



# Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision-making through generative AI

David Uriel Socol de la Osa<sup>1</sup> and Nydia Remolina<sup>2,3</sup> 

<sup>1</sup>Hitotsubashi University, Hitotsubashi Institute for Advanced Study, Graduate School of Law, Tokyo, Japan

<sup>2</sup>Singapore Management University, Yong Pung How School of Law, Singapore

<sup>3</sup>SMU Centre for AI and Data Governance, Singapore

**Corresponding author:** Nydia Remolina; Email: [nydiarl@smu.edu.sg](mailto:nydiarl@smu.edu.sg)

**Received:** 27 November 2023; **Revised:** 22 May 2024; **Accepted:** 04 June 2024

**Keywords:** access to justice; accountability; AI ethical principles; AI regulation; artificial intelligence; fairness; generative AI; judicial decision-making; transparency; AI in the courtroom

## Abstract

Generative artificial intelligence (GenAI) has gained significant popularity in recent years. It is being integrated into a variety of sectors for its abilities in content creation, design, research, and many other functionalities. The capacity of GenAI to create new content—ranging from realistic images and videos to text and even computer code—has caught the attention of both the industry and the general public. The rise of publicly available platforms that offer these services has also made GenAI systems widely accessible, contributing to their mainstream appeal and dissemination. This article delves into the transformative potential and inherent challenges of incorporating GenAI into the domain of judicial decision-making. The article provides a critical examination of the legal and ethical implications that arise when GenAI is used in judicial rulings and their underlying rationale. While the adoption of this technology holds the promise of increased efficiency in the courtroom and expanded access to justice, it also introduces concerns regarding bias, interpretability, and accountability, thereby potentially undermining judicial discretion, the rule of law, and the safeguarding of rights. Around the world, judiciaries in different jurisdictions are taking different approaches to the use of GenAI in the courtroom. Through case studies of GenAI use by judges in jurisdictions including Colombia, Mexico, Peru, and India, this article maps out the challenges presented by integrating the technology in judicial determinations, and the risks of embracing it without proper guidelines for mitigating potential harms. Finally, this article develops a framework that promotes a more responsible and equitable use of GenAI in the judiciary, ensuring that the technology serves as a tool to protect rights, reduce risks, and ultimately, augment judicial reasoning and access to justice.

## Policy Significance Statement

As the judicial system contemplates the incorporation of generative AI technologies, courts, policymakers, and regulators find themselves at a crossroads, seeking a balanced approach to harness these technologies for enhancing judicial decision-making while ensuring ethical integrity and responsible use. The lack of consensus on active utilization and implementation methodologies has allowed judges in some jurisdictions to use generative AI (GenAI) in the judicial decision-making process without comprehensively or holistically considering the systemic risks and potential pitfalls of the technology. Unless prudential frameworks for generative AI in judicial decision-making are utilized, this could harm the rule of law, access to justice, and create negative externalities for society as a whole. This paper contributes to the evolving discourse by proposing a framework



that guides jurisdictions in formulating regulations that govern the use of GenAI in the courtroom. This framework aims to navigate the complexities of technological integration while upholding the principles of justice, fairness, and transparency that are foundational to judicial processes.

## 1. Introduction

In November 2022, Open AI launched ChatGPT (Marr, 2023). Soon after, generative artificial intelligence (GenAI) rapidly ascended in popularity in the field of artificial intelligence (AI), and its potential integration across industries took the spotlight for academics, policymakers, investors, and generally the public and private sectors alike (Marr, 2023). The ascent to prominence of GenAI can be attributed to a convergence of improved computational power, vast amounts of available training data, and significant advancements in machine learning algorithms (Leslie, 2023; Orchard and Tasiemski, 2023). The democratization of this technology through publicly available platforms and the popularization of Large Language Models (LLMs) and cloud-based services has further propelled GenAI's accessibility, allowing small companies and individuals to leverage its vast potential, and use these technologies at no (or minimal) cost (Leslie, 2023; Orchard and Tasiemski, 2023).

The technology's versatility has found disruptive applications in numerous fields. AI is being used to create art and music, advance scientific research, and even revolutionize financial services (Remolina, 2024). As computational costs continue to decline while data availability scales, GenAI applications are being woven into the fabric of daily digital interactions, ranging from personalized consumer experiences to sophisticated problem-solving. Indeed, widespread use and expansion of the technology seem to be backed by strong market-confidence in the technology, highlighted by the 425% increase in venture capital investments in GenAI from 2020 to mid-2023 (Marr, 2023).

As is the case in other sectors, the capability of GenAI tools to aid lawyers and judges in their tasks seems to hold transformative promise. GenAI may be able to make legal processes more efficient with its ability to analyze massive amounts of legal language, provide fast answers to complex research questions, and create legal summaries and documents accordingly.<sup>1</sup> These and other possible applications have made the technology appealing to lawyers and judges. However, the legal profession is characterized by heightened responsibilities given the inherent potential to significantly impact individuals' rights, business deals and practices, and the economy and society generally.<sup>2</sup> Given these heightened risks, the incorporation of GenAI into law requires a nuanced approach, prioritizing ethical considerations and robust safeguards—especially inside of the courtrooms, where fundamental rights and freedoms are at stake (Marr, 2023; Gutierrez, 2024).

While the legal profession recognizes the need for caution, there is currently a lack of comprehensive safeguards for AI deployment in courtrooms. Early case studies on the present state of GenAI in judicial decision-making processes from Colombia, Mexico, Peru, and India show that the technology is being applied on an ad hoc case-by-case and judge-by-judge approach that has not comprehensively or adequately addressed the risks of GenAI.<sup>3</sup> In contrast to the lack of regulatory coverage and guidance in these regulatory environments, other countries have taken a more proactive approach in developing guidelines for the use of GenAI in court decisions.<sup>4</sup> Amongst nations taking a targeted regulatory approach, the United Kingdom and New Zealand have published guidelines for judicial office-holders to instruct them in the use of GenAI. In jurisdictions such as the European Union, general guidelines have been developed regarding the use of AI and automation in courts, though without addressing the newest iterations of GenAI specifically. Other jurisdictions like Canada have taken a more cautious approach: The Federal Court of Canada explicitly stated it will not deploy AI to aid in judgments or orders without

<sup>1</sup> See generally discussion *infra* Section 2.

<sup>2</sup> See generally discussion *infra* Section 2.2.

<sup>3</sup> See discussion *infra* Section 3.2.

<sup>4</sup> See generally, discussion *infra* Section 3.1.

engaging in public consultations first. Lastly, other proactive jurisdictions, as is the case in Singapore, have decided to develop and deploy GenAI models for automated judicial decision-making in specific cases (Marr, 2023).

This article maps the variety of these approaches and explores the potential applications of GenAI by judges and other judicial office-holders within the context of their decision-making processes. Structurally, the article will (1) review the unique characteristics of GenAI, examining the promise and perils associated with integrating such technology into the judicial system; (2) analyze various approaches for integrating GenAI into the courtroom, drawing valuable risk-mitigation insights from case studies in Colombia, Peru, Mexico, and India, as well as from the proactive approaches taken by other jurisdictions; in order to finally (3) propose a set of strategies to ensure the safe and responsible integration of AI into the courtroom. Generally, these solutions encompass (i) safety standards and requirements for GenAI systems to be used by judges and (ii) general guidance for the utilization and implementation of GenAI in judicial decision-making. Specifically, this framework focuses on ex-ante control through capacity assessments for system deployment, stakeholder engagement, verification processes for all GenAI systems to be utilized in courts, trusted datasets, clear allocation of responsibility, enhancing legal education, case-based risk assessments for specific or generalized uses of GenAI in courts, human-in-the-loop verification systems, specific procedural rights (like information and consent rights relating to the use or scope of the use of AI), and ongoing iterative audits. By implementing these comprehensive safeguards targeting the GenAI systems to be deployed, as well as focusing on the practical application of the tools, the legal system can harness the potential of GenAI while mitigating and managing its risks, ultimately promoting a more accessible and efficient, yet responsible, system of justice.

## 2. The promise (and pitfalls) of generative AI

While the use of machine learning has permeated many aspects of modern life, in order to discuss the use of GenAI in courtrooms, we need to start with a shared understanding of some key terms. Artificial intelligence is an umbrella term defined initially as “the science and engineering of making intelligent machines” (Manning, 2020). There are multiple ways that intelligent machines can be designed, and many functionalities that define AI systems. Key AI models include predictive, prescriptive, and generative AIs—the latter being the object of this article, given its present integration into judicial decision-making, and the specific risks its deployment raises.

Predictive analytics models apply statistical algorithms and machine learning to historical data to identify patterns and make projections and predictions (Kumar and Garg, 2018). Examples of predictive AI models include credit score generation and the detection of fraudulent financial transactions through network monitoring (Kumar and Garg, 2018). Typically, models used in predictive analytics assign a score to reflect the likelihood of a future event occurring with a higher score suggesting a greater probability of the event happening (Kumar and Garg, 2018). Another relevant AI model is prescriptive AI, which suggests the best course of action for a desirable outcome (Lepenioti et al., 2020). To achieve this, the approach integrates outputs from predictive analytics while employing AI processes and optimization algorithms alongside expert systems in a probabilistic framework—aiming to facilitate adaptive, automated, and optimized decision-making that is subject to contextual constraints and sensitive to time (Lepenioti et al., 2020). Prescriptive AI examples include personalized healthcare treatment plans or automatically changing airline ticket prices based on customer demand, gas prices, and weather, among other variables (Segal, 2022).

Lastly, GenAI works by learning from large amounts of data using complex algorithms and deep learning in order to create novel content, such as text, images, music, videos, computer code, and more, based on user inputs or prompts (Harvard University Information Technology, 2023). Examples of GenAI include Large Language Models like OpenAI’s ChatGPT (OpenAI, 2024) and Google’s Gemini (Google, 2024), which are designed to understand and generate human-like text; image creators like OpenAI’s Dall-e, which create images from textual descriptions (OpenAI, 2021); and GitHub’s Copilot, which assists users in code generation, offering suggestions based on the context of the existing code (<https://>

[github.com/](https://github.com/)). These tools are capable of handling and responding to complex prompts, and in the case of language-based models, are adept at providing convincingly human responses (Chan and Hu, 2023).

In the context of judicial decision-making, while predictive AI might suggest a case's likely outcome based on past iterations, and prescriptive AI could allow for systemic optimization in resource management allocation for case resolution, GenAI could draft arguments, identify relevant statutes, and summarize complex documents—all tasks with the potential to influence a judge's research and understanding of law and its application. The following section will delve into the potential of GenAI to revolutionize the judicial system, exploring both the opportunities and the challenges that need to be addressed for responsible integration.

### ***2.1. The promises of generative AI in judicial systems***

GenAI holds significant promise for judicial decision-making, particularly in its potential to increase case-resolution efficiency, which can help alleviate resource management pressures on judicial systems—a notion that is especially relevant for under-resourced jurisdictions (Grossman et al., 2023). The core appeal of GenAI in this context lies in its ability to analyze large volumes of legal documents, precedents, and case law, thereby providing judges (and legal practitioners) with insights and synthesized information that might otherwise be overlooked due to resource, time, and other processing constraints; as well as provide complex argumentation and aid in building the foundations of legal reasoning, inclusive of drafting case-related documents. This could not only streamline judicial decision-making and drafting processes, but could also ensure that judgments are well-informed and grounded in the broader legal context, potentially making legal analysis more comprehensive and providing legal reasoning with increased efficiency and documental depth (Chien et al., *forthcoming*). Further, for GenAI models trained on internal decisions, GenAI could help judges maintain internal jurisprudential consistency—where desired—leading to enhanced precedential clarity, and the long-term benefits of judicial predictability (Chien et al., *forthcoming*). In short, GenAI could expand access to justice by making it more agile, and even provide it with increased certainty and non-subjectivity, thus potentially reinforcing the rule of law. However, as will be discussed in the following section, many of these promises come with significant ethical and legal perils that must be addressed and mitigated in order to realize the true potential of GenAI in the courtroom.

### ***2.2. Ethical and legal perils of using GenAI in court decisions***

While GenAI holds promise for the legal system, its implementation in the courtroom raises significant ethical and legal concerns. These concerns echo challenges already explored by relevant scholarship on AI governance—particularly in predictive models—around transparency, fairness, bias, accountability, and data protection (Ferrara, 2023; Cantens, 2024). However, GenAI introduces new risks, such as the generation of misleading or fabricated information (“hallucinations”),<sup>5</sup> a phenomenon by which AI models produce responses that are factually incorrect or entirely fictitious despite appearing plausible; as well as the potential to undermine judicial independence, for example, by relying on AI recommendations that could influence a judges' independent decision-making processes. Building on existing scholarship on AI in law, this article delves deeper into GenAI integration in courtrooms. This section will examine how GenAI may exacerbate existing concerns and introduce new risks in the context of judicial decision-making, explore perils such as algorithmic bias and hallucinations, and how they could impact adjudication and procedural fairness, judicial independence, and access to justice.

#### ***2.2.1. Transparency, Public Trust, and Acceptance***

Many GenAI systems used today are often considered “black boxes.” (Cheng and Liu, 2024) This is because their decision-making processes are complex and not easily interpretable, with the systems giving

<sup>5</sup> See discussion *infra* Section 2.2.3.

responses based on an amalgamation of vast amounts of often unidentifiable sources, leading to the unexplainability of results (Yampolskiy, 2019). As explained by Yampolskiy, “[t]o explain...a decision, which relies on literally billions of contributing factors, AI has to either simplify the explanation and so make the explanation less accurate/specific/detailed or to report it exactly but such an explanation elucidates nothing by virtue of its semantic complexity, large size and abstract data representation” (Yampolskiy, 2019).

This is particularly problematic in judicial decision-making processes, which require clear thought pathways that consider ethics, reasonability, and fairness as core foundations. The lack of transparency in current GenAI-supported judicial decision-making can undermine accountability and due process, as judges, legal experts, and parties to judicial processes, may find it puzzling to understand and challenge how a specific decision was reached, and whether biases or errors were introduced into the final rationale provided.

Due to the complexity of some human circumstances and the nuanced understanding of the context required for fair and ethical legal judgments, GenAI systems may struggle to reliably inform decisions with significant legal consequences. Human oversight or involvement may remain crucial to ensure both the fairness of the outcome and the public’s perception of fairness (Morison and McInerney, 2024). Relying on GenAI systems for judicial decisions, either in whole or without proper safeguards, could reduce such human articulable intercession in the legal process, potentially reducing the transparency of justice served, thus bringing a potential consequent reduction in fairness, quality, and trust in the judicial system. The “black box” feature of GenAI’s processing is therefore impactful towards both the due transparency of legal reasoning that judicial processes owe to its parties, as well as to public trust. In modern democracies and jurisdictions with a strong rule of law, the legitimacy of a judiciary is dependent on public will, understanding, and in many ways, approval. Undermining this would jeopardize the integrity and foundations of the system itself.

GenAI implementation into judicial decision-making could therefore lead to a lack of explainability in process and outcome, which could (i) undermine transparency, reasonability, and fairness and (ii) erode public trust and acceptance of legal decisions, having adverse effects on how justice is administered, delivered, and perceived (Bühlmann and Kunz, 2011; Versteeg and Ginsburg, 2017). Indeed, as U.S. Supreme Court Chief Justice Roberts noted, “[a]t least at present, studies show a persistent public perception of a ‘human-AI fairness gap,’ reflecting the view that human adjudications, for all of their flaws, are fairer than whatever the machine spits out” (Roberts Jr., 2023).

### 2.2.2. *Fairness, bias amplification, and precedent interpretation*

GenAI systems can inadvertently perpetuate biases when these are present in the training data (Wachter et al., 2021). The problem of bias undermining fairness in GenAI stems from the way these systems are trained and the data they learn from. GenAI systems are trained on vast datasets retrieved from various sources—which can be as indiscriminate as the internet, like it is for ChatGPT (the GenAI language model utilized in the case studies discussed below in this article); or they can be specific and targeted datasets, like it would be if a jurisdiction were to incorporate a GenAI that included only a select group of cases and decisions from a specific court or tribunal (Wachter et al., 2021) (Wachter, 2022; Starke et al., 2022). These foundational databases may contain inherent biases that can manifest in the form of racial, gender, ethnic, or cultural leanings, and can extend to other biases concerning socioeconomic status, religion, and more (Juzgado Civil Transitorio de San Juan de Miraflores, n.d.). Additionally, GenAI models and LLMs can learn patterns and associations from the data during training, which can establish a direct-learning and reiterative amplification of specific biases.<sup>6</sup> Even if the data were to be collected solely from previous judicial decisions, those decisions could have still been made with biased leanings—whether contemporary or from historical inertia—that may not be fully captured, understood, or disclosed by an LLM when using GenAI to support judicial decision-making.

<sup>6</sup> See discussion generally *supra* Section 2.



If the training data contains biased data points, the models inadvertently amplify, propagate, and even obfuscate those biases in their generated outcome, leading to AI systems producing content that may reflect and reinforce societal biases (Floridi and Cows, 2022). Moreover, GenAI lacks a true understanding of the outcomes they generate: Responses are generated based on patterns in the data the models were trained on, and the model tuning performed by the developers, but these systems lack genuine comprehension or critical thinking abilities.<sup>7</sup> As the GenAIs to be deployed in judicial decision-making are in some ways limited to language emulation, this can result in the systems generating biased content without recognizing its implications. Additionally, the use of GenAI could exacerbate negative feedback loops given that the content that models create is increasingly difficult to distinguish from real data or non-synthetic data (Remolina, 2024), even if the systems were to be adjusted for case validity or precedential authority. For example, should content produced by AI be recycled into the training datasets for subsequent models (a process that takes place in open-sourced networks such as ChatGPT) when there is a risk of perpetuating and amplifying existing biases, this could result in a cascading effect, where each new generation of AI is increasingly more prejudiced than the last (Nicoletti and Bass, 2023).<sup>8</sup>

These inherent challenges in GenAI systems raise ethical concerns about the fairness of decisions, especially in cases involving sensitive issues such as race, gender, socioeconomic status, or other historical sources of bias and discrimination. Ensuring fairness and addressing bias in AI-generated content is a critical ethical consideration in judicial decision-making. It requires stakeholders to consider whether solutions in terms of data selection and processing, algorithmic audits, fairness metrics, user feedback, content moderation, or other novel solutions should be implemented and adjusted in the context of GenAI-driven judicial decisions.

While the use of GenAI in the courtroom holds potential to improve efficiency and consistency, its effectiveness hinges on the quality of the data it is trained on, and the tuning performed by its developers. If this data and tuning harbors biases inherent in historical legal or non-legal practices (depending on the training data chosen), the GenAI system could unintentionally and without understanding, perpetuate and amplify these biases in its decision-making, potentially leading to discriminatory outcomes and undermining the principles of fairness and equality within the justice system. It is crucial to ensure that the training data and tuning for AI-powered legal tools are rigorously evaluated and unbiased to avoid perpetuating historical injustices and safeguard the fundamental rights of all individuals.

### 2.2.3. *Data quality and hallucinations*

The quality of the responses provided by GenAI is largely dependent on the quality of the data the system is trained on—whether legal or otherwise. When a GenAI system is trained on vast amounts of non-legal data, but then applied in a legal context (as is the case in the case studies discussed below in this article, in which the judges used ChatGPT), the tool can generate content based on non-legal sources like blogs, news reports, and academic papers. While these sources may be helpful in certain contexts, they do not carry binding weight like statutes, regulations, and case law, and cannot be relied upon as such. Even if GenAI systems are trained on legal data, there are still substantial risks. These risks include bias embedded in the data, as historical legal decisions may have favored certain demographics or socioeconomic groups, with the GenAI system inheriting and amplifying those biases. Other risks include incompleteness of the database that will inform the content, leading to inaccurate or misleading results; or inaccuracy of the data,

<sup>7</sup> Generative AI models involve “stitching together sequences of linguistic forms it has observed in vast training data, according to probabilistic information about how they combine, but without any reference to meaning”. Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell ‘On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?’ in Conference on Fairness, Accountability, and Transparency FAccT 21, Virtual Event March 2021. <https://doi.org/10.1145/3442188.3445922>; cited in Morison, John, and Tomás McInerney. “When should a computer decide? Judicial decision-making in the age of automation, algorithms and generative artificial intelligence.” *Research Handbook on Judging and the Judiciary (Elgar-Routledge, Law and Society Series)* S. Turenne and M. Moussa Eds (2024).

<sup>8</sup> Nicoletti and Bass (2023) explain this problem in the context of generative AI creating images.

which may contain factual errors or outdated information. Under these conditions, GenAI could provide unreliable legal reasoning or misinterpret the relevance or legal context of the precedents.

In addition to risks associated with the underlying data quality, there are also risks in how this data gets processed, and then returned to the user. This is particularly concerning because of the ability of GenAIs to confidently generate content that appears truthful despite factual inaccuracies, due to how these systems function, learn, and structure responses. For example, LLMs (a type of GenAI that generates text based on vast amounts of data)<sup>9</sup> operate by statistically and algorithmically predicting the most likely sequences of words, rather than based upon the “truth” of a statement (Garry et al., 2024).<sup>10</sup> This, in addition to opaque and aggregate processing and response capabilities, can lead to a phenomenon known as “AI hallucination,” where the LLM generates seemingly coherent but factually inaccurate or misleading statements, potentially creating a false narrative (del Campo and Leach, 2022). These hallucinations degrade the reliability of the content GenAI tools create.<sup>11</sup>

Even the creators of these tools recognize and are open about the potential failures in the veracity of the technology’s outputs. Exemplifying this, Meta’s 2022 release of its LLM Galactica, lasted only 3 days online before being pulled down for, among other reasons, “not being able to distinguish truth from falsehood.” (Heaven, 2022) Further, OpenAI admitted that its tools should not be used as a primary decision-making tool (Kirchner et al., 2023), and Google’s senior vice president stated recently that “[a]s we’ve said from the beginning, hallucinations are a known challenge with all LLMs—there are instances where the AI just gets things wrong.” (Raghavan, 2024)

In the context of judicial decision-making, the fallibility of these systems carries an elevated risk, given that there is an inherent potential jeopardy of fundamental rights and freedoms. Because the confidence level associated with any particular response from GenAI tool is variable, a single unchecked instance of the AI “getting things wrong” may result in loss of property, imprisonment, or even separation from loved ones when a judge relies on a GenAI to inform their decision-making. This level of systemic error risk requires an abundance of caution when considering the integration of AI in the courtroom, requiring rigorous testing and implementation within robust ethical, legal, and supervisory frameworks. This inconsistency in content accuracy can also undermine public trust in the use of GenAI, which may have significant consequences in the judicial system and the rule of law. Further development is crucial to ensure LLMs and other AI systems are not only demonstrably unbiased, but also accurate and grounded in factual information, or at the very least, transparent when unsure as to the accuracy of its outputs.

#### 2.2.4. Accountability and legal liability

Legal liability may become a contentious issue when AI systems are involved in judicial decision-making. This determination of liability refers to establishing who is responsible for defective GenAI outcomes in a judicial process—whether it is the AI developer, the judge-user, or any other party. Liability and the chain of responsibility is a legal space that currently requires significant clarification. For instance, there is a need for determining whether legal professionals and/or judges have a duty to review the accuracy and legality of the outcomes that algorithms produce, to what extent, and in which ways. There is a need to ascertain who, how, and through which processes different responsible offices, individuals, or judicial decision-makers may be held professionally responsible if they do not exercise due diligence (and to which level) in reviewing and verifying AI-generated content that leads to, informs, or supports, a judicial decision. Additionally, questions about product liability may also arise—a subject currently under discussion in some jurisdictions, such as the European Union.<sup>12</sup> For example, if a third-party vendor or developer supplies a GenAI model used in legal decisions, they could be held liable for defects or damages caused by the system

<sup>9</sup> See discussion *supra* Section 2.

<sup>10</sup> See discussion *supra* Section 2.

<sup>11</sup> See discussion *infra* Section 3 (generally referring to the Avianca case).

<sup>12</sup> On 28 September 2022, the European Commission released the proposal for an Artificial Intelligence Liability Directive (“AI Liability Directive”), that deals with claims for harm caused by AI systems, or the use of AI, adapting non-contractual civil liability rules to artificial intelligence.

they produced. Particularly, developers may be held liable if their products are found to be defective or if they fail to meet any specific metrics or standards of performance claims.

The lack of a clear attribution of responsibility in AI-driven judicial decisions raises critical questions about accountability. Without a designated party responsible for any judicial errors in process or adjudication, ranging from material mistake to bias amplification, the potential for miscarriages of justice increases. This ambiguity could erode accountability, public trust in the legal system, and infringe upon individuals' rights to due process, fair adjudication, and review of grievances such as in appeal processes.

#### 2.2.5. *Data protection challenges*

GenAI systems are trained on vast datasets that often contain personal data input by developers or users (Gupta et al., 2023). This data includes personal identifiers such as race, gender, and age; it can even contain information provided by users when interacting with the GenAI through prompts. This has raised significant privacy concerns for individuals, industry, and government alike. In fact, a 2024 report by Cisco from 2600 security and privacy professionals in 12 countries (5 Europe, 4 Asia, and 3 Americas) found that many businesses are limiting the use of GenAI and 27% have temporarily banned the technology's use given privacy concerns (Cisco, 2024).

The incorporation of GenAI into judicial decision-making, therefore, raises data privacy and protection challenges due to its need to collect, store, and analyze sensitive personal data in order to provide its responses and increase its efficiency and accuracy. These information processes, necessary to make any legal GenAI functional, require robust safeguards to ensure general compliance with data privacy regulations, and to protect individuals' rights to privacy and data protection.

More specifically, these risks exist at various stages of the data processing progression. First, regarding the introduction of information into the system, the input of vast judicial and extra-legal data into a GenAI LLM raises concerns regarding the sourcing and handling of this data, including whether informed consent was obtained for its use and whether relevant privacy rights were respected. In the U.S. Supreme Court's Year End Report, Chief Justice Roberts notes hesitations from legal scholars regarding "whether entering confidential information into an AI tool might compromise later attempts to invoke legal privileges." (Roberts Jr., 2023) Second, the storage and safeguard of data used in training and fine-tuning these models is critical for both individuals' rights and the efficiency of the relevant GenAI. Data breaches or unauthorized access to these datasets can result in the exposure of sensitive information, leading to potential data protection and privacy violations. Third, GenAI systems may inadvertently generate content that discloses personal or confidential information. This could occur due to the inherent limitations of GenAI LLMs in understanding the nuances of context and privacy. For example, a GenAI system used to analyze pre-trial documents might unintentionally generate summaries that reveal sensitive details about individuals involved in a case. These privacy violations may also be connected to human-machine interactions, as when users—for instance, judges or other judicial office-holders—interact with a GenAI system, they may provide sensitive information such as personal identifiers or confidential details, which the system would then incorporate into its database and responses. The lack of transparency in data processing and results<sup>13</sup> may make it challenging to pinpoint specific data origins for iterated generative output. This opaqueness could lead to difficulties in addressing potential data privacy breaches or ensuring individuals' right to the elimination of personal data or rectification of inaccurate information.<sup>14</sup>

To mitigate the data protection risks posed by GenAI in judicial decision-making, and ensure the responsible and ethical use of AI data retrieval, processing, and storage, it is imperative to establish robust data management protocols within the judiciary that prioritize transparency and implement comprehensive data governance strategies.

<sup>13</sup> See generally discussion *supra* Section 2.2.1.

<sup>14</sup> For an example of these rights see Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1, Arts. 16–18 (giving individuals the right to be "forgotten" via the elimination of their data, for the data to be rectified, and to restrict its processing).



### 2.2.6. Judicial independence

GenAI poses a potential threat to judicial independence by introducing the risk of undue influence on legal reasoning. Judges who rely on GenAI-generated content without critical analysis may inadvertently allow underlying biases, historical data patterns, or specific ideological perspectives embedded within the algorithms or the training data to shape their decisions in ways that could undermine their independent judgment. This could lead to undue influence and subconscious bias in decision-making, undermining the fundamental principle of independent judgment and potentially resulting in rulings influenced by algorithmic inertia rather than objective legal interpretations (Gautam et al., 2024).

Furthermore, the reliance on GenAI in judicial decision-making, particularly within the current trend of using non-legal technologies in the courtroom (i.e., ChatGPT use by judges),<sup>15</sup> raises concerns about the potential for an arbitrary homogenization of legal outcomes. Though non-arbitrariness is a core principle to be pursued in the courtroom, it is necessary to address each case independently and impartially, as well as according to its uniqueness—and the uniformity of AI-generated content could overshadow the richness of judicial thought that arises from diverse perspectives and context-analysis. As previous scholarship has pointed out, not everything in the law could and should be automated (Deakin and Markou, 2020). Uniformity and non-legality of sources and GenAI responses have the potential to undermine the core principles of independent, case-by-case adjudication, and judicial discretion, which relies heavily on nuanced interpretations and impartial consideration of unique circumstances, as well as on very precise information sourcing (like jurisdictional hierarchies and precedent considerations) to be the basis of decision-making (Re and Solow-Niederman, 2019). These challenges to rationale independence and case individuality are raised when predictive AI is used in judicial decision-making. However, based on the case studies detailed in this paper, we argue that, without a proper framework and guidance to use AI in the courtroom, judges may fall in the trap of *technological solutionism* (Morozov, 2013) given the user-friendly features of tools such as ChatGPT and the promises of the technology.

### 2.2.7. Access to justice

Intuitively, we tend to think that the use of advanced information technologies increases efficiency across sectors. In the context of AI, this has been advocated by consultants, investors, user companies, and even proven by academics in different contexts (Tarafdar et al., 2019). Regarding GenAI in judicial decision-making, however, these theoretical benefits of cost reduction and efficiency need to be analyzed under a different perspective, one that is inclusive of more metrics than processing efficiency—as the desired output of a judicial process is not just delivering an outcome, but a fair one both formally (in process) and materially (in adjudication). The increased use of AI by judges and attorneys can increase disparities among and within legal systems (Grossman et al., 2023), threatening access to formal and material justice, with the potential to have macro and micro-scale impacts. At a macro-scale, nations with more resources can develop and integrate GenAI tools at a faster rate, increasing their efficiency, and therefore, also increasing the gap between different jurisdictions' factual access to the system. On a micro-scale, individuals accessing the justice system without GenAI tools can be hindered in contrast to those with resources to access these tools and utilize them at the service of their legal arguments. Contrary to perceived ideals of increased access to justice, jurisdictions struggling with poor access to justice may suffer even more due to these increased competitive costs or lack of technical expertise. For instance, the state usually provides defense attorneys at no cost for defendants who lack defense counsel. However, the increased hidden costs (Grossman et al., 2023) of deploying GenAI could frustrate efforts to build the necessary capacity for providing technological aid to defendants who need it. Hence, jurisdictions that use GenAI unsystematically, without full consideration of the plethora of challenges posed by GenAI, are likely to face these challenges in the near future.

For GenAI to have a positive impact on access to justice, it is necessary to take a conscious and unified strategy towards deploying GenAI within the context of each jurisdiction. However, this also raises

<sup>15</sup> See discussion *infra* Section 3.

concerns about differences in capability for GenAI implementation being paired with resource-capabilities. For instance, Singapore is testing the use of GenAI in the small claims tribunal first—which sees about 10,000 cases a year—with an eye to expanding it to other areas, such as divorce maintenance and civil claims (Ocampo, 2023). Singapore is adopting a unified strategy on the use of GenAI in courts, experimenting with the technology while pondering the potential risks it could create, and developing its own tailored GenAI system under a two-year memorandum of understanding with Harvey AI, an American legal tech start-up (Ling, 2023). The model being developed will advise users on their rights, guide them through the processes of a legal trial, and even estimate how much a user may be able to claim (Ling, 2023).

Another jurisdiction actively exploring the use of AI in the administration of justice is Estonia, the second-fastest country in Europe in terms of court proceedings (Invest in Estonia, 2022). The Estonian Ministry of Justice has expressed keen interest in exploring AI projects to identify areas where AI can be beneficial. While AI-powered judicial decisions or the use of GenAI tools are not currently deployed in Estonian courts, the Ministry has expressed its openness to exploring these possibilities in the future to better support judges and court officials.

Well-resourced countries like Singapore and Estonia are taking a proactive, cohesive, and state-led approach. They are investigating and investing in GenAI development and implementation, ensuring robust data quality control, and tailoring the technology for the legal system. This approach has the potential to improve efficiency and fairness in the legal processes. On the other hand, the case studies analyzed below, localized in Mexico, India, Colombia, and Peru, exhibit a less cohesive, ad-hoc approach in their GenAI implementation strategies. In these jurisdictions, judges have deployed GenAI on a case-by-case basis, often without considering the potential risks and with little regulatory guidance. This disparity raises concerns about resource availability and its impact on access to justice: Wealthier nations, like Singapore, can invest in robust data quality control and develop tailored GenAI systems, potentially leading to fairer and more efficient legal processes. However, resource-constrained countries might struggle to implement all relevant safeguards, hindering their ability to integrate GenAI responsibly, or at all. This creates a worrying scenario where access to justice becomes increasingly dependent on a nation's financial resources, potentially widening the gap among jurisdictions' judicial efficiency, and for those already struggling within the legal system.

### 3. GenAI at the bench: Different Approaches Across Jurisdictions

There are several approaches to integrating AI into the courtroom, each reflecting differing regulatory strategies and levels of preparedness for technological advancements in the judicial system. This section explores this responses by (1) examining the current state of regulatory approaches to GenAI in judicial decision-making across various jurisdictions, and (2) delving into real-world examples where judges have begun to integrate GenAI into their decision-making processes, despite the absence of formal guidelines. Initially, the section highlights various approaches to regulating AI, including tailored regulatory frameworks, general non-targeted regulations, integration within existing legal structures, and prohibitions. This overview aims to elucidate global strategies for managing AI development and implementation. Subsequently, the section delves into cases of unregulated deployment of GenAI in the courtroom. By examining the selected case studies of Colombia, Peru, Mexico, and India, where judges have used GenAI lacking specific regulatory frameworks, we can gain valuable insights into the evolving applications of GenAI within the judicial system. This exploration not only sheds light on the potential benefits of GenAI in legal proceedings, but also highlights the challenges and risks that can arise when its use precedes the establishment of clear guidelines.

The potential of GenAI to revolutionize the legal practice generally and judicial decision-making specifically, has captured the imagination of legal scholars and technologists alike. It is, therefore, unsurprising that applications of this new technology are already permeating through the courts. An early application of the technology in 2023, in the case of *Mata v. Avianca, Inc.* (FindLaw, 2023) highlighted potential pitfalls associated with using GenAI within a judicial context. An attorney in this

case used GenAI to draft legal documents that were subsequently identified as containing case citations and legal opinions that were either incorrect or fabricated (FindLaw, 2023). In 2023, sanctions were imposed on the two lawyers involved in this process (Merken, 2023). This case raised serious concerns regarding the ethical use of GenAI in legal practice—at least in the present developmental stage of the technology and before protective guidelines and standards are developed. It is not only lawyers who have been keen to experiment with GenAI systems and their potential benefits in the legal arena, but judges are also commencing to use these technologies. Despite the informal integration of GenAI into legal practice, jurisdictions have remained ununified in their responses to the technology's use in judicial processes.

### 3.1. *Proactive regulatory approaches across jurisdictions*

Some jurisdictions have taken a proactive and unified pathway towards developing guidelines for the use of GenAI—or AI more generally—in court decisions. These jurisdictions can be generally divided into four strategic approaches: First, those that have adopted a targeted approach to GenAI through specific regulation. Second, jurisdictions that have adopted a general AI guideline-based approach, not specific to judicial decision-making, but that will exert legal influence over the space. Third, jurisdictions that are developing and deploying their own GenAI systems to assist judges. And fourth, jurisdictions that have preferred to ban the use of AI to assist courts altogether or as a temporary measure.

#### 3.1.1. *The specialists: A targeted approach to the use of Gen AI*

One approach to integrating GenAI into the courtroom involves creating targeted regulations specifically designed to address the unique risks and challenges the technology presents within the judicial system. To this effect, jurisdictions such as the United Kingdom (UK) and New Zealand have published guidelines for judicial office-holders to counsel and instruct them in the use of AI.

The UK's "Artificial Intelligence Guidance for Judicial Office Holders" (December 2023) approves the use of AI in courts, overall encouraging caution (Courts and Tribunals Judiciary of the United Kingdom, 2023). The document provides general direction for judges and judicial office-holders on the potential uses and risks of AI in Courts and Tribunals, while specifically advising against AI use in certain cases, including legal research and analysis, due to limitations in verification and inferential thinking capabilities, as these systems "do not produce convincing analysis or reasoning" (Courts and Tribunals Judiciary of the United Kingdom, 2023).

New Zealand's "Guidelines for use of generative artificial intelligence in Courts and Tribunals" (December 2023) takes a more targeted approach. Addressing judges and judicial staff, these guidelines focus on risks associated with GenAI chatbots, such as ChatGPT, Bing Chat, or Google Gemini (Courts of New Zealand, 2023). The guidelines acknowledge severe limitations of publicly available GenAI chatbots, stating that "[e]ven with the best prompts, the output may be inaccurate, incomplete, misleading, or biased" (Courts of New Zealand, 2023). To mitigate these risks, the guidelines restrict judges from feeding data into an AI chatbot that is not already in the public domain, preventing them from "enter[ing] any information that is private, confidential, suppressed or legally privileged" (Courts of New Zealand, 2023). Notably, these guidelines do not currently require disclosure of GenAI use to the public when utilized in judicial decision-making.

#### 3.1.2. *The generalists: A guidelines-based approach to the use of AI*

A number of jurisdictions have introduced broad guidelines for the use of AI and automation tools in courts, but these documents often do not specifically address the unique risks posed by recent advancements in GenAI. This generalist approach to AI regulation avoids focusing solely on judicial applications, opting instead for broader frameworks that apply across sectors. While this may encourage innovation and flexibility, it also risks overlooking the specific challenges and ethical concerns raised by GenAI, particularly in the context of sensitive judicial decision-making.

The EU for example, has taken a generalist approach to regulating AI in judicial decision-making through the risk-based regulatory framework applicable to AI systems. The EU's Artificial Intelligence Act (AIA) (European Commission, 2024) regulates high-risk AI systems across different industries. While the AIA does not explicitly target courts, its categorization of AI based on risk could indirectly impact how courts utilize this technology. For instance, if a court uses an AI system for risk assessment in criminal sentencing, it might need to comply with stricter regulations under the AIA, such as robust data quality control or ensuring human oversight in the decision-making process (European Commission, 2021).

It is important to note the existence of the European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment (The European Commission for the Efficiency of Justice, 2018). This soft law instrument, issued by the Council of Europe, outlines ethical principles for AI use in courts. This document provides foundational ethical principles that are consistent with the AI Governance principles published by several jurisdictions, agencies, and the private sector globally (Habuka & Socol, 2024).<sup>16</sup> The document also explains the need for an in-depth public debate on these tools prior to the implementation of public policies for their development (The European Commission for the Efficiency of Justice, 2018). Even though GenAI may not have been as popular or developed as it is presently when the charter was published, the text encourages the use of chatbots to facilitate access to the various existing sources of information and also the use of AI to generate document templates like court applications or lease agreements (The European Commission for the Efficiency of Justice, 2018). However, the charter is not legally binding and predates the recent surge in GenAI technology. While this document is a valuable starting point for ethical discussions, it may not fully address the specific considerations surrounding GenAI in judicial settings. Therefore, a generalist approach relying solely on existing regulations might need to be supplemented with additional guidance or regulations specifically tailored to address the unique risks exacerbated and opportunities presented by GenAI in judicial decision-making, particularly as the technology continues to evolve.

### 3.1.3. *The deployers of GenAI*

One approach to AI governance goes beyond regulation, emphasizing the creation of clear guidelines alongside the proactive development and integration of responsible AI systems within specific sectors. Notably, as aforementioned,<sup>17</sup> Singapore is in the process of developing its own customized GenAI system through a two-year memorandum of understanding with Harvey AI. This model incorporates built-in features designed to mitigate risks and enhance rights protection, including advising users on their rights, guiding them through legal processes, and even estimating the value of potential claims (Ling, 2023). Given the short timeline since the announcement of these measures, which were made public in 2024, the full impact of these developments on the administration of justice and the effectiveness of risk-mitigation strategies remain to be seen at the time of this article publication.

### 3.1.4. *The deterrent approach*

Some jurisdictions have opted for a more cautious strategy, advocating for a temporary ban or discouragement on the use of GenAI in courts until a more robust regulatory framework, or certifiable precautionary measures, can be established. In December 2023, the Canadian Federal Court banned the use of AI in making “judgments and orders” until a public consultation on the matter occurred (Federal Court of Canada (Cour fédérale), 2023). In its reasoning, the Court acknowledged “the potential for AI to impact adversely on judicial independence” (Federal Court of Canada (Cour fédérale), 2023). Further, the Court recognized the risk that “public confidence in the administration of justice might be undermined by some uses of AI” (Federal Court of Canada (Cour fédérale), 2023). This cautious approach, exemplified by the Canadian case, prioritizes mitigating potential risks to principles like judicial independence and public trust, over swiftly

<sup>16</sup> Transparency, fairness, accountability, human in the loop, and explainability.

<sup>17</sup> See discussion supra Section 2.2.7.

embracing GenAI's promise. It also underscores the importance of public consultation, ensuring civil society has a voice in shaping how, or even if, GenAI will be integrated into the judicial system.

### 3.2. *Unregulated approach to generative AI in judicial decision-making*

While some jurisdictions have begun to explore the use of GenAI in judicial settings, many have yet to establish clear regulatory frameworks to govern its implementation. The jurisdictions examined in the case studies below have not issued specific guidance—whether binding or otherwise—on the development and deployment of GenAI in judicial decision-making. In Colombia,<sup>18</sup> Peru (Juzgado Civil Transitorio de San Juan de Miraflores, *n.d.*), Mexico,<sup>19</sup> and India,<sup>20</sup> where there have been documented instances of judges using GenAI in the courtroom, there are currently no targeted safeguards in place to ensure its fair use. These uses of GenAI have emerged in the absence of regulations tailored to the technology, as well as the lack of comprehensive information technology laws that outline safe use standards or establish clear guidelines for GenAI within the judicial context.

#### 3.2.1. *Colombia*

Colombia has had two high-profile cases involving the use of GenAI in judicial decision-making. The first of these cases revolves around a court document released on January 30, 2023, in the case of *Salvador Espitia Chávez v. Salud Total EPS (2023)*. In this case, Judge Juan Manuel Padilla issued a ruling on whether a child with autism should be eligible for medical treatment coverage. Believed to be the first documented case of a judge utilizing GenAI in the courtroom, the judge openly acknowledged that he used ChatGPT to guide his decision rationale and published the questions and answers of the interaction that he had with the GenAI LLM. The judge asked ChatGPT four questions related to the following items:

- (1) Related to whether paying moderating fees is exonerated in the case of therapy for a child with autism.

*Response summary: According to Law 1753 from 2015, minors with autism are exonerated from moderating fees, binding private and public entities.*

- (2) Related to granting tutelage processes.

*Response summary: Tutelage processes may be granted so that the entity providing services may be compelled to provide therapy.*

- (3) Related to whether the moderating fee may be a barrier to access health services.

*Response summary: Moderating fees may be a barrier for minors with autism in accessing health services.*

- (4) Related to jurisprudence of the constitutional court in similar cases.

*Response summary: There have been favorable cases from the constitutional court (one of the sources of legitimate decision-making in this regard) made in favor of minors with autism within this context.*

<sup>18</sup> Note that while Colombia Law 2213 of 2022, which was developed in response to the COVID-19 pandemic, permits the inclusion technological innovation into judicial decision-making with few limitations on technologies that may be utilized, and methods of deployment, provided it makes judicial processes nimbler and increases access to justice, Colombia does not have any law or guidance related specifically to AI or GenAI in judicial decision-making. [Research complete, formal citation to be provided].

<sup>19</sup> [Research complete, formal citation to be provided].

<sup>20</sup> [Research complete, formal citation to be provided].



While the judge emphasized that the LLM did not replace his own judgment, and that it was deployed to optimize time spent writing sentences with previous corroboration of the information provided by the GenAI system, the unregulated use of the technology still raises concerns. The primary concern lies in the potential for biased or unreliable data to influence critical judicial decisions. Without clear legal or ethical guidelines, GenAI tools could perpetuate existing biases or generate flawed outputs based on inaccurate information. This is especially true in instances with heightened sensitivity such as this process, as it involves the medical care of a minor with autism. Especially in cases where minors, persons with disabilities, and other marginalized or vulnerable groups are concerned, the use of GenAI that may affect ruling outcomes, potentially signaling that they should be especially scrutinized. Additionally, using a GenAI LLM like ChatGPT, which does not disclose the origins of its responses and that is connected to extra-legal sources (i.e., encompassing information that extends beyond laws and previous cases), can introduce inaccurate or non-legal information into a response, potentially undermining the legality and accuracy of the response. This is a concern that permeates throughout these case studies, as all judicial office-holders alluded to utilizing ChatGPT. Furthermore, Judge Padilla's case highlights a secondary risk—the erosion of public trust. The media's criticism surrounding the use of GenAI (Gutierrez, 2023), even for efficiency purposes, underscores the public's desire for transparency and established safeguards when such technology is integrated into the justice system.

Ten days later, in a second Colombian case, Magistrate María Victoria Quiñones from the Administrative Tribunal of the Magdalen issued a court order in a case where Union Temporal de Servicios Integrados y Especializados de Tránsito y Transporte de Santa Marta—SIETT filed a direct remedy claim (*reparación directa*) against Colombia's National Police. As in *Salvador Espitia Chávez v. Salud Total EPS*, ChatGPT prompts were again transcribed (2023). In this case, Magistrate María Victoria Quiñones used ChatGPT to address technical questions that assisted in planning and conducting a judicial hearing within the Metaverse. The AI's responses helped in understanding certain concepts related to the metaverse and in organizing the hearing. The questions asked were the following:

1. What is an avatar?
2. What is the most effective method to verify the authenticity of those connecting to a meeting and/or virtual hearing?
3. Methods for verifying the authenticity of the avatar in the metaverse.

On February 15, 2023, the relevant hearing was held through Meta's *Horizon Workroom* (Meta, 2024), and was livestreamed through YouTube (Victoria en línea-Vicky Quiñones, 2023). In this case, despite serious concerns over privacy and authenticity when conducting legal proceedings in the Metaverse, there is no indication of vetting or verifying the questions directed to ChatGPT. To date, no other judge in Colombia has replicated this approach, and no other audiences have taken place in the *Metaverse* (Ritterbusch and Teichmann, 2023), or other virtual reality (VR) platforms.

The judge's reliance on ChatGPT to navigate the complexities of a virtual courtroom raises potential risks, including the LLM's limited understanding of procedural nuance, its potential focus on efficiency over security, and the overlooking of other technical legal implications. First, LLMs excel at processing vast amounts of information, but they may struggle with the legal system's inherent subtleties and the ethical preservation of rights. The complexities of legal systems and due process could be misinterpreted by an LLM, leading the judge to overlook crucial steps for ensuring a fair trial. Second, LLMs are often optimized for efficiency and may prioritize completing tasks over rigorously evaluating security protocols. For example, an LLM might suggest avatar verification methods that seem efficient but lack robustness, leaving the court vulnerable to identity theft or manipulation. Finally, an LLM might not anticipate secondary potential legal risks of using avatars, including risks in evidence presentation, adversarial procedural issues, or witness independence, comprehensive treatment, and potential intimidation, among others. These risks highlight the danger of relying on LLMs for judicial decision-making, even in focused procedural contexts, as these may have significant consequences on procedural rights, adjudicatory outcomes, and the overall fairness of the judicial process.

### 3.2.2. Peru

On March 27, 2023, Judge Frank Paul Flores Garcia used GenAI to support his sentencing rationale and legal basis in a civil law case related to child support obligations for a mother and father with respect to their daughter (Juzgado Civil Transitorio de San Juan de Miraflores, [n.d.](#)).

To determine the proportion of child support that each parent should bear according to their capacity, the judge consulted ChatGPT for guidance on the relevant mathematical operations. The judge did not share any specific details about the prompt(s) used or the answer(s) he received from the system. He merely noted that he was assisted by “ChatGPT,” and that it was “corresponding to apply the technique of mathematical proportion to the effects of establishing which is the amount corresponding to each parent, according to their income, to satisfy the expenses of maintenance of their daughter” (Juzgado Civil Transitorio de San Juan de Miraflores, [n.d.](#)). The judge alluded to ChatGPT as a foundation for his legal reasoning, and footnoted a transcribed URL (Juzgado Civil Transitorio de San Juan de Miraflores, [n.d.](#)). Judge Flores Garcia may not have been aware that this footnote provided no insight into his consultation, given that readers cannot access his personal record with ChatGPT.

This raises concerns about transparency and visibility in decision-making mechanisms, alongside legal education and technological competency matters. Integrating AI into judicial decision-making without a proper understanding of its inner workings and limitations can lead to opaque processes that infringe upon individuals’ rights to due process and fair adjudication. The potential for the manifestation of risks such as unintended bias amplification, judicial independence, data protection, or simply judicial error within such AI systems, necessitates robust frameworks to safeguard the fundamental rights and liberties of all individuals that interact with the justice system—requiring solutions that include enhancing legal education surrounding GenAI, its uses in the courtroom, and standard transparency and use-case disclosures.

### 3.2.3. Mexico

During a hearing on March 29, 2023, Magistrate Reyes Rodríguez Mondragón, president of the Superior Chamber of the Electoral Tribunal of the Mexican Judiciary, shared that he had consulted ChatGPT on his phone on the matters concerning the hearing discussion (Tribunal Electora TEPJF, [2023](#)). Magistrate Reyes disclosed his consultation with ChatGPT and read some of the outputs that he obtained to illustrate (i) the material substance of his point, and (ii) how ChatGPT could be of help in being exhaustive and complete in legal arguments.

In this appeal, the Electoral Tribunal was discussing the use of the expression “you know who” (in Spanish, “*ya sabes quién*”) (López Ponce, [2023](#)) when used within the context of an advertisement forming part of an electoral pre-campaign of the Morena political party. The Electoral Tribunal was deciding whether the use of the expression “you know who” could be interpreted as a signal of support by President Andrés Manuel López Obrador. Magistrate Rodríguez noted that a cursory analysis on ChatGPT had quickly allowed him to ascertain the relationship between the expression and its political references. The Magistrate argued that these new technologies may be a useful tool in satisfying certain judicial duties—particularly in pursuing seemingly comprehensive and thorough legal research. In fact, he noted that the decision subject to appeal had a judicial defect and should be revoked because it was incomplete, given that there was no contextual analysis of the “you know who” expression and thereby failed to be legally “complete.”

Magistrate Rodríguez stated that “the Electoral Tribunal of the State of Mexico reiteratively is not complying with the principle and duty of exhaustiveness (...) it is not providing sufficient reasons (...) when even technology now facilitates a series of information, obviously processing databases and all the knowledge that is available to the courts.” This implies that, in order to (i) satisfy the duty of exhaustiveness and (ii) provide sufficient reasoning, Magistrates may need to utilize modern technologies in their adjudication process and legal reasoning substantiation.

However, in this same hearing, Magistrate Vargas contested Magistrate Reyes with regard to the inclusion of ChatGPT as a foundation for legal completion. Magistrate Vargas noted: “I would like to think

that what you have just told us is simply an isolated example and is not a forecast of what will be the jurisprudence of this court (...) because I would be concerned that now our resolutions will be taken based on what ChatGPT says.” He also inquired as to whether, in the future, courts will remain relevant by that logic since “ChatGPT [could] tell us everything” (Tribunal Electora TEPJF, 2023). Magistrate Vargas further noted how ChatGPT may contradict the principles of presumed innocence, presentation of proof, and valuation of evidence, all of which are central to the fair functioning of courts (Tribunal Electora TEPJF, 2023).

While AI in legal reasoning promises to be a resource for exhaustive information retrieval and thorough research, it raises ethical concerns regarding fundamental legal principles, potential bias, the erosion of judicial discretion, and the displacement of the human element in interpreting complex legal nuances. This case from the Mexican Judiciary highlights these concerns in sharp relief. Magistrate Rodríguez’s use of ChatGPT during a hearing, while intended to pursue comprehensiveness, underscores the potential for such tools to undermine core legal principles. Magistrate Vargas’ critique raises valid concerns about relying on AI outputs as a substitute for critical analysis and independent judicial reasoning.

#### 3.2.4. India

In 2023, High Court judge Anoop Chitkara refused bail to a man accused of rioting, criminal intimidation, criminal conspiracy, and murder (Times of India, 2023). Judge Chitkara recognized that he had turned to ChatGPT to help guide his reasoning, though he noted that he did not rely on ChatGPT to decide his ruling. Instead, in substantiating his legal basis, he consulted with the GenAI tool to help determine if he was relying too heavily on his own “consistent view” developed from past cases, concerning allegations involving an unusually high level of cruelty, and whether they should count against granting bail (Smith et al., 2023). In this regard, it seems that there is a claim made that by using GenAI, the Judge was not only supporting his rationale but also enhancing its impartiality, ensuring that his decision-making was more balanced and less influenced by personal bias.

Judge Chitakara’s disclosed his consultation with ChatGPT, to which he presented the following prompt: “What is the jurisprudence on bail when the assailant assaulted with cruelty?” ChatGPT’s transcribed response stated that “If the assailants have been charged with a violent crime that involves cruelty, they may be considered a danger to the community and a flight risk. In such cases, the judge may be less inclined to grant bail or may set the bail amount very high to ensure that the defendant appears in court and does not pose a risk to public safety.” (Suryam, 2023) The LLM added that presumption of innocence is also a fundamental principle of the justice system and that bail is often granted if the judge is sure the accused does not pose a risk to public safety. Judge Chitkara decided that the accused acted with cruelty, subsequently denying the bail request.

This Court made it clear that any reference to ChatGPT was only intended to present a broader picture on bail jurisprudence where cruelty is a factor. Much like the case studies above, this approach on the use of AI in judicial decision-making has not been in accordance with any GenAI guidelines of judicial AI deployment or broader legal architecture in the country, nor is it consequent with a comprehensive strategy for all judges to consider its implications and correct use of the technology: GenAI was deployed, as in the above-noted cases, circumstantially and on an ad hoc basis. This ad hoc method of deployment of GenAI in judicial decision-making, without a clear or unified regulation for risk mitigation, a comprehensive framework, or guiding principles, poses numerous risks, including unequal application of the law due to a lack of standardization, the undermining of judicial discretion and potentially accuracy through reliance on non-legal algorithmic tools, the exacerbation of existing biases embedded within the algorithms themselves, and the creation of opaque decision-making processes—both at the judicial and algorithmic levels.

In addition to these concerns, this case, like the ones above, used ChatGPT, an open network that accesses all information on the internet. This raises concerns about data quality and the legal competence of the GenAI to present legal opinions. ChatGPT and similar open-source AI tools are not specifically

designed for legal research or reasoning. They may process vast amounts of information, but they lack the ability to critically evaluate the legal relevance and accuracy of that information. This can lead to misleading or unreliable outputs, potentially influencing judicial decisions in unintended ways. Judge Chitkara's reliance on ChatGPT to check his own "consistent view" exemplifies this risk. The GenAI's response, while seemingly relevant, does not delve into the nuances of specific legal precedents in India generally or the specific court, or consider the unique circumstances or laws applicable to the case. This highlights the need for GenAI tools specifically designed for the legal domain, trained on high-quality legal data and capable of providing contextually relevant legal reasoning. The ad-hoc use of such tools, as seen in the case studies presented, further underscores the need for comprehensive frameworks. These frameworks should establish clear guidelines for data quality, address potential biases within the AI, and ensure transparency in GenAI application.

#### 4. Risk mitigation strategies: Early lessons and future direction

While the potential benefits of GenAI in judicial decision-making are undeniable, its integration necessitates a multifaceted approach to risk mitigation. As seen in the case studies presented above, utilizing GenAI on an ad hoc case-by-case and judge-by-judge basis, without due consideration of the potential pitfalls and risks associated with the technology, should be cause for concern as it could threaten certain aspects of the rule of law and infringe upon fundamental liberties that should be protected throughout the legal process. While GenAI integration in courtrooms promises increased efficiency, streamlining the legal process must not come at the expense of fairness and transparency. The judicial system, in this regard, serves a dual purpose: It should provide timely rulings, but also ensure its rulings are fair and delivered, upholding the principles of due process. GenAI tools should be evaluated not just by their ability to expedite decisions, but also by their contribution to a just and equitable legal system.

Therefore, in order to harness the promise of GenAI's potential while mitigating its risks and protecting against its limitations, this section explores two key areas for safeguarding the integrity and fairness of the judicial process. The section proposes a dual-prong approach, based on (i) integrating foundational precautions for GenAI systems that will be utilized in judicial decision-making, and (ii) providing guiding principles for responsible GenAI deployment in courtrooms. The article first proposes ex-ante essential design features that must be woven into the fabric of GenAI systems intended for judicial use, such as verification and licensing frameworks, data-set standards, clear allocation of responsibility, explainability of outcome, and algorithmic fairness; and then provides ex-post practical considerations for implementing and utilizing GenAI models in the courtroom throughout the lifecycle of the technology, such as enhancing legal education, deploying GenAI on a risk-tiered basis, establishing disclosure and accuracy-verification standards, protecting specific procedural rights (including information and consent rights related to the use or scope of the use of AI), and conducting ongoing iterative audits and review processes. By addressing the implementation of GenAI in the courtroom at both the development and deployment stages, we aim to establish a comprehensive framework for responsible and trustworthy GenAI integration within the judiciary.

##### 4.1. Foundational standards

This section outlines core ex-ante standards to consider for the development and deployment of GenAI systems intended for judicial use. These safeguards represent a baseline for jurisdictions to consider when integrating GenAI technologies into their courtrooms.

###### 4.1.1. Capacity assessment

The potential of GenAI in courts is undeniable, but its effective integration hinges on a crucial factor: capacity. Before embracing GenAI, jurisdictions must undergo an assessment to determine their readiness for this complex technology. Key considerations for capacity assessment include infrastructure, expertise, and overall resource analysis to determine the capabilities met—and required—for responsible GenAI

implementation. Utilizing GenAI necessitates a nuanced comprehension of the technology's inherent risks, alongside a jurisdiction's capability to not only factually implement the technology but mitigate these perils. Jurisdictions must assess whether they possess the requisite expertise within their judicial systems or if more investments in training, recruitment, or regulatory or infrastructural development are necessary. Furthermore, as implementation progresses, jurisdictions should conduct evaluations to ascertain the compatibility of their existing technology with GenAI systems, identifying necessary upgrades or modifications for seamless integration.

This evaluation process is of particular relevance for developing economies grappling with resource constraints in GenAI deployment. As part of capacity-building efforts, emerging markets could pursue strategies including in-house development, or fostering public-private partnerships aimed at driving digital government reforms and surmounting implementation challenges. Unfortunately, such collaborations are infrequent: According to the UN eGovernment Surveys covering 193 countries, only 42 countries participated in co-creating education e-services, while a small minority of 22 countries engaged in collaborative development of justice-related e-services (OECD, 2022). Capacity assessments and consequent development toward covering personnel, technology, or regulatory and infrastructure needs can help effectively integrate GenAI into the courtroom.

#### 4.1.2. *Stakeholder engagement*

A proactive approach to stakeholder engagement in developing guidelines and AI use cases for courts offers significant long-term benefits. In fact, it can be considered a foundational element for implementing GenAI in the judicial system inclusively. Stakeholder engagement brings a valuable diversity of perspectives to the table. Judges, lawyers, court staff, litigants, GenAI developers, and the public all have unique insights into how GenAI could impact legal processes and outcomes. The importance of stakeholder engagement goes beyond mere practicality: The justice system relies heavily on ongoing public trust and legitimacy. Actively engaging stakeholders can help ensure that the implementation of GenAI fosters acceptance amongst those it affects and upholds the integrity of the judicial process. Stakeholder engagement strategies could include opening public forums and educational workshops, with sessions informing citizens about GenAI's potential applications in court and addressing concerns about fairness, transparency, and data privacy; consulting with legal professionals and bar associations, engaging judges, lawyers, and legal experts who can provide valuable insights on how to integrate GenAI responsibly while maintaining ethical legal practices; and collaborating with civil society organizations such as academia, NGOs, and advocacy groups, ensuring a diverse range of voices are heard and potential negative impacts on marginalized communities are considered. This emphasis on stakeholder engagement is not limited to GenAI in courts; it extends to broader AI adoption. The call for public participation in AI governance aligns strongly with a public value management approach to public administration. This approach prioritizes transparency, accountability, and citizen participation, potentially influencing the consolidation of robust and inclusive AI governance frameworks and more informed decision-making by public authorities (Wilson, 2022).

#### 4.1.3. *Licensing and verification*

To ensure only qualified GenAI systems enter courtrooms, this article presents an approach based upon an ex-ante verification process guaranteeing that any GenAI utilized in judicial decision-making meets a set of basic standards—both for third-party developers integrating a GenAI into the judicial system, as well as for GenAI developed in-house within the judiciary structure. While this proposal could be applicable to any use of AI by courts, it is particularly relevant in the context of GenAI given the extra-legal nature of the data sources in current models being utilized by judges, as explored in the case studies explained above.<sup>21</sup>

<sup>21</sup> See generally discussion *supra* Section 3.2, making reference to judges and magistrates using ChatGPT to guide their legal and process reasoning.



A licensing and verification regime should be established to assess GenAI systems. This ensures a high bar for functionality, adherence to foundational legal principles, and minimizes potential risks in judicial processes and outcomes. The integration of a GenAI into judicial decision-making would require meeting the standards of elements such as the following:

*Data quality:* A core focus should be on the quality and comprehensiveness of the data used to train GenAI algorithms. High standards for data inclusion are crucial, ensuring that GenAIs which will be used in court are trained on accurate, unbiased, and complete datasets that reflect the real-world situations these AI systems will encounter in judicial decision-making. This includes both legal data and any relevant extra-legal data (if any) that will be introduced into the algorithmic system and may inform its output.

*Algorithmic fairness and bias mitigation strategies:* The verification process should scrutinize the algorithms for potential biases and ensure developers have implemented techniques like fairness testing and de-biasing methods to minimize discriminatory outcomes. For example, data gathered and processed to be utilized in assessing risk of recidivism should be representative and mindful of diverse populations and avoid perpetuating historical biases by means such as not algorithmically integrating biometrics within the context of judicial decisions, enhancing unbiased data, and further preserving data privacy.

*Explainability of the AI's reasoning process:* Judges and parties involved in legal proceedings need to understand how GenAI arrives at its conclusions. This could involve providing clear explanations of the factors, precedent cases, or laws and regulations considered by the relevant GenAI and the weight assigned to each factor in reaching a decision.

*Security and privacy safeguards for sensitive data:* Robust security measures are essential to protect sensitive data used by GenAI systems from unauthorized access or manipulation. Additionally, privacy safeguards should be in place to ensure personal information is obtained, stored, handled, and processed according to legal data privacy requirements.

*Adherence to ethical guidelines for judicial AI use:* A clear set of ethical guidelines should be established to govern the development and deployment of judicial GenAI. These guidelines could address issues like transparency, accountability, fairness, and the potential for human oversight at certain stages of training and deployment.

This licensing process should be overseen or ratified by a qualified independent body with expertise in AI, law, and ethics. Developers should be required to demonstrate their GenAI systems meet these standards before they can be used in courtrooms.

Additionally, courts that develop their own GenAI models should be subject to rigorous internal verification processes mirroring the external licensing regime that this article proposes. This internal verification procedure would ensure the in-house GenAI systems adhere to the same high foundational standards as those developed by third parties.

#### 4.1.4. *Trusted datasets and explainability*

Integrating GenAI into judicial decision-making necessitates high-quality datasets, as a GenAI system's potential is made possible, as well as limited, by the information it is trained on (Wilson, 2022). Biased or incomplete data can lead the system to perpetuate existing societal biases or generate inaccurate legal reasoning, potentially undermining the fairness and accuracy that are cornerstones of the justice system (Wilson, 2022). As such, when integrating GenAI into the courts, the following factors must be considered:

- *Data Quality:* GenAI systems are only as good as the data on which they are trained. Robust measures must be implemented to ensure that the data used for judicial GenAI is accurate, unbiased, representative, and complete. This requires data quality monitoring and mitigation strategies for

potential biases, comprehensive historical tracking accounting for court hierarchy and current legal developments, and weighted analytics on legal interpretation progressions. Examples of such strategies could include employing diverse data sources, using procedural and material fairness metrics during training, extricating biometric data from judicial decisions, and implementing human-in-the-loop review processes. Further measures may include establishing closed-network datasets, instead of utilizing open-source GenAIs in the courtroom—meaning, deploying only GenAIs trained on a defined and finite amount of trustworthy data (whether legal or extra-legal) that could be jurisdiction or even court-specific, and not the comprehensive universe of the internet.<sup>22</sup> This pre-emptive control must be focused on preventing the “garbage in, garbage out” phenomenon (Lew et al., 2020; Bell et al., 2022), by which low-quality input results in low-quality output. In essence, the success of integrating GenAI into judicial decision-making requires the review and implementation of high-quality datasets. By prioritizing high-quality data and implementing robust safeguards at the informational origin of a GenAI’s content, we can ensure that GenAI in the courtroom augments legal reasoning, reducing the risk of undermining the quality and accuracy of judicial processes and adjudications.

- *Data Access, Explainability, and Origin Visibility:* Transparency and fairness demand clear protocols for data access. Judges and relevant parties should be able to understand the data sources used by the GenAI system and how it may have influenced its outputs. This could involve providing summaries of the data used, allowing for limited access to anonymized data for specific reasons such as tracing purposes, or requiring GenAIs to disclose the data sources and characteristics, inclusive of precedents, laws, and regulations used at arriving at a conclusion. Explainability empowers judges and litigants to understand the reasoning behind a GenAI’s outputs, fostering trust and transparency in the decision-making process, and ensuring a fair legal basis of judicial reasoning, while data access allows for clear and directed audits and verification of the data used to train the GenAI, mitigating potential biases and promoting fairness in the application of the technology.

#### 4.1.5. Allocation of responsibility

A clear and consistent understanding of who bears responsibility for potential GenAI-related damages to the judicial process, its outcome, or its integrity, is critical when introducing GenAI into the courtroom. Responsibility allocation is crucial to account for, mitigate, or remedy potential miscarriages of justice, judicial errors, or any form of harm to judicial systems and processes caused by GenAI use. Accompanying its potential, GenAI includes high-risk features, including hallucination, inferential and reasoning mistakes, or the outputting of biased information, among other potential disruptions to the fairness of the judicial process. Striking a balance between incentivizing responsible GenAI, development, providing reliable and robust accountability networks, and empowering judges to make decisions effectively and efficiently, is key in this space. Overly stringent measures against developers—provided by, for example, a responsibility system that focuses solely on a GenAI’s algorithmic setup and its consequent input to judicial reasoning—could stifle innovation and deter companies from entering the field due to the high risk of extensive liability. Conversely, placing the burden of responsibility solely on judges—through for example, a responsibility system that focuses on the judge’s ultimate decision-making control on a GenAI’s output and its application—might disincentivize them from utilizing GenAI, or produce an overly burdensome review process for GenAI-created content, hindering potential efficiency benefits. Determining the exact level of responsibility for each party can be complex, and clear legal frameworks and guidelines are needed to navigate these challenges. Some considerations to strike this court-and-jurisdiction-dependent balance include:

<sup>22</sup> It is worth noting that all uses in the case studies presented in Section 3 utilized ChatGPT, an open-source LLM, that may have difficulty distinguishing legal and binding input from contextual and non-binding information. See generally discussion *supra* Section 3.

- *Developer Accountability*: Developers should be held accountable for the algorithmic design and functionality of their GenAI systems. This requires clear disclosures of their AI's modeling and reasoning process, enabling scrutiny for potential bias or flaws in the algorithms. However, liability allocation networks must avoid overly punitive developer liability frameworks. These could create high barriers to access this sector, not presenting enough of an incentive for developers to pursue stepping into this space. Clear balancing of responsibility is crucial: Developers should be held accountable for functionality and failure to meet certain measurable pre-agreed upon standards, as well as demonstrably flawed algorithms that lead to harm, but not for unforeseen errors or damages in complex legal scenarios—especially those attributable to judges, or in which the judges would be responsible for accuracy or fairness review. A crucial tool to help balance incentivizing development and maintaining safe judicial thresholds is the ex-ante certification process. Under the certification model developed in this article, developers should be required to undergo a certification or licensing process that assesses their development practices and adherence to certain legal and ethical guidelines for judicial GenAI deployment. This would prevent latent unforeseen liabilities during deployment, and be able to “weed out” potential systemic shortcoming in GenAIs prior to liability scenarios.
- *Judicial Discretion*: Judges retain ultimate decision-making authority, and thus bear responsibility for the process and outcome of their judicial processes. However, absolute responsibility for GenAI outcomes could discourage the judiciary from using GenAI, or slow down judicial processes with overly demanding review processes to GenAI content, thus missing out on leveraging the technology's potential efficiency gains. To address this, judges should be equipped with comprehensive training and resources for critically evaluating GenAI outputs. Instead of full responsibility allocation, liability distribution should focus on a tiered approach integrating (i) training, allowing judges to exercise informed discretion when utilizing GenAI and understand its limitations; and (ii) risk-based responsibility. A risk assessment framework should be implemented, establishing that high-risk cases involving significant potential harm would necessitate a more rigorous review process, whereas low-risk cases would require lesser scrutiny from the relevant judicial office-holders. With this elevated review process and scrutiny obligations on GenAI's output for high-risk cases, there should be an accompanying higher degree of responsibility for the judge, both in (a) the diligence exercised in the examination process of reviewing the GenAI output, and (b) the responsibility allocated in the relevant tiered-threshold. This may even involve discarding the use GenAI in high-risk cases altogether, or at least in specific portions of cases (e.g., not utilizing GenAI in sentencing, but potentially utilizing it in precedent review and summarization), and include semi-automatized or automatized-with-redress review for low-risk cases.<sup>23</sup>
- *Shared Responsibility Model*: This article proposes shared responsibility models, where both developers and the court system bear some degree of liability depending on the specific circumstances of the case and the content of the algorithmic output. This could involve factors like judges' GenAI training, the cases' complexity and risk profile, and the transparency of the GenAI system used. Ultimately, the goal is to establish a framework that encourages responsible development and GenAI integration into the judicial system, empowers judges and allows for increased efficiency, while protecting the integrity of judicial decision-making.

#### 4.1.6. Prompt engineering in the judicial system

A prompt is a natural language text that requests the GenAI system to perform a specific task (Marvin et al., 2023). In all the case studies, it is judges that instruct the system to provide an output, interacting with LLMs via prompts. For GenAI to operate in a desirable manner, it is important to give context and detailed information through these interactions with the systems to produce relevant and effective outcomes. Therefore, prompt engineering is crucial to refine instructions until the judicial officer gets

<sup>23</sup> See generally discussion *supra* Section 2.2.7.

the desired outcomes from the system (Knoth et al., 2024). For instance, in the healthcare sector, prompt engineering has emerged as a promising approach to guiding model generation by providing targeted information-seeking mechanisms. This field proposes prompt design methods for different medical tasks. Namely, it explores how to choose and design prompt elements, how to use prompts to guide a model generation of text that meets medical requirements, and how to evaluate the impact of different prompt designs on model performance (Wang et al., 2023).

In all case studies, all judges used human-designed prompts. Also, all guidelines and policies mapped in this paper assumed that a human will be the one giving the prompt to the GenAI system. Human-designed prompts depend on the human's expertise in selecting relevant information and constructing prompts in a way that will produce valuable outcomes by the model (Wang et al., 2023). According to the emerging literature on prompt engineering, human prompts are particularly effective for tasks with well-defined input data and structured output requirements (Wang et al., 2023). However, designing effective human prompts requires the expertise of someone who understands the specific field where the GenAI model is being used, as well as human-machine interaction. Even though we assume that judges and judicial officers are experts, levels of competency, experience, and subject-matter expertise, as well as requirements to access the judicial profession, vary across jurisdictions and across fields.

Therefore, ensuring the quality of human-designed prompts could be a challenging task. Additionally, the parties affected by the judicial decision are currently not in a procedural position to assess such quality in order to challenge the decision as part of the due process. Due to these limitations of human-designed prompts, researchers are exploring automated prompt design methods to improve the efficiency and adaptability of prompt-based approaches. Automated prompts are generated using various algorithms and techniques, eliminating the need for human intervention, a technique that is already being deployed in sectors such as the medical field (White et al., 2023). Ultimately, the choice of prompt structure and overall systemic design will depend on the specific task and available resources—an area that needs further exploration in the context of judicial systems and the use of GenAI.

## 4.2. *Application principles*

This section delves into the practical principles that should guide the implementation of GenAI within the courtroom, as well as the practical application once a system is deployed. Building upon the established foundational standards, these conventions provide a roadmap for responsible and effective integration of GenAI tools for use in judicial proceedings. Through the application of these principles, and the provision of GenAI-specific procedural rights, the article intends to ensure that GenAI augments human decision-making while safeguarding the core tenets of fairness, transparency, and accountability within the legal system.

### 4.2.1. *Updating ethical standards and enhancing continuous legal education*

With the potential benefits of integrating GenAI into the courtrooms, also come risks—and with these risks, increased duties and responsibilities for judges.<sup>24</sup> The inclusion of GenAI in the judicial system requires a review of ethical standards of the legal profession, alongside holistic legal training to educate current and future judges (and lawyers) on the use and misuse of GenAI in the courtroom. Such training should entail enabling practitioners to manage the existence, capabilities, and limitations of these technologies.

First, legal ethics requirements should be updated to include requirements regarding the use of GenAI. Such requirements could address issues like transparency and disclosure duties in AI use, data privacy considerations, and potential bias management and review responsibilities in AI outputs. Secondly, lawyers and judges should be mandatorily required to learn about these technologies under jurisdiction-wide strategies employed to bridge the gap in the tech savviness of stakeholders in the legal system.

<sup>24</sup> See discussion *supra* Section 4.1.3.

Continuous and updated legal education requirements could be a valuable tool for this purpose. However, this would necessitate building those capabilities from the ground for jurisdictions like Colombia—one of the court systems analyzed in the case studies above—which currently lack mandatory CLE programs.<sup>25</sup>

#### 4.2.2. Case-based risk assessment

Not all legal cases are created equal when it comes to GenAI integration in the courtroom, and likewise, not all potential uses of GenAI by judicial office-holders create the same risks. There are multiple uses for GenAI that hold different risks. For example, using GenAI to draft decision texts after the legal reasoning has been decided has a different risk profile than assisting a judge in the decision-making process, or gathering context and determining the weight of certain proof around a general factual claim. A crucial aspect of developing a responsible judicial GenAI use is establishing a risk assessment framework that guides judges in (i) determining whether the application of the technology is warranted or appropriate, and (ii) what level of review or scrutiny a GenAI output may warrant. This framework should allow judges to exercise discretion based on a calculated risk analysis.

- *Risk-Based Deployment and Modular Accuracy Thresholds*: For cases with significant potential consequences for a defendant's rights, such as those involving criminal sentencing or loss of liberty, a heightened level of caution for the deployment of GenAI is warranted. In these high-risk use cases, stricter accuracy thresholds for GenAI outputs should be mandated, and a higher level of responsibility should be provided to the reviewing process within the judicial system. These thresholds could be based on metrics relevant to the specific legal domain, such as the predictive accuracy of risk assessment algorithms, or a more precise and narrow database sourcing of information to generate the system's output. This should also affect the scope that GenAI can have on the specific case—from being excluded from deployment altogether given the high risk of infringing upon an individual's rights; to limited application, ranging from reviewing case law or weighing evidence; to potentially providing comprehensive adjudication and decision for a process, to be reviewed by judges or judicial system staff.
- *A Graduated Approach to Review and Responsibility*: The level of review and scrutiny for GenAI outputs, and the level of responsibility mandated, should also be commensurate with the case's risk. Low-risk cases might require a basic review of the GenAI output by the judicial system. However, high-risk cases would necessitate a more thorough examination by the judge, clerks, or a specific organization dedicated to these reviews. These increased review requirements could include legal professionals delving deeper into the GenAI's reasoning process, the data used (e.g., case precedents, laws, and extra-legal information), and potential biases. Correspondingly, as the risk of infringing upon rights and freedoms increases, the responsibility borne by judges (and developers) should be heightened. Judges must evaluate the GenAI's outputs and ensure they align with their own legal reasoning before incorporating them into decisions. Developers, on the other hand, bear accountability for the algorithmic design and functionality of their GenAI systems in these high-risk scenarios.

By implementing a risk-based deployment model with escalating levels of accuracy thresholds, appropriately allocated review requirements, and shared responsibility, the judicial system can leverage the potential benefits of GenAI while prioritizing the protection of fundamental rights and freedoms.

<sup>25</sup> Some jurisdictions, such as the District of Columbia and Israel, recommend, but do not require, that attorneys complete a minimum yearly hours of continuing legal education courses. See 'Minimum CLE Hours by State' (*Lorman*) < <https://www.lorman.com/Minimum-CLE-Hours-by-State> (accessed 3 April 2024).



4.2.3. *Disclosure to the parties.* When GenAI is deployed in a judicial process, this article proposes mandatory disclosure of its use to all relevant parties. This transparency serves several key purposes:

- *Empowering Parties, Ensuring Fairness:* By knowing GenAI is involved, parties can make informed decisions about their legal strategy. For instance, they might choose to challenge the GenAI output or request a more traditional review process. Ultimately, transparency promotes fairness in the proceedings by ensuring all parties have a clear understanding of the tools being employed and the ability to shape their participation accordingly.
- *Promoting Trust, Procedural Awareness:* Transparency builds trust in the judicial process. Knowing GenAI is being used allows parties to have awareness of their procedural status, the rights available to them, and with these, have the confidence that their case is being considered within the boundaries of due process.

To effectively fulfill these purposes, the disclosure of GenAI should follow robust and transparent methods. These information distribution methods should prioritize clarity, accessibility, and comprehensiveness. Information disclosed should be presented in a way that is understandable to all parties involved, regardless of their technical or legal expertise. Additionally, the disclosure should be comprehensive, detailing, among other items, the specific type of GenAI tool used, the stage(s) in the process where it will be employed, the limitations of the technology, and redress rights.

- *Methods and Content of Disclosure:* The specific method of disclosure can be tailored to the risk profile of the case, the type of GenAI used, and the means and resources of the court. Some due inclusions and potential approaches follow:
  - i. **Timeline and Consent:** The timing of disclosure can be adjusted to the case, court, and risk level associated with the case.
    - a. *A Priori Disclosure for Informed Consent:* Early “a priori” disclosure should be considered preferred practice. Informing the parties of GenAI involvement prior to its deployment maintains procedural integrity and user awareness. Knowing how their case is being managed fosters a sense of agency and allows parties to, for example, challenge the GenAI output, advocate for limiting GenAI use to specific aspects of the case, or challenge the technology’s deployment in the proceedings altogether. Consent to the use of GenAI by the parties is a variable that should respond to the specific risk and case, as well as to other considerations that may inform a jurisdiction’s judicial policy.
    - b. *A Posteriori Disclosure for Urgent Situations:* For unforeseen circumstances where GenAI becomes necessary, specifically in low-risk cases, “a posteriori” disclosure could be permitted. However, even in these situations, prompt disclosure after the fact is still essential. This ensures fairness and allows parties to understand how GenAI may have influenced the proceedings.
    - c. *Risk-Adjusted Disclosure Strategy:* The timing of disclosure should be part of a risk-adjusted disclosure strategy. This strategy considers the potential consequences of the case and tailors the level of transparency accordingly. High-risk situations demand the utmost transparency, with early disclosure and detailed explanations being paramount. Conversely, low-risk cases allow for more flexibility in the timing of disclosure, as long as information is promptly distributed, and in some cases, consent is ultimately obtained. By implementing a risk-adjusted disclosure strategy, the judicial system can strike a balance between transparency and efficiency where parties are empowered with the information they need to participate in the process effectively and with awareness of their status, while the legal process maintains its momentum.
  - ii. **Method:** The chosen method for disclosing GenAI use should be flexible and responsive to the specific court and case. Court orders might be appropriate for formal notification, while pre-trial motions could be a tool for parties to raise concerns and trigger disclosure discussions.

Ultimately, the method should ensure timely and comprehensive information reaches all relevant parties.

- iii. **Level of scrutiny:** The disclosure information should outline the level of scrutiny the judge will apply to the GenAI output. This could range from a cursory review to a more in-depth analysis depending on the case's risk and the GenAI's role.
  - iv. **General or Specific Involvement of GenAI:** The disclosure should clarify the extent of GenAI involvement in the case. This may include disclosures on whether GenAI will be deployed (a) generally throughout the process, or (b) for specific purposes (e.g., weighing evidence, identifying relevant case law, or structuring attenuating circumstances). This transparency allows parties to understand the potential impact of GenAI on their case, and exercise their rights accordingly.
  - v. **Explanatory Documents:** The court could provide parties with explanatory documents outlining the nature and limitations of the GenAI system used.
  - vi. **Escalation and Review Rights:** Disclosure should inform parties of their rights, including the right to challenge the use of GenAI in the case comprehensively, modify the scrutiny and risk basis, request a review of the decision-making process, or generally challenge for process or accuracy of outcome (among others); as well as secondary post-adjudication motions, if they believe the GenAI output is flawed or biased.
- *Balancing Disclosure with Efficiency:* While transparency is crucial, it should be balanced with the efficiency advantages GenAI provides to the court system. A court-and-case-specific balance must be pursued, ensuring parties understand the role of GenAI, and their rights associated with said use, without hindering the enhanced efficiency of the legal process.

By establishing clear disclosure protocols, the judicial system can ensure transparency and fairness when GenAI is integrated into court proceedings. This fosters trust in the system and empowers parties to navigate the legal process effectively.

#### 4.2.4. Verification system

While GenAI holds promise for enhancing judicial efficiency, ensuring the veracity, reliability, and non-biased nature of its outputs is paramount. As such, it is crucial to establish a verification network certifying that GenAI-generated content is reviewed for fairness and precision, sometimes requiring human oversight.

- *Scrutiny Tailored to Risk:* The level of scrutiny applied to GenAI content should be commensurate with the risk associated with the case. High-risk cases, such as those involving criminal sentencing or significant impacts on rights and freedoms, necessitate a more rigorous verification process. Information generated by GenAI, particularly in high-risk scenarios, should be subject to independent verification by court personnel or qualified third parties. This could involve cross-referencing data sources, verifying legal citations, and ensuring factual accuracy. Conversely, low-risk cases may require a reduced review process, tailored to the needs and rights affected by each specific process.
- *Flexible Review Network:* The verification process should be adaptable to the specific court and case. For instance, courts with limited resources could leverage standardized verification protocols and tools. Courts handling complex cases might require the formation of specialized review teams with expertise in AI and relevant legal domains. Essentially, this article proposes an adaptive review network, the standards of which can be decided by the relevant body of government of judges, and deployed via its network of clerks, staff, the judges themselves, potentially operating in collaboration with automated review technology.
- *GenAI as an Enhancer, not a Bottleneck:* Verification should not become cumbersome to the extent it slows down the judicial system. The goal is to utilize GenAI as an efficiency enhancer while

mitigating potential risks. Streamlined verification procedures and leveraging technology-assisted verification tools can help achieve this balance.

#### 4.2.5. *Specific procedural rights: Scope, consent, and escalation procedures*

Integrating GenAI into the courtroom necessitates establishing clear procedures for individuals to safeguard their rights and challenge potential biases or errors. Some key rights and consequent considerations include:

- *Right to Opt-Out or Modify AI Use:*
  - **Refusal of AI Integration:** In certain situations, particularly those with the highest level of risk to infringe upon fundamental freedoms, individuals should have the right to refuse GenAI integration into their case entirely. This method would allow parties to opt for a traditional judicial process if they have concerns about fairness or transparency.
  - **Modification of AI Scope:** Individuals should have the right to influence the scope of GenAI use within their case. This could involve choosing between:
    - **Full Integration:** GenAI is used for the entire trial, potentially analyzing evidence, suggesting relevant case law, and aiding in judicial reasoning.
    - **Limited Application:** GenAI use is restricted to specific tasks, such as analyzing historical precedents or weighing the strength of evidence. This allows for a more targeted approach and potentially mitigates risks associated with broader AI involvement.
- *Escalation and Appeals:*
  - **Right to Appeal:** Just as with traditional judicial procedures, individuals should retain the right to appeal. Within the context of GenAI application specifically, individuals should have the right to challenge (i) adjudicative or procedural decisions made with the aid of GenAI, and (ii) specifically appeal the results of GenAI-generated content. This ensures individuals have recourse if they believe the GenAI output has significantly impacted the outcome in a way that is unfair or inaccurate.
  - **Transparency in Appeals:** Appeals processes involving GenAI-aided decisions should be transparent. Individuals should have access to explanations of how the GenAI was used and how it may have influenced the decision. This allows for a more informed, fair, and targeted appeal process.

#### 4.2.6. *Ongoing audits and evaluation*

Algorithmic audits should also become a common tool of GenAI practice and a staple of regulation (Bandy, 2021). Even outside of the implementation of this regulatory proposal, AI models like ChatGPT do go through various forms of auditing and evaluation, notwithstanding the fact that the specifics of these processes are not fully disclosed to the public for proprietary and even reputational reasons.<sup>26</sup> It is not currently clear how AI audits should be conducted, and the private sector has developed an unstandardized approach, which undermines interoperability. This is still a nascent field in which regulators need to think about the basic components of such audits before they can reliably address algorithmic harms (Goodman and Trehu, 2023). Nonetheless, an adequate, iterative, and transparent AI audit framework for GenAI that is used in courts could increase the transparency and interpretability of these AI systems and overall enhance fairness and public trust.

<sup>26</sup> For instance, it has been documented by journalists that some Generative AI companies underpay Kenyan workers to serve as content moderators for ChatGPT to screen out descriptions of violence and sexual abuse during run-up to OpenAI's hit chatbot. Karen Hao and Deepa Setharaman, *Cleaning Up ChatGPT Takes Heavy Toll on Human Workers*, The Wall Street Journal (24 July 2023). Available at <https://www.wsj.com/articles/chatgpt-openai-content-abusive-sexually-explicit-harassment-kenya-workers-on-human-workers-cf191483>.

## 5. Conclusion

The use of AI has the potential to significantly increase efficiency and performance, making AI-driven automation appealing for many industries. Agile data processing capabilities and digitization have the potential to reduce costs, enhance effectiveness and innovation, and optimize system strategies across diverse sectors (Tarafdar et al., 2019). This potential is also present in judicial decision-making, where AI holds the promise of increasing court systems' responsiveness, providing stable and non-arbitrary outcomes, and expanding access to justice as well as strengthening the rule of law. However, the unmethodical and arbitrary use of GenAI by courts creates legal and ethical risks that may affect the legitimacy of the courtrooms' process and outcome. These risks could exacerbate challenges such as maintaining public trust in the judicial system, reducing explainability of court decisions, and potential bias amplification; thus, undermining the promise of using GenAI to establish itself as a legitimate tool to safely streamline judicial processes, expedite case resolution, improve judicial reasoning, enhance legal research, and promote outcome fairness.

To address the gap between enhancing the efficiency of courts through GenAI and safeguarding the integrity of judicial processes, this article has explored how courts around the world are using GenAI in judicial decision-making, as well as provided recommendations to safely integrate AI into the courtroom. The article has analyzed the ethical challenges of the use of GenAI in the courtroom, examined specific use-cases by various jurisdictions, and proposed a framework to adequately mitigate and manage the risks created by GenAI while leveraging its potential. This research offers valuable insights for a wide range of jurisdictions, including those currently lacking a clear and unified approach to integrating GenAI in courtrooms, as well as those looking to enhance ongoing regulatory efforts. The proposed recommendations focus on two key areas: (i) Specific safety standards for the GenAI systems judges will utilize, and (ii) general guidance for the technology's implementation within courtrooms. This framework provides and expands upon various regulatory tools, including ex-ante control through capacity assessments for system deployment, stakeholder engagement, certification processes for all GenAI systems to be utilized in courts, trusted dataset building, clear allocation of responsibility; as well as deployment principles such as enhancing legal education, case-based risk assessments for specific or generalized uses of GenAI in courts, human-in-the-loop verification systems, specific procedural rights (like information and consent rights related to the use or scope of the use of AI), and ongoing iterative audits. By implementing these safeguards on the GenAI systems to be deployed, as well as establishing safe practices for the practical application of the tools, the legal system can harness the potential of GenAI while mitigating its risks, ultimately promoting a more accessible and efficient, yet responsible, system of justice.

### Provenance

This article was accepted and presented as a full paper at the Data for Policy Conference 2024. It has been accepted in Data & Policy on the basis of the integrated review process that the Conference and Journal share for full papers. It is included in the journal as part of the Data for Policy Proceedings 2024.

**Data availability statement.** Data availability is not applicable to this article as no new data were created or analyzed in this study.

**Author contribution.** Conceptualization: N.R.; D.S. Methodology: N.R.; D.S. Desk Research about the Colombian, Mexican, UK, Canadian, Estonian, and Peruvian case studies: D.S. Desk Research about the Indian, Singaporean, Estonian, UK, Canadian approaches to GenAI: N.R. Writing draft: N.R.; D.S. All authors approved the final submitted draft.

**Funding statement.** NA.

**Competing interest.** The authors declares no competing interests exist.

## References

**Bandy J** (2021) Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1. <https://doi.org/10.1145/3449148>.

- Bell F, et al.** (2022) AI decision-making and the courts: A guide for judges, tribunal members, and court administrators. *Australasian Institute of Judicial Administration Incorporated*, June 2022. Available at [https://aija.org.au/wp-content/uploads/woocomerce\\_uploads/2022/06/AI-DECISION-MAKING-AND-THE-COURTS\\_Report\\_V5-2022-06-20-1lzkls.pdf](https://aija.org.au/wp-content/uploads/woocomerce_uploads/2022/06/AI-DECISION-MAKING-AND-THE-COURTS_Report_V5-2022-06-20-1lzkls.pdf) (accessed 2 April 2024).
- Bühlmann M and Kunz R** (2011) Confidence in the judiciary: Comparing the independence and legitimacy of judicial systems. *West European Politics* 34(2), 317–345.
- Cantens T** (2024) *How will the state think with ChatGPT? The challenges of generative artificial intelligence for public administrations.* *AI & Society* 1–12.
- Chan CKY and Wenjie Hu** (2023) Students’ voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20, 43. Available at <https://doi.org/10.1186/s41239-023-00411-8> (accessed 3 April 2024).
- Cheng L and Liu X** (2024) Unravelling power of the unseen: Towards an interdisciplinary synthesis of Generative AI regulation. *International Journal of Digital Law and Governance*.
- Chien C, et al.** (forthcoming) How LLMs Can Help Address the Access to Justice Gap through the Courts. *Loyola of Los Angeles Law Review*.
- Cisco** (2024) Privacy as an Enabler of Customer Trust: Data Privacy Benchmark Study. *Cisco*, 22. Available at [https://www.cisco.com/c/dam/en\\_us/about/doing\\_business/trust-center/docs/cisco-privacy-benchmark-study-2024.pdf?CCID=cc000160&DTID=odidc000016&OID=rptsc032067](https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-privacy-benchmark-study-2024.pdf?CCID=cc000160&DTID=odidc000016&OID=rptsc032067) (accessed 3 April 2024).
- Courts and Tribunals Judiciary of the United Kingdom** (2023) Artificial Intelligence (AI) Guidance for Judicial Office Holders, 12 December 2023. Available at <https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf> (accessed 2 April 2024).
- Courts of New Zealand** (2023) Guidelines for use of generative artificial intelligence in Courts and Tribunals, 7 December 2023. Available at <https://www.courtsofnz.govt.nz/going-to-court/practice-directions/practice-guidelines/all-benches/guidelines-for-use-of-generative-artificial-intelligence-in-courts-and-tribunals/> (accessed 18 March 2024).
- Deakin S and Markou C** (eds) (2020) *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence*. Hart.
- del Campo M and Leach N** (eds) (2022) *Machine Hallucinations: Architecture and Artificial Intelligence*. John Wiley & Sons.
- European Commission** (2021) Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. Available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206> (accessed 3 April 2024).
- European Commission** (2024) Shaping Europe’s digital future—AI act, 6 March 2024. Available at <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (accessed 20 March 2024).
- Federal Court of Canada (Cour fédérale)** (2023) Interim Principles and Guidelines on the Court’s Use of Artificial Intelligence, 20 December 2023. Available at <https://www.fct-cf.gc.ca/en/pages/law-and-practice/artificial-intelligence> (accessed 3 April 2024).
- Ferrara E** (2023) Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Science* 6 (1), 3.
- FindLaw** (2023) *Roberto Mata v. Avianca, Inc.*, No. 1:22-cv-01461-PKC (S.D.N.Y. June 22, 2023). Available at <https://caselaw.findlaw.com/court/us-dis-crt-sd-new-yor/2335142.html>.
- Floridi L and Cows J** (2022) A unified framework of five principles for AI in society. *Machine Learning and the City: Applications in Architecture and Urban Design* 535–545.
- Garry M, Chan WM, Foster J, Linda A** (2024) Henkel, Large language models (LLMs) and the institutionalization of misinformation, Trends in Cognitive Sciences, 2024.
- Gautam S, Venkit PN, and Ghosh S** (2024) From melting pots to misrepresentations: Exploring harms in generative AI. arXiv preprint [arXiv:2403.10776](https://arxiv.org/abs/2403.10776).
- Goodman EP and Trehu J** (2023) Algorithmic auditing: Chasing Ai accountability algorithmic auditing. *Santa Clara High Technology Law Journal* 39, 290. Available at <https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=1689&context=chtlj>.
- Google** (2024). Available at <https://gemini.google.com/> (accessed 3 April 2024).
- Grossman MR, Grimm PW, Brown DG, and Xu MY** (2023) The GPTJudge: Justice in a generative AI world. *Duke Law & Technology Review* 23(1), 1–34.
- Gupta M, et al.** (2023) From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2023.3300381>.
- Gutiérrez JD** (2023) Hablemos sobre el uso de ChatGPT para redactar decisiones judiciales, 26 February 2023. Available at <https://www.lasillavacia.com/red-de-expertos/red-de-ciencia-e-innovacion/hablemos-sobre-el-uso-de-chatgpt-para-redactar-decisiones-judiciales/> (accessed 20 March 2024).
- Gutiérrez JD** (2024) “Chapter 24: Critical appraisal of large language models in judicial decision-making”. Handbook on Public Policy and Artificial Intelligence. Cheltenham, UK: Edward Elgar Publishing, 2024.
- Hiroki Habuka, David U** (2024) “Socol de la Osa. Shaping Global AI Governance: Enhancements and Next Steps for the G7 Hiroshima AI Process”. Center for Strategic & International Studies, May 24, 2024. Available at: <https://www.csis.org/analysis/shaping-global-ai-governance-enhancements-and-next-steps-g7-hiroshima-ai-process>.
- Harvard University Information Technology** (2023) Generative Artificial Intelligence. Available at <https://huit.harvard.edu/ai> (accessed 26 November 2023).



- Heaven WD (2022) Why Meta's latest large language model survived only three days online. *MIT Technology Review*, November 18, 2022, available at: <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/>.
- Invest in Estonia (2022) Introducing Salme, Estonian courts' speech recognition assistant. Available at <https://investinestonia.com/introducing-salme-estonian-courts-speech-recognition-assistant/> (accessed 18 March 2024).
- Juzgado Civil Transitorio de San Juan de Miraflores (n.d.) Expediente 00052-2022-18-3002-JP-FC-01.
- Kirchner JH, et al. (2023) New AI classifier for indicating AI-written text. *OpenAI*, 31 Jan 2023. Available at <https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text> (accessed 23 November 2023).
- Knonth N, Tolzin A, Janson A and Leimeister JM (2024) AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence* 6, 100225.
- Kumar V and Garg ML (2018) Predictive analytics: A review of trends and techniques. *International Journal of Computer Applications* 182, 31–37. [10.5120/ijca2018917434](https://doi.org/10.5120/ijca2018917434).
- Lepenioti K, et al. (2020) Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management* 50, 57–70. <https://doi.org/10.1016/j.ijinfomgt.2019.04.003>.
- Leslie D (2023) Does the sun rise for ChatGPT? Scientific discovery in the age of generative AI. *AI and Ethics* 1–6.
- Lew G, et al. (2020) Garbage in, garbage out: Doing a disservice to AI. *AI and UX: Why Artificial Intelligence Needs User Experience* 87–108.
- Ling LY (2023) Small Claims Tribunals to roll out AI program to guide users through legal processes. *The Straits Times*, Singapore, 27 September 2023. Available at <https://www.straitstimes.com/singapore/small-claims-tribunal-to-roll-out-ai-program-to-guide-users-through-legal-processes> (accessed 26 November 2023).
- López Ponce J (2023) TEPJF valida que Morena use la frase “ya sabes quién” en Edomex. *Milenio Política*, 29 March 2023. Available at <https://www.milenio.com/politica/tepfj-valida-morena-use-frase-edomex> (accessed 3 April 2024).
- Manning C (2020) Artificial intelligence definitions. *Human-Centered Artificial Intelligence*. Available at <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf> (accessed 2 April 2024).
- Marr B (2023) A Short History Of ChatGPT: How We Got To Where We Are Today. *Forbes*, 19 May 2023. Available at <https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/?sh=7b916f0d674f> (accessed 24 November 2023).
- Marvin G, Hellen N, Jjingo D and Nakatumba-Nabende J (2023) Prompt engineering in large language models. In *International Conference on Data Intelligence and Cognitive Informatics*. Singapore: Springer Nature Singapore, pp. 387–402.
- Merken S (2023) New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief. Reuters. Available at <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/> (accessed 2 April 2024).
- Meta (2024) Meta Horizon Workrooms. Available at <https://www.oculus.com/experiences/quest/2514011888645651/> (accessed 3 April 2024).
- Morison J and McInerney T (2024) When should a computer decide? Judicial decision-making in the age of automation, algorithms and generative artificial intelligence. In Turenne S and Moussa M (eds), *Research Handbook on Judging and the Judiciary: Law and Society Series*. Elgar-Routledge.
- Morozov E (2013) To save everything, click here: Technology, solutionism, and the urge to fix problems that don't exist. *Allen Lane* 5.
- Nicoletti L and Bass D (2023) Humans are biased: Generative AI is even worse. *Bloomberg Technology + Equality*. Available at <https://www.bloomberg.com/graphics/2023-generative-ai-bias/> (accessed 23 June 2023).
- Ocampo Y (2023) Embracing AI in Singapore's Courts. *OpenGov Asia*, 6 October 2023. Available at <https://opengovasia.com/embracing-ai-in-singapores-courts/> (accessed 26 November 2023).
- OECD (2022) *EMnet Working Group on Digital Transformation in Emerging Markets. The future of digital transformation in emerging markets. EM net OECD Emerging Market Network*. Available at [https://www.oecd.org/dev/Key\\_Messages\\_EMnet\\_WorkingGroup\\_DigitalTransformation\\_EmergingMarkets.pdf](https://www.oecd.org/dev/Key_Messages_EMnet_WorkingGroup_DigitalTransformation_EmergingMarkets.pdf) (accessed 19 March 2024).
- OpenAI (2021) DALL·E: Creating images from text. *OpenAI Research*, 5 January 2021. Available at <https://openai.com/research/dall-e> (accessed 2 March 2024).
- OpenAI (2024). Available at <https://openai.com/> (accessed 3 April 2024).
- Orchard T and Tasiemski L (2023) The rise of Generative AI and possible effects on the economy. *Economics and Business* 9(2), 9–26.
- Raghavan P (2024) Gemini image generation got it wrong. We'll do better. *Google Gemini*, 23 February 2024. Available at <https://blog.google/products/gemini/gemini-image-generation-issue/> (accessed 3 April 2024).
- Re MR and Solow-Niederman A (2019) Developing artificially intelligent justice. 22 *Stanford Technology Law Review* 242.
- Remolina N (2024) Generative AI in finance: Risks and potential solutions. *Journal of Law, Ethics and Technology*.
- Ritterbusch GD and Teichmann MR (2023) Defining the metaverse: A systematic literature review. *IEEE Access* 11. <https://doi.org/10.1109/ACCESS.2023.3241809>.
- Roberts Jr. JG. (2023) 2023 Year-End Report on the Federal Judiciary. U.S. Supreme Court. Available at <https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf>.
- Salvador Espitia Chávez v. Salud Total EPS (2023) File No. 13001410500420220045901, 1st Labor Court of the Circuit, Cartagena, Judgment No. 032, January 30, 2023. Available at <https://pupilacdn.nyc3.cdn.digitaloceanspaces.com/diariojudicial-public/documentos/000/106/904/000106904.pdf>.

- Segal T** (2022) What is prescriptive analytics? How it works and examples. *Investopedia*, last modified 31 December 2022). Available at <https://www.investopedia.com/terms/p/prescriptive-analytics.asp> (accessed 2 April 2024).
- Smith A, Moloney A and Asher-Schapiro A** (2023) Are AI chatbots in courts putting justice at risk? *Thomson Reuters Foundation Newsroom*, 04 May 2023. Available at <https://www.context.news/ai/are-ai-chatbots-in-courts-putting-justice-at-risk> (accessed 23 November 2023).
- Starke C, et al.** (2022) Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society* 9(2), 20539517221115189.
- Suryam S** (2023) Punjab and Haryana High Court uses ChatGPT in bail order. *Bar and Bench*, 28 March 2023. Available at <https://www.barandbench.com/news/litigation/punjab-haryana-high-court-uses-chatgpt-bailorder> (accessed 3 April 2024).
- Tarafdar M, Beath CM and Ross JW** (2019) Using AI to enhance business operations. *MIT Sloan Management Review* 60(4).
- Times of India** (2023) *In a first, Punjab and Haryana high court uses Chat GPT to decide bail plea. The Times of India, India*, 28 March 2023. Available at <https://timesofindia.indiatimes.com/india/in-a-first-punjab-and-haryana-high-court-uses-chat-gpt-for-deciding-upon-bail-plea/articleshow/99070238.cms> (accessed 3 April 2024).
- Tribunal Electora TEPJF** (2023) Sesión Pública—Miércoles 29 Marzo 2023—TEPJF. *Youtube*, 30 March 2023. Available at <https://www.youtube.com/watch?v=OwaZg3quyls&t=3679s> (accessed 3 April 2024).
- Unión Temporal de Servicios Integrados y Especializados de Tránsito y Transporte de Santa Marta—SIETT v. Nación—Ministerio de Defensa Nacional—Policía Nacional** (2023) Rad. No. 47-0012333-000-2020-00014-00, Tribunal Administrativo del Magdalena, Despacho 01, February 10, 2023. Available at <https://forogpp.files.wordpress.com/2023/02/2020-014-siETT-vs-nacion-policia-nacional-solicitud-audiencia-en-el-metaverso-1.pdf> (accessed 3 April 2024).
- Versteeg M and Ginsburg T** (2017) Measuring the rule of law: a comparison of indicators. *Law & Social Inquiry* 42(1), 100–137.
- Victoria en línea -Vicky Quiñones** (2023) Audiencia Inicial en el Metaverso. Reparación Directa Rad. 2020–014 SIETT VS POLICIA NACIONAL. *Youtube*, 15 February 2023. Available at <https://www.youtube.com/watch?v=LXi2TX9OBmQ> (accessed 3 April 2024).
- Wachter S** (2022) The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law. *Tulane Law Review* 97, 149.
- Wachter S, Mittelstadt B and Russell C** (2021) Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review* 41, 105567.
- Wang J, Shi E, Yu S, Wu Z, Ma C, Dai H, Yang Q, et al.** (2023) Prompt engineering for healthcare: Methodologies and applications. arXiv preprint [arXiv:2304.14670](https://arxiv.org/abs/2304.14670) (accessed May 22, 2024).
- White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J and Schmidt DC** (2023) A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint [arXiv:2302.11382](https://arxiv.org/abs/2302.11382).
- Wilson C** (2022) Public engagement and AI: A values analysis of national strategies. *Government Information Quarterly* 39(1), 101652.
- Yampolskiy RV** (2019) Unexplainability and incomprehensibility of artificial intelligence. *arXiv preprint*, [arXiv:1907.03869](https://arxiv.org/abs/1907.03869) (accessed 3 April 2024).