithmic transparency instruments adopted by the public sector worldwide, see GPAI, 'Algorithmic Transparency: A State-of-the-Art Report' (Global Partnership on Artificial Intelligence (GPAI) 2024) <https://www.gpai.ai/projects>.

26 For a thorough discussion on what does "accountability" means in the context of AI governance, see Claudio Novelli, Mariarosaria Taddeo and Luciano Floridi, 'Accountability in Artificial Intelligence: What It Is and How It Works' [2023] AI & SOCIETY <https://doi.org/10.1007/s00146-023-01635-y>.

27 See AI HLEG, 'Ethics Guidelines for Trustworthy AI' (European Commission 2019) AI High-Level Expert Group on Artificial Intelligence (AI HLEG) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419>. For an overview of the scope and process of algorithmic audits, see Ada Lovelace Institute and DataKind UK, 'Examining the Black Box: Tools for Assessing Algorithmic Systems' (2020) <https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/>.

28 For an explanation of the principle of "transparency and explainability", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

29 Ibid.

30 See AI HLEG (n 28) 17.

31 AI HLEG (n 28).

32 For an explanation of the principle of "human oversight", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence. See also ibid.

33 See ibid.

34 For an explanation of the principle of "Responsibility and accountability", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence.

35 For an explanation of the principle of "Multi-stakeholder and adaptative governance and collaboration", refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence. See also, AI HLEG (n 28).

36 For an overview of algorithmic impact assessment tools and processes, see Dillon Reisman and others, 'Algorithmic Impact Assessments Report: A Practical Framework for Public Agency

Accountability' (AI Now Institute 2018) <https://ainowinstitute.org/publication/algorithmic-impact-assessments-report-2>; Ada Lovelace Institute and DataKind UK (n 28).

[37] Further, the potential implication of AI tools for the information ecosystem may require a more specific risk assessment based on UNESCO, 'Guidelines for the Governance of Digital Platforms: Safeguarding Freedom of Expression and Access to Information through a Multistakeholder Approach' (The United Nations Educational, Scientific and Cultural Organization (UNESCO) 2023).

[38] Alessandro Mantelero, 'Human Rights Impact Assessment and AI' in Alessandro Mantelero (ed), *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI* (TMC Asser Press 2022) <https://doi.org/10.1007/978-94-6265-531-7_2>; Alessandro Mantelero and Maria Samantha Esposito, 'An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems' (2021) 41 Computer Law & Security Review 105561; Government of the Netherlands, 'Fundamental Rights and Algorithms Impact Assessment (FRAIA)' (2021) <https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms>; Government of Canada, 'Algorithmic Impact Assessment Tool' (2023) <https://canada-ca.github.io/aia-eia-js/>.

[39] AI HLEG (n 28).

[40] "Model cards are short documents accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions, such as across different cultural, demographic, or phenotypic groups (e.g., race, geographic location, sex, […] skin type […] and intersectional groups (e.g., age and race, or sex and Fitzpatrick skin type) that are relevant to the intended application domains. Model cards also disclose the context in which models are intended to be used, details of the performance evaluation procedures, and other relevant information." Margaret Mitchell and others, 'Model Cards for Model Reporting' (2019). See also: GPAI (n 26); Matías Valderrama, María Paz Hermosilla and Romina Garrido, 'State of the Evidence: Algorithmic Transparency' (Open Government Partnership and GobLab (Universidad Adolfo Ibáñez) 2023) <https://www.opengovpartnership.org/documents/state-of-the-evidence-algorithmic-transparency/>; Open Government Partnership, 'CHAPTER 8: Algorithmic Transparency', *The Skeptic's Guide to Open Government* (2022).

[41] UNESCO, 'Global Toolkit on AI and the Rule of Law for the Judiciary' (n 1) 42.

[42] AI HLEG (n 28) 31.

[43] ibid.

[44] For a review of diverse algorithmic transparency instruments, with an emphasis on repositories or registers of public algorithms, that aim at disclosing information about the AI systems adopted in the public sector, see GPAI (n 26).

[45] For an overview of algorithmic impact evaluations, see Ada Lovelace Institute and DataKind UK (n 28).

[46] AI HLEG (n 28) 17.

[47] See Terzidou, 'Automated Anonymization of Court Decisions: Facilitating the Publication of Court Decisions through Algorithmic Systems' (n 6).

[48] Impact evaluation instruments, such as UNESCO's "Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence" and the human rights impact assessment framework proposed by Mantelero, may be used to evaluate the impacts on human rights of tools that are used by judicial operators. See, Mantelero (n 41).

[49] Juan David Gutiérrez, '¿Están Los Jueces En Capacidad de Usar Modelos de Lenguaje a Gran Escala (LLMs)?' (2023) 7 Revista EXCEJLENCIA 10; David Bawden, 'Origins and Concepts of Digital Literacy' in Colin Lankshear and Michele Knobel (eds), *Digital literacies: concepts, policies and practices* (Peter Lang 2008) <https://search.ebscohost.com/login.aspx?direct=true&db=edshlc&AN=edshlc.011566630.3&site=eds-live> accessed 17 April 2023.

[50] Mantelero (n 41); Mantelero and Esposito (n 41).

[51] For general recommendations on how judicial operators can assess matters of privacy and data protection, see UNESCO, 'Guidelines for Judicial Actors on Privacy and Data Protection' (The United Nations Educational, Scientific and Cultural Organization (UNESCO) 2022) <https://unesdoc.unesco.org/ark:/48223/pf0000381298>.

[52] UK Courts and Tribunals Judiciary (n 3) 3.

[53] See Courts of New Zealand (n 3); UK Courts and Tribunals Judiciary (n 3) 6.

[54] UK Courts and Tribunals Judiciary (n 3) 3.

[55] Courts of New Zealand (n 3) 2.

[56] The *Guidelines* do not allude to the term "hallucination" because this use contributes to anthropomorphizing AI systems, particularly chatbots.

[57] BSI (n 24) 9.

[58] "LLMs generally generate linguistically correct and convincing text and are capable of making statements on a wide variety of topics. This can create the impression of a human-like performance, leading to excessive trust in the statements and the performance of the models (so-called automation bias). For users, this may result in drawing incorrect conclusions from the generated texts or accepting statements without questioning them" ibid 10.

[59] ibid 9.

[60] For example, for guidelines on how to cite the output of AI chatbots with the APA citation system, see Timothy McAdoo, 'How to Cite ChatGPT' (*APA Style*, 2024).