unesco

GUIDELINES FOR THE

USE OF AI SYSTEMS IN COURTS AND TRIBUNALS

Published in 2025 by the United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France

ISBN 978-92-3-100830-6
https://doi.org/10.58338/LIEY8089

The Guidelines were prepared by:
Juan David Gutiérrez, PhD, Associate Professor, University of Los Andes, Colombia.

UNESCO Reviewers:
Cedric Wachholz, Elvis Michel Kenmoe, Guilherme Canela De Souza, Ikran Abdirahman, Jaco Du Toit, Mehdi Benchelah, Prateek Sibal, Kamel El Hilali.

A list of external reviewers and consultations is available in annex (p.31).

Project management and coordination:
Prateek Sibal, Programme Specialist, Digital Policies and Digital Transformation, UNESCO.
Ikran Abdirahman, Consultant, Digital Policies and Digital Transformation, UNESCO.
Gustavo Fonseca Ribeiro, Consultant, Digital Policies and Digital Transformation, UNESCO.
Mei Jie Chen, Trainee, Digital Policies and Digital Transformation, UNESCO.
Dr. Kamel El Hilali, AI & Rule of Law Specialist, Digital Policies and Digital Transformation, UNESCO.

Cover photo - Graphic design - Cover design – Illustrations – Typeset : Abracadabra Estudio de Diseño, Mexico
Printed by: UNESCO

Printed in 2025. Printed in France.

How to cite: UNESCO, Guidelines for the Use of AI Systems in Courts and Tribunals, 2025

Funded by the European Union

The Guidelines have been developed as part of the European Commission funded project “Supporting Member States in Implementing UNESCO’s Recommendation on the Ethics of AI through Innovative Tools.”

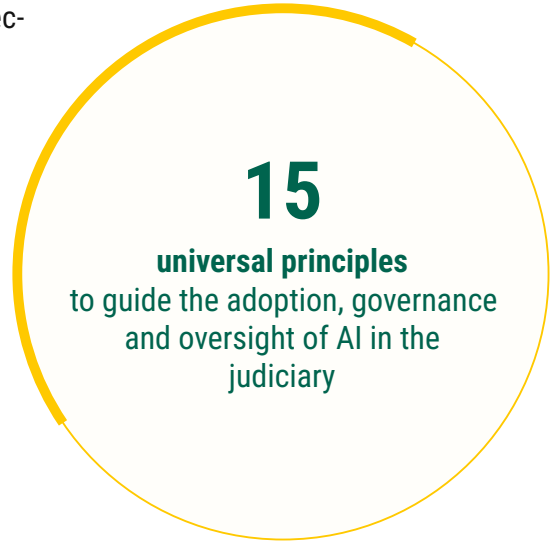
SHORT SUMMARY

ARTIFICIAL INTELLIGENCE IN COURTS: FROM IDEA TO REALITY

Artificial Intelligence is transforming justice systems worldwide, offering new possibilities to improve access to justice, streamline case management, and assist judicial decision-making. Yet, these innovations also bring complex ethical and human rights challenges. UNESCO's Guidelines for the Use of AI in Courts and Tribunals present the first global ethical and operational framework to ensure that AI serves justice while upholding the rule of law and fundamental rights.

Built around fifteen universal principles—from transparency, accountability, and human oversight to human rights protection and multistakeholder governance—the Guidelines provide practical orientation for judges, court administrators, and policymakers exploring AI adoption. They advocate for AI as an assistive, not substitutive, tool—used responsibly and always under meaningful human supervision.

Complemented by a capacity-building toolkit, the Guidelines draw on consultations with experts from over 160 countries, engaging more than 36,000 judicial actors. As AI becomes integral to judicial systems, this publication offers timely guidance to ensure technology strengthens justice, transparency, and trust—constructing the defences of peace in the minds of men and women.



unesco

Since wars begin in the minds of men and women, it is in the minds of men and women that the defenses of peace must be constructed

unesco

GUIDELINES FOR THE

USE OF AI SYSTEMS IN COURTS AND TRIBUNALS

Table of Contents

The Guidelines in One Page

Artificial Intelligence (AI) is increasingly integrated into judicial systems worldwide, creating opportunities to enhance access to justice, improve case management, and support judicial decision-making. Its use, however, also raises significant ethical, legal, and human rights challenges.

To address these, UNESCO has developed the *Guidelines for the Use of AI in Courts and Tribunals*—the first globally oriented ethical and operational framework for AI in judicial contexts. The Guidelines aim to ensure that AI applications serve the cause of justice while safeguarding the rule of law and fundamental rights.

"The Guidelines develop 15 universal principles to guide the adoption, governance and oversight of AI in the judiciary"

They are anchored in fifteen universally applicable principles that guide the adoption, governance, and oversight of AI in the judiciary: the protection of human rights, proportionality, feasibility of benefits, safety, information security, accuracy and reliability, explainability, auditability, transparency, awareness and informed use, responsibility, accountability and contestability, human oversight and decision-making, human centric and participatory design, and multi stakeholder governance. Together, these principles provide practical direction for court leaders, judges, and judicial staff considering AI solutions. AI is envisaged not as a replacement for judicial reasoning, but as an assistive tool to expand access to information, enhance efficiency, and support equitable justice—always under meaningful human supervision and ethical review.

The Guidelines are complemented by a comprehensive capacity building toolkit. Through UNESCO's programme for the judiciary on freedom of expression and the safety of journalists, and the AI & the Rule of Law programme, more than 36,000 judicial actors in over 160 countries have been engaged, strengthening awareness of both AI's potential benefits and its risks. The development of the Guidelines was informed by global stakeholder consultations, ensuring that their recommendations reflect diverse legal traditions, technological capacities, and societal needs.

As AI gains prominence in courtrooms, legal systems stand at a pivotal juncture. Decisions made today on its deployment will shape the administration of justice for decades to come. Whether you are a judicial professional, legal scholar, technology developer, or engaged citizen, the *Guidelines* offer a valuable resource to inform and support the responsible integration of AI into justice systems worldwide.

Introduction

Digital transformation is reshaping how all branches of government operate. Governments worldwide are adopting artificial intelligence (AI) with varying degrees of intensity and focus.[1] For instance, public bodies in Latin America and the Caribbean are implementing or piloting around 700 AI tools to carry out various government functions. The deployment of AI systems can help improve public service delivery and automate administrative processes, enhance citizen engagement, monitor public infrastructure, optimise resource allocation, and support evidence-based policy decisions across education, healthcare, mobility, agriculture, security, and environmental sectors.

The adoption of AI tools in the public sector demands governance frameworks that balance innovation with human rights protection, ensuring technological advancement serves democratic values and contributes to sustainable development.

Courts and tribunals worldwide are increasingly interested in adopting AI tools (Box No. 1) to support a variety of judicial tasks, such as large-scale document search, review and discovery; automated online dispute resolution; prediction of future events (e.g., outcomes of litigations); decision-making and ruling processes; anonymisation and pseudo-anonymisation of court decisions; classification of filed documents; transcription (voice-text) and translation (both voice and text).

Box 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 that translates thousands of documents from English into ten local languages and vice versa.
- In the European Union (EU), law firms and judges use Natural Language Processing tools to predict judicial decisions (for example, from the Court of Justice of the EU).
- 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), CEPEJ (2024), Kufakwababa (2021), Medvedeva et al. (2020), Ministro do Superior Tribunal de Justiça (2020), and UNESCO (2023).

These applications of AI systems in the judiciary have the potential to bring improvements in efficiency, effectiveness, and access to justice. Moreover, the use of trustworthy AI tools could reduce the "noise" in the judicial system, which refers to unwanted variability in judgements that should be identical, as defined by the late Nobel prize winner Daniel Kahneman and his collaborators Olivier Sibony and Cass R. Sunstein.

More recently, judges, judicial support staff, prosecutors, and lawyers around the world have started using chatbots powered by Large Language Models (LLMs) to draft legal documents, judicial decisions, and elaborate arguments in court hearings (Box No. 2). Moreover, these chatbots have been used by judges and tribunals in diverse areas of law: constitutional, administrative, privacy, criminal, and family, among others.

Box 2

USE CASES OF GENERATIVE AI BY JUDGES AND LAWYERS

- In Argentina, judges have used LLMs to summarise their decisions in plain and accessible language.
- Law firms, legal service companies, and universities have developed generative AI systems based on LLMs, independently or with tech companies, to conduct legal research and litigation work, add context to a case, summarise legal texts, and draft emails and contracts. These tools are not limited to general-purpose commercial LLMs; judges and lawyers can also use LLMs designed exclusively to carry out legal activities and other open-source LLMs.
- Nevertheless, reports have surfaced from Australia, Brazil, Canada, South Africa, Spain, and the United States regarding judges and lawyers who have issued judicial decisions or submitted legal documents that contained references to non-existent rulings due to the inappropriate use of AI chatbots.

Sources: Adams (2023), Ambrogi (2023), Benetts et al. (2023), CEPEJ (2024), Gutiérrez (2024), Harden (2024), LexisNexis (2023), Merken (2025), Morcillo (2025), Taylor (2025), Terzidou (2025), Weiser (2023), Weiser and Bromwich (2023), and Witten (2023).

However, formal guidance for individuals or organisations on the adequate use of AI tools in the justice sector is scarce. Only a few judicial authorities have published principles, rules, or guidelines on the ethical and responsible use of AI tools.

These include Argentina (2024), Australia (2024), Brazil (2020, 2025), Canada (2023, 2024), Colombia (2024), New Zealand (2023), Singapore (2024), the United Kingdom of Great Britain and Northern Ireland (2023), and the United States (2024). In addition, several courts have issued rulings establishing principles and rules for the judiciary's use of AI tools (Box No. 3).

Box 3

RULING OF THE COLOMBIAN CONSTITUTIONAL COURT ON THE USE OF AI SYSTEMS BY THE JUDICIARY

- In August 2024, the Constitutional Court of Colombia decided on the constitutionality of a second instance ruling involving a minor's fundamental right to health. The Court examined whether the second-instance judge had breached the fundamental due process right by incorporating text produced with ChatGPT in his ruling. The content created with the chatbot consisted of answers to four legal questions that the judge transcribed as part of the motivation of his ruling.
- In its ruling, the Court cited UNESCO's Recommendation on the Ethics of Artificial Intelligence, described its values and principles, and recognised that its ethical framework "one of the most consulted and widely distributed on the subject of AI worldwide". The ruling also cited UNESCO's Global Toolkit on AI and the Rule of Law for the Judiciary and acknowledged that it "provides valuable information on the international human rights laws, principles, standards and case law relevant to the ethical use of AI, in order to mitigate its potential risks to human rights."
- The Court ruled that the use of AI systems by the judiciary is constitutional as long as the use of the machine does not replace human judgement, especially regarding judicial decision-making. AI systems can be used by the judiciary to support managerial tasks (e.g., notetaking and reports), documentary tasks (e.g., classification of documents), judiciary

functions (e.g., identify pertinent case law), and to correct and synthesize text with the supervision of the user.

- Moreover, the Court ordered that the members of the judiciary who wished to use AI systems to support their judicial activities had to follow "principles of (i) transparency, (ii) accountability, (iii) privacy, (iv) non-substitution of human rationality, (v) seriousness and verification, (vi) risk prevention, (vii) equality and equity, (viii) human control, (ix) ethical regulation, (x) compliance with good practices and collective standards, (xi) continuous monitoring and adaptation, and (xii) suitability." (Source: Colombia, Constitutional Court, Ruling T – 323 / 2024).

- In the aftermath of this ruling, Colombia decided to implement the UNESCO Draft Guidelines on the use of AI in courts and tribunals. Colombia's Guidelines for the Responsible and Safe Use of Generative AI in the Judicial Branch were adopted in December 2024.

New AI laws increase the need for guidance. 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". This classification triggers obligations for the deployers of these AI systems, such as implementing a risk management system and ensuring human oversight of the tools.

Additionally, AI systems integrated into decision-making processes may have cascading effects on the whole judicial system. On the one hand, such impacts may constitute negative externalities and human rights violations. For instance, the deployment might contribute to systematically discriminating against vulnerable or marginalized individuals or groups, or restricting access to the digital ecosystem and relevant devices. On the other hand, AI systems for court management in the judiciary may benefit court users by ensuring the trials take place 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., anonymising court decisions). Hence, new measures at the organisational and individual levels are required to prevent and mitigate adverse effects and improve the administration of justice.

Furthermore, AI systems are a plurality of technologies that can potentially transform and disrupt the practice of law and judicial adjudication. The resulting changes in the justice sector must not jeopardize human rights. Resolution 78/265 adopted by the General Assembly of the United Nations (UN) on 21 March 2024 emphasized that "human rights and fundamental freedoms must be respected, protected and promoted throughout the life cycle of artificial intelligence systems". Therefore, the judiciary, and more generally, UN Member States, should "refrain from or cease the use of artificial intelligence systems that are impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights, especially of those who are in vulnerable situations".

Finally, the use of AI systems should not hinder the competency, independence, and impartiality of the judiciary because, as stated by the recitals of the Bangalore Principles of Judicial Conduct, these attributes are essential for protecting human rights and because "the implementation of all the other rights ultimately depends upon the proper administration of justice."

The Guidelines propose principles and recommendations for adopting, deploying, and using AI tools to support the administration of justice while respecting human rights.

Objectives of the Guidelines

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* propose principles and recommendations for adopting, deploying, and using AI tools to support the administration of justice. The *Guidelines* provide direction at both the organisational (court) and individual levels (judges, judicial officers) (NB: from now on, we will refer collectively to these organisations and individuals as the judiciary).

Although the *Guidelines* are specifically designed to orient the judiciary, they could be pertinent to the legal profession at large. In this sense, most of the principles and recommendations proposed could be adapted for prosecutors, arbitrators, lawyers, bar associations, civil servants, and researchers. In this vein, civil society organisations such as professional associations and universities have recently published similar guidance and recommendations.

Box 4 **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).
- 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.
- Language Models (LMs): 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).

- Lifecycle of AI systems: Cycle that includes the stages of pre-design, design, development, evaluation, testing, deployment, use, sale, procurement, operation and decommissioning (UN General Assembly, resolution 78/265).
- Machine Learning (ML): "a set of techniques that enables machines to learn automatically using patterns and deductions rather than direct instructions from a person" (UNESCO 2023, 21).
- Natural Language Processing (NLP): "a ML technique that analyses vast amounts of human text or speech data (transcribed or acoustic) for specific properties, such as meaning, content, intention, attitude, and context" (UNESCO 2023, 21).
- Organisations of the judiciary: Bodies that govern the judiciary, courts, and tribunals.
- Safe, secure and trustworthy AI systems: AI systems that "are human-centric, reliable, explainable, ethical, inclusive, in full respect, promotion and protection of human rights and international law, privacy preserving, sustainable development oriented, and responsible" (UN General Assembly, resolution 78/265).

Additionally, the *Guidelines* 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, automating tasks, and supporting decision-making processes. Moreover, LLMs can facilitate the interaction between computers and users through natural language.

However, the misuse of AI systems may undermine society's trust in the judicial system. The potential negative impacts require that the judiciary implement accountability frameworks before and after deploying AI systems.

Therefore, the *Guidelines* propose specific guidance to ensure the judiciary's integration of AI systems that are ethical and respectful of human rights.

The Guidelines are a living document that will require periodic updates to address the rapid advancement of AI technology and its growing applications within justice systems.

Methodology

The *Guidelines* are based on UNESCO's extensive work with judicial operators over the past decade. They rely on background documents produced by UNESCO, a public consultation process, then a peer review process of the draft guidelines by legal AI experts worldwide, secondary sources and specialised literature.

First, since 2013, UNESCO has trained over 36,000 judicial operators across 160 countries on diverse topics such as freedom of expression, access to information and on the application and impact of AI justice systems. These judicial actors received training through a "series of Massive Open Online Courses (MOOCs), on-the-ground training and workshops for Supreme Court judges, and the publications of several toolkits and guidelines."

Furthermore, over 5,000 persons participated in the UNESCO global MOOC on AI and the Rule of Law that was offered to strengthen "capacities of judicial operators worldwide to protect human rights, democratic principles, and the rule of law in relation to the use of AI." Additionally, a global survey on the use of AI systems by judicial operators was conducted in late 2023. The results offered insights on how judicial operators access and use AI systems, and highlighted their needs regarding training and guidance.

Overall, this practical experience revealed judges' expectations and perceptions regarding the governance of AI systems in courtrooms, thereby providing substantial input for drafting the Guidelines.

Second, the documents published by UNESCO on AI ethics were also important in laying down the main principles, values and practices included in the Guidelines. This is the case of UNESCO's:

- "Recommendation on the Ethics of Artificial Intelligence"
- "Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence"; and
- The "Global Toolkit on AI and the Rule of Law for the Judiciary", which provided a framework on the applications and risks of AI systems in judicial contexts.

Third, the public consultation of the draft guidelines in August and September 2024 allowed UNESCO to gather over a hundred contributions sent by individuals and organisations from 41 countries, along with collective feedback from events organised with judicial operators across different regions.

Fourth, twenty external reviewers participated in revising the draft guidelines. This diverse global panel included judges, legal scholars, attorneys, and technology specialists from Africa, Asia, Europe, North Africa & Middle East, North and Latin America. Their complementary expertise —spanning courtroom practice, legal technology research, AI governance, and digital transformation of legal systems— ensured the guidelines were critically evaluated from multiple perspectives, enhancing accuracy and practicality.

Fifth, the Guidelines considered the Bangalore Principles of Judicial Conduct, diverse secondary sources and specialised literature, cited in the endnotes, including principles and guidelines published by judiciary authorities, courts, and civil society organisations of diverse jurisdictions, including bar associations and law societies.

Finally, the Guidelines are a living document that will require periodic updates to address the rapid advancement of AI technology and its growing applications within justice systems.

The Guidelines *aim to advance universal respect for justice and the rule of law while ensuring the protection and promotion of human rights and fundamental freedoms.*

Guidelines for the use of AI systems in Courts and Tribunals

The *Guidelines* aim to advance universal respect for justice and the rule of law while ensuring the protection and promotion of human rights and fundamental freedoms.

1. Principles

The Guidelines offer principles and specific recommendations for judicial organisations and individuals implementing AI systems, including generative AI. In this regard, "principles" refer to general prescriptions intended to orient organisations and individuals in developing, acquiring, and using AI systems in an ethical and human rights-respecting manner. The Guidelines develop 15 principles to meet this goal (Figure No. 1).

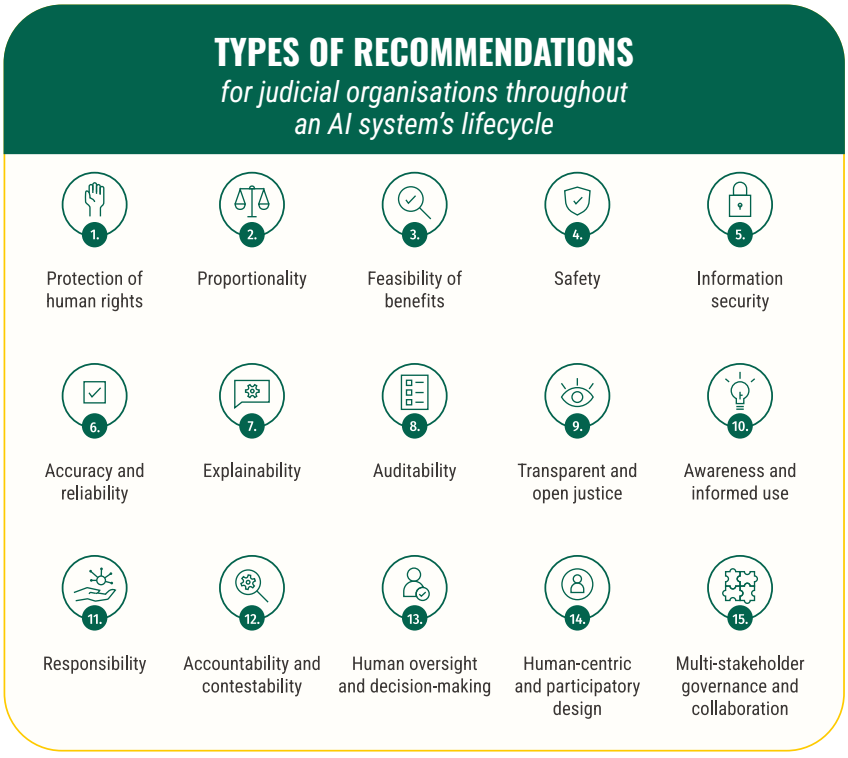


Figure No. 1. Principles for the development, acquisition, adoption, deployment and use of AI systems by the judiciary

Organisations and individual members of the judiciary should follow these principles throughout the AI systems' entire lifecycle:

1.1. Protection of human rights: Take the necessary steps to respect, protect, and promote human rights and the rule of law throughout the AI systems' lifecycle. There must be effective safeguards regarding AI systems that might represent a high risk to human rights, such as implementing the accountability assessments described in section 2 of the Guidelines. The deployment of AI systems should be particularly protective of the rights of women and girls, persons with disabilities, children, refugees, migrants, Stateless persons, minorities, and other populations in situation of marginalization and/or vulnerability. While all human rights may be potentially affected by certain AI systems along the life cycle of these tools, the following four are particularly important in the context of judicial proceedings:

a. **Non-discrimination:** Adopt, deploy, and use AI systems that aim to achieve goals through processes that safeguard fairness and ensure inclusive access to technology. Prevent biased development and applications of AI systems and outcomes that may aggravate, reproduce, reinforce, or perpetuate discrimination based on race, colour, descent, gender, age, language, religion, political opinion, national, ethnic, social or economic condition, or disability and any other grounds.

b. **Equality:** Ensure equal access and treatment for all before the courts regardless of digital capability. For instance, if AI tools mediate access to justice, additional barriers may arise for those without digital resources or skills. Therefore, judicial implementations of AI tools should consider those who cannot afford these technologies, lack digital literacy, or face other technological barriers. For that purpose, following the Leave No One Behind (LNOB) principle developed as part of the UN Agenda for Sustainable Development to address digital exclusion is pertinent.

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

d. **Right to privacy and personal data protection:** Adopt, deploy, and use AI systems that protect privacy and personal data processed for the administration of justice. The judiciary should avoid using AI tools in ways that generate risks of disclosing such data or enable unauthorised access by third parties.

e. **Liberty and security:** Ensure that individuals are not detained based on decisions assisted by AI that the individual is not aware of, is unable to understand how the AI system reached its decision or is unable to challenge or appeal the decision-making process or the decision itself.

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

1.3. Feasibility of benefits: Before adopting AI systems, the judiciary should assess both the specific potential benefits (for the judiciary, users, and the public) and its institutional capacity to realise these benefits.

1.4. Safety: Adopt, deploy, and use AI systems that avoid, address, prevent, mitigate, or eliminate unwanted harm for the parties, individual judges, the courts to which they belong, the judiciary's reputation and authority, and for the public.

1.5. Information security: Adopt, deploy, and use AI systems that protect confidential information in line with international standards for access to information and establish safeguards against cyber threats and ensure the judiciary's continued control over its data.

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

1.7. Explainability: Adopt, deploy, and use AI systems that can explain the rationale behind their outputs and decisions and how they use inputs for such purposes. Explainability refers to making intelligible and providing insight –to deployers and users– into the inputs, reasoning, outcomes, and functioning of AI systems, making these elements understandable, interpretable, and traceable for humans.

1.8. Auditability: Undertake administrative, legal, and human measures to ensure and enable that the AI systems may be audited during deployment.

1.9. Transparent and open justice: Adopt, deploy, and use AI systems that are transparent in terms of how the system was developed, how it operates, its training data, its limitations (including its margin of error), its capabilities, and the purpose of the systems. Inform in a proper and timely manner when and how AI systems are deployed, acquired, and 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. Ensure that individuals are made aware that they may challenge such decisions.

1.10. 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, know the intended purpose for using a specific AI system to carry out judicial activities and stay up to date on technological advancements.

1.11. Responsibility: Organizations and individuals that develop and 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.

1.12. Accountability and contestability: Ensure accountability by informing and explaining how and why certain AI tools were adopted by the judiciary and the results achieved (or not) with their deployment. Additionally, introduce mechanisms that allow affected parties to contest and cross-examine any output produced by an AI system that may influence the outcome of the case at hand, as well as challenge the decisions and hold judicial operators accountable for their errors.

1.13. Human oversight and decision-making: Judges and magistrates should not delegate any part of their mandate or rely exclusively on AI systems to adopt decisions or automate entire processes that may negatively impact the rights of individuals or communities. They should preserve the possibility of human intervention in all decision-making processes where AI systems are deployed. AI tools must not replace the decision-maker's independent analysis of facts, law, and evidence.

1.14. Human-centric and participatory 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. The design process should include members of the judiciary and the public to ensure that the process is inclusive.

1.15. Multi-stakeholder governance and collaboration: Judicial organisations should consult diverse stakeholders, hold public consultation and implement citizen-participation approaches throughout the AI system's life cycle, especially those that may be directly or indirectly affected by its deployment, such as women, persons with disabilities, juveniles, and children. 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.

2. Specific guidance for judicial organisations

These Guidelines define "recommendations" as suggested pathways of action that stem from the "principles" that help both organisations acquire and develop safe, secure and trustworthy AI systems, and individuals use AI systems in an ethical, responsible, and human rights-respecting manner.

This section's recommendations are divided into three categories (Figure No. 2) and are directed to the judiciary authorities, courts, and tribunals that intend to assess, develop, adopt, use and evaluate AI systems.

Figure No. 2. Types of recommendations for judicial organisations throughout an AI system's lifecycle.

2.1. On the development, acquisition, use, and evaluation of AI systems

The following recommendations provide guidance on actions that should be taken by judicial organisations and authorities in different stages of the AI lifecycle.

2.1.1. Protect the independence and impartiality of the Judiciary. Ensure that the judiciary has an oversight role and responsibility in the development, deployment and use of AI tools and that there are regulations to protect its independence. AI tools should not essentially alter or annul the fundamental features of the impartiality of judges and tribunals.

2.1.2. Consider cultural and contextual specificities of the population affected by the deployment of AI systems and consider the barriers to access justice of individuals without the knowledge or skills to interact with the legal system, e.g., older, underrepresented, and underserved populations.

2.1.3. Establish multi-stakeholder governance of AI tools. Draft an AI policy or strategy that diagnoses the main barriers for the effective adoption of AI systems, establishes objectives and metrics, and formulates action plans. Furthermore, explore the establishment of a multi-stakeholder advisory board for continuous review and updates of the AI policy or strategy. Individual members of the judiciary should be invited to participate in the continuous improvement of the AI policy and any other instrument adopted by the organisation, for example, by providing feedback based on their experiences with AI systems and the rules and orientations established for their use. Moreover, 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 organisation should consult with relevant stakeholders about the opportunities, potential risks, and impacts of introducing such tools. The judiciary should

proactively involve marginalised communities in the development processes, ensuring their perspectives are considered.

2.1.4. **Adopt algorithmic transparency and explainability measures from the outset**. Start by determining if you should opt for open-source or proprietary systems. Before committing to adopt an AI tool, the judicial organisation should assess the consequences of adopting either option in terms of access to information about the system, explainability, and auditability. Then, consider transparency measures at the early stages of AI development and procurement of AI tools. For example, develop, acquire and obtain information from the AI system's developers and providers about how the model was developed (including the data used to train it) and how it works, the tool's adequate uses, and the limits and risks associated with its adoption.

2.1.5. **Necessity, proportionality, suitability, and alignment**. The necessity, proportionality, and suitability 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).

2.1.6. **Fairness by design**. When the judicial organisation commissions the development or acquisition of AI tools, steps should be taken to ensure that the developers or providers prevent biases in data and in the design of the algorithms. The latter entails that the development and testing teams are multidisciplinary and diverse.

2.1.7. **Phased implementation of AI tools.** Consider the possibility of piloting new AI tools prior to their full-scale deployment and plan a phased implementation. Piloting allows to assess the functionality, impact, and potential risks of AI in a controlled environment. A phased implementation strategy would allow for adjustments based on initial findings and stakeholder feedback, thereby promoting a more informed, responsible, and ethically sound deployment process.

2.1.8. **Sustainability and environmental impact of AI systems**. The development, deployment, and usage of certain AI systems may have a substantial environmental footprint that the judiciary should consider.

2.1.9. **Establish internal rules** and guidelines for the development, deployment and use of AI tools based on the human rights framework. Determine which AI systems can be used, 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 organisation. The guidelines should emphasize that members of the judiciary take responsibility for any materials they produce using AI. For that purpose, the guidelines could include an assessment of potential legal liabilities and the corresponding mechanisms to prevent them. Furthermore, the guidelines should offer detailed scenarios and examples to orient judiciary members on what constitutes negligent use and the steps to be taken when they fail to adhere to established AI usage protocols.

2.1.10. **Ensure human intervention.** Ensure that the system always 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.

2.1.11. **Implement accountability frameworks before deploying AI systems.** Before deploying an AI system to support decision-making processes, define an accountability framework to prevent and mitigate negative impacts and risks and enhance positive ones. The judiciary can implement an "algorithmic impact assessment" to identify the likely effects (positive and negative), an "algorithmic risk assessment" to map the potential risks (such as system failures), or a "fundamental rights impact assessment". Ensure the developer or provider of the AI

system agrees to allow and collaborate with algorithmic audits carried out by the organisation or independent and external parties. If the AI system was developed internally, facilitate the development of 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." The judiciary should consider the role of civil society and independent auditors, particularly in cases where AI tools can generate high risks of undermining fundamental rights. Such reports may inform whether the systems should be deployed and recommend risk prevention, mitigation, redressal, and monitoring measures. The algorithmic assessments should be available for public consultation. 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, and privacy. These assessments should be done in consultation with the relevant stakeholder groups.

2.1.12. Conduct algorithmic audits and impact evaluations after deploying AI systems. To that end, the judiciary should explore the creation of guidelines for monitoring, auditing and evaluating AI systems in the judicial sector. The assessments should assess whether these systems comply with regulations or more targeted issues such as bias, fairness, transparency, explainability, security, or performance. They can also evaluate the effects of the 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. Comprehensive human rights impact assessments should be public. Publish periodic reports on the impacts and performance (e.g., effectiveness and efficiency) of the adopted AI systems in meeting the organisation's objectives. Furthermore, "establish processes for third parties (e.g., suppliers, consumers, distributors/vendors) or workers to report potential vulnerabilities and risks. Additionally, the organisation that governs the judiciary should establish a risk management system that enables the organisation to identify, detect, monitor, classify, diagnose, control, prevent risks and mitigate harm. The management system should assign clear roles, responsibilities, and procedures along the AI lifecycle.

2.1.13. Cease using AI systems that negatively affect human rights. Implement systems developed and offered in compliance with international human rights law, national constitution, and local laws, such as data protection and privacy, non-discrimination, 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 organisational standards of the organisations that govern the judiciary. The judiciary should refrain from adopting AI systems that have the potential to adversely impact human and/or civil rights, e.g., by perpetuating discriminatory outcomes and other injustices. In the same vein, as recommended by the Report of the UN Secretary-General "Human rights in the administration of justice", the judiciary should refrain from or cease from using AI tools "that are impossible to operate in compliance with international human rights law", and these include predictive applications "in judicial decisions in criminal cases and in recidivism risk assessments for bail and parole decisions, unless and until the responsible authorities can demonstrate that such applications are in compliance with the right to a fair trial, including an independent judiciary, and the rights to liberty and security, non-discrimination, freedom from torture and ill-treatment and privacy, as well as other affected human rights, and that the authorities have addressed the disproportionate negative impact that the use of these technologies can have on specific groups." When the judiciary has preliminary evidence that the use of an AI tool poten-

tially produces significant negative effects or human rights violations, the deployment and use of the system should be suspended until a thorough investigation is conducted. When algorithmic assessments or impact evaluations find that specific AI systems violate human rights, the judiciary will discontinue their use and decommission the tool.

2.1.14. Accessibility of AI tools. Ensure that the members of the judiciary who could benefit from the AI tools have adequate access to the systems. The latter includes ensuring that individuals can access the necessary hardware to operate the tools and the skills to use them effectively. For that purpose, it is pertinent to periodically diagnose, monitor, and address the barriers to adopting AI systems, especially regarding the barriers faced by individuals who are part of the judiciary.

2.1.15. Security and cybersecurity-enhancing measures. AI systems are prone to adversarial attacks, which can shift their outcomes. For example, generative AI systems may be vulnerable to prompt injections, data extraction, evasion, and poisoning attacks. Hence, before deploying a tool, the judiciary should test the resiliency of and secure it before deployment and maintain it throughout its lifecycle. Adopt technical, managerial, and human measures to prevent, control, and mitigate cybersecurity risks and incidents. When deploying and using the AI system entails accessing cloud services, ensure that the level of protection offered by the provider matches local legal standards and organisational security standards.

2.1.16. Enhance personal data and privacy protections. Establish a robust data governance framework and infrastructure to protect personal data, promote responsible data-sharing practices for personal and non-personal data and ensure that the data is inclusive. The organization should also establish "data protocols governing data access. These protocols should outline who can access data and under what circumstances. Only duly qualified personnel with the competence and need to access individual's data should be allowed to do so". 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 minimisation. 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 judicial organisations 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 anonymisation 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). 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 court documents and anonymisation rules derived from data protection and privacy laws.

2.1.17. Proactively disclose key information about the AI systems used by the judiciary. For example, judiciary authorities can publish an online repository with key information about the AI systems adopted to administer justice. The repository can include information about the training data, the model (e.g. variables, weights), whether it is proprietary or open source, how the AI systems work, who uses the tools, how they are used, the implications for judicial decision-making processes, and whether the decisions adopted with the support of the system are challengeable. This information should be periodically updated, at least yearly, and should indicate the date of the last update.

Moreover, the creation of public repositories of AI systems used by courts and tribunals could contribute towards the openness of data. For that purpose, assess to what extent the transparency instruments can fulfil the so-called FAIR principles regarding open data: findable, accessible, interoperable and reusable by anyone or for any purpose. UNESCO's guidelines for establishing and implementing open data policies may be considered for this purpose. In parallel to this effort, the judiciary could also undertake public awareness campaigns about how AI systems are used in judicial processes.

2.1.18. **Ensure the contestability of decisions supported or sustained by AI systems**. Implement the necessary measures to guarantee the right to oppose decisions supported or sustained by AI tools. This requires meaningful insights into how the outputs were produced and the right for the parties to audit and examine the underlying AI model to identify shortcomings that may affect their case.

2.2. Human talent

The following recommendations provide guidance on the actions that should be taken by the judiciary to support individuals who will deploy and use AI systems.

2.2.1. **Revise the curriculum of Judicial Schools**. Consider incorporating AI topics into the permanent curriculum of Judicial Schools, such as fundamentals of AI, applications of AI technologies across justice systems, and AI's impact on judges' independence, the administration of justice, particularly concerning human rights, ethics, and governance issues.

2.2.2. **Provide access to training**. Offer regular 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 accordance with international, regional, and national human rights law, and assess their outputs critically. The organisations of the judiciary should consider whether it is feasible and pertinent to establish mandatory training programs for judicial staff to maintain up-to-date knowledge and competency.

2.2.3. **Research on the usage and effects of AI tools**. Judiciary authorities should periodically study how their members are using AI systems. The research could document case studies to showcase good practices and potential pitfalls to avoid. The latter would help individual members of the judiciary to report on their experiences and challenges and share the lessons with their peers. This should be done in full respect of the independence of the judiciary (see supra, 2.1.1).

2.2.4. **Assess the need to update human resources policies and practices**. Depending on the AI tools being used, the judiciary may need to review its workforce planning by either upskilling existing personnel and/or hiring new personnel. This may require updates to the recruitment policies and processes for performance evaluation to ensure that appropriate uses of the AI tools are being evaluated and incentivised.

2.3. Use of generative AI systems

The following recommendations provide guidance on what actions should be taken by judiciary authorities with regards to the adoption, deployment and use of generative AI systems.

2.3.1. **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 or 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).

2.3.2. Ensure content authenticity and integrity. To achieve this, several measures should be taken. (i) All AI-generated legal documents, evidence presentations, or judicial opinions should be clearly labelled 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 and each jurisdiction's legal standards for accuracy and reliability. An example of this could be a certification process for an AI tool used to analyse evidence, ensuring that it is accurate and reliable enough to be used in a court of law.

2.3.3. 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, in line with international standards**.** Specific applications of AI should be banned or limited, considering their impact on human rights, especially privacy and data protection. For example, when the terms of use of a generative AI tool indicate that the provider will use the user's prompts 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 the judiciary from losing 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 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.

3. Specific guidance for individual members of the judiciary

The recommendations of this section are for individuals who are part of the judiciary, including magistrates, judges, justices, prosecutors, judicial officers, and judicial support staff who intend to deploy and use AI systems. The recommendations are divided into three categories (Figure No. 3).

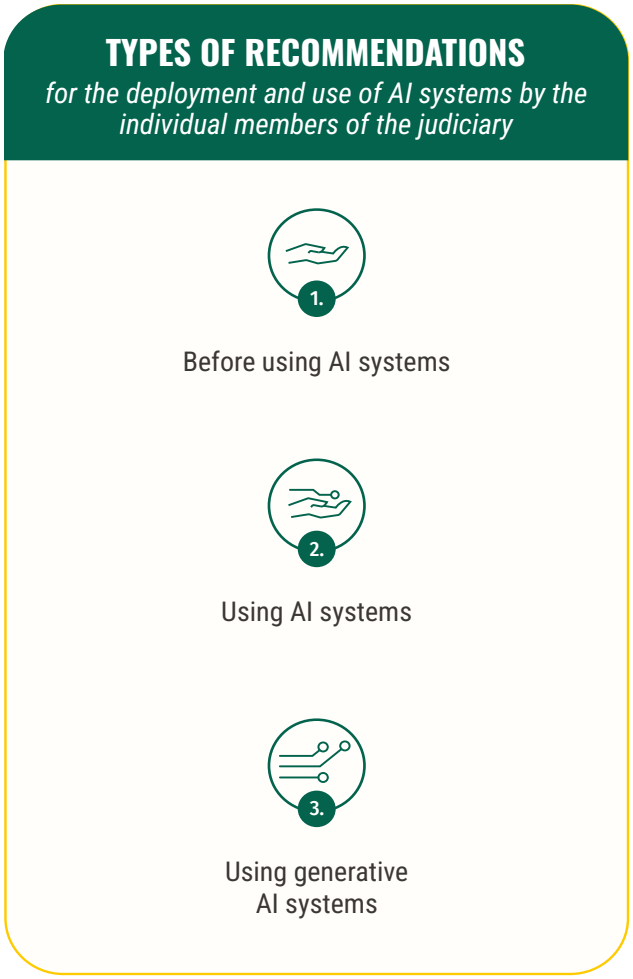


Figure No. 3. Types of recommendations for the deployment and use of AI systems by the individual members of the judiciary.

3.1. Before using AI systems

The following recommendations provide guidance on what actions should be taken by the individuals who are part of the judiciary before deploying and using AI systems.

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. Judiciary organisations should ensure that individuals have access to education programs and ongoing training to critically assess the outputs and performance of the tools.

3.1.2. Use AI tools that have been tested through algorithmic assessments and follow the prescriptions of human rights due diligence reports. As previously explained, judicial organisations should implement accountability frameworks before and after the deployment of AI systems to identify implications and risks for access to justice and human rights. 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. Moreover, follow the prescriptions of human rights due diligence reports of AI systems which should be conducted throughout the life cycle of AI systems by judiciary authorities.

3.2. Using AI systems

The following recommendations provide guidance on what actions should be taken by the individuals who are part of the judiciary during the deployment and use of AI systems.

3.2.1. Avoid overreliance on AI systems to make substantive decisions. Judges and magistrates should 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. The judicial decision is not a mere output, but a decision that is taken in a process.

3.2.2. 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 and prohibited uses of the tools, 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 the developer of any liability.

3.2.3. Proactive 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.

3.2.4. Responsibility and accountability. Adhere to standards 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.

3.2.5. Opportunity to review decisions and contestability. Make parties aware of the use of AI systems and provide adequate opportunities to challenge and contest decisions taken with or supported by AI systems, 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. Ensure that individuals have a right to an effective remedy in circumstances where AI systems have an adverse impact on their human rights, including during and after the completion of criminal proceedings against them. The judiciary should ensure that any individual who wishes to challenge these decisions has access to legal aid services, informational resources, or assistance in navigating the appeal process.

3.2.6. Proactive reports for preventing harm. Inform the organisation 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 harm to human rights.

3.3. Using generative AI systems

The following recommendations provide guidance on what actions should be taken by the individual members of the judiciary with regards to the adoption, deployment and 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. 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.

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, summarising, 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). However, all the above 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, but that is currently unavailable. On the contrary, these tools are ineffective for conducting research of information that cannot be independently assessed or verified. See suggested uses for generative AI tools at the end of this section.

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."

3.3.4. Do not replace expert witness testimonies with LLMs. Expert witnesses' testimony encompasses the processes and methodologies by which the expert reports are produced and the descriptions, reasons, arguments, analyses, evaluations and conclusions laid out in the text. LLMs should not be used to produce the types of outputs that would be required from expert witnesses, not only due to the unreliability of the output but also due to the opacity with which the system infers from its training data.

3.3.5. 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. Moreover, you 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." 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.

3.3.6. 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. "For various reasons, LLMs offer no guarantees regarding the factuality, quality, and desired formatting (e.g., specific code format) of their outputs." Therefore, avoid excessive trust in AI tools by always ensuring that the output of generative AI systems is verified and contrasted with reliable sources of information before its use in legal materials and documents. Human critical thinking must be central in all interactions with generative AI systems.

3.3.7. 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. When text produced by the chatbot is included in a judicial decision, the entire interaction leading to the final result –both the user's prompts and the chatbot's answers– should be documented. Depending on the length and complexity of the interaction, this documentation can be included in the main body of the text, in a footnote, or an annexe.

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

3.3.9. Responsibility. Take responsibility for using the output produced by generative AI systems to draft rulings and judicial decisions and inform your participation in court hearings. Ensure accuracy of the output before using it.

FINALLY, HERE ARE TEN EXAMPLES OF USES FOR GENERATIVE AI SYSTEMS:

1. Correct or improve initial drafts of administrative reports, meeting minutes, e-mails, speaking notes, judicial decisions and rulings. Case-use examples:

    - Improve grammar, spelling and structure of the texts.
    - Editing the text's tone, for example, making it more accessible for different target audiences or for plain language.
    - Preparing draft translations of passages.

2. Reorganise references according to a given source citation system.

3. Comparing texts and sources.

4. Design presentations and creating visualisations.

5. Generation and adjustment of computer code.

6. Making initial explorations of topics and pursuing subsequent search stages from reliable sources.

7. Brainstorming ideas.

8. Structuring of arguments.

9. Explore alternative or critical explanations about points of law.

10. Generate summaries: always verify the accuracy of the output.

AI tools are not a substitute for qualified legal reasoning, human judgment, or tailored legal advice.

Conclusion

As underlined at the beginning of this document, AI tools can support the core objectives of the judicial sector. Still, 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.

The *Guidelines* offer principles and recommendations for maximising the benefits of AI tools in the justice system and preventing potential harms, especially human rights violations. While the judiciary can directly implement the principles and recommendations for organisations and individuals, they may also be used as **a template for developing context-specific national and subnational guidelines.**

However, effective implementation of AI tools in the justice sector requires more than adopting rules and guidelines. It is also essential to ensure that human talent is prepared to use AI tools ethically and responsibly and that AI tools are accessible, trustworthy, and human-rights-compliant.

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". These AI tools cannot substitute for qualified legal reasoning because they have no rationale or contextual understanding of a legal problem. Non-lawyers' use of LLMs for legal purposes 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.

In sum, the judiciary should aim to use AI tools to enhance, rather than replace, human judgment.

Annex

External Reviewers (in alphabetical order):

Karine Caunes (Global Program Director, EU Law, Center for AI and Digital Policy, United States), Giulia Gentile (Assistant Professor, Essex Law School, United Kingdom of Great Britain and Northern Ireland), Felipe Guzmán (Consultant on regulatory instruments for the digital transformation of the Judicial Branch, Superior Council of the Judiciary of Colombia), Ameen Jauhar (Team Lead, Centre for Applied Law & Tech, Vidhi Centre for Legal Policy), Amanda Leal (AI Governance and Policy Specialist at HealthAI), Martin Okumu Masiga (Secretary General, Africa Jurists and Judges Forum, Johannesburg, South Africa), Grace Mutung'u (Researcher and Advocate of the High Court of Kenya), Judge Jean Aloise Ndiaye (Supreme Court of Senegal), Alfonso Peralta Gutiérrez (Judge of the First Instance and Criminal Investigation, Granada, Spain), Matthieu Quinou (Doctor of Private Law and Attorney at Law at the Paris Bar, France), Anca Radu (Researcher, European University Institute, Department of Law, Florence, Italy), Dr. Krishna Ravi Srinivas (Adjunct Professor at NALSAR University of Law, India), Dr. Catherine Regis (Professor, Faculty of Law, University of Montreal, Canada), Dr. Dory Reiling (Retired Senior Judge, International Court and ICT expert, Amsterdam, Kingdom of the Netherlands), 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), Colin van Noordt (Researcher, Kingdom of the Netherlands), Carolina Villadiego (Legal Counsel for Latin America, International Commission of Jurists, Geneva, Switzerland), Anthony Wong (President of IFIP, International Federation for Information Processing, Austria).

Beyond the expert consultations, public consultations on the draft guidelines were organised between August and September 2024. The final guidelines have been enriched by over a hundred contributions sent by individuals and organizations in 41 countries, along with collective feedback from events organized by UNESCO with judicial operators in different parts of the world.

UN OHCHR, World Health Organization, Access Now (United States), Access to Justice Centre for Excellence, ACT (Canada): The App Association (United States), Advocacy and Awareness Centre Africa (AAC Africa) Initiative (Kenya), African IGF & African Digital Rights Network (Ethiopia, contributor from Egypt), Anticorruption Research Centre (Kenya), ARTICLE 19 (Brazil, and Office for Mexico and Central America), Auticon (France), Cambridge Existential Risk Initiative (United Kingdom of Great Britain and Northern Ireland), Center for

AI and Digital Policy (United States), Centre for Health Equity, Law & Policy (C-HELP, India), Centre for Internet and Society (CIS, India), Centre for Space, Cyberspace & Data Law (Bond University, Australia), Centre for the Study of Law and Governance (Jawaharlal Nehru University, India), Centro de Capacitación y Perfeccionamiento 'Ricardo C. Núñez' del Poder Judicial de la Provincia de Córdoba (Argentina), Centro Federal De Conciliacion Y Registro Laboral (Mexico), Centro para la Investigación del Cibercrimen y la Ciberseguridad (Center for CIC, Colombia), Centre for Intellectual Property and Information Technology Law (CIPIT, Kenya), Consejo de la Judicatura Federal (Mexico), Council for Research on International Economic Relations (ICRIER, India), Court of Appeal of Cairo (Egypt), Cuevas Mella (Dominican Republic), Daksh Society (India), Rights Nepal, El Veinte (Colombia), Empatia Lab (Peru), Escuela Judicial Electoral (Electoral Judicial School, Mexico), European Institute of Public Administration (Luxembourg) in collaboration with Maastricht Law & Tech Lab of Maastricht University (Kingdom of the Netherlands), Faculdade de Tecnologia de Teresina (Brazil), Faculty of Law, University of Technology Sydney (Australia), Immanuel Kant Baltic Federal University (Russian Federation), Hellenic Institute of International & Foreign Law (Greece), Ikigai Law (India), Indian Institute of Foreign Trade Kolkata (India), Inno Canyon Consulting (Nigeria), International Association of Prosecutors (Austria), Kictanet (Kenya), Laboratorio de Innovación Tecnológica de la Facultad de Derecho de la Universidad Nacional de Lomas de Zamora UNLZ-LAB (Argentina), Lawyers without Borders (Canada, El Salvador), Learned Hand (United States), Lithuanian National Commission for UNESCO (Lithuania), Maseru Magistrate Court (Lesotho), Ministerio Publico (Panama), Ministry of Foreign Affairs of Spain, Ministry of Law, Justice and Parliamentary Affairs (Nepal), National Academy of Legal Studies and Research (NALSAR, India), National Judicial Academy (Nepal), Nepal Bar Association, New Zealand Law Society, Osh Technology & Artificial Intelligence Industries (Egypt), Pan African Lawyers Union (Tanzania), Poder Judicial Asunción (Paraguay), Priva Sapien (India), Rama Judicial (Colombia), Rencontre Africaine pour la Défense des Droits de l'Homme (RADDHO, Senegal), S.B Law firm (Islamabad, Pakistan), Saffi technologies/Kenya internet action network (Kenya), School of Law, Trinity College Dublin (Ireland), Sofia Regional Court (Bulgaria), Supreme Court of Paraguay, Supremo Tribunal de Justicia del Estado de Aguascalientes (Mexico), Taiz University (Yemen), TeachSomebody (Kingdom of the Netherlands), The Civil Law Team of Association of Polish Judges IUSTITIA (Poland), The Supreme Court of Nepal, SGH Warsaw School of Economics (Poland), UBA – IALAB (Argentina), Universidad Americana (Paraguay), Universidad Galileo (Guatemala), University of Bologna (Italy), University of Minas Gerais (UFMG, Brazil), University of New Brunswick, Faculty of Law (Canada), University of Strasbourg (France), Western Sydney University, IFIP, ACS (Australia).

Individual contributions from judges (Argentina, Costa Rica, Egypt, Guatemala, Paraguay, Tanzania), court staff (Nepal), lawyers and legal professionals (Cameroon, Colombia, Guatemala, Kenya, Mexico), civil society organisations (Nepal), academics (Belgium, Costa Rica, Philippines, Switzerland), and journalists (El Salvador).

Endnotes

1 GPAI, *Algorithmic Transparency: A State-of-the-Art Report* (Global Partnership on Artificial Intelligence (GPAI), November 2024) 🔗; Nestor Maslej and others, *The AI Index 2025 Annual Report* (AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, April 2025), p. 456 🔗; Sarah Muñoz-Cadena and Juan David Gutiérrez, 'Repositorios de Algoritmos Públicos En El Mundo' (Sistemas de Algoritmos Públicos, Universidad de los Andes, March 2025) 🔗.

2 Sarah Muñoz-Cadena and others, 'Sistemas de IA En El Sector Público de América Latina y El Caribe' (Universidad de los Andes, March 2025) 🔗.

3 Yannis Charalabidis, Rony Medaglia, and Colin van Noordt, *Research Handbook on Public Management and Artificial Intelligence* (Edward Elgar Publishing, 2024), doi:10.4337/9781802207347; M. Manzoni and others, *AI Watch Road to the Adoption of Artificial Intelligence by the Public Sector: A Handbook for Policymakers, Public Administrations and Relevant Stakeholders* (Publications Office of the European Union, 2022), doi:10.2760/288757 (online),10.2760/693531 (print); Regine Paul, Emma Carmel, and Jennifer Cobbe, *Handbook on Public Policy and Artificial Intelligence* (Edward Elgar Publishing, 2024), doi:10.4337/9781803922171.

4 For an overview of different tasks carried out with AI systems in the justice sector, see F. Bell and others, *AI Decision-Making and 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) 🔗.

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', ed. by Lexing Xie, *PeerJ Computer Science*, 2 (2016), pp. 1-19, doi:10.7717/peerj-cs.93; Urvashi Aneja and Dona Mathew, *Artificial Intelligence in India's Judicial System: A Case of Organised Irresponsibility?* (Digital Futures Lab, March 2023) 🔗; Elsa Estevez, Sebastián Linares, and Pablo Fillottrani, *PROMETEA: Transformando La Administración de Justicia Con Herramientas de Inteligencia Artificial* (Banco Interamericano de Desarrollo, 2020) 🔗; 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) 🔗; Goretty Carolina Martinez, 'La Inteligencia Artificial y Su Aplicación al Campo Del Derecho', *Alegatos*, no. 82 (2012), pp. 827-46; Masha Medvedeva, Michel Vols, and Martijn Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights', *Artificial Intelligence and Law*, 28.2 (2020), pp. 237-66, doi:10.1007/s10506-019-09255-y; Ministro do Superior Tribunal de Justiça, *Artificial Intelligence: Technology Applied to Conflict Resolution in the Brazilian Judiciary* (Fundação Getulio Vargas, 2020) 🔗; Marion Oswald and others, 'Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and "Experimental" Proportionality', *Information & Communications Technology Law*, 27.2 (2018), pp. 223-50, doi:10.1080/13600834.2018.1458455; Víctor Saavedra and Juan Carlos Upegui, *PretorIA y La Automatización Del Procesamiento de Causas de Derechos Humanos* (Derechos Digitales América Latina y Dejusticia, March 2021) 🔗.

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)) 🔗 [accessed 27 May 2024].

5 For example, the European Commission for the Efficiency of Justice of the Council of Europe (CEPEJ) has documented more than one hundred AI tools that contribute to the management of judicial processes and to the decision-making tasks of judges and magistrates. See CEPEJ, *Resource Centre Cyberjustice and AI* (European Commission for the efficiency of justice (CEPEJ), 6 May 2024) 🔗.

6 Kahneman and his collaborators refer to "noise" as an error in human judgment –distinct from bias– that consists of unwanted variability in decision-making that should have been identical, causing random inconsistency in decisions. In this sense, "judgments are noisy, as people who are expected to agree end up at very different points". Moreover their book analysed how machine-learning algorithms could reduce noise in the judicial decision-making scenarios. Daniel Kahneman, Olivier Sibony, and Cass R. Sunstein, *Noise. A Flaw in Human Judgement*, 1st edn (Little, Brown and Company, 2021), p. 4.

7 Nathan-Ross Adams, 'Parker v Forsyth NO: Lessons for Using AI for Legal Research', *Michalsons*, 11 July 2023 🔗 ; 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 🔗 ; Francesca Bennetts and others, 'Responsible AI: Navigating the Risks and Embracing the Possibilities', *Allen Overy*, 12 November 2023 🔗 ;

Juan David Gutiérrez, 'Judges and Magistrates in Peru and Mexico Have ChatGPT Fever', *Tech Policy Press*, 19 April 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 🔗; Juan David Gutiérrez, 'AI Technologies in the Judiciary: Critical Appraisal of Large Language Models in Judicial Decision-Making', in *Handbook on Public Policy and AI*, ed. by Regine Paul, Emma Carmel, and Jennifer Cobbe (Edward Elgar Publishing, 2014) 🔗; LexisNexis, 'LexisNexis Launches Lexis+ AI, a Generative AI Solution with Linked Hallucination-Free Legal Citations', *LexisNexis*, 25 October 2023 🔗;

Benjamin Weiser, 'Here's What Happens When Your Lawyer Uses ChatGPT', *The New York Times*, 27 May 2023, section New York 🔗 ; 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 🔗 ;

Mark Witten, 'Applying Generative AI to Law: Opportunities and Risks | Queen's Law', *Queen's University*, 23 October 2023 🔗 ;

Kieran Woboditsch-Velasco, 'The Quiet Revolution: Generative AI's Rise in Canadian Law Firms', *MyOpenCourt*, 25 November 2023 🔗 ;

Nuria Morcillo, 'Sentencias falsas, leyes extranjeras y filtrado de datos: los riesgos de usar ChatGPT se cuelan en los despachos de abogados', *El País*, 18 February 2025, section Economía 🔗 ;

Sara Merken, 'Lawyers in Walmart Lawsuit Admit AI "hallucinated" Case Citations', *Reuters*, 10 February 2025, section Legal Industry 🔗 ;

Sam Harden, 'You'd Think Lawyers Would Know by Now', *Team Do Something*, 14 February 2025 🔗 ;

Kalliopi Terzidou, 'Generative AI Systems in Legal Practice Offering Quality Legal Services While Upholding Legal Ethics', *International Journal of Law in Context*, 2025, pp. 1-22, doi:10.1017/S1744552325000047.

8 Juan David Gutiérrez, 'Chapter 24: Critical Appraisal of Large Language Models in Judicial Decision-Making', in *Handbook on Public Policy and Artificial Intelligence* (Edward Elgar Publishing, 2024), pp. 323-38, doi:10.4337/9781803922171.00033.

9 For guidelines and policies issued by the bodies that govern the judiciary and by courts, see:

UK Courts and Tribunals Judiciary, *Artificial Intelligence (AI) - Guidance for Judicial Office Holders* (United Kingdom of Great Britain and Northern Ireland, 12 December 2023) 🔗 ;

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), 21 August 2020) 🔗;

Courts of New Zealand, *Guidelines for Use of Generative Artificial Intelligence in Courts and Tribunals* (New Zealand, 7 December 2023) 🔗 ;

Federal Court, *Interim Principles and Guidelines on the Court's Use of Artificial Intelligence* (Federal Court, 20 December 2023) 🔗 ;

Federal Court, *Notice to the Parties and the Profession: The Use of Artificial Intelligence in Court Proceedings* (Federal Court, 7 May 2024) 🔗 ;

Queensland Courts, *The Use of Generative Artificial Intelligence (AI) Guidelines for Responsible Use by Non-Lawyers*, 14 May 2024 🔗 ;

Supreme Court of Victoria, *Guidelines for Litigants: Responsible Use of Artificial Intelligence in Litigation* (Supreme Court of Victoria, May 2024) 🔗 ;

Canadian Judicial Council, *Guidelines for the Use of Artificial Intelligence in Canadian Courts* (Canadian Judicial Council, September 2024) 🔗 ;

Supreme Court of the State of Delaware, *Interim Policy on the Use of GenAI by Judicial Officers and Court Personnel* (Supreme Court of the State of Delaware, 21 October 2024) 🔗 ;

Supreme Court of the Republic of Singapore, *Guide on the Use of Generative Artificial Intelligence Tools by Court Users* (Supreme Court of the Republic of Singapore, 23 September 2024) 🔗 ;

Consejo Superior de la Judicatura, *Lineamientos Para El Uso y Aprovechamiento Respetuoso, Responsable, Seguro y Ético de La Inteligencia Artificial En La Rama Judicial* (Consejo Superior de la Judicatura, de diciembre de 2024) 🔗 ;

Conselho Nacional de Justiça (CNJ), *Resolução N° 615, de 11 de Março de 2025. Estabelece Diretrizes Para o Desenvolvimento, Utilização e Governança de Soluções Desenvolvidas Com Recursos de Inteligência Artificial No Poder Judiciário.* (Conselho Nacional de Justiça (CNJ), 11 March 2025) 🔗 .

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.

Likewise, the Supreme Court of Singapore, in September 2024, issued its "Guide on the Use of Generative Artificial Intelligence Tools by Court Users" that applies to court users ("any person who is involved in a Court case, including prosecutors, lawyers, Self-Represented Persons, or witnesses") in "all matters in the Supreme Court, the State Courts and the Family Justice Courts".

In October 2024, the Superior Court of Justice of Rio Negro in Argentina approved its "Protocol of Good Practices for the use of Generative Artificial Intelligence (IAGen)" for magistrates, officials and employees.

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); 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), 12 February 2024) 🔗.

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

11 See sections 2 - 5 of the Chapter III of the European Union's AI Act

12 Bell and others, *AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators*; Juan David Gutiérrez, 'Retos Éticos de La Inteligencia Artificial En El Proceso Judicial.', in *Derecho Procesal. #NuevasTendencias. XLI Congreso Colombiano de Derecho Procesal*, ed. by ICDP (Instituto Colombiano de Derecho Procesal (ICDP) y Universidad Libre., 2020) 🔗; 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) 🔗 ; 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) 🔗 ; Kalliopi Terzidou, 'The Use of Artificial Intelligence in the Judiciary and Its Compliance with the Right to a Fair Trial', *Journal of Judicial Administration*, 31.3 (2022), pp. 154-68; Giulia Gentile, 'AI in the Courtroom and Judicial Independence: An EU Perspective', *EUIdeas*, 22 August 2022 🔗 .

13 Terzidou, 'The Use of Artificial Intelligence in the Judiciary and Its Compliance with the Right to a Fair Trial'; Kalliopi Terzidou, 'Automated Anonymization of Court Decisions: Facilitating the Publication of Court Decisions through Algorithmic Systems', in *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, ICAIL '23 (Association for Computing Machinery, 2023), pp. 297-305, doi:10.1145/3594536.3595151.

14 For a framework of different types and classifications of AI tools, see OECD, 'OECD Framework for the Classification of AI Systems' (OECD Publishing, 2022) 🔗 .

15 The UN General Assembly's resolution 78/265 is available here: 🔗

16 The Bangalore Principles establish standards for ethical conduct of judges. "In July 2006, the United Nations Economic and Social Council (ECOSOC) adopted a resolution recognizing the Bangalore Principles as representing a further development of, and as being complementary to, the 1985 United Nations Basic Principles on the Independence of the Judiciary. ECOSOC invited States to encourage their judiciaries to take into consideration the Principles when reviewing or developing rules with respect to judicial conduct." The Bangalore Principles are available here: 🔗

17 On the utility and acceptability of using generative AI tools by arbitrators and the potential risks of adopting these technologies, see João Ilhão Moreira and Jiawei Zhang, 'ChatGPT as a Fourth Arbitrator? The Ethics and Risks of Using Large Language Models in Arbitration', *Arbitration International*, 2024, pp. 1-14, doi:10.1093/arbint/aiae031; David M. den Bakker, 'Arbitration and AI: The Role of LLMs in Drafting Awards' (unpublished Master Thesis, Radboud University Nijmegen, 2024), doi:10.5281/zenodo.14601385.

18 For guidelines and recommendations issued by universities, law societies, bar associations, and professional associations, see:

Bell and others, *AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators*;

James E. Baker, Laurie Hobart, and Mathew Mittelsteadt, *AI for Judges: A Framework* (Center for Security and Emerging Technology, December 2021) 🔗 ;

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, 24 November 2023) 🔗 ;

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, 25 July 2023) 🔗 ;

The State Bar of California, *Recommendations from Committee on Professional Responsibility and Conduct on Regulation of Use of Generative AI by Licensees*, 16 November 2023;

The Bar Council, *Considerations When Using ChatGPT and Generative Artificial Intelligence*

Software Based on Large Language Models, 30 January 2024 ;

Queensland Law Society, *Guidance Statement No 37 Artificial Intelligence in Legal Practice*, 6 June 2024 ;

New South Wales Bar Association, *Issues Arising from the Use of AI Language Models (Including ChatGPT) in Legal Practice*, 12 July 2023 ;

CIARB**,** *Guideline on the Use of AI in Arbitration* (The Chartered Institute of Arbitrators (CIARB), March 2025) .

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: .

19 UNESCO, 'Training Judicial Operators', 15 October 2024 .

20 UNESCO, 'UNESCO Global MOOC on AI and the Rule of Law Engaged Thousands of Judicial Operators', 20 April 2023 .

21 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) .

22 UNESCO, 'Recommendation on the Ethics of Artificial Intelligence' (2022) . 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) .

23 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) .

24 UNESCO, *Global Toolkit on AI and the Rule of Law for the Judiciary*.

25 In the public consultation process of the draft guidelines, between August and September 2024, UNESCO received individual feedback from 99 individuals and organizations based in Argentina, Australia, Belgium, Brazil, Bulgaria, Cameroon, Canada, Colombia, Costa Rica, Denmark, Dominican Republic, Egypt, El Salvador, France, Guatemala, India, Ireland, Italy, Kenya, Kingdom of the Netherlands, Lesotho, Liberia, Lithuania, Mexico, Nepal, New Zealand Nigeria, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Senegal, Spain, Switzerland, Tanzania, United Kingdom of Great Britain and Northern Ireland, United States of America, and Yemen. Additionally, UNESCO received contributions from members of the judiciary in events organized in India, Kenya, Madagascar, and Nepal.

26 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, *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*.

27 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.

28 *Ibid.* See also, UN, *Human Rights in the Administration of Justice: Report of the Secretary-General* (United Nations (UN), 7 August 2024) ; on gender bias, see UNESCO, Challenging Systematic Prejudices: An Investigation into Gender Bias in Large Language Models, 2024, .

29 The Bangalore Principles propose the following applications of the principle of "equality" that are pertinent for this Guidelines: "5.1. A judge shall be aware of, and understand, diversity in society and differences arising from various sources, including but not limited to race, colour, sex, religion, national origin, caste, disability, age, marital status, sexual orientation, social and economic status and other like causes ("irrelevant grounds"). 5.2. A judge shall not, in the performance of judicial duties, by words or conduct, manifest bias or prejudice towards

any person or group on irrelevant grounds. 5.3. A judge shall carry out judicial duties with appropriate consideration for all persons, such as the parties, witnesses, lawyers, court staff and judicial colleagues, without differentiation on any irrelevant ground, immaterial to the proper performance of such duties." *The Bangalore Principles of Judicial Conduct* (United Nations, 2018), pp. 14-15 .

30 UNSDG, 'Leave No One Behind', *United Nations Sustainable Development Group (UNDSG)* [accessed 26 March 2025].

31 *Ibid*.

32 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. See also, UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*.

33 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.

34 An analogous principle termed "justification principle" was proposed by Dan Jerker B. Svantesson, 'Cybercrime and the Adoption of Artificial Intelligence Systems for Judicial Decision-Making in Criminal Justice Systems', *Commonwealth Cybercrime Journal*\, 1.1 (2023), pp. 152-79.

35 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.

36 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), 4 April 2024) . For more information about access to information laws, see UNESCO's webpage: .

37 See AI HLEG, *Ethics Guidelines for Trustworthy AI* (European Commission, 8 April 2019), p. 17 .

38 AI HLEG, *Ethics Guidelines for Trustworthy AI*.

39 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. Similarly, as it is argued by the guidelines published by the Canadian Judicial Council: "Explainability is the need for such AI tools to provide clear, understandable explanations for their output, making it easier for users (and for those affected) to interpret, trust, contest, or accept AI output in critical workflows. Explainability is akin to the requirement for judges to provide reasoned explanations for their decisions in law". See, Canadian Judicial Council, *Guidelines for the Use of Artificial Intelligence in Canadian Courts*, p. 9.

40 See AI HLEG, *Ethics Guidelines for Trustworthy AI*. 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*, 29 April 2020 . UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*.

41 UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*.

41 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*. For an overview of the principle of "open justice" in the context of AI systems, see Brian M Barry, 'AI for Assisting Judicial Decision-Making: Implications for the Future of Open Justice', *Australian Law Journal*, 98.9 (2024), pp. 656-69.

43 For an explanation of the principle "Awareness and literacy, refer to section III.2 of UNESCO's Recommendation on the Ethics of Artificial Intelligence. See also, UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*.

44 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.

45 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', *AI & SOCIETY*, 2023, doi:10.1007/s00146-023-01635-y.

46 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 AI HLEG, *Ethics Guidelines for Trustworthy AI*.

47 See AI HLEG, *Ethics Guidelines for Trustworthy AI*.

48 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, *Ethics Guidelines for Trustworthy AI*; UNESCO and i4Policy, *Multi-stakeholder AI Development: 10 Building Blocks for Inclusive Policy Design* (United Nations Educational, Scientific and Cultural Organization (UNESCO), Innovation for Policy Foundation (i4Policy), 2022) 🔗 ; Anri Van der Spuy, *What If We All Governed the Internet? Advancing Multi-stakeholder Participation in Internet Governance* (UNESCO, 2017) 🔗 .

49 The Bangalore Principles propose the following applications of the principle of "judicial independence" that are pertinent for this Guidelines: "1.1. A judge shall exercise the judicial function independently on the basis of the judge's assessment of the facts and in accordance with a conscientious understanding of the law, free of any extraneous influences, inducements, pressures, threats or interference, direct or indirect, from any quarter or for any reason. 1.2. A judge shall be independent in relation to society in general and in relation to the particular parties to a dispute that the judge has to adjudicate. 1.3. A judge shall not only be free from inappropriate connections with, and influence by, the executive and legislative branches of government, but must also appear to a reasonable observer to be free therefrom." *The Bangalore Principles of Judicial Conduct*, pp. 8-9.

50 The Bangalore Principles propose the following applications of the principle of "impartiality" that are pertinent for this Guidelines: "2.1. A judge shall perform his or her judicial duties without favour, 2.2. A judge shall ensure that his or her conduct, both in and out of court, maintains and enhances the confidence of the public, the legal profession and litigants in the impartiality of the judge and of the judiciary." *The Bangalore Principles of Judicial Conduct*, p. 10.

51 The judiciary should consider the trade-offs of adopting tailored-made tools versus off-the-shelf tools. The latter may be less costly and faster to implement but less transparent and have limited customization. The former may be more expensive (although costs can be lowered when built on open-source code) and require more time for implementation (including development and testing), but more transparent and aligned with the organization's needs and objectives.

52 AI HLEG, *Ethics Guidelines for Trustworthy AI*.

53 For example, human-induced errors during the selection or labeling of data in supervised learning machine learning tools; statistical lack of input, representation, or raw data that can reinforce bias in unsupervised learning machine learning tools; environment-driven errors in reinforcement machine learning tools, etc.

54 "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' (presented at the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 2019), pp. 220-29, doi:10.1145/3287560.3287596. See also: GPAI, *Algorithmic Transparency: A State-of-the-Art Report*; Matías Valderrama, María Paz Hermosilla, and Romina Garrido, *State of the Evidence: Algorithmic Transparency* (Open Government Partnership and GobLab (Universidad Adolfo Ibáñez), May 2023) 🔗 ; Open Government Partnership, 'CHAPTER 8: Algorithmic Transparency', in *The Skeptic's Guide to Open Government*, 2022.

55 UNESCO, *Global Toolkit on AI and the Rule of Law for the Judiciary*, p. 42.

56 See, Brian Calvert, 'AI Already Uses as Much Energy as a Small Country. It's Only the Beginning.', *Vox*, 28 March 2024 ; Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat, 'Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model', *Journal of Machine Learning Research*, 24.253 (2023), pp. 1-15; Sasha Luccioni, Yacine Jernite, and Emma Strubell, 'Power Hungry Processing: Watts Driving the Cost of AI Deployment?', in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24 (Association for Computing Machinery, 2024), pp. 85-99, doi:10.1145/3630106.3658542.

57 UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*.

58 For an overview of different accountability frameworks that may be implemented before the deployment of AI systems, see Dillon Reisman and others, *Algorithmic Impact Assessments Report: A Practical Framework for Public Agency Accountability* (AI Now Institute, 9 April 2018) ;

Ada Lovelace Institute and DataKind UK, *Examining the Black Box: Tools for Assessing Algorithmic Systems*;

Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template', *Computer Law & Security Review*, 54 (2024), p. 106020, doi:10.1016/j.clsr.2024.106020;

UNESCO, *Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence*;

Government of the Netherlands, *Fundamental Rights and Algorithms Impact Assessment (FRAIA)* (Kingdom of the Netherlands, 31 July 2021) .

59 An "algorithmic impact assessment" is a process that assesses future impacts –at the individual, community, and societal level– that may be associated with the deployment and use of an AI system and identifies how to address potential negative consequences. The assessment may focus social and ethical implications, economic effects, impact on human rights, and environmental consequences. Reisman and others, *Algorithmic Impact Assessments Report: A Practical Framework for Public Agency Accountability*.

60 An "algorithmic risk assessment" is a process that maps specific risks that may be generated with the deployment and use of an AI system, analyzes vulnerabilities and potential harms and identifies the measures required to prevent and address risks. The assessments focus on potential system errors or failures, security and safety vulnerabilities, and regulatory breaches. Ada Lovelace Institute and DataKind UK, *Examining the Black Box: Tools for Assessing Algorithmic Systems*.

61 A "fundamental rights impact assessment" is a process that identifies, understands, and assesses the potential adverse impacts of AI systems regarding fundamental rights and prescribes how to address and prevent negative impacts. Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template'.

62 UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*, p. 19.

63 AI HLEG, *Ethics Guidelines for Trustworthy AI*, p. 31.

64 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).

65 Ada Lovelace Institute and DataKind UK, *Examining the Black Box: Tools for Assessing Algorithmic Systems*.

66 Alessandro Mantelero, 'Human Rights Impact Assessment and AI', in *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, ed. by Alessandro Mantelero (T.M.C. Asser Press, 2022), pp. 45-91, doi: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', *Computer Law & Security Review*, 41 (2021), p. 105561, doi:10.1016/j.clsr.2021.105561; Government of the Netherlands, *Fundamental Rights and Algorithms Impact Assessment (FRAIA)*; Government of Canada, *Algorithmic Impact Assessment Tool* (Canada, 12 April 2023) 🔗 .

67 Ada Lovelace Institute and DataKind UK, *Examining the Black Box: Tools for Assessing Algorithmic Systems*.

68 A "human rights impact assessment" is a process that identifies the impact of AI systems on human rights and social rights-holders and provides recommendations on how to prevent or mitigate them. It is usually conducted in response to critical situations. Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template'. See also, UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*.

69 AI HLEG, *Ethics Guidelines for Trustworthy AI*, p. 31.

70 Alessandro Mantelero, 'Human Rights Impact Assessment and AI', in *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, ed. by Alessandro Mantelero (T.M.C. Asser Press, 2022), pp. 45-91, doi: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', *Computer Law & Security Review*, 41 (2021), p. 105561, doi:10.1016/j.clsr.2021.105561; Government of the Netherlands, *Fundamental Rights and Algorithms Impact Assessment (FRAIA)*; Government of Canada, *Algorithmic Impact Assessment Tool* (Canada, 12 April 2023) 🔗.

71 For example, on 21 June 2022, the Court of Justice of the European Union (CJEU) issued a judgment prohibiting specific "automated predictive threat detection" instruments (tools that "automatically sift through massive amounts of data in order to predict potential threats to public security"). The CJEU found that machine learning-based AI tools "capable of modifying without human intervention or review the assessment process" were incompatible with the rights to privacy and personal data protection. The CJEU argued that "the opacity which characterises the way in which artificial intelligence technology works … may deprive the data subjects also of their right to an effective judicial remedy".The CJEU's judgement is available here: 🔗 For summary and analysis of the case, see Christian Thönnes and Niovi Vavoula, 'Automated Predictive Threat Detection after Ligue Des Droits Humains: Implications for ETIAS and CSAM (Part I)', *VerfBlog*, 8 May 2023 🔗.

72 UN, *Human Rights in the Administration of Justice: Report of the Secretary-General*, pp. 18-19.

73 For a detailed account of generative AI tools' vulnerabilities see BSI, *Generative AI Models. Opportunities and Risks for Industry and Authorities*.

74 AI HLEG, *Ethics Guidelines for Trustworthy AI*, p. 17.

75 See Terzidou, 'Automated Anonymization of Court Decisions: Facilitating the Publication of Court Decisions through Algorithmic Systems'.

76 For example, the European Commission for the Efficiency of Justice of the Council of Europe (CEPEJ) and the Brazilian judiciary published repositories that document the AI systems adopted in their jurisdictions.

77 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, *Algorithmic Transparency: A State-of-the-Art Report*.

78 UNESCO, Open Data for AI: What Now? (United Nations Educational, Scientific and Cultural Organization (UNESCO), 2023) 🔗.

79 See for example, the curriculum of the UNESCO Global MOOC on AI and the Rule of Law. UNESCO, 'UNESCO Global MOOC on AI and the Rule of Law Engaged Thousands of Judicial Operators'.

80 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, 'Human Rights Impact Assessment and AI'.

81 In this vein, the Bangalore Principles propose the following applications of the principle of "competence and diligence" that are pertinent for this Guidelines: "6.3. A judge shall take reasonable steps to maintain and enhance the judge's knowledge, skills and personal qualities necessary for the proper performance of judicial duties, taking advantage for that pur pose of the training and other facilities that should be made available, under judicial control, to judges." *The Bangalore Principles of Judicial Conduct*, p. 16.

82 Juan David Gutiérrez, '¿Están Los Jueces En Capacidad de Usar Modelos de Lenguaje a Gran Escala (LLMs)?', *Revista EXCEJLENCIA*, 7 (2023), pp. 10-15; David Bawden, 'Origins and Concepts of Digital Literacy', in *Digital Literacies: Concepts, Policies and Practices*, ed. by Colin. Lankshear and Michele. Knobel (Peter Lang, 2008), pp. 17-32 🔗.

83 Mantelero, 'Human Rights Impact Assessment and AI'; Mantelero and Esposito, 'An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems'.

84 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) 🔗.

85 UK Courts and Tribunals Judiciary, *Artificial Intelligence (AI) - Guidance for Judicial Office Holders*, p. 3.

86 See Courts of New Zealand, *Guidelines for Use of Generative Artificial Intelligence in Courts and Tribunals*; UK Courts and Tribunals Judiciary, *Artificial Intelligence (AI)-Guidance for Judicial Office Holders*, p. 6.

87 UK Courts and Tribunals Judiciary, *Artificial Intelligence (AI) - Guidance for Judicial Office Holders*, p. 3.

88 Courts of New Zealand, *Guidelines for Use of Generative Artificial Intelligence in Courts and Tribunals*, p. 2.

89 For a similar guidance regarding the generation of evidence aimed at court users, see Supreme Court of the Republic of Singapore, *Guide on the Use of Generative Artificial Intelligence Tools by Court Users*.

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

91 BSI, *Generative AI Models. Opportunities and Risks for Industry and Authorities*, p. 9 ; See also, UNESCO, Challenging Systematic Prejudices: An Investigation into Gender Bias in Large Language Models, 2024, 🔗.

92 "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" BSI, *Generative AI Models. Opportunities and Risks for Industry and Authorities*, p. 10.

93 BSI, *Generative AI Models. Opportunities and Risks for Industry and Authorities*, p. 9.

94 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.

95 For a similar guidance, see Supreme Court of the State of Delaware, *Interim Policy on the Use of GenAI by Judicial Officers and Court Personnel*.

96 Emily M. Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21 (Association for Computing Machinery, 2021), pp. 610-23 (p. 616), doi:10.1145/3442188.3445922.

97 Matthew Dahl and others, 'Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models', 2024 🔗; Varun Magesh and others, 'Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools', 2024.

AI has entered the courtroom. Judges and legal professionals have started using AI systems to save time and focus on important tasks. This technological shift can lead to significant improvements for human rights and the delivery of justice.

Yet, how can we ensure that the benefits of AI outweigh its risks (bias, discrimination, errors and hallucinations)?

Courts have the power to control and shape how AI systems are developed and used to safeguard judicial independence, the rule of law, and human rights.

Justice stakeholders have called on UNESCO to develop these Guidelines to provide key recommendations and principles on the necessary steps to consider before, during and after deploying AI systems.

These Guidelines will assist court leaders to design policy frameworks relevant to the local context. They will help judges, court staff, and legal professionals, regardless of their digital skills, to develop a professional use of AI.

The AI & the Rule of Law programme at UNESCO equips judges and legal professionals around the globe with the skills they need to uphold justice, the rule of law and human rights in the age of AI.

GUIDELINES FOR THE USE OF AI SYSTEMS IN COURTS AND TRIBUNALS

9 789231 008306

unesco