

# ALGORITHMIC JUSTICE AND HUMAN RIGHTS: STRUCTURAL RISKS AND EMERGING REGULATORY FRAMEWORKS

GEOFREDO ANGULO LÓPEZ<sup>1</sup>

**Abstract:** The incorporation of artificial intelligence (AI), particularly large language models (LLMs), into judicial processes poses unprecedented challenges for the protection of human rights, especially regarding privacy, informational self-determination, and algorithmic transparency. This article introduces the concept of the “Hermes judge” as a normative and interpretative model capable of articulating the technical rationality of automated systems with the principles of the constitutional rule of law. Through a critical analysis of international regulatory frameworks—such as the GDPR, the European Union's AI Act, and the FAIR principles—regulatory gaps, structural biases, and decision-making opacity are identified. In light of these risks, ethical and regulatory guidelines are proposed to ensure that AI functions as a cognitive aid, supporting tasks such as argument generation, precedent identification, and case summarization, without replacing hermeneutic judgment or compromising human dignity. From a functional perspective, Judge Hermes critically mediates between automated decision-making and fundamental principles of law, safeguarding judicial deliberation and the human dimension of justice.

**Keywords:** Artificial intelligence, Judge Hermes, Human rights, Data protection, Algorithmic governance, Judicial automation, Informational self-determination, Human-in-the-loop, Algorithmic bias.

## 1. INTRODUCTION

The crisis of the postmodern state and the welfare state has given rise to new ways of conceiving legal and social reality, which is increasingly globalized, interconnected, and cybernetic. In this context, the emergence of disruptive technologies, particularly AI and algorithmic systems, has profoundly transformed the models of interpretation, application, and legitimation of law, surpassing the categories of modern theoretical-legal artifice. As François Ost points out, law, as a linguistic sign, requires interpretation by its addressees, and as an expression of will, it must be internalized and accepted (Ost, F., 1993). When legal subjects mentally reconstruct the normative message and mediate its application thru acts of will, law is configured as a necessarily unfinished work, always in suspense and in constant re-elaboration. This ontological perspective, that is, one grounded in a fundamental shift in the nature of legal reality rather than merely in the ways law is known or interpreted, is especially relevant in contemporary legal systems, where technological mediation is redefining forms of justice. Such a shift poses the challenge of integrating artificial intelligence with its logics of technical rationality, decisional automation, and algorithmic encoding of behavioral expectations without fracturing the constitutional framework of human rights. In this context, the rights to privacy, personal intimacy, and personal data protection can no longer be understood as ancillary guarantees, but must be recognized as essential pillars of human dignity in the digital age.

---

<sup>1</sup> Facultad de Derecho de la Universidad Autónoma de Yucatán (México), ORCID <https://orcid.org/0000-0002-8367-9906> (geofredo.angulo@correo.uady.mx).

To this complexity is added a new layer of structuring power: information technology, the Internet, and, in general, digital technologies. Beyond the classic division of functions among public authorities, organized civil society, regional and international organizations, political parties, and the media, algorithmic power has emerged, capable of shaping legal and democratic structures. Since the Internet took shape as a “network of networks,” a cyber paradigm was inaugurated that not only transformed communication and the economy but, as Roszak (1994) warned, also introduced unprecedented ethical dilemmas by concentrating political power and generating new forms of opacity and social domination. Along the same lines, Sartori (1998) anticipated that the home minicomputer would reconfigure social and political practices, a prophecy now realized in the ubiquity of smartphones, which function as nodes of control, consumption, and data production. In this context, the threat posed by AI is not merely technical or regulatory but structural, as decision-making about personal data is increasingly delegated to non-state actors, such as digital corporations, thereby weakening the guarantor role of the constitutional rule of law. This transformation highlights the urgent need for a robust legal response capable of balancing technological innovation, algorithmic transparency, and the protection of fundamental rights (Coeckelbergh, 2021, pp. 91–93; Cáceres Nieto, 2024, pp. 77–78).

Understanding AI is essential to equip the legal framework with the tools necessary to address these structural transformations. From an epistemological perspective, AI has been defined as “the branch of computer science that studies the software and hardware necessary to simulate human behavior and understanding” (Albert-Márquez, 2021, p. 211). Its ultimate goal, in what is known as “strong” AI—using Searle’s terminology (1980, cited in Albert-Márquez, 2021, p. 211)—is to simulate human intelligence by creating artificial agents endowed with understanding, awareness of that understanding, and, ultimately, awareness of their own existence. From other operational perspectives, Russell and Norvig define it as “the study and design of rational agents capable of acting to achieve the best possible outcome in their environment, given their goals and the information available to them” (2020, p. 1), emphasizing the autonomy and adaptive capacity of systems. Chollet (2021, pp. 1–7) describes it as “the effort to automate intellectual tasks normally performed by humans,” highlighting its link to complex cognitive processes and not merely digital editing tools. Ignoring these dimensions, as in the definition validated by the Supreme Court, reduces AI to automated visual editing programs, weakening criminal law’s capacity to provide coherent, technically informed regulatory responses that respect the principle of specificity. This case underscores the urgent need to equip the Mexican legal framework with concepts that are both technologically operational and scientifically consistent, thus avoiding decisions that increase uncertainty in one of the most sensitive and controversial areas of digital constitutionalism.

These conceptual clarifications are not merely academic. They provide the foundation to understand the challenges that arise when legal decisions are delegated to automated systems, whose opacity and complexity can undermine transparency, traceability, and accountability in judicial reasoning. The opacity of machine learning algorithms, particularly those based on deep neural networks, not only hinders the understanding of their outputs but also complicates the attribution of responsibility when such decisions prove harmful. In a rule-of-law system, this lack of clarity endangers the principle of

transparency in decision-making and undermines public trust in the institutions responsible for administering justice (Coeckelbergh, 2021, pp. 91–93). Information technologies have never been mere tools; they are actors with autonomous power. This transformation places the law before an unprecedented challenge: operating in a scenario where norms are no longer confined to regulating human action but must also confront algorithmic architectures and automated processes that perform quasi-normative function and automated processes that perform quasi-normative functions, meaning that they influence or constrain human behavior in ways similar to legal norms, even though they are not formal laws. For example, algorithms that automatically filter content, assign credit scores, or manage access rights can shape decisions and behavior, acting as if they were regulatory rules.

In this context, law subjective, fragmented, and a constantly shifting set of discourses (Zagrebel'sky, 1995), emerges as a fluctuating and fragmented order, whose ductility makes it adaptable but also vulnerable to dispersion and loss of effectiveness. As François Ost argues, the key lies not in centralized command but in a circulation of legal discourse that provides coherence to this network. From this perspective, the figure of the Hermes judge stands as a paradigm called to articulate a form of law capable of balancing technological innovation with the protection of human dignity. As Recaséns Siches (1956, as cited in Feito Torrez, 2020, pp. 112–117), emphasized in proposing a *logos* of the reasonable, law cannot be reduced to a closed formalism but must remain open to human experience. In the algorithmic era, this notion becomes especially relevant: the Hermes judge embodies that practical rationality, not as a passive guarantor of constitutional order but as a normative reconfigurator who balances technological innovation and human rights. In particular, since personal data has become the input, medium, and outcome of computational power, the Hermes judge represents a judiciary committed to the essential principles of the rule of law: algorithmic transparency, accountability, prohibition of profiling-based discrimination, and comprehensive protection of privacy. These dimensions—fundamental today—mark the difference between a digital justice centered on the person and a justice reduced to the opaque logic of machines.

This article critically examines the structural risks posed by algorithmic justice in relation to human rights, with a particular focus on privacy, equality, and non-discrimination. Specifically, the expansion of AI systems in various fields such as the administration of justice and public safety has reignited an old debate with new nuances: can the law delegate decisions to algorithms without compromising its ethical and protective principles? The central question guiding this work is precisely to what extent AI can be integrated into the judicial sphere without replacing the ethical and hermeneutic deliberation of human judges. From this perspective, the objective is to analyze the limits and possibilities of AI in highly sensitive legal contexts and to evaluate its ethical and regulatory implications. Methodologically, the article adopts a hermeneutic and critical approach, aimed at reflecting on the regulatory, ethical, and judicial challenges posed by AI in the field of human rights, particularly with regard to privacy, informational self-determination, and algorithmic transparency. Using the model of Judge Hermes—a conceptual figure developed by François Ost—it examines how a judiciary capable of balancing technological innovation and constitutional guarantees can be articulated. The analysis seeks to highlight tensions, identify regulatory gaps, and propose interpretive

criteria to move toward algorithmic governance centered on human dignity, fundamental rights, and democratic control.

## **2. FROM PROFILING TO STRUCTURAL BIAS: PREDICTIVE JUSTICE AND ALGORITHMIC RISKS**

Based on this premise, it is essential to analyze how AI, by operating through profiling and prediction processes, can reproduce structural biases and generate adverse impacts on human rights, such as the right to work, privacy, and due process. This critical dimension requires deep reflection on predictive justice and the role of Judge Hermes as a guarantor of fairness in automated environments. In this sense, the regulatory and technical deficit not only increases citizens' exposure to opaque automated decisions, but also compromises other fundamental areas of the legal system. Artificial intelligence is increasingly present in judicial processes, including predictive algorithms for recidivism, automated case management, and decision-support systems for judges. While these technologies can improve efficiency, they also introduce significant risks, such as algorithmic bias, lack of transparency, and erosion of procedural safeguards, which can threaten fundamental rights. From this perspective, not only privacy but also essential guarantees like the right to work, access to justice, and due process are at stake, highlighting the growing need for a judiciary capable of mediating between technological innovation and the protection of human rights.

The OECD (Organization for Economic Cooperation and Development) and studies such as McKinsey's "Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation" warn of the risk of massive job losses as a result of automation; in some developed countries, it is predicted that up to 50% of jobs will be in the hands of automated systems by 2035 (Guijosa, 2017). At the same time, a "data market" is emerging, in which personal information is traded using micro-segmentation algorithms, and cases of misuse by cybercriminals and criminal networks have increased (Cáceres Nieto, 2024, p. 105). As Shoshana Zuboff (2019, p. 14) has argued, this "surveillance capitalism" consolidates an economic model based on the extraction of human experience as raw material for behavioral predictions. Incorporating this perspective into the legal debate highlights that these practices not only threaten privacy but also autonomy, freedom of choice, and democratic balance, underscoring the need for a judiciary capable of protecting human rights in contexts shaped by algorithmic power.

This problem is not limited to the theoretical context but is already manifesting itself in concrete practices that directly affect people's working lives and human rights. Recent studies have shown that the adoption of AI in the Mexican labor market is undergoing profound transformations that carry risks. The Federal Telecommunications Institute (IFT, 2022; 2023) has documented that a significant number of companies that use AI in hiring or performance evaluation processes have structural biases that can lead to discriminatory decisions (Medina Romero and Torres Chávez, 2025). These findings coincide with those of international studies, such as that of the World Economic Forum (2018), which predicts the loss of 1.4 million jobs in the United States by 2026 as a result of automation (Fredin, 2018). Research warns of the adverse effects of poorly designed

algorithms or those trained with biased data, especially in sensitive processes such as personnel selection or performance evaluation. In the Latin American context, various studies agree on the urgency of establishing solid regulatory and ethical frameworks that respond effectively to these risks, particularly in countries such as Mexico, where weak regulation in this area still prevails.

Automation and digital technologies are transforming labor structures worldwide, which in turn affects social conditions that courts must consider when protecting fundamental rights. For instance, as routine tasks are automated, issues of employment displacement, inequality, and precarious work emerge, which may require judicial oversight and legal frameworks that safeguard workers' rights in increasingly automated environments (Manyika et al., 2017; Rivas-Vallejo, 2021). Incorporating this perspective highlights that AI in the judiciary does not operate in isolation but intersects with broader social and economic transformations that influence access to justice and human rights protection.

This warning not only highlights the risks of opacity and automatism in algorithmic management but also invites us to reconsider the role of legal actors as guarantors of critical and humanistic rationality. The structural transformation of employment requires economic and legal responses that accompany technological change with guarantees of equity. In this context, Judge Hermes is a judicial figure designed to mediate between automated decision-making systems and human-centered legal reasoning. He addresses the challenges posed by algorithmic management, including the reproduction of structural inequalities under the guise of neutrality, opacity, and over-reliance on automation. By constructing legal meanings in these contexts, Judge Hermes resists the dehumanization of law and upholds a justice that is situated, deliberative, and committed to principles of equity, freedom, equality, and non-discrimination.

This structural transformation of employment and the algorithmic management of labor decisions cannot be analyzed in isolation from the legal principles that protect human dignity. In fact, these arguments coincide with Ronald Dworkin's thinking on the primacy of rights as protective principles. While AI tends to treat data and decisions instrumentally, Dworkin emphasizes that individual rights—and not just rules of efficiency—must prevail. Human rights should not be conceived as mere concessions of the legal system or sacrificed in the name of collective utility. On the contrary, Dworkin argues that rights function as moral principles that limit even legally authorized decisions, which is opposed to any model that prioritizes efficiency over justice. The examples provided by Cáceres Nieto—such as technological unemployment, commercial profiling, and criminal risks—together with the forecasts of the World Economic Forum and the McKinsey study, highlight the fragility of the social fabric described by Zygmunt Bauman in *Liquid Modernity* (Bauman, 2000, p.70). In this context, surveillance and profiling, driven by new forms of social control, respond to the need to manage these risks in a context where both social networks and traditional legal institutions are becoming less effective in ensuring security, cohesion, and the protection of human rights.

In this context, the proliferation of AI technologies that facilitate profiling and mass surveillance aligns with this trend: they offer a sense of security and protection, but at the



expense of individual privacy and freedom. As Bauman warns, these dynamics can lead to a “surveillance society,” in which the protection of human rights is compromised by the restriction of the right to privacy and personal autonomy, in an attempt to manage fears linked to insecurity in an increasingly liquid world and a system far from equilibrium (Bauman, 2000, p.50). The idea of human dignity must be the ethical and legal anchor of algorithmic design: the person is an end and not a means, in Kantian terms. When personal data is treated as commercial information, AI threatens to turn the subject into an object, stripping them of their capacity for action and reducing their existence to a variable in a calculation.

### 3. INFORMATIONAL SELF-DETERMINATION AND PRIVACY IN THE ALGORITHMIC ERA

The present paper not only highlights the structural risks of algorithmic profiling and predictive justice, but also underscores the need to strengthen the legal safeguards that protect personal autonomy from data exploitation. In this context, it is essential to recover the concept of informational self-determination as a regulatory and ethical axis to address the challenges posed by AI in the digital age. Recognized by the German Constitutional Court since 1983, this principle has established itself as a key instrument for legally empowering citizens against automated processing systems characterized by their opacity and lack of democratic control. In this regard, the jurisprudence of the Spanish Constitutional Court has made significant normative contributions. In this regard, the case law of the Spanish Constitutional Court has made significant regulatory contributions. Judgment STC 94/1998 constitutes a milestone in the consolidation of the right to personal data protection as an autonomous and independent fundamental right, although related to the rights to dignity and personal freedom, honor, privacy, and personal image. In this ruling, the Constitutional Court interprets Article 18.4 of the Spanish Constitution as a reinforced guarantee of personal dignity and freedom, and emphasizes that the data subject has the right to control the use of their data and to oppose its processing for purposes other than those for which it was initially obtained (Tribunal Constitucional de España, 1993). The Spanish Court emphasizes that the use of personal data must be guided by the principles of consistency and rationality, to ensure that the processing of information is in line with legitimate interests and does not violate human dignity (Adinolfi, 2007, p.15).

Along the same lines, the Supreme Court of Justice of the Nation (SCJN) has developed substantial criteria regarding the right to personal data protection as a direct expression of informational self-determination, especially in the context of the use of digital technologies.<sup>2</sup> In Unconstitutionality Action 82/2021, the Full Court invalidated the provisions of the Federal Telecommunications and Broadcasting Law that created the National Registry of Mobile Phone Users (PANAUT), as it considered that the obligation to collect biometric data from users without their consent and without judicial control violated the rights to privacy, personal data protection, and human dignity. The Court emphasized that the massive and indiscriminate collection of data without safeguards, such as the application of interpretive guidelines such as necessity, suitability, proportionality, and legitimate

---

<sup>2</sup> For a systematic analysis of the jurisprudential criteria issued by federal courts regarding the right to personal data protection—including its scope in relation to digital technology—please refer to the institutional work: *Criteria of the Federal Judiciary on personal data protection* (Consejo de la Judicatura Federal, 2018).

purpose, is incompatible with the constitutional principles governing the processing of personal information (SCJN, 2021). This approach reflects an evolution in case law in which informational self-determination is understood not only as the right to control personal data, but also as a mechanism for protecting autonomy against its possible exploitation or misuse, especially in environments of increasing digitization and use of technologies such as AI.

In this context, one of the most alarming phenomena illustrating the risks of algorithmic exploitation of personal data is the use of AI technologies to generate synthetic content, known as deepfakes<sup>3</sup>. While their misuse already threatens privacy, intimacy, and personal autonomy, these technologies are also increasingly relevant to judicial processes: manipulated audiovisual content can compromise the integrity of evidence, challenge due process, and require judges to critically assess the validity of AI-generated information. Their proliferation thus underscores the need for legal and judicial responses that safeguard human rights and uphold the principle of informational self-determination.

Currently, unconstitutionality action 66/2024, resolved by the SCJN, is a case that highlights the regulatory challenges surrounding AI. In this case, the Plenary analyzed the validity of Article 185 Bis C of the Penal Code of the State of Sinaloa, which reformulated the crime of violation of sexual privacy to include conduct committed through the use of artificial intelligence, such as the manipulation of intimate images, audio, or videos with a realistic appearance, without consent. The Court validated both this expansion and the legal definition of artificial intelligence incorporated into the law, which is understood as “applications, programs, or technology that analyzes photographs, audio, or video and offers automatic adjustments to make alterations or modifications.” Although the decision was based on the need for clarity accessible to the average citizen and the difficulty of establishing unambiguous concepts in the face of rapidly evolving technologies, various academic voices have warned that this definition is deficient from a technical and legal point of view. By limiting itself to describing a subset of tools, such as those associated with deepfakes, and omitting structural elements such as machine learning, operational autonomy, or synthetic content generation, there is a risk of serious conceptual fragmentation, criminal ambiguity, and future vulnerability. Far from reducing uncertainty, an imprecise decision such as the one validated by the Court can undermine the principle of specificity and hinder public and judicial understanding of the true scope and limits of AI in essentially controversial and sensitive contexts (SCJN, 2024).

#### **4. ALGORITHMIC GOVERNANCE: REGULATORY CHALLENGES AND COMPARATIVE MODELS**

In Mexico, there is a significant regulatory gap in the area of AI, which exposes citizens to a growing risk of privacy violations and discriminatory automated decision-making. This regulatory vacuum cannot be explained solely from a technical or legislative perspective, but

---

<sup>3</sup> For a broader view of the risks of non-consensual synthetic content and its impact on women's rights, see: *The State of Deepfakes: Landscape, Threats, and Impact* (Ajder et al., 2019). As well as the United Nations thematic update on the protection of human rights and the integrity of information in the era of generative artificial intelligence (Konwar, 2025).

rather responds to a more profound transformation of the digital ecosystem. As Éric Sadin argues, drawing on Heidegger's concept of *aletheia*—understood as the uncovering of reality or being—certain algorithmic systems have acquired a disturbing ability to “tell the truth”: they define what is real based on data patterns, without transparency or democratic control. This trend is evident, for example, in the medical field, where algorithms dictate diagnoses without taking into account human singularities, leading to a shift in clinical responsibility and a weakening of the ethics of care (Sadin, 2020). This dynamic jeopardizes fundamental principles such as human judgment, sovereignty, and institutional accountability (Sadin, 2020, pp.127-129). To address these regulatory gaps and the risks of technological power concentration in the state, the approach adopted by the European Union is illustrative. It already has a regulatory framework in place, such as the General Data Protection Regulation (GDPR), which establishes requirements for consent, transparency, and accountability in the processing of personal data, including in the context of AI. This situation highlights the urgent need to create a comprehensive regulatory framework that specifically addresses the ethical and legal challenges posed by AI, especially with regard to data protection, privacy, and human rights (Pérez-Ugena, 2024, p. 8).

A prime example of the judicial application of AI in Mexico is the ruling issued by the Second Collegiate Court for Civil Matters of the Second Circuit, in which AI tools were used to calculate the amount of a guarantee in an *amparo* trial. This precedent, a first in the national judicial context, establishes that the ethical and responsible use of AI requires compliance with minimum principles such as proportionality, safety, personal data protection, transparency, explainability, oversight, and human decision-making. In the absence of specific regulations, the court proposes a guide for judicial self-restraint based on international standards, such as the Ethical Guidelines for Trustworthy AI, the UNESCO Recommendation, and the European AI Regulation. This ruling underscores the need for algorithmic governance with a human rights perspective, in which AI acts as a technical aid without replacing the judge's hermeneutic judgment (SCJN, 2025b). In this sense, the incorporation of AI into auxiliary tasks in the judicial process is justified by advances in new technologies and the consolidation of digital justice. Thus, the justification for this criterion lies in the fact that certain calculations, such as updating values, applying interest rates, and weighing procedural deadlines, are essential for establishing guarantees but are not part of the judge's core decision-making process. AI is therefore used to reduce human error, ensure transparency and traceability, promote consistency and standardization of precedents and amounts, and improve procedural efficiency, freeing up time for substantive analysis and strengthening the reasoning behind judicial decisions. In this way, the court ensures that AI functions as a cognitive aid, preserving the essential core of the jurisdictional function and aligning itself with the principles of digital justice and the standards of reasoning set forth in Article 16 of the Federal Constitution, which constitutes a strategic recommendation for courts seeking to adopt best practices in the administration of justice (SCJN, 2025a).

The absence of an explicit reference to algorithmic governance reflects a regulatory gap that is repeated in other jurisdictions facing the challenges of digital constitutionalism. In contrast to the judicial self-restraint effort mentioned above, in March 2025, an initiative to reform the Telecommunications Law and the Federal Criminal Code was presented in



the Chamber of Deputies. This initiative assigns mobile phone companies the obligation to prevent and report illegal activities on their networks, but fails to establish clear principles of algorithmic governance, limiting their ability to ensure transparency and technological accountability (Cámara de diputados, 2025). This regulatory gap cannot be understood in light of the philosophical background of contemporary technical power. As Éric Sadin argues, we live in the age of coercive power: an algorithmic authority that replaces human judgment and action with automated protocols that prescribe behavior. In this context, algorithmic governance transcends the technical and legal spheres to become an essential anthropological question: who decides, with what criteria, and under what legitimacy? (Sadin, 2020, pp. 20-21). Technological singularity, understood not only as a threshold of technical development, but also as a philosophical frontier that redefines what it means to be human, poses unprecedented risks. The possibility of artificial systems surpassing human intelligence and the hypothesis of transferring consciousness to digital media creates a power that not only automates decisions, but also threatens to replace human judgment with self-referential computational rationality (Elena Ortega, 2019, pp. 83-93).

Therefore, an institutional response that goes beyond regulation and comprehensively addresses its ethical, legal, and social implications is essential. In this regard, the European Union has taken a pioneering role in building a robust regulatory framework that balances technological innovation with the protection of fundamental rights. Key documents such as the Ethical Guidelines for Trustworthy Artificial Intelligence (Comisión Europea, 2019), the European Parliament Resolution on Global Industrial Policy in Artificial Intelligence and Robotics (Parlamento Europeo, 2019), the White Paper on Artificial Intelligence (Comisión Europea, 2020), and the Draft Regulation on Artificial Intelligence propose harmonized standards for safe and trustworthy AI. Added to this is progress towards a code of good practice for the management and auditing of AI systems that incorporates governance, transparency, and traceability procedures throughout the entire technological lifecycle (Comisión Europea, 2021).

This strategy, known as Euro-regulationism, should not be interpreted as excessive regulation, but rather as an ethical commitment to ensuring that AI is centered on human dignity. At the same time, it constitutes a preventive response to narratives such as “technological singularity,” which is defined as a hypothetical point in scientific and technological development at which AI would surpass human intelligence and cause unpredictable transformations in civilization. This concept, which has been addressed by various authors, is linked to the convergence of emerging technologies and the acceleration of technical change (Elena Ortega, 2019, pp. 95-96).<sup>4</sup> In this imaginary, singularity constitutes an emerging form of algorithmic power that threatens to replace legal and political deliberation with the dictates of automated systems that are opaque and devoid of democratic debate. In this sense, the legal system runs the risk of being absorbed by a technical rationalism that dispenses with both human judgment and justice as practical reason (Albert-Márquez, 2021, pp. 213, 215, 223).

---

<sup>4</sup> One of the foundational texts on the concept of “technological singularity” is Vinge (1993) *“The Coming Technological Singularity: How to Survive in the Post-Human Era”*.

A prime example of this vision is the set of Ethical Guidelines for Trustworthy Artificial Intelligence (Comisión Europea, 2019), promoted by the Council of Europe, which is based on the premise that AI is not an end in itself, but a means to improve human well-being. These guidelines are based on three essential pillars—lawfulness, ethics, and robustness—and are specified in seven fundamental requirements: human intervention and oversight; technical robustness and security; privacy and proper data management; transparency, diversity, non-discrimination, and fairness; social and environmental well-being; and accountability (Barona Vilar, 2024, p. 97). Although not legally binding, they represent an ethical roadmap that guides the development of intelligent systems from a person-centered perspective and in accordance with the principles of European constitutionalism.

However, normative recognition of rights does not always translate into effective guarantees. The deployment of algorithmic tools has led to significant setbacks, ranging from the reproduction of structural biases with discriminatory consequences based on gender, race, age, or socioeconomic status, to serious errors in automated medical diagnoses based on models trained without adequate clinical supervision. Added to this are progressive interferences in personal, work, and emotional spheres, particularly on digital platforms, which directly affect procedural rights such as effective judicial protection, the presumption of innocence, and the right to an adequate defense. These risks call into question the protective nature of contemporary law and require a renewed interpretative framework capable of addressing algorithmic governance based on the principles of fairness, transparency, and meaningful human control (Barona Vilar, 2024, p. 98). In this regard, the difference between the regulatory progress of the European Union and the delay of the Mexican State is not only evident in the legislative sphere, but also in the practical application of algorithmic tools in the judicial system. Countries such as Estonia, Germany, and the United Kingdom have begun to incorporate automated mechanisms to resolve small claims, draft judgments, or assist in administrative proceedings, while in China, AI courts are already operating, analyzing more than a hundred crimes, generating draft judgments, and resolving simple legal disputes under an integrated digital justice model (Ester Sánchez, 2025, p. 319).

The European approach has established a pioneering model for AI regulation, focusing on controlled innovation and the protection of rights. A key component of this model is the creation of regulatory sandboxes, i.e., supervised testing environments where companies, particularly SMEs, can experiment with innovative AI systems under the supervision of the competent authorities. These regulatory tools are essential to prevent the premature exclusion of new business models from the market due to non-compliance with existing regulatory frameworks, as they allow companies to demonstrate that they can offer the necessary levels of protection for users (Pošćić and Martinović, 2022, p. 79). At the same time, the European AI Regulation provides for the creation of a European AI Committee to ensure harmonized implementation and the development of common standards, consolidating the European Union as a global benchmark in ethical and trustworthy artificial intelligence. However, the success of this model in Europe does not guarantee its automatic replicability. Implementing such robust structures can pose a significant challenge for other regions with different legal frameworks and economic resources. In this regard, in the Ibero-American sphere, the regulatory response reflects proactive adaptation rather than a simple copy of

the European model. The Ibero-American Charter on Artificial Intelligence (CLAD, 2023) establishes fundamental principles such as human autonomy, transparency, accountability, and security. Although these principles are not binding, they do offer relevant guidance for the construction of regulatory frameworks in the region. Countries such as Chile and Brazil have begun to include these guidelines in their draft legislation, indicating a regional trend toward risk management and the protection of the rights of people affected by AI (Salcido Ledezma, 2025, pp. 34-35). These initiatives demonstrate that, while European influence is evident, each country is adapting these principles to its own legislation and social reality, rather than importing a complete model.

The global landscape of AI regulation is becoming increasingly complex and diverse. The European approach contrasts with models in other regions, such as the United States, which prioritizes innovation and market self-regulation, and China, which focuses on state control and national security. Understanding these differences is crucial to assessing each region's position in the global debate. Regulation is not a “one size fits all” approach; each model has its own advantages and disadvantages that must be analyzed to promote responsible innovation and the protection of human rights. In the judicial context, this means developing legal and institutional frameworks that specifically address the challenges posed by AI in courts, including decision-making transparency, accountability, and the protection of fundamental rights. While other Ibero-American countries have adopted proactive measures in this area, Mexico remains on the sidelines, as serious regulatory and institutional limitations persist in ensuring that AI in the judiciary operates in a manner consistent with human rights protection. The absence of specific legislation on the matter, together with the fragility of personal data protection mechanisms, has been exacerbated by the disappearance of the National Institute for Transparency, Access to Information, and Protection of Personal Data (INAI), which constitutes an institutional regression and a critical fracture in the Mexican State's digital safeguards. The lack of a specialized authority to oversee the algorithmic processing of personal data is compounded by the absence of judicial protocols to ensure transparency, traceability, and meaningful human control over automated decisions (Guzmán García, 2025, pp. 20-23).

The progressive reduction of the state's regulatory capacity in technological environments has led to the responsibility for defining the operational content of rights such as privacy and data deletion falling disproportionately on private corporations. This delegation creates a paradox: while legal systems formally recognize these rights, the mechanisms for ensuring their exercise on transnational platforms remain ineffective. The associated risks are not merely theoretical, but derive from the actions or omissions of multiple actors: data controllers and processors, developers of algorithms and software that design opaque and difficult-to-audit decision-making structures, end users (both public and private), and data protection authorities, which are called upon to regulate, supervise, and sanction (REDIPD, 2020). The complex interaction between these actors poses unprecedented challenges in terms of accountability, transparency, and the development of verifiable and effective regulatory mechanisms.

In contrast to the proactive trend in other Ibero-American countries, the growing technological dependence of countries such as Mexico on foreign solutions generates a form of “data colonialism,” characterized by the systematic extraction of national data

without local governance mechanisms or tangible benefits for citizens (Salcido Ledezma, 2025, pp. 36-38). This lack of governance is exacerbated by the structural invisibility of the algorithmic processes that manage data, as Rodríguez Amat argues when proposing an interpretive model that allows for the de-automation of the perception of AI systems. This model identifies three levels (superficial, intermediate, and deep) and three moments (production, invisible processing, and recognition) in which data acquires meaning, revealing that behind every automatism there are human, ideological, and programmed decisions. At the superficial level, data is presented as visible discourse; at the intermediate level, it is algorithmically recombined for specific purposes; and at the deep level, it is standardized for exchange, concealing its origin and purpose (Rodríguez-Amat, 2022, pp. 58-61). This design makes it possible to visualize the critical points of human intervention and requires ethical, discursive, and contextualized regulation of data.

As a matter of fact, ethical concerns surrounding AI have prompted international regulatory responses. Among these, the Toronto Declaration on Equality and Non-Discrimination in Machine Learning AI Systems, promoted by Human Rights Watch, Access Now, and Amnesty International (2018), stands out. This transfers international human rights standards to the algorithmic sphere and points out that protection against discrimination is a binding legal obligation and not merely a recommendation or ethical aspiration. The Declaration warns that AI systems can amplify historical biases and deepen structural inequalities if not properly supervised. The Declaration proposes three areas of action: the state's duty to prevent discrimination and ensure transparency; corporate responsibility through external audits and data publication; and the right to redress for those affected. This framework reaffirms that AI ethics must be aligned with human rights as a normative core, becoming a key reference point for algorithmic governance centered on human dignity (Grigore, 2022, pp. 169-172). In contexts with weak regulatory frameworks and no effective oversight of platforms, the Declaration is particularly relevant for highlighting areas where human intervention is necessary and for demanding transparency in the attribution of algorithmic meaning.

The absence of a national AI strategy coordinated between public authorities, academia, and civil society hinders the development of local capacities and the establishment of standards adapted to the context. This omission compromises digital sovereignty and limits democratic control over technologies that directly affect daily life, the administration of justice, and public management. Therefore, the technological and regulatory gap requires not only solid regulatory frameworks, but also public policies aimed at the ethical and technical training of legal authorities and operators. If not addressed, the transformative potential of AI in the judicial sphere could become a factor of institutional dehumanization and a direct threat to procedural rights and guarantees. These demands are particularly relevant given the growing sophistication of AI systems and what José Manuel Elena Ortega (2019, p.96) calls the “God Equivalence Hypothesis”: the hypothesis of a distributed artificial mind with global control capabilities from a digital network. Although this scenario may seem speculative, it requires a clear legal response that reaffirms the centrality of human dignity in the face of any form of algorithmic depersonalization, establishing regulatory frameworks capable of addressing both the technical complexity and the social impacts of AI.

## 5. ALGORITHMIC JUSTICE AND HUMAN JUDGMENT: ETHICAL AND REGULATORY LIMITS OF AUTOMATION

In today's digital environment, AI can be defined as a set of information systems designed to emulate human cognitive functions, such as reasoning, learning, and decision-making, using algorithms and machine learning models trained with large volumes of data (Delgadillo and González, 2023). Its rapid expansion has been made possible by the massive availability of information, which is the essential raw material for these systems. However, this data collection dynamic does not always comply with clear legal standards, as practices of data extraction without informed consent proliferate, even feeding opaque markets with little oversight. This reality poses urgent ethical and legal challenges, especially with regard to privacy, recognized as a fundamental human right that requires new regulatory frameworks and interpretive guidelines capable of responding to the structural risks of the algorithmic era (Sánchez Díaz, 2024, pp. 182-184).

In this regard, AI has profoundly transformed data ethics and privacy in the judicial domain, posing specific challenges for courts, such as the protection of sensitive personal information, the integrity of evidence, and the accountability of automated decision-making systems. From a philosophical perspective, Adela Cortina emphasizes the need to protect social rights in the digital society through redistributive measures, such as universal basic income or robot taxation, which ensure trust and equity in technological deployment (Cortina Orts, 2019, p. 392). In the healthcare field, Blázquez Ruiz (2022, p. 259) warns about the violation of patient privacy, noting that patients must retain the right to be informed and to control access to their data, especially when algorithmic diagnostic or clinical management tools are employed. These examples reflect a common requirement: to address the implications of AI from the design stage, integrating both technical and ethical-legal criteria to ensure free and informed decision-making. For this reason, recent ethical frameworks, such as AI4People from the Atomium European Institute,<sup>5</sup> are essential for adapting the classical principles of beneficence, non-maleficence, autonomy, and justice, while incorporating explainability and accountability as essential prerequisites for ensuring reliable, transparent, and traceable AI.

In the criminal justice system, AI raises significant ethical and legal challenges. A prime example is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)<sup>6</sup> system, used in the United States to calculate the risk of recidivism. Although its designers assert that it does not incorporate explicit racial variables, research

---

<sup>5</sup> This document, prepared by the AI4People scientific committee under the direction of Luciano Floridi et al. (2018), presents five fundamental ethical principles for the development and adoption of AI: beneficence, non-maleficence, autonomy, justice, and explainability. It also offers 20 specific recommendations to promote a "Good AI Society."

<sup>6</sup> This software evaluates individuals based on variables such as criminal history, age, social environment, and consumption habits, and generates a risk score that influences key judicial decisions, such as the granting of parole. However, it has been heavily criticized for reproducing and amplifying racial and socioeconomic biases, as many of its variables are correlated with membership in historically marginalized groups (Angwin, et al., 2016).



by (Angwin et al., 2016), based on more than 7,000 cases in Florida courts, revealed that the system generated structural biases: only 20% of those classified as “high risk” subsequently committed violent crimes, and individuals of African descent were nearly twice as likely to be misclassified as repeat offenders compared to white individuals. These biases were not merely statistical but translated into concrete legal consequences, as exemplified in the case of Glen Rodriguez, who was denied parole due to an incorrect automated rating (Wexler, 2017).<sup>7</sup>

As Mark Coeckelbergh points out, AI not only reproduces historical inequalities but can also amplify them on a structural scale. A paradigmatic example in the judicial context is COMPAS, an algorithm used to assess recidivism risk, which generates false positives that disproportionately affect racialized populations. Such algorithmic bias poses critical challenges for courts, undermining fairness, due process, and the protection of fundamental rights. Added to this are facial recognition systems used in criminal investigations, which can identify individuals without consent, violating privacy and personal integrity. Cases like Microsoft’s Tay chatbot (BBC, 2016) illustrate how AI systems can rapidly amplify social biases, highlighting the urgent need for secure regulatory frameworks, human oversight, and ethical-legal criteria to guide judicial use of AI (Coeckelbergh, 2021, pp. 18–19).

Beyond these examples, evidence demonstrates that AI not only reproduces but structurally amplifies inequalities. As Coeckelbergh warns, the apparent objectivity of data conceals historical biases embedded in institutional practices, rendering AI an opaque reflection of entrenched injustices (Coeckelbergh, 2021, pp. 18-19). This paradox compels a critical examination of the purported neutrality of algorithms and underscores the need to strengthen ethical oversight of their use in the judicial sphere. Rather than ensuring impartiality, these systems may reinforce discriminatory patterns and violate fundamental rights, such as the presumption of innocence and the right to a fair trial (Muñoz Rodríguez, 2020, pp. 702-705).

To address this problem, it is essential to preserve “meaningful human intervention” that prevents the uncritical delegation of judicial decision-making to opaque algorithms. This principle positions Judge Hermes at the core of the algorithmic era: an active guarantor of constitutional oversight over technologies that, under the guise of neutrality, may perpetuate structural injustices. Consequently, the judicial function is redefined as a hermeneutic and ethically grounded exercise, vigilant in the face of the risks posed by automation. As Dworkin (2000, pp. 13-15, 157) argues, even in “difficult cases,” the judge does not decide on the basis of discretion, but rather through interpretations consistent with constitutional principles, thereby humanizing the law and safeguarding human dignity even within a transformed technological context.

Contemporary law is dynamic and constantly evolving, which renders automated systems vulnerable to operating with outdated or decontextualized information. While

---

<sup>7</sup> See: UNESCO (2023), *Kit de herramientas global sobre IA y el Estado de derecho para el poder judicial*, p.104.

earlier AI systems reduced legal reasoning to rigidly formalized syllogistic schemes, the emergence of large language models (LLMs) introduces more flexible and context-sensitive capabilities, such as generating legal arguments, summarizing case law, and suggesting interpretive reasoning. Nevertheless, LLMs still face critical limitations, including opacity, potential bias, and challenges in accountability, meaning that human judicial oversight remains essential to ensure fairness, interpretive rigor, and respect for fundamental rights. Although judges employ deductive schemes, the selection of inferential rules and the evaluation of principles necessitate ethical deliberation that cannot be replicated by any algorithm. In this context, the model of Judge Hermes assumes particular significance, as it embodies the need for critical and contextual human judgment that AI cannot substitute, given its lack of social sensitivity, ethical discernment, and integrative capacity—dimensions that constitute the very essence of justice in complex societies (Martínez Bahena, 2012, pp. 836-837). While AI can help systematize information and streamline processes, it cannot replace the judicial function as an interpretive and humanizing act. Ultimately, reducing the law to computational efficiency ignores its ethical and reflective dimension (Velázquez Fernández, 2019, pp. 239-241). The judge's role is not merely to apply rules mechanically; it is primarily to ensure that fundamental rights and human dignity are protected in contexts where automated systems and algorithms influence judicial decision-making.

In this regard, it is essential to highlight the structural limitations of algorithmic adjudication. A “robot judge” does not seek a fair decision, but rather an accurate one, based on the automated application of normative sources through predefined procedures. This logic recalls the paradigm of the automaton judge in classical formalism, where judicial activity was reduced to the syllogistic application of the law, disregarding context and the values at stake. A return to this mode would effectively strip the law of its ethical and argumentative content, reducing it to a mere computational operation (Solar Cayón, 2019, cited in Tirso & Sánchez, 2025, p. 333). While AI can be applied to quantitative tasks, such as the settlement of convictions or enforcement proceedings, its design is inadequate for disputes that require qualitative interpretation and the weighing of principles. In these scenarios, the Hermes judge reaffirms their role as an irreplaceable hermeneutic guarantor, tasked with safeguarding the dialogical, ethical, and pluralistic character of justice in societies transformed by digitalization. In this new legal construct, Judge Hermes advocates a hermeneutic approach that does not strictly follow the original intention of the legislature, but rather takes into account pluralistic narratives and transdisciplinary knowledge in the context of the increasing hybridization of the judiciary, in which algorithms not only support lawyers in their decision-making, but increasingly replace them (Barona Vilar, 2024, pp. 83-84). Given the danger that the judiciary will become technologized and function as a self-referential and hermetic mechanism, Judge Hermes has the task of formulating interpretations that preserve the central importance of human rights. His mission is to ensure that fundamental principles such as dignity, privacy, and non-discrimination are not subordinated to the algorithmic logic of automated systems, which tend to function like “black boxes.”

In the context of a risk society marked by structural uncertainty, Judge Hermes plays an essential role as the guarantor of fundamental rights in the face of judicial

automation. The figure of the robot judge, conceived by Gómez Colomer as a machine that issues judgments by applying algorithms to specific facts, excludes key dimensions of human judgment, such as ethical, ideological, and emotional considerations. As Miraut states, the judicial function is not limited to the mechanical application of positive law, but includes subjective elements that cannot be translated into algorithmic formulas (Miraut Martín, 2023, cited in Tirso & Sánchez, 2025, p. 333).

A recent working paper by Posner and Saran (2025) comparing GPT-4o with experienced judges in a simulated judicial setting finds that GPT-4o adheres closely to legal precedent and is unaffected by factors such as defendant sympathy, whereas human judges in the original experiment displayed sensitivity to such extra-legal influences. This highlights a fundamental difference in decision-making styles and underscores the limitations of relying solely on AI for judicial reasoning. Even when instructed to consider contextual elements, AI responds with formalistic logic. This tension demonstrates that, although AI provides efficiency and regulatory consistency, as in the Catalan pilot project to draft simple commercial judgments, it lacks the ability to integrate moral judgments and axiological assessments (Morell, 2025).<sup>8</sup> This is not a matter of defending emotional decision-making, but rather of reaffirming that justice is a situated hermeneutic practice. Judge Hermes interprets, weighs, and contextualizes from the perspective of human rights and human dignity, preserving the deliberative and humanistic nature of law in the face of the technical rationality of automated systems.

This methodological contrast poses a crucial ethical and legal dilemma, particularly concerning so-called neuro-rights and the imperative of preserving meaningful human intervention in highly technological environments (Yuste et al., 2021, pp. 154-164; De Asís Roig, 2022, p. 63, cited in Tirso & Sánchez, 2025, p. 333). Although AI has been promoted as a solution to the inefficiency of many judicial systems, the deployment of large language models (LLMs) in decision-support roles raises significant structural risks. Unlike rigid algorithms, LLMs can generate legal reasoning and text, which introduces new challenges for accountability, transparency, and the protection of fundamental rights. AI should be conceived as a tool to aid judicial reasoning, but never as a substitute for ethical and deliberative judgment. As Battelli points out, the real challenge is not only to streamline processes, but also to ensure that justice retains its hermeneutic, ethical, and deeply human character (Battelli, 2021, cited in Tirso & Sánchez, 2025, p. 333).

Indeed, while traditional AI replicates rational faculties such as calculation, deduction, and classification, large language models (LLMs) extend these capacities by simulating aspects of human-like reasoning in language, generating arguments, summarizing case law, and contextualizing information. Nevertheless, they remain fundamentally different from genuine human understanding. In the legal sphere, this distinction is crucial: judgment requires interpretation, deliberation, and sensitivity to context (Velázquez Fernández, 2019, pp. 249-252). For this reason, Judge Hermes

---

<sup>8</sup> For a more technical analysis of the study, see Posner and Saran (2025), *Judge AI: Assessing Large Language Models in Judicial Decision-Making*.

cannot be replaced without undermining the axiological foundations of law. His role as guarantor of deliberative, situated justice goes beyond the mechanical application of preprogrammed rules, reconstructing the meaning of law based on case complexity and the primacy of human dignity. Even if a machine passed the Turing test, it would not possess consciousness or moral judgment. Justice, understood as a situated hermeneutic practice, cannot be reduced to artificially intelligent simulations (Elena Ortega, 2019, pp. 90-91). Judge Hermes thus not only interprets rules but also resists delegating law's interpretation to entities that, while efficient, lack ethical sensibility.

To gauge the scale of the problem, consider that each algorithmic decision operates like a mathematical formula that produces a technically correct result, but one that is indifferent to whom it affects or in what context it is applied. The law, however, cannot function in this way: it is not enough to obtain an answer; it is necessary to understand the values at stake and the people involved in each case. Hence, before blindly trusting the technical accuracy of an algorithm, it is essential to ask who reviews that decision and who ensures that it does not violate human rights. These questions lead to a central conclusion: every automated decision in the judicial sphere requires human validation and control. Consequently, the principle of “human in the loop”<sup>9</sup> must be reinforced, understood as a strategy of epistemic verification and ethical assurance against the risks of unsupervised automation. This approach implies that humans actively participate in different stages of the design, implementation, and use of AI systems, especially in the validation of the knowledge they produce (Rojas-Contreras et al., 2025, pp. 1-3). In the judicial sphere, this means that judges should not be replaced by technical efficiency, but rather that any recommendation generated by AI should be reviewed, contextualized, and checked against constitutional principles by trained legal practitioners (UNESCO, 2023, pp. 43-45)<sup>10</sup>. This reflection not only warns of the risks of judicial automation, but also offers a normative horizon for rethinking justice as a human-centered, deliberative, and ethically grounded practice in the digital age.

Furthermore, AI is still far from replicating the complexity of human judgment in social, legal, and emotional contexts. While algorithms outperform humans in specific pattern recognition and optimization tasks, they lack an understanding of the social environment and the subjective dimension that accompanies each case. This gap is especially critical in the judicial sphere, as data never tells the whole story and decisions involve values, emotions, and contextual justice. AI can serve as a predictive tool, but it should not replace the judgment of legal professionals or assume responsibility for making final decisions that affect fundamental rights. Therefore, it is necessary to move toward dynamic interpretive formulas, based on functional, contextual, and axiological criteria, that allow judges to adapt the meaning of the law to technological challenges.

---

<sup>9</sup> The Human-in-the-Loop (HITL) approach is a fundamental principle in the design and responsible use of artificial intelligence systems. It refers to the active intervention of humans at a critical stage of the automated process, whether in training, supervision, validation, or decision-making.

<sup>10</sup> See: Stankovich (2022), who emphasizes the importance of ethical impact assessments and public certification of algorithms, tools also advocated by UNESCO as essential mechanisms to ensure that AI deployment respects the rule of law and human rights.

Only then will they be able to evaluate the reasonableness of each decision from a complex perspective and, above all, one centered on human rights.

Based on the discussion above, the greatest ethical challenge lies in ensuring that AI truly assists human judgment rather than replaces it. In the context of large language models (LLMs), assistance may include generating legal arguments, summarizing case law, identifying relevant precedents, or highlighting potential inconsistencies in reasoning. While LLMs can enhance human decision-making and mitigate cognitive biases or information overload, they do not possess moral understanding or the ability to weigh values in context. That responsibility remains with critical human agents such as judges, authorities, educators, and researchers who serve as epistemic and ethical guarantors in technologically mediated decision-making. The digital era offers powerful tools that evolve rapidly, but it is essential to ensure that human perspective, deliberation, and ethical judgment remain central to the administration of justice.

AI can be a powerful ally in education, research, and justice, particularly through tools like large language models that assist in reasoning, summarizing information, and identifying relevant precedents. However, it will never replace the deep understanding, ethical judgment, and human sensitivity that only we can provide. Therefore, the real challenge is not merely technical, but ethical: ensuring that technology serves people, respects their dignity, and protects their rights. Only in this way can we build a future in which knowledge is not only automated, efficient, or innovative, but also equitable, inclusive, and responsible; a future in which technology supports critical thinking while human presence and judgment remain central. In this sense, the principle of keeping the human in the loop is indispensable: human beings must never be removed from decision-making, as their critical judgment and ethical responsibility are the only guarantees that technological development will remain at the service of justice and human rights.

## REFERENCES

- ADINOLFI, G. (2007). 'Autodeterminación informativa, consideraciones acerca de un principio general y un derecho fundamental', *Cuestiones Constitucionales. Revista Mexicana de Derecho Constitucional*, 1(17), pp. 1–29. Available at: <https://doi.org/10.22201/ijj.24484881e.2007.17.5807>
- AJDER, H., PATRINI, G., CAVALLI, F. and CULLEN, L. (2019). *The State of Deepfakes: Landscape, Threats, and Impact*. Amsterdam: Deepttrace. Available at: [https://regmedia.co.uk/2019/10/08/deepfake\\_report.pdf](https://regmedia.co.uk/2019/10/08/deepfake_report.pdf)
- ALBERT-MÁRQUEZ, M. (2021). 'Posthumanismo, inteligencia artificial y Derecho', *Persona y Derecho*, 84, pp. 207–230. Available at: <https://doi.org/10.15581/011.84.010> [Accessed: 22 september 2025]
- ANGWIN, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias*. ProPublica. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [Accessed: 26 september 2025]



- BARONA VILAR, S. (2024). ‘Justicia con algoritmos e Inteligencia Artificial, ¿acuerpando garantías y derechos procesales o liquidándolos?’, *Derechos y Libertades: Revista de Filosofía del Derecho y Derechos Humanos*, 51, pp. 83–115. Available at: <https://doi.org/10.20318/dyl.2024.8584> [Accessed: 9 July 2025]
- BAUMAN, Z. (2000). *Modernidad líquida*. Madrid: Fondo de Cultura Económica.
- BBC (2016). *Tay, la robot racista y xenófoba de Microsoft*. Available at: [https://www.bbc.com/mundo/noticias/2016/03/160325\\_tecnologia\\_microsoft\\_tay\\_bot\\_adolescente\\_inteligencia\\_artificial\\_racista\\_xenofoba\\_lb](https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb) [Accessed: 13 June 2025].
- BLÁZQUEZ RUIZ, F. J. (2022). ‘Riesgos para la privacidad en la aplicación de la inteligencia artificial al ámbito biosanitario. Implicaciones éticas y legales’, *Anales de la Cátedra Francisco Suárez*, (56), pp. 245-268. Available at: <https://doi.org/10.30827/acfs.v56i.21677> [Accessed: 13 July 2025]
- CÁCERES NIETO, E. (2024). ‘Reflexiones sobre la inteligencia artificial aplicada al derecho y el derecho de la inteligencia artificial: ¿Vamos hacia el mundo de Black Mirror?’, *Revista del Posgrado en Derecho de la UNAM*, 12(20), pp.75-118. Available at: <https://doi.org/10.22201/ppd.26831783e.2024.20.419> [Accessed: 16 July 2025]
- CÁMARA DE DIPUTADOS (2025). *Iniciativa que reforma y adiciona diversas disposiciones de la Ley Federal de Telecomunicaciones y Radiodifusión y del Código Penal Federal*, LXVI Legislatura. Available at: [https://sil.gobernacion.gob.mx/Archivos/Documentos/2025/03/asun\\_4858151\\_20250319\\_1742510064.pdf](https://sil.gobernacion.gob.mx/Archivos/Documentos/2025/03/asun_4858151_20250319_1742510064.pdf) [Accessed: 29 May 2025].
- CENTRO LATINOAMERICANO DE ADMINISTRACIÓN PARA EL DESARROLLO. CLAD, (2023). *Carta Iberoamericana de Inteligencia Artificial en la Administración Pública*. pp. 13-15. Available at: <https://rinedtep.edu.pa/server/api/core/bitstreams/e04dfadf-8fd9-45f8-b075-9fd70456c6db/content> [Accessed: 22 May 2025].
- CHOLLET, F. (2021). *Deep Learning with Python*, 2nd ed. New York: Manning Publications Co. LLC, pp. 1–7. Electronic version available at: <https://sourestdeeds.github.io/pdf/Deep%20Learning%20with%20Python.pdf> [Accessed: 1 June. 2025].
- COECKELBERGH, M. (2021). *Ética de la inteligencia artificial*, translated by L. Álvarez Canga, 1st ed. Madrid: Cátedra, Colección Teorema. Available at: <https://www.marcialpons.es/media/pdf/9788437642123.pdf> [Accessed: 19 May 2025].
- COMISIÓN EUROPEA (2019). *Directrices Éticas para una Inteligencia Artificial Confiable*, Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, Bruselas, April. Available at: <https://digitalstrategy.ec.europa.eu/es/library/ethics-guidelines-trustworthy-ai> [Accessed: 29 May 2025].
- COMISIÓN EUROPEA (2020). *Libro Blanco sobre la Inteligencia Artificial: un enfoque europeo hacia la excelencia y la confianza*, Bruselas. Available at: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52020DC0065> [Accessed: 29 May. 2025].

- COMISIÓN EUROPEA (2021). *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de IA)*, Bruselas. Available at: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206> [Accessed: 30 May. 2025].
- CONSEJO DE LA JUDICATURA FEDERAL (2018). *Criterios del Poder Judicial de la Federación en materia de protección de datos personales*, 2nd ed. México: Suprema Corte de Justicia de la Nación. Available at: <https://www.cjf.gob.mx/transparencia/resources/proteccionDatos/criteriosComiteTranparencia-13042023.pdf> [Accessed: 22 May 2025].
- CORTINA ORTS, A. (2019). 'Ética de la inteligencia artificial', *Anales de la Real Academia de Ciencias Morales y Políticas*, (96), pp. 379-394. Ministerio de Justicia. Available at: [https://www.boe.es/biblioteca\\_juridica/anuarios\\_derecho/abrir\\_pdf.php?id=ANU-M-2019-10037900394](https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-M-2019-10037900394) [Accessed: 13 May 2025].
- DELGADILLO, A. and GONZÁLEZ, C. (2023). 'Inteligencia artificial generativa: ¿qué es? ¿es un riesgo o ventaja?', *Revista Conecta*. Available at: <https://conecta.tec.mx/es/noticias/guadalajara/educacion/inteligencia-artificial-generativa-que-es-es-un-riesgo-o-ventaja> [Accessed: 11 June. 2025].
- DWORKIN, R. (2000). *Los derechos en serio*, translated by J.R. Capella, Barcelona: Ariel Derecho.
- ELENA ORTEGA, J. M. (2019). 'Singularidad tecnológica: ¿mito o nueva frontera de lo humano?', *Naturaleza y Libertad. Revista de Estudios Interdisciplinarios*, 12, pp. 87-103. Available at: <https://doi.org/10.24310/NATyLIB.2019.v0i12.6269>
- ESTER SÁNCHEZ, A. T. (2025). 'La inteligencia artificial en la justicia: Desafíos y oportunidades en la toma de decisiones judiciales', *Anales de la Cátedra Francisco Suárez*, 59, pp. 317–340. Available at: <https://doi.org/10.30827/acfs.v59i.31404>
- FEITO TORREZ, M. V. (2020). 'El juez Hermes y el logos de lo razonable: Por qué la aplicación silogística del Derecho no es suficiente', *Derecho y Ciencias Sociales*, 23, pp.112–117. Available at: <https://doi.org/10.24215/18522971e079>
- FLORIDI, L., COWLS, J., BELTRAMETTI, M., CHATILA, R., CHAZERAND, P., DIGNUM, V., LUETGE, C., MADELÍN, R., PAGALLO, U., ROSSI, F., SCHAFER, B., VALCKE, P. and VAYENA, E. (2018). 'AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations', *Minds and Machines*, 28(4), pp. 689–707. Available at: <https://doi.org/10.1007/s11023-018-9482-5>
- FREDIN, E. (2018). 'Keep calm and reskill – Reporte del Foro Económico Mundial plantea estrategias para aumentar el empleo frente a la automatización', *Edu News / Instituto para el Futuro de la Educación*. Available at: <https://goo.su/VUUYZy> [Accessed: 27 May. 2025].
- GRIGORE, A. E. (2022). 'Derechos humanos e inteligencia artificial', *IUS ET SCIENTIA*, 8(1), pp. 164–175. Available at: <https://doi.org/10.12795/IETSCIENTIA.2022.i01.10>

- GUIJOSA, C. (2017). ‘¿Los robots robarán nuestros trabajos? Estudio muestra cómo la automatización puede transformar el empleo en el mundo y en México’, *Edu News/Instituto para el Futuro de la Educación*. Available at: <https://goo.su/o35zK> [Accessed 27 May. 2025].
- GUZMÁN GARCÍA, M. Á. (2025). ‘Retos en la garantía de derechos humanos frente a la desaparición del INAI’, *Anuario de Filosofía y Teoría del Derecho*, 19, e19511. Available at: <https://doi.org/10.22201/ijj.24487937e.2025.19.19511>
- HUMAN RIGHTS WATCH, ACCESS NOW AND AMNESTY INTERNATIONAL (2018). ‘The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems’. Available at: <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems> [Accessed: 9 June. 2025].
- INSTITUTO FEDERAL DE TELECOMUNICACIONES [IFT] (2022). Anuario Estadístico 2022. Available at: <https://www.ift.org.mx/estadisticas/anuario-estadistico-2022>
- IFT (2023). Anuario Estadístico 2023. Available at: <https://www.ift.org.mx/sites/default/files/contenidogeneral/estadisticas/anuarioestadistico2023.pdf>
- KONWAR, P. (2025). ‘La protección de los derechos humanos y la integridad de la información en la era de la IA generativa’, *Crónica ONU*, 7 de enero. Available at: <https://www.un.org/es/cr%C3%B3nica-onu/la-protecci%C3%B3n-de-los-derechos-humanos-y-la-integridad-de-la-informaci%C3%B3n-en-la-era-de-la> [Accessed: 23 May 2025].
- MANYIKA, J., LUND, S., CHUI, M., BUGHIN, J., WOETZEL, L., BATRA, P., KO, R., and SANGHVI, S. (2017). ‘Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages’, *McKinsey Global Institute*. Available at: <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages> [Accessed: 27 May. 2025].
- MARTÍNEZ BAHENA, G. C. (2012). ‘La inteligencia artificial y su aplicación al campo del Derecho’, *Revista Alegatos*, (82), pp. 827–846. Available at: <https://alegatos.azc.uam.mx/index.php/ra/article/view/205>
- MEDINA ROMERO, M. Á. and TORRES CHÁVEZ, T. H. (2025). ‘Regulación de la Inteligencia Artificial: Desafíos para los Derechos Humanos en México’, *Revista Iberoamericana de Derecho Electrónico*, Vol. 15 Núm. 30 enero –junio 2025, e833 Available at: <https://doi.org/10.23913/ride.v15i30.2291>
- MORELL, J. (2025). ‘Los “jueces IA” son más legalistas que los jueces humanos’, *Blog de Innovación Legal – Abogacía Española*. Available at: <https://www.abogacia.es/publicaciones/blogs/blog-de-innovacion-legal/los-jueces-ia-son-mas-legalistas-que-los-jueces-humanos/> [Accessed: 25 June. 2025].
- MUÑOZ RODRÍGUEZ, A. B. (2020). ‘El impacto de la inteligencia artificial en el proceso penal’, *Anuario de la Facultad de Derecho. Universidad de Extremadura*, 36, pp. 695–728. Available at: <https://doi.org/10.17398/2695-7728.36.695>

- OST, F. J. (1993). “Hércules, Hermes: tres modelos de juez”, *Academia, Revista sobre la Enseñanza del Derecho*, Universidad de Buenos Aires, año 4, n° 8, 2007, p. 117. Publicado originalmente en *Doxa*, n° 14. Disponible en: <http://historico.juridicas.unam.mx/publica/librev/rev/acdmia/cont/8/teo/teo5.pdf>
- PARLAMENTO EUROPEO (2019). *Resolución de 12 de febrero de 2019 sobre una política industrial global en materia de inteligencia artificial y robótica* (2018/2088(INI)). Available at: [https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081\\_ES.html](https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_ES.html) [Accessed: 29 May. 2025].
- PÉREZ-UGENAC, M. (2024). ‘Análisis comparado de los distintos enfoques regulatorios de la inteligencia artificial en la Unión Europea, EE. UU., China e Iberoamérica’, *Anuario Iberoamericano de Justicia Constitucional*, 28(1), pp. 129–156. Available at: <https://doi.org/10.18042/cepc/aijc.28.05>
- POŠČIĆ, A. and MARTINOVIĆ, A. (2022). ‘Regulatory sandboxes under the draft EU Artificial Intelligence Act: An opportunity for SMEs?’, *InterEULawEast*, 9(2), pp. 71–117. Available at: <https://doi.org/10.22598/iele.2022.9.2.3>
- POSNER, E.A. and SARAN, S. (2025). ‘Judge AI: Assessing Large Language Models in Judicial Decision-Making’, *University of Chicago Law School, Coase-Sandor Institute for Law & Economics Research Paper* 25(03). Available at: <https://doi.org/10.2139/ssrn.5098708>
- RED IBEROAMERICANA DE PROTECCIÓN DE DATOS [REDIPD] (2020). ‘Recomendaciones generales para el tratamiento de datos en inteligencia artificial’. Available at: <https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf> [Accessed: 25 June. 2025].
- RIVAS-VALLEJO, P. (2021). ‘Discriminación algorítmica: detección, prevención y tutela’, *Jornadas Catalanas de Derecho Social 2021*. Available at: [https://www.researchgate.net/publication/354860623\\_Discriminacion\\_algoritmica\\_deteccion\\_prevencion\\_y\\_tutela](https://www.researchgate.net/publication/354860623_Discriminacion_algoritmica_deteccion_prevencion_y_tutela) [Accessed: 28 May. 2025].
- RODRÍGUEZ-AMAT, J. R. (2022). ‘Hacia una gobernanza de los datos de las plataformas. Explorando los desajustes entre los datos y el sentido’, *RAE-IC. Revista de la Asociación Española de Investigación de la Comunicación*, 9(18), pp. 45–74. Available at: <https://doi.org/10.24137/raeic.9.18.4>
- ROJAS-CONTRERAS, M., ORJUELADUARTE, A. and SANTOS JAIMES, L.M. (2025). ‘Human-in-the-loop (HITL) as a Verification and Validation Strategy for Knowledge Generated by Generative Artificial Intelligence’, *Proceedings of the 23rd LACCEI International Multi-Conference for Engineering, Education and Technology*, pp. 1–10. Available at: <https://doi.org/10.18687/LACCEI2025.1.1.1903>
- ROSZAK, T. (1994). *The cult of information: A neo-Luddite treatise on high-tech, artificial intelligence, and the true art of thinking* (2nd ed.). University of California Press.
- RUSSELL, S. and NORVIG, P. (2020). ‘What Is AI?’, in *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson Education, pp. 1–5. Available at: <https://luismejias21.>



- wordpress.com/wp-content/uploads/2017/09/inteligencia-artificial-un-enfoque-moderno-stuart-j-russell.pdf [Accessed: 30 May 2025].
- SADIN, É. (2020). *La inteligencia artificial o el desafío del siglo: Anatomía de un antihumanismo radical*, translated by M. Martínez, Buenos Aires: Caja Negra.
- SALCIDO LEDEZMA, M. A. (2025). ‘Reflexiones sobre la regulación en materia de inteligencia artificial: un análisis desde la perspectiva del derecho constitucional’, *ERITRONIO. Revista Iberoamericana de Derecho y Tecnología*, 1(1). Available at: <https://eritronio.org/index.php/revista/article/view/12>
- SÁNCHEZ DÍAZ, M. F. (2024). ‘Inteligencia artificial generativa y los retos en la protección de los datos personales’, *Revista Estudios en Derecho a la Información*, (18), pp. 179-205. Instituto de Investigaciones Jurídicas, UNAM. Available at: <https://doi.org/10.22201/ijj.25940082e.2024.18.18852>
- SARTORI, G. (1998). *Homo videns: Televisione e post-pensiero* [Homo videns: Television and post-thinking]. Laterza.
- SUPREMA CORTE DE JUSTICIA DE LA NACIÓN [SCJN] (2021). *Acción de Inconstitucionalidad 82/2021 y su acumulada 86/2021*. Ciudad de México: SCJN. Available at: [https://www2.scjn.gob.mx/juridica/engroses/cerrados/Publico/Proyecto/AI82\\_2021y86\\_2021acumuladaPL.pdf](https://www2.scjn.gob.mx/juridica/engroses/cerrados/Publico/Proyecto/AI82_2021y86_2021acumuladaPL.pdf) [Accessed: 29 May. 2025].
- SCJN (2024). *Acción de Inconstitucionalidad 66/2024, promovida por el Ejecutivo Federal contra el artículo 185 Bis C del Código Penal del Estado de Sinaloa*, ponencia del Ministro Juan Luis González Alcántara Carrancá’, *El Juego de la Corte*. Available at: <https://eljuegodelacorte.nexos.com.mx/la-suprema-corte-inicia-mal-la-discusion-sobre-inteligencia-artificial> [Accessed: 20 May 2025].
- SCJN (2025A). *Inteligencia artificial aplicada en procesos jurisdiccionales. Constituye una herramienta válida para calcular el monto de las garantías que se fijen en los juicios de amparo* (Tesis aislada, Registro digital: 2031009, Tesis: II.2o.C.8 K (11a.)). Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito. *Semanario Judicial de la Federación*, Undécima Época, Materia Común. Available at: <https://sjfsemanal.scjn.gob.mx/detalle/tesis/2031009> [Accessed: 29 May 2025].
- SCJN (2025B). *Inteligencia artificial aplicada en procesos jurisdiccionales. Elementos mínimos que deben observarse para su uso ético y responsable con perspectiva de derechos humanos* (Tesis aislada, Registro digital: 2031010, Tesis: II.2o.C.9 K (11a.)). Segundo Tribunal Colegiado en Materia Civil del Segundo Circuito. *Semanario Judicial de la Federación*, Undécima Época, materias: constitucional, común. Available at: <https://sjf2.scjn.gob.mx/detalle/tesis/2031010> [Accessed: 29 May 2025].
- STANKOVICH, M. (2022). ‘AI and Big Data Deployment in Health Care: Proposing Robust and Sustainable Governance Solutions for Developing Country Governments’, *UNDP Human Development Report 2021/22: Special Report on Human Security*. Available at: <https://hdr.undp.org/system/files/documents/background-paper-document/2021-22hdrstankovich.pdf> [Accessed: 19 June. 2025].



- TIRSO, A., and Sánchez, E. (2025). “La inteligencia artificial en la justicia: Desafíos y oportunidades en la toma de decisiones judiciales”. *ACFS. Revista de Filosofía Jurídica y Política*, Universidad de Granada, núm. 59, 2025, p. 333. Available at: <https://revistaseug.ugr.es/index.php/acfs/article/view/31404> [Accessed: 21 June. 2025].
- TRIBUNAL CONSTITUCIONAL DE ESPAÑA (1993). *Sentencia 238/2012*. Available at: <https://hj.tribunalconstitucional.es/es-ES/Resolucion/Show/238>
- UNESCO (2023). *Kit de herramientas global sobre IA y el Estado de derecho para el poder judicial*. Available at: [https://unesdoc.unesco.org/ark:/48223/pf0000387331\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000387331_spa) [Accessed: 26 June. 2025].
- VELÁZQUEZ FERNÁNDEZ, H. (2019). ‘¿Qué tan natural es la inteligencia artificial? Sobre los límites y alcances de la biomimesis computacional’, *Naturaleza y Libertad. Revista de Estudios Interdisciplinarios*, (12), pp. 237-258. Available at: <https://doi.org/10.24310/NATyLIB.2019.v0i12.6277>
- VINGE, V. (1993). ‘The coming technological singularity: How to survive in the post-human era’, *NASA Technical Reports Server*, Document ID 19940022856. Available at: <https://ntrs.nasa.gov/citations/19940022856>
- WEXLER, R. (2017). ‘Cuando un programa informático te mantiene en la cárcel’, *The New York Times*, 13 June 2017. Available at: <https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html> [Accessed: 25 June. 2025].
- WORLD ECONOMIC FORUM. (2018). *Towards a reskilling revolution: A future of jobs for all*. World Economic Forum, pp. 7-8 Available at: [https://www3.weforum.org/docs/WEF\\_FOW\\_Reskilling\\_Revolution.pdf](https://www3.weforum.org/docs/WEF_FOW_Reskilling_Revolution.pdf) [Accessed: 16 June. 2025].
- YUSTE, R., GENSER, J. and HERRMANN, S. (2021). ‘It’s Time for Neuro-Rights’, *Horizons: Journal of International Relations and Sustainable Development*, (18), pp. 154–165. Available at: <https://www.jstor.org/stable/48614119>
- ZAGREBELSKY, GUSTAVO, (1995). *El derecho dúctil*. Ley, derechos, justicia, 1ª ed., Editorial Trotta, (translated by Marina Gascón), pp. 115-135.
- ZUBOFF, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

Received: 3<sup>rd</sup> October 2025Accepted: 8<sup>th</sup> January 2026