Automation as Delegation of Power: Constitutional Constraints on AI Systems for the Administration of Justice*

Irina Carnat

Table of contents

1. Introduction

Among the most influential scholarship addressing the impact of technology-driven automation on public administrations is D. K. Citron's landmark 2008 article Technological Due Process, which argues that automated systems function as *de facto* delegations of rulemaking power, fundamentally challenging traditional notions of administrative accountability[1]. Nearly two decades later, this consideration still holds, further empowered by even more advanced technological artifacts and a greater risk of public accountability erosion. As instances of ChatGPT systems being used for judicial decision-making have already been documented across several jurisdictions[2], the concerns around said *de facto* delegation of power are

* Peer-reviewed article.
This work was supported by the SMaRT COnSTRUCT project (CUP J53C24001460006, as part of FAIR, PE0000013, CUP B53C22003630006, Italian National Recovery and Resilience Plan funded by NextGenerationEU).

[1] D.K. Citron, *Technological Due Process*, in *Washington University Law Review*, 85, 2008, 1249 ff.

[2] See, generally, D.U. Socol de la Osa – N. Remolina, *Artificial Intelligence at the Bench: Legal and Ethical Challenges of Informing—or Misinforming—Judicial Decision-Making through Generative AI*, in *Data & Policy*, 6, 2024, e59 ff. The author presents a series of cases from Colombia, Mexico, Peru, and India, featuring the judicial use of ChatGPT for different decision-making purposes, ranging from substantive legal research on autism therapy coverage and bail jurisprudence to procedural guidance for conducting virtual reality hearings and mathematical calculations for child support determinations, all conducted without established regulatory frameworks or specialized legal artificial intelligence tools. Also note that these represent only the documented and established cases, while many more are still being reported. See, for instance, in Europe, R. Pascoe, *Dutch Judge Uses ChatGPT to Help Reach a Verdict*, in *dutchnews.nl*, 5 August 2024. See also L. Taylor, *Colombian Judge Says He Used ChatGPT in Ruling*, in *The*

more than ever worth addressing.

The introduction of Artificial Intelligence (AI) systems in courtrooms is not just any form of automation that, in Citron's words, can jeopardize due process values[3]. Instead, it is a form of automated decision-making (ADM) process that can further disrupt traditional governance models due to its autonomy feature[4]. Moreover, considering the recent leap in the state-of-the-art legal automation thanks to the advent of Large Language Models (LLMs)[5], Generative AI (GenAI) systems pose additional challenges to ensuring meaningful human control and oversight due to their seemingly agentic properties[6]. In fact, unlike more traditional algorithmic ADM system, which process large volumes of data according to pre-established rules, GenAI systems can be deployed for a wider variety of tasks, with greater autonomy and with even lesser scrutiny over the training data.

However, if democratic societies do not wish to rule out altogether the potential benefits that AI-driven automation can bring to the administration of justice, we must carefully balance the efficiency argument with rigorous yet pragmatic risk assessment[7]. This balancing act requires understanding how established legal frameworks regulate AI development and deployment in high-stake decision-making processes. This entails, in other words, bridging the constitutional and administrative law principles aiming at preserving the principles of democratic governance as well as individual fundamental rights protection with the risk-based approach promoted by the recently approved EU Regulation on Artificial Intelligence (AI Act), which arguably follows a product safety logic rooted in private law tradition.

While existing scholarship on ADM in public sectors is both copious and prominent, often addressing the accountability gap, opaqueness, lack of explainability, and power asymmetry, these contributions typically rely on high-level considerations of due process, rule of law, and democratic oversight of algorithmic decision-making. Therefore, such scholarship remains largely ungrounded in the interpretation of the applicable legal framework, except for the GDPR, prior to the entry into force of the AI Act. Also, many authors cover traditional ADM through AI systems as machine-learning computational systems, but fewer contributions target LLM-based AI systems, despite their tangible impact on the legal profession and the documented cases in the judiciary. Therefore, while some considerations from previous scholarship may still hold, adaptation to the

Guardian, 3 February 2023; J.D. Gutiérrez, *ChatGPT in Colombian Courts: Why We Need to Have a Conversation about the Digital Literacy of the Judiciary*, in *verfassungsblog.de*, 2023.

[3] D.K. Citron, *Technological Due Process*, cit., 1300.

[4] O. Pollicino – F. Paolucci, *Regulating AI Autonomy: A Constitutional Framework for the Digital Era*, in M. Durante – U. Pagallo (eds.), *The De Gruyter Handbook on Law and Digital Technologies*, Berlin, 2025, 353 ff.

[5] H. Surden, *ChatGPT, Artificial Intelligence (AI) Large Language Models, and Law*, in *Fordham Law Review*, 92, 2024, 1941 ff.

[6] N. Kolt, *Governing AI Agents*, in *Notre Dame Law Review*, 101, forthcoming.

[7] D.K. Citron, *Technological Due Process*, cit., 1298.

recent change in the technological state of the art is nonetheless required. Against this brief conceptual background, the present article builds on existing scholarship on ADM in the public sector, limitedly to the European Union's legal order, and further aims to operationalize said constitutional and administrative law principles for AI-assisted public decision-making processes within the AI Act's risk-based legal framework.

To achieve this goal, the article first provides an account of the change in the state of the art of legal automation and its implications for ADM in public sectors. More specifically, section 2 showcases how the advent of GenAI shifts the focus from data governance to human-computer interaction. By focusing on the technical features of LLMs and the organizational specificities of GenAI systems, it presents additional risks, beyond bias and privacy violations, that jeopardize their safe deployment in high-stakes decision-making processes.

Section 3 then identifies the core constitutional values ensuring judicial accountability under the aegis of the rule of law and how these might affect and be affected by the factual delegation of judicial decision-making power to AI systems. It conceptually starts from the fundamental right to an effective judicial remedy, complemented by the right to good administration, to unfold the founding principles of due process, judicial independence, duty to state reasons, transparency, explainability, etc. However, faced with a proliferation of said principles of constitutional value from the EU administrative law discourse, it emphasizes the difficulty of translating them into actionable points for AI system development and deployment.

Finally, section 4 proceeds with an in-depth analysis of the AI Act's provisions relevant to addressing the risks of AI systems for the administration of justice. Particular attention is paid to the potential influence that a GenAI system may have on judicial decision-making, along with the requirements for fundamental right impact assessment and the right to an explanation. By a close analysis of such requirements, this article devises an algorithmic accountability framework that accounts for the constitutional constraints established for the democratically safe deployment of (Gen)AI systems for judicial decision-making.

The proposed algorithmic accountability framework bears the potential to procedurally and substantially operationalize the constitutional principles for accountable judicial decision-making into actionable risk management measures. This allows for *ex ante* risk mitigation and *ex post* remedies in case of violation of the fundamental right to an effective judicial remedy.

2. Automation as delegation of power: the advent of agentic generative AI

AI-driven Natural Language Processing (NLP) has had its most significant impact in the legal field due to language's fundamental role in establishing the rule of law, as legal systems depend entirely on linguistic expression to embody core principles of justice and governance. As M. Hildebrandt explains, the rule of law functions as an institution that ensures equal

application of legal standards while providing both predictability and the ability to challenge decisions[8]. This requires legal norms that rely on the inherent flexibility and interpretive capacity of natural language. As such, any meaningful advancement in legal automation must understand such a unique power that language has in the legal domain[9].

2.1. A shift in the state of the art of ADM

Previous applications of NLP depended on the state of the art of machine-learning algorithms to extrapolate patterns from the training dataset to automate the performance of a certain task[10] or to make a prediction on a new case.[11] Such AI capabilities have been deployed in the legal profession for automated document assembly, contract review,[12] prediction of the likelihood of legal outcomes,[13] etc. In the context of the administration of justice, algorithms were used for purely administrative or organizational purposes,[14] while in other cases, they were used to predict whether judges will hear a case and how they will decide it.[15] Predictive analytics models use statistical algorithms and machine learning on historical data to identify patterns and assign probability scores for future events, with applications including, for example, credit scoring and fraud detection[16]. This latter case is commonly referred to as 'predictive justice', claimed to improve the consistency and predictability of judicial decisions.[17] However, these traditional machine-learning models are technically more limited

[8] M. Hildebrandt, *The Adaptive Nature of Text-Driven Law*, in *Journal of Cross-Disciplinary Research in Computational Law*, 1, 2020, 10 ff.

[9] Surden, *Machine Learning and Law: An Overview*, in R. Vogl (ed.), *Research Handbook on Big Data Law*, Cheltenham, 2021, 171 ff.

[10] E. Filtz – M. Navas-Loro – C. Santos – A. Polleres – S. Kirrane, *Events Matter: Extraction of Events from Court Decisions*, in S. Villata – J. Harašta – P. Křemen (eds.), *Legal Knowledge and Information Systems*, Amsterdam, 2020, 33 ff.

[11] See generally N.A.K. Rosili – R. Hassan – N.H. Zakaria – S. Kasim – F.Z.C. Rose – T. Sutikno, *A Systematic Literature Review of Machine Learning Methods in Predicting Court Decisions*, in *IAES International Journal of Artificial Intelligence (IJ-AI)*, 10, 2021, 1091 ff. See also I. Chalkidis – I. Androutsopoulos – N. Aletras, *Neural Legal Judgment Prediction in English*, in arXiv, 5 June 2019.

[12] See, for instance, eBrevia by DFIN, available at dfinsolutions.com/products/ebrevia, 13 October 2025.

[13] D.M. Katz, *Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry*, in *Emory Law Journal*, 62, 2013.

[14] S.F. Schwemer – L. Tomada – T. Pasini, *Legal AI Systems in the EU's proposed Artificial Intelligence Act*, in Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2021), 21 June 2021.

[15] F. Pasquale – G. Cashwell, *Prediction, Persuasion, and the Jurisprudence of Behaviorism*, in *University of Toronto Law Journal*, 68, 2018.

[16] V. Kumar – M.L. Garg, *Predictive Analytics: A Review of Trends and Techniques*, in *International Journal of Computer Applications*, 182, 2018, 31 ff.

[17] F. Bex – H. Prakken, *Can Predictive Justice Improve the Predictability and Consistency of Judicial Decision-Making?*, in E. Schweighofer (ed.), *Legal Knowledge and Information Systems*, Amsterdam, 2021, 207 ff.

in their capabilities and potential deployment, requiring intense activities of task definition and manual labelling for supervised learning.[18] Moreover, their lack of understanding of the underlying legal semantics makes them more limited for applications in the legal domain[19].

Conversely, while still based on machine-learning algorithms from a technical perspective, LLMs are trained on immense amounts of data, amounting approximately to the entire corpus of textual data available online, facilitated by increased computational capacity in the recent years[20]. Among these, the Generative Pre-trained Transformer (GPT) model uses large-scale artificial neural networks to learn from text corpora, where weights are tuned to predict good probability distributions between an input token sequence and the next token in the output sequence[21].

A prominent example that gained global attention is ChatGPT, developed by OpenAI, which uses a generative model specifically trained to interact with users conversationally. It can not only generate text in natural language but also answer follow-up questions, coherently add information to previous requests, correct previous responses, admit potential errors, adopt corrective measures to adapt to user requests, and refuse to provide answers to requests considered inappropriate[22].

Similar systems currently available on the market include Claude by Anthropic[23], Google's Deepresearch, powered by the LLM Gemini[24], xAI's Grok[25], etc. From an organizational perspective, it is worth noting that the majority of the LLMs are US-based, except for the Chinese Deepseek model that has recently gained global attention[26], while the European counterparts, after the French company Mistral AI partnered with Microsoft[27], cannot propose a flagship model yet. An exception to this trend may be the recently released Swiss open-source model Apertus, claimed to be built with a "compliance-by-design" approach[28]. This anticipates a few

---

[18] H. Surden, *ChatGPT, Artificial Intelligence (AI) Large Language Models, and Law*, cit. The author also explores the earlier limitation of NLP, *ibid*., 5 ff.

[19] U. Pagallo – M. Durante, *The Pros and Cons of Legal Automation and Its Governance*, in *European Journal of Risk Regulation*, 7, 2016, 323 ff.

[20] A. Vaswani – N. Shazeer – N. Parmar – J. Uszkoreit – L. Jones – A.N. Gomez – Ł. Kaiser – I. Polosukhin, *Attention Is All You* Need, in arXiv, 5 December 2017.

[21] See, generally, T. Kaufmann – P. Weng – V. Bengs – E. Hüllermeier, *A Survey of Reinforcement Learning from Human Feedback,* in arXiv, 2023. On how ChatGPT was trained, see OpenAI, *Proximal Policy Optimization,* in openai.com, 20 July 2017. Cf. J. Schulman – F. Wolski – P. Dhariwal – A. Radford – O. Klimov, *Proximal Policy Optimization Algorithms,* in arXiv, 28 August 2017. See also OpenAI, *Aligning Language Models to Follow Instructions,* in openai.com, 14 February 2024.

[22] OpenAI, *Introducing ChatGPT*, in openai.com, 2021.

[23] Anthropic, Overview | Claude, in claude.com, 2025.

[24] Google, Gemini Deep Research – Your Personal Research Assistant, in gemini.google, 2025.

[25] X, Welcome, in x.ai, 2025.

[26] Deepseek, DeepSeek, in deepseek.com, 2025.

[27] Y. Malik – K. Hu, *Microsoft Partners with OpenAI's French Rival Mistral,* in Reuters, 26 February 2024.

[28] Swiss AI, Apertus, in swiss-ai, 2025.

concerns related to their private ownership and proprietary nature[29], along with the data processing outside the EU.
Given the wide adoption of these models, documented also by instances of usage in judicial settings, caution about their impact on the decision-making process is required.

2.2. (Over)relying on GenAI systems for judicial decision-making

Despite being an extraordinary tool, ChatGPT is fundamentally a chatbot. The difference from its traditional predecessors is that it is not programmed based on if-then programming rules but uses machine learning techniques with a generative model. Therefore, rather than providing predefined responses limited to a series of possible cases, as more traditional ADM systems do, tools like ChatGPT can autonomously generate responses with more complex syntactic structures that consider context, language register, and previously provided answers, simulating natural conversation.
The headline «GPT-4 passes the Bar exam» may be perceived as an ominous prophesy for the future of legal practice[30]. It creates expectations that LLM-powered AI systems can perform legal work at least as competently as newly qualified lawyers, but more efficiently in terms of costs and computation power. Besides the business opportunities that this technological advancement presents, particularly for private sector organizations, it simultaneously generates concerns about potential risks that researchers continue to investigate and assess.
In striking the balance between opportunities and risks, the efficiency argument weighs considerably[31]. LLMs demonstrate remarkable capabilities for legal interpretation, question answering, case prediction, and legal text generation, positioning them as potentially disruptive forces in cognitive legal automation[32]. Integrated into custom AI systems[33], these models can automate cognitive tasks traditionally requiring human intelligence[34], in-

[29] O. Pollicino – G. De Gregorio, *Constitutional Law in the Algorithmic Society*, in H.-W. Micklitz – O. Pollicino – A. Reichman – A. Simoncini – G. Sartor – G. De Gregorio (eds.), *Constitutional Challenges in the Algorithmic Society*, Cambridge, 2021, 11 ff.

[30] D.M. Katz – M.J. Bommarito II – S. Gao – P.D. Arredondo, *GPT-4 Passes the Bar Exam,* in arXiv, 15 March 2023.

[31] J. Zeleznikow, *Can Artificial Intelligence and Online Dispute Resolution Enhance Efficiency and Effectiveness in Courts*, in *International Journal for Court Administration*, 8, 2017, 30 ff.

[32] See, among others, F. Yu – L. Quartey – F. Schilder, *Legal Prompting: Teaching a Language Model to Think Like a Lawyer,* in *arXiv*, 8 December 2022; I. Chalkidis*, ChatGPT May Pass the Bar Exam Soon, but Has a Long Way to Go for the LexGLUE Benchmark*, in *arXiv,* 10 March 2023; J. Kaddour – J. Harris – M. Mozes – H. Bradley – R. Raileanu – R. McHardy, *Challenges and Applications of Large Language Models,* in *arXiv,* 19 July 2023.

[33] See for instance CoCounsel at *casetext.com*, DeepJudge at *deepjudge.ai*, or Aptus.ai at *aptus.ai.*

[34] A recent and interesting study conducted in China, on the judges' use of LLMs in their decision-making processes identified three relatively stable and recognizable patterns: 1) judges make initial decisions; 2) the large language model generates reasoning based on

cluding reasoning, research, and decision-making processes[35].

Computer science scholarship has identified various risk profiles associated with LLM usage[36], whereby special consideration was given to the so-called "hallucination" of generative LLMs, which refers to instances where language models produce factually incorrect, unfaithful, or nonsensical outputs that nonetheless appear fluent and plausible[37]. This phenomenon stems from various factors including data inconsistencies, training methodology limitations, and the absence of reliable sources for generated content[38]. Research demonstrates an alarming prevalence of hallucinations in legal applications[39], as it is very unlikely that LLMs were trained on accurate legal data. This purely technological limitation becomes particularly impactful when coupled with the anthropomorphic tendencies that users exhibit toward LLMs[40], which refer to the act of attributing human-like traits such as expertise, reasoning capabilities, and trustworthiness to these systems based on their conversational interfaces and sophisticated language generation[41].

As argued elsewhere, the intersection of anthropomorphism and hallucinations creates a compounded risk profile[42]: users may place unwarranted trust in AI-generated legal content that appears authoritative but contains fundamental inaccuracies[43]. This phenomenon may be also referred to as

---

the judges' decisions; and 3) judges revise the reasoning generated by AI to make the final judgment. J.Z. Liu – X. Li, *How Do Judges Use Large Language Models? Evidence from Shenzhen*, in *Journal of Legal Analysis*, 16, 2024, 235 ff.

35 H. Surden, *Artificial Intelligence and Law: An Overview*, in Georgia State University Law Review, 35, 2019; A. Agrawal – J.S. Gans – A. Goldfarb, *Do We Want Less Automation?*, in *Science*, 381, 2023, 155 ff.

36 L. Weidinger – J. Mellor – M. Rauh – C. Griffin – J. Uesato – P.-S. Huang – M. Cheng – M. Glaese – B. Balle – A. Kasirzadeh – Z. Kenton – S. Brown – W. Hawkins – T. Stepleton – C. Biles – A. Birhane – J. Haas – L. Rimell – L.A. Hendricks – W. Isaac – S. Legassick – G. Irving – I. Gabriel, *Taxonomy of Risks Posed by Language Models*, in *ACM Computing Surveys*, 2022.

37 Z. Ji – N. Lee – R. Frieske – T. Yu – D. Su – Y. Xu – E. Ishii – Y.J. Bang – A. Madotto – P. Fung, *Survey of Hallucination in Natural Language Generation*, in ACM Computing Surveys, 2022; S. Curran – S. Lansley – O. Bethell, *Hallucination Is the Last Thing You Need,* in arXiv, 20 June 2023.

38 OpenAI, *Introducing ChatGPT*, cit.

39 M. Dahl – V. Magesh – M. Suzgun – D.E. Ho*, Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models,* in arXiv, 2 January 2024.

40 A. Salles – K. Evers – M. Farisco, *Anthropomorphism in AI*, in *AJOB Neuroscience*, 11, 2020, 88 ff.

41 K.R. McKee – X. Bai – S.T. Fiske*,* Humans Perceive Warmth and Competence in Artificial Intelligence, in iScience, 26, 2023; Y. Kim – S.S. Sundar, *Anthropomorphism of Computers: Is It Mindful or Mindless?*, in *Computers in Human Behavior*, 28, 2012, 241 ff.; M. Natarajan – M. Gombolay, *Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction*, *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020.

42 I. Carnat, *Human, All Too Human: Accounting for Automation Bias in Generative Large Language Models*, in *International Data Privacy Law*, 2024.

43 M. Kneer – M.T. Stuart, *Playing the Blame Game with Robots*, Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 2021.

"agency laundering"[44], whereby responsibility is shifted from developers and deployers to the technology itself[45]. The seemingly increased agency attributed to such LLM-based GenAI systems may potentially obscure accountability mechanisms essential for judicial decision-making processes: a challenge that current constitutional frameworks may be ill-prepared to address[46].

2.3. The agentic properties of AI systems

Viewing AI systems as agents rather than tools has several implications for effective algorithmic governance.

The first aspect relates to the level of autonomy of these systems, compared to more traditional automation tools. Conventional ADM systems operate under pre-established rules and data inputs and are designed to essentially assist human decision-makers within their oversight boundaries[47], albeit with various degrees of automation[48]. In contrast, AI systems fundamentally differ in their capacity to learn, adapt, and evolve decision-making patterns over time, thus representing a qualitative and tangible shift from structured automation to increased agentic autonomy that fundamentally challenges the predictability and controllability conditions necessary to exercise effective human oversight for accountability purposes[49].

The second aspect proves relevant in terms of risk management: as a tool, the AI system serves predetermined human objectives and presents risks primarily related to their deployment and usage contexts, whereas as an agent, it may exhibit unpredictable behavior and may generate outcomes that diverge from human intentions[50]. This distinction also reflects different risk mitigation strategies: the former requires specific restrictions concerning the context and purpose of deployment, while the latter demands additional emphasis on technological development and integrated safety mechanisms[51].

---

[44] A. Rubel – A. Pham – C. Castro, *Agency Laundering and Algorithmic Decision Systems*, in N.G. Taylor – C. Christian-Lamb – M.H. Martin – B. Nardi (eds.), *Information in Contemporary Society*, Cham, 2019.

[45] This phenomenon can also be viewed through the lens of Nissenbaum's four barriers to accountability. See H. Nissenbaum, *Computing and Accountability*, in Communications of the ACM, 37, 1994, 72 ff. See also S. Nyholm, *Responsibility Gaps, Value Alignment, and Meaningful Human Control over Artificial Intelligence*, in *Risk and Responsibility in Context*, London, 2023.

[46] O. Pollicino – F. Paolucci, Regulating AI Autonomy: *A Constitutional Framework for the Digital Era*, cit., 356.

[47] *Ibid.*

[48] F. Palmiotto, *When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis*, in *German Law Journal*, 25, 2024, 210 ff.

[49] O. Pollicino – F. Paolucci, *Regulating AI Autonomy: A Constitutional Framework for the Digital Era*, cit., 356-361.

[50] C. Prunkl, *Human Autonomy at Risk? An Analysis of the Challenges from AI*, in *Minds and Machines*, 34, 2024, 26 ff.

[51] *Ibid.*, 26.

The third, and perhaps most important, aspect concerns the very delegation of decision-making power. As AI systems become more capable and thus are delegated more important tasks, they are perceived as agents and not as mere tools. In this regard, recent scholarship examines AI agency through established principal-agent framework[52]. However, despite their merit in raising concerns around power asymmetries, authority over delegated tasks and the potentiality of such AI agents to act in their self-interest rather than aligning with the principal's values[53], it can be critiqued from one fundamental stance: «the economic theory of principal-agent problems and the common law of agency was developed around a particular type of agent: human beings»[54], but the standard of care for AI systems cannot be assessed on the same grounds as human performance[55]. It follows that potential issues arising from such a delegation of decision-making power are surely complicated but not entirely cause by AI's agentic properties. The appropriate analytical focus therefore shifts from AI agency per se to the allocation of responsibility among human actors who design, deploy, and oversee these systems.

Moving beyond the instrumentalist theory of technology[56], while also rejecting the technological neutrality argument[57], a relational understanding of AI acknowledges that this technology inherently involves some loss of control due to its autonomy feature, but the ultimate impact stems not from its agentic properties exclusively but from how individuals interact with it[58]. In other words, AI merely alters the associated power dynamic, but delegating a task to AI does not eliminate the associated responsibility. Rather, it transforms how that responsibility is allocated between humans and the algorithmic components, a decision ultimately made by the AI developers, data scientists, and the entire cohort of actors involved[59]. As such, any delegation of decision-making to AI systems, accounting for its level of autonomy and perceived agency, must ensure that accountability is preserved among all the actors involved.

---

[52] N. Kolt, Governing AI Agents, cit., 16. Cfr. H.C.H. Hofmann, *Assessing Cyber Delegation in European Union Public Law*, in H.C.H. Hofmann – F. Pflücke (eds.), *Governance of Automated Decision-Making and EU Law*, Oxford, 2024, 42 ff.

[53] On alignment in AI, see, among others, Z. Kenton – T. Everitt – L. Weidinger – I. Gabriel – V. Mikulik – G. Irving, *Alignment of Language Agents*, in arXiv, 26 March 2021; S. Nyholm, *Responsibility Gaps, Value Alignment, and Meaningful Human Control over Artificial Intelligence*, cit.; I. Gabriel, *Artificial Intelligence, Values, and Alignment*, in *Minds and Machines*, 30, 2020, 411 ff.

[54] N. Kolt, *Governing AI Agents*, cit., 30.

[55] G. Comandé, *Multilayered (Accountable) Liability for Artificial Intelligence*, in S. Lohsse – R. Schulze – D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*, Oxford-Baden-Baden, 2018, 180 ff.

[56] M. Heidegger, *The Question Concerning Technology, and Other Essays*, New York, 1977.

[57] M. Kranzberg, *Technology and History: "Kranzberg's Laws"*, in *Technology and Culture*, 27, 1986, 544 ff.

[58] M. Hildebrandt – L. Tielemans, *Data Protection by Design and Technology Neutral Law*, in *Computer Law & Security Review*, 29, 2013, 509 ff.

[59] K. Martin, Ethical Implications and Accountability of Algorithms, in *Journal of Business Ethics*, 160, 2019, 835 ff.

3. Constitutional constraints on AI systems for the judiciary

The analysis in the previous section challenged prevailing assumptions about AI agents, redirecting focus from their technical autonomy to the accountability mechanisms required when delegating decision-making authority[60]. When such delegation involves judicial decision-making power, however, the complexity increases substantially. Judicial decision-making represents a distinctive form of exercise of public power that serves as the ultimate guarantor of individual rights and provides ex post review of administrative decisions, including those employing ADM systems[61]. Therefore, judicial decision-making cannot be formally delegated, and judicial accountability must always be preserved[62].

In this regard, Hofmann's analysis of the so-called "cyber-delegation" proves particularly insightful[63]. He points out that as any limitations on fundamental rights resulting from ADM must be predetermined by recognizable law under art. 52(1) of the Charter of Fundamental Rights. The procedural steps and criteria that automated systems are designed to assist or replace must clearly illustrate the chosen decision-making criteria, requiring transparency regarding both informational inputs and their processing methods. Furthermore, Hofmann's analysis of the Meroni doctrine's[64] limitations on delegation of discretionary powers[65] emphasizes that acts conferring powers to be exercised through ADM must circumscribe essential decision-making elements and cannot delegate fundamental balancing decisions concerning basic values to AI tools[66].

If delegating powers to ADM systems must adhere to such rigorous procedures for protecting fundamental rights, delegation in the judicial context demands even more stringent constraints, accounting for its fundamental role for democratic governance. Such constraints are represented, on the one hand, by the overarching principle of the rule of law that

[60] S. Demková, *The Horizontalisation of Fundamental Rights as a Part of EU Digital Constitutionalism?*, cit., 78.: «[…] clear allocation of responsibilities can be said to exist where the legislative arrangement of the obligations pertaining to specific decision-making conduct is concrete, precise, and transparent to enable the affected person to know where they can bring their claim against potential violations of their EU rights, and to enable the courts or other competent supervisory authorities to enforce protection of such rights».

[61] M. Artigot, *A Call for an Urgent and Unavoidable Democratic Debate: AI & ADM in Courts and its Impact on Rights and Justice*, in G. Gentile – P. Meszaros – N. Passalacqua – D. Penserini (eds.), *Artificial Intelligence and Courts*, 2024.

[62] C. Coglianese – D. Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, in *The Georgetown Law Journal*, 105, 2017, 33: «[…] if government actions should be undertaken by humans, then delegation to autonomously learning machines could potentially transfer governmental power outside the bounds that the Constitution permits».

[63] H.C.H. Hofmann, *Assessing Cyber Delegation in European Union Public Law*, cit.

[64] Case 9/56, *Meroni v ECSC High Authority* (1958).

[65] For empirical studies on how automation affects street-level bureaucratic activities, see N. de Boer – N. Raaphorst, *Automation and Discretion: Explaining the Effect of Automation on How Street-Level Bureaucrats Enforce*, in *Public Management Review*, 25, 2023, 42 ff.

[66] H.C.H. Hofmann, *Assessing Cyber Delegation in European Union Public Law*, cit.

governs judicial decision-making power, and on the other hand, by the fundamental right to an effective judicial remedy, as briefly analyzed below.

3.1. The rule of law as constitutional safeguard for an effective judicial remedy

The EU firmly grounds the exercise of public power within the boundaries of the rule of law, which serves as a foundational principle for the entire EU legal order, and it is enshrined among the values in art. 2 of the Treaty on European Union (TEU), alongside human dignity, freedom, and democracy[67]. Within this framework, the rule of law has evolved beyond a mere theoretical concept to become an enforceable constitutional principle with both formal and substantive components that is intrinsically linked to respect for democracy and for fundamental rights[68].

Under the aegis of the rule of law, the right to an effective judicial remedy and to a fair trial, guaranteed as a fundamental right under art. 47 of the Charter of Fundamental Rights (CFR), is a hallmark of EU law. The very existence of a right to an effective judicial review is regarded as the "right of rights", expressing the essence of the rule of law[69], and safeguarding its democratic facet[70]. As elaborated by the Court of Justice of the European Union (CJEU), this right is essential for ensuring both the rule of law within the Union and the protection of fundamental rights[71]. In other words, it serves the double function of both controlling the exercise of public administrative power and ensuring the application of EU law[72]. The CJEU has further consolidated this principle by affirming that art. 19(1) TEU gives concrete expression to the value of the rule of law stated in art. 2 TEU by imposing on Member States a justiciable obligation to «provide remedies sufficient to ensure effective legal protection in the fields covered by Union law»[73].

At this point, a preliminary clarification is warranted as to the use of the term "constitutional" throughout this article. While the EU does not have a formal constitution, the CJEU has long characterised the Treaties as the

[67] See Recital 6 of the AI Act, explicitly mentioning the values in art. 2 of the TUE.

[68] L. Pech, *The Rule of Law as a Well-Established and Well-Defined Principle of EU Law*, in *Hague Journal on the Rule of Law*, 14, 2022, 113.

[69] In a comparative perspective, see R.H. Fallon, *"The Rule of Law" as a Concept in Constitutional Discourse*, in *Columbia Law Review*, 97, 1997.

[70] For an extensive study on the right to effective judicial remedies in the context of automated decision-making, see S. Demková, *Automated Decision-Making and Effective Remedies: The New Dynamics in the Protection of EU Fundamental Rights in the Area of Freedom, Security and Justice*, Cheltenham, 2023.

[71] H.C.H. Hofmann, *The Right to an Effective Remedy and to a Fair Trial - Article 47 of the Charter and the Member States*, in S. Peers – T. Harvey – A. Ward (eds.), *The Charter of Fundamental Rights of the European Union*, 2nd ed., Oxford-London, 2019, 18.

[72] M. Hildebrandt, *Algorithmic Regulation and the Rule of Law*, in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376, 2018, 6.

[73] Case C-64/16, *Associação Sindical dos Juízes Portugueses* (2018), §32, cited in L. Pech, *The Rule of Law as a Well-Established and Well-Defined Principle of EU Law*, cit., 121.

Union's own constitutional charter. In the landmark Les Verts judgment, the Court held that the then EEC was a «Community based on the rule of law, inasmuch as neither its Member States nor its institutions can avoid a review» of the conformity of their measures with «the basic constitutional charter, the Treaty»[74]. This characterization was subsequently reinforced in Opinion 1/91 and Opinion 2/13, where the Court described the EU as possessing «its own constitutional framework and founding principles»[75]. Accordingly, references to "constitutional" constraints in this article designate the EU's autonomous constitutional order as enshrined in the Treaties, rather than the constitutional frameworks of individual Member States.

This clarification proves essential to understanding the subsequent case law on arts. 2 and 19 TEU, which has developed well beyond the foundational *Associação Sindical dos Juízes Portugueses* ruling. In that judgment, the Court established that art. 19(1) TEU applies to «the fields covered by Union law» irrespective of whether Member States are implementing EU law within the meaning of art. 51(1) CFR[76]. The scope of this provision reaches any national court that may potentially rule on EU law questions, which means practically every Member State court[77]. As Spieker explains, this produces a mechanism of «mutual amplification»: art. 19 TEU translates art. 2 TEU into a judicially applicable obligation, while art. 2 TEU extends the scope of application of art. 19 TEU beyond any other link to EU law[78].

This framework was subsequently consolidated through multiple Grand Chamber judgments. In *Commission v Poland*, the Court found, for the first time, a Member State in violation of art. 19(1) TEU, establishing the irremovability of judges as a component of judicial independence[79]. In *A.K. and Others*, the Court provided concrete criteria for assessing independence and mandated national courts to disapply incompatible national rules[80]. In Repubblika, it extended the doctrine beyond Poland and established a principle of non-regression, holding that a Member State «cannot […] amend

---

[74] Case 294/83, *Les Verts v Parliament* (1986), §23.

[75] Opinion 2/13 (Accession of the EU to the ECHR) (2014), §158. See also Opinion 1/91 (EEA Agreement) (1991), §21.

[76] Case C-64/16, *Associação Sindical dos Juízes Portugueses*, cit., §29.

[77] See L.D. Spieker, *Breathing Life into the Union's Common Values: On the Judicial Application of Article 2 TEU in the EU Value Crisis*, in German Law Journal, 20, 2019, 1182 ff., 1200; M. Bonelli – M. Claes, *Judicial Serendipity: How Portuguese Judges Came to the Rescue of the Polish Judiciary*, in European Constitutional Law Review, 14, 2018, 622 ff., 637.

[78] L.D. Spieker, *Defending Union Values in Judicial Proceedings. On How to Turn Article 2 TEU into a Judicially Applicable Provision*, in A. von Bogdandy – P. Bogdanowicz – I. Canor – C. Grabenwarter – M. Taborowski – M. Schmidt (eds.), Defending Checks and Balances in EU Member States, Heidelberg, 2021, 237ff., 250–251.

[79] Case C-619/18, *Commission v Poland (Independence of the Supreme Court)* (2019), §§76, 82. See M. Schmidt – P. Bogdanowicz, *Ascertaining the 'Guarantee of Guarantees': Recent Developments Regarding the Infringement Procedure in the EU's Rule of Law Crisis*, in A. von Bogdandy et al. (eds.), Defending Checks and Balances in EU Member States, Heidelberg, 2021, 207 ff.., 211.

[80] Joined Cases C-585/18, C-624/18, C-625/18, *A.K. and Others* (2019), §§120–154. See M. Leloup, *An Uncertain First Step in the Field of Judicial Self-government*, in European Constitutional Law Review, 16, 2020, 145 ff.

its legislation in such a way as to bring about a reduction in the protection of the value of the rule of law»[81]. As Lenaerts has articulated, judicial independence has thereby become an «autonomous concept of primary EU law» and national judges function as the «arm of EU law» within a single EU judicial architecture[82]. National judges are, accordingly, not only judges of domestic law but simultaneously judges of EU law, and the constitutional standards here invoked are those of the Union's own legal order.

3.2. Preserving judicial independence against undue algorithmic influence

Essential to effective remedies is judicial independence – both internal, understood as freedom from hierarchical pressure, and external, understood as freedom from other governmental branches – as an integral part of the fundamental democratic principle of the separation of powers[83]. As such, judges should not be subject to political influence or manipulation[84]. The CJEU has established that judicial independence is indispensable, clarifying that a "court" is always to be understood as meaning an "independent court"[85]. Following the Latin maxim *ubi ius, ibi remedium*, rights without enforcement mechanisms remain hollow, requiring both accessible remedial avenues and effective remedies themselves[86], manifested in procedural guarantees such as the duty of public bodies to provide reasoned decisions enabling individuals to exercise their rights meaningfully,[87] as reflected also in other Member States constitutional frameworks[88].
Against this background, the introduction of AI systems in judicial processes creates significant challenges for effective remedies and judicial independence by potentially fragmenting decision-making authority[89]. These

[81] Case C-896/19, *Repubblika v Il-Prim Ministru* (2021), §48.

[82] K. Lenaerts, *On Checks and Balances: the Rule of Law Within the EU*, in Columbia Journal of European Law, 29, 2023, 25ff., 35. See also C. O'Neill, *Defending Judicial Independence in Court: A Subjective Right to Independence in EU Law*, in Liverpool Law Review, 2025, 65 ff.

[83] R.H. Fallon, *"The Rule of Law" as a Concept in Constitutional Discourse*, cit., 9.

[84] European Commission for Democracy through law, *Adopted by the Venice Commission at Its 86th Plenary Session*, Venice, 25-26 March 2011.

[85] Case C-619/18, *Commission v Poland (Independence of the Supreme Court)* (2019), cit., § 48, cited in P. Bárd, *In Courts We Trust, or Should We? Judicial Independence as the Precondition for the Effectiveness of EU Law*, in *European Law Journal*, 27, 2021, 194-195.

[86] G. De Gregorio – S. Demková, *The Constitutional Right to an Effective Remedy in the Digital Age: A Perspective from Europe*, in C. van Oirsouw – J. Temperman – T. Vandamme – S. Touschner (eds.), *European Yearbook of Constitutional Law 2023: Constitutional Law in the Digital Era*, The Hague, 2024, 226.

[87] H.C.H. Hofmann, *The Right to an Effective Remedy and to a Fair Trial*, cit., 25.

[88] For example, art. 111 of the Italian Constitution enshrines the right to a fair trial before an impartial judge, mandates reasoned judicial decisions and guarantees access to the Court of Cassation for violations of law, see G. Amore – M.M. Lazzara, *Causes of Reflection on the Use of AI in Civil Justice*, in *The Italian Law Journal*, 10, 2024, 543.

[89] H.C.H. Hofmann, *Assessing Cyber Delegation in European Union Public Law*, cit., 34; see also S. Greenstein, *Preserving the Rule of Law in the Era of Artificial Intelligence (AI)*, in *Artificial*

challenges manifest across three critical dimensions that directly concern constitutional safeguards.

First, the deployment of AI raises important concerns regarding judicial independence and the exercise of judicial discretion[90], potentially undermining the constitutional architecture of separation of powers. Especially in criminal law proceedings, maintaining judicial independence is crucial for the right to a fair trial[91], but since GenAI systems may exercise undue external influence on any kind of judicial reasoning[92], it is relevant also in administrative and civil law proceedings[93]. As such, judicial independence can only be preserved through effective human oversight over the AI system's output[94], ensuring that judges retain the ultimate decision-making authority rather than becoming mere validators of algorithmic recommendations[95]. This novel allocation of responsibility between the AI system's developers and the judicial authority require distinct design principles for agent oversight in public administration contexts[96], where the stakes for individual rights are particularly high.

Second, the constitutional requirement to state reasons is jeopardized by algorithmic opacity. The duty to provide reasoned decisions enables individuals to understand the basis for judicial determinations and to exercise their rights to appeal or challenge those decisions effectively[97]. However, this requirement is increasingly more difficult to satisfy due to the intrinsic opacity of many advanced AI models, especially LLMs, creating barriers to explaining how certain inputs generated specific outputs. In this regard,

---

Intelligence and Law, 30, 2022, 291 ff.; A.Z. Huq, *Artificial Intelligence and the Rule of Law*, Univerity of Chicago, Public Law Working Paper No. 764, 2021; P.M. Nowotko, *AI in Judicial Application of Law and the Right to a Court*, in *Procedia Computer Science*, 192, 2021, 2220 ff.

[90] Amore – M.M. Lazzara, *Causes of Reflection on the Use of AI in Civil Justice*, cit., 544. Cfr. D.U. Socol de la Osa – N. Remolina, *Artificial Intelligence at the Bench*, cit., 21.

[91] S. Quattrocolo, *Fair Trial, Fair Judge: A Constitutional Framework for Digital Criminal Justice*, in G. De Gregorio – O. Pollicino – P. Valcke (eds.), *The Oxford Handbook of Digital Constitutionalism*, Oxford, 2024, 14.

[92] As it is explained infra in section 4.1, the degree of influence exercised by the AI system on the judicial decision-making process constitutes a factor that determines the risk classification of the AI system either as high-risk under art. 6, para. 2, or falling under the derogation in art. 6, para. 3.

[93] O. Mir, *Algorithms, Automation, and Administrative Procedure at EU Level*, in H.C.H. Hofmann – F. Pflücke (eds.), *Governance of Automated Decision-Making and EU Law*, Oxford, 2024.

[94] K. Härmand, *Translating Procedural Principles into Algorithms*, in G. Gentile – P. Meszaros – N. Passalacqua – D. Penserini (eds.), *Artificial Intelligence and Courts*, 2024, 12.

[95] D.U. Socol de la Osa – N. Remolina, *Artificial Intelligence at the Bench*, cit., 21.

[96] C. Schmitz – J. Rystrøm – J. Batzner, *Oversight Structures for Agentic AI in Public-Sector Organizations*, in E. Kamalloo – K. Xu – A. McCallum – Y. Liu – D. Reiter – D. Stoyanov (eds.), *Proceedings of the 1st Workshop for Research on Agent Language Models (REALM 2025)*, 2025.

[97] M. Fink – M. Finck, *Reasoned A(I)Dministration: Explanation Requirements in EU Law and the Automation of Public Administration*, in *European Law Review*, 47, 2022, 382. Accordingly, art. 47 CFR has been interpreted as to include the duty to state reasons, see K. Gutman, *The Essence of the Fundamental Right to an Effective Remedy and to a Fair Trial in the Case-Law of the Court of Justice of the European Union: The Best Is Yet to Come?*, in *German Law Journal*, 20, 2019, 884.

meaningful transparency requires not only access to the code itself, which may not be readily interpretable, but also to the predictive variables and formulae used to generate outputs in comprehensible terms[98]. Failure to provide this information may result in contestability of the decision, as well as grounds for appeal[99]. Moreover, this algorithmic opacity becomes especially problematic when judicial officers lack the technical expertise to assess the AI system's reasoning process. Thus, judges necessitate additional training beyond basic AI literacy to include knowledge on how the AI influences legal reasoning.

Third, judicial accountability itself is placed at risk when effective oversight measures cannot be implemented. The emphasis on human agency serves a purpose beyond merely monitoring automation as it essentially justifies legal decisions[100]. This means, in practical terms, imposing meaningful limits on automation in public decision-making processes[101], accounting not only for the AI's disruptive features, demanding increased transparency and explainability when deployed in high-stake contexts, but also directly regulating the permissible uses of AI systems in judicial settings[102].

These constitutional requirements converge on a single imperative: any AI system deployment in judicial contexts must preserve both the structural independence of judicial authority and the substantive capacity of affected individuals to meaningfully contest decisions.

4. (Gen)AI systems for the administration of justice under the AI Act

In line with the previous analysis, ss Hofmann suggests, accountability frameworks governing the delegation of decision-making power must be grounded in legislative acts or regulatory rulemaking.[103]. Arguably, the AI Act does represent such a legislative source, though its risk-based approach requires careful coordination with the constitutional framework governing public authority. Integrating these frameworks requires recognizing the rise of algorithmic accountability as a new set of substantive and procedural rights and remedies designed to address the lack of transparency and power imbalances, including rights to explanation and rights

98 T. Dancy – M. Zalnieriute, *AI and Transparency in Judicial Decision-Making*, in *Oxford Journal of Legal Studies (forthcoming),* 2025, 26-27.

99 A. Kouroutakis, *Rule of Law in the AI Era: Addressing Accountability, and the Digital Divide*, in *Discover Artificial Intelligence*, 4, 2024, 4.

100 R. Koulu, *Human Control over Automation: EU Policy and AI Ethics*, in *European Journal of Legal Studies*, 12, 2020, 9.

101 M. Coeckelbergh, *AI for Climate: Freedom, Justice, and Other Ethical and Political Challenges*, cit., 2056. See also U. Pagallo, *From Automation to Autonomous Systems: A Legal Phenomenology with Problems of Accountability*, Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017.

102 T.M. Lechterman, *The Concept of Accountability in AI Ethics and Governance*, in J.B. Bullock – Y.-C. Chen – J. Himmelreich – V.M. Hudson – A. Korinek – M.M. Young – B. Zhang (eds.), *The Oxford Handbook of AI Governance*, Oxford, 2022, 15.

103 H.C.H. Hofmann, *Assessing Cyber Delegation in European Union Public Law*, cit., 50.

to obtain information about algorithmic decision-making processes[104].
The following paragraphs proceed with a *de lege lata* analysis of select provisions within the AI Act that, interpreted in concert, prove most relevant for operationalizing constitutional constraints in judicial contexts.

4.1. The risk classification rules – and exceptions – under the AI Act

A risk-based approach to AI regulation entails that the compliance requirements for AI systems are commensurate to their risk profile[105]. As such, it is quantitative in nature, as risk is defined as «the combination of the probability of an occurrence of harm and the severity of that harm»[106]. The AI Act mainly regulated high-risk AI systems, to be determined according to the risk classification rules in art. 6. AI systems for the administration of justice fall under the classification of high-risk under art. 6, para. 2, by explicit reference to Annex III, point 8(a). More specifically, to be classified as high-risk, the AI system must have the specific intended purpose of being used «by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution». While this wording aligns with the state-of-the-art capabilities of generative AI systems mentioned *supra*, it rules out other scenarios that might nonetheless have important applications for judicial decision-making.
The first scenario worth analyzing is presented by the provision in art. 6, para. 3, according to which the same AI systems listed in Annex III, including those intended to be deployed in the judiciary, are subject to a derogation to such a presumption of high-risk if they «[do] not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making». Intuitively, the materiality of influence on the judicial decision-making process emerges as a critical risk determining factor. In this regard, Recital 53 specifies that «[a]n AI system that does not materially influence the outcome of decision-making should be understood to be an AI system that does not have an impact on the substance, and thereby the outcome, of decision-making, whether human or automated». As such, it is reasonable to infer that the materiality of such influence is determined by the role that the AI system plays in the decision-making process, proportionally linked to the degree of system automation, as argued elsewhere[107]. To corroborate this, art. 6 para. 3 further illustrates certain conditions for the derogation to the presumption of high-risk to

[104] O. Pollicino – G. De Gregorio, *Constitutional Law in the Algorithmic Society*, cit., 21.

[105] M. Ebers, *Truly Risk-Based Regulation of Artificial Intelligence How to Implement the EU's AI Act*, in *European Journal of Risk Regulation*, 2024,

[106] Art. 3(2) AI Act.

[107] I. Carnat, *Addressing the Risks of Generative AI for the Judiciary: The Accountability Framework(s) under the EU AI Act*, in *Computer Law & Security Review*, 55, 2024.

apply: specifically, the AI system must serve a limited, supportive role – performing narrow procedural tasks, improving completed human work, detecting decision-making patterns without replacing human assessment, or preparing materials for subsequent evaluation – rather than autonomously making decisions in high-risk contexts, although this derogation never applies where the AI system performs profiling of natural persons. This mechanism heavily relies on the provider's self-assessment that these conditions are met, creating a regulatory gap whereby AI systems with potentially substantial influence on judicial reasoning may evade high-risk classification, lacking third-party verification prior to deployment. However straightforward such conditions may appear, their practical implication in judicial contexts raises considerable legal uncertainty. The notion of "material influence" proves inherently difficult to operationalize, given the multifaceted and still understudied ways in which GenAI outputs can shape judicial reasoning. This interpretive challenge is exacerbated by the self-assessment framework under art. 6, para. 4, whereby providers must document their non-high-risk classification before market placement while remaining subject to the sword of Damocles of subsequent reassessment by national competent authorities under art. 80. This regulatory architecture thus creates substantial compliance uncertainty for providers of AI systems intended to support judicial decision-making while simultaneously presenting potential threats to judicial independence by permitting the deployment of systems that lack appropriate safeguards commensurate with their actual risk profile.

The second scenario concerns specifically LLM-based GenAI systems. Grounding the risk classification rules in the notion the intended purpose of use – apparently – excludes general-purpose AI (GPAI) systems, such as ChatGPT. Faced with the increased use of such tools in the judiciary, potential uncertainty around the risk classification of GPAI systems may have relevant consequences for their risk management framework. In fact, under the AI Act, only GPAI models, of which LLMs are an example, are subject to specific risk classification and compliance requirements laid down in Chapter V. It establishes a tiered regulatory framework whereby providers of GPAI models bear primary responsibility for transparency obligations, including the provision of technical documentation and training data summaries, with models presenting systemic risks facing heightened requirements for adversarial testing, risk mitigation, incident reporting, and cybersecurity protection[108].

However, as specified in Recital 101, these obligations primarily address the relationship between GPAI model providers and downstream providers, recognizing that GPAI model providers occupy a particular position in the AI value chain as their models form the basis for diverse downstream systems. As such, their compliance obligations are mainly aimed at enabling downstream providers of AI systems to comply with the AI Act's requirements.

[108] For the risk classification criteria for GPAI models presenting systemic risks and general compliance requirements, see Art. 51-55 AI Act. European Commission, *Guidelines on the Scope of the Obligations for General-Purpose AI Models Established by Regulation (EU) 2024/1689 (AI Act)* C(2025) 5045 Final.

As for GPAI systems themselves, defined in art. 3(66) as systems «based on a general-purpose AI model and which [have] the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems», they present an apparent regulatory lacuna[109]. Such systems appear to elude the risk classification framework established in art. 6, for which the high-risk profile is dependent on the intended purpose of use as defined by the AI system's provider. The inherent characteristic of GPAI systems, namely their capability to serve multiple, diverse purposes, renders the application of intended purpose-based classification problematic. This creates substantial legal uncertainty regarding whether and under what conditions a GPAI system used in judicial contexts would trigger the high-risk compliance requirements, particularly when such use occurs through direct deployment by judicial authorities rather than through integration into purpose-specific systems explicitly intended for judicial decision-making support.
Notably, the formal high-risk classification is highly relevant as it determines most of the compliance requirements for AI systems, including all those mentioned in the following paragraphs, along with the corresponding risk management measures.

4.2. The responsibility allocation along the AI value chain between providers, deployers, and surveillance authorities

Despite such an apparent regulatory gap for GPAI systems, the AI Act provides certain mechanisms for the effective allocation of responsibility along the AI value chain. The distinctive roles of providers and deployers of AI systems may also help operationalize the necessary constitutional safeguards in the judicial context, where individual fundamental rights are at stake.
Drawing upon previous works on establishing accountability duties along the AI value chain for high-risk AI systems[110], it is worth focusing here on specific provisions that effectively allocate the responsibility between providers and deployers with respect to the risk to fundamental rights posed by the AI system.
The first provision is art. 25, para. 1(c), which addresses situations where a deployer, such as a judge, uses an AI system, including a GPAI system, not formally classified as high-risk under art. 6 in a manner that effectively modifies its intended purpose[111]. This applies, for instance, when a GPAI system such as ChatGPT, or an AI system benefiting from the art. 6, para. 3 derogation, is deployed to automate decision-making in ways that materially influence judicial outcomes. In such circumstances, the de-

[109] C. Boine – D. Rolnick, *Why The AI Act Fails to Understand Generative AI*, in *Minnesota Journal of Law, Science & Technology*, 26, 2025.

[110] I. Carnat, *Addressing the Risks of Generative AI for the Judiciary*, cit.

[111] See also Recital 84 AI Act. Cfr. P. Hacker – M. Holweg, *The Regulation of Fine-Tuning: Federated Compliance for Modified General-Purpose AI Models,* 11 June 2025.

ployer assumes the legal status of provider and becomes subject to the corresponding obligations under the AI Act. Although it is very unlikely that the individual judge will be able to comply with all the obligations intended for providers[112], it is reasonable that the concept of deployer refers here to the public authority introducing the AI system into their decision-making process[113], i.e. the Ministry of Justice.

The second provision concerns again deployers, who bear important post-market monitoring obligations under art. 26, para. 5, as they must continuously monitor the operation of the AI system in accordance with the instructions of use and immediately inform the provider or distributor and the relevant market surveillance authority upon identifying any risks to fundamental rights, whether potential future violations or serious incidents constituting infringements of Union law that have already occurred. Upon such identification, deployers must suspend system use without undue delay.

To complete this accountability framework, market surveillance authorities assume a supervisory role that proves particularly significant for GPAI systems. Art. 75, para. 2 establishes that where market surveillance authorities have sufficient reason to consider that GPAI systems capable of direct use by deployers for at least one high-risk purpose are non-compliant with the Regulation's requirements[114], they must cooperate with the AI Office to conduct compliance evaluations and inform the Board and other market surveillance authorities accordingly.

This reporting obligation creates a continuous feedback loop between deployers and providers, enabling the incorporation of real-world experience and timely implementation of corrective actions, duly supervised by the market surveillance authority.

4.3. Fundamental rights impact assessment

This collaborative accountability framework is further operationalized through the fundamental rights impact assessment (FRIA) mandated under art. 27 for public bodies deploying high-risk AI systems[115]. Given that judicial authorities constitute public bodies governed by public law, the FRIA represents a critical proactive safeguard mechanism for AI systems intended for judicial decision-making contexts.

---

[112] C. Boine – D. Rolnick, *Why The AI Act Fails to Understand Generative AI*, in *Minnesota Journal of Law,* cit.

[113] I. Koivisto – R. Koulu – S. Larsson, *User Accounts: How Technological Concepts Permeate Public Law through the EU's AI Act*, in *Maastricht Journal of European and Comparative Law*, 31, 2024, 412.

[114] Recital 161 AI Act.

[115] The FRIA emerged as a significant achievement by the European Parliament during the AI Act's legislative process, addressing the Commission's initial failure to adequately implement fundamental rights protection within its proposed risk-based regulatory model, despite emphasizing human-centric principles. See A. Mantelero, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template*, in *Computer Law & Security Review*, 54, 2024.

The FRIA follows a methodologically grounded approach rooted in the traditions of both Data Protection Impact Assessment and Human Rights Impact Assessment, though adapted specifically for AI contexts[116]. This expert-based assessment comprises three essential phases: planning and scoping to contextualize the AI system and identify affected categories of rightsholders; risk analysis to assess potential impacts on fundamental rights by evaluating both likelihood and severity of adverse impacts; and risk management to prevent or mitigate identified risks through appropriate measures. The assessment maintains four defining characteristics: an *ex ante* approach that precedes deployment, a rights-based focus on risk assessment, a circular iterative structure following the system throughout its lifecycle, and an expert-based nature requiring specialized knowledge of fundamental rights doctrine and case law[117].

Under art. 27, para. 1, deployers must assess and document six key elements: the implementation process in accordance with instructions for use; the timeframe and frequency of system use; categories of affected persons; specific risks of harm to those persons; human oversight implementation measures; and arrangements for risk materialization including governance and complaint mechanisms. This comprehensive pre-deployment assessment enables deployers to anticipate and mitigate potential fundamental rights impacts before the system becomes operational[118].

Art. 27, para. 2 allows deployers to rely on previously conducted FRIAs carried out by providers, provided those assessments remain valid during deployment. However, when changes are identified in the post-deployment phase, deployers must update the FRIA accordingly. The implementation of FRIA in practice reveals significant interconnections with DPIA obligations, as answering questions in one assessment framework often provides essential information for the other, highlighting the complementary nature of data protection and broader fundamental rights considerations in AI deployment[119].

The FRIA under art. 27 represents a specific type of fundamental rights impact assessment applicable to deployers, yet it complements the broader obligation for providers to assess fundamental rights impacts during risk assessment under art. 9. These two assessments follow common risk man-

[116] *Ibid.*, 3-4.

[117] For practical guidance on implementing this four-phase FRIA methodology, encompassing purpose definition and responsibility allocation, systematic risk assessment using socio-technical factors throughout the AI system's development lifecycle, proportionality justification through balancing organizational interests against fundamental rights impacts, and iterative implementation of technical and organizational mitigation measures, see H. Janssen – M. Seng Ah Lee – J. Singh, *Practical Fundamental Rights Impact Assessments*, in *International Journal of Law and Information Technology*, 30, 2022, 200 ff.

[118] In this context, the notion of risk materialization must be properly understood as distinct from harm materialization and ex post remedies, instead referring to the emergence of foreseeable risks during system design and deployment that require proactive management measures, see A. Mantelero, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cit., 9.

[119] A. Thomaidou – K. Limniotis, *Navigating Through Human Rights in AI: Exploring the Interplay Between GDPR and Fundamental Rights Impact Assessment*, in *Journal of Cybersecurity and Privacy*, 5, 2025, 7 ff.

agement methodology and maintain a crucial linkage: deployer expertise in contextual deployment can reduce provider burden by addressing residual risks specific to implementation scenarios, while providers must supply deployers with adequate information and documentation to enable the proper execution of a FRIA[120]. Indeed, in line with the previously analyzed feedback loop, deployers must monitor system operation and inform providers of potential fundamental rights risks while taking into account FRIA results, thus creating a continuous assessment cycle.
This integration acknowledges that isolated accountability mechanisms prove insufficient for comprehensive oversight. Rather, the AI Act establishes what has been termed "aggregate accountability" by designing interconnected mechanisms across three governance stages: rule-setting through standardized assessment requirements, implementation through deployment guidelines and human oversight measures, and enforcement through reporting obligations and market surveillance[121], thus contributing to an effective accountability-based governance structure.

4.4. The right to an explanation

Having established *supra* that constitutional constraints demand transparent and reasoned judicial decision-making, the AI Act operationalizes the administrative duty to state reasons through art. 86's right to an explanation of individual decision-making. Art. 86, para. 1 establishes that affected persons have the right to obtain «clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken» when decisions based on high-risk AI systems produce legal effects or similarly significantly affect their health, safety, or fundamental rights, provided such persons consider the decision to have an adverse impact.
The scope of art. 86 extends beyond the GDPR's framework under arts. 15 and 22[122], creating what Kaminski and Malgieri term a "hydraulic effect": the broader the GDPR's right to explanation, the narrower art. 86's scope; conversely, if the GDPR's protections are limited, art. 86 assumes a gap-filling role[123]. Critically, while the GDPR's art. 22 covers solely automated decision-making, including scenarios where humans act as mere rubber-stampers of algorithmic outputs[124], art. 86 encompasses semi-automated decisions where AI systems play a substantial role alongside

[120] A. Mantelero, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act*, cit., 6.

[121] M.E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, in *Southern California Law Review,* 92, 6, 2019, 1529.

[122] A. Häuselmann, *Déjà vu? An Analysis of Explanations Concerning Decision-Making Under the GDPR and the AI Act*, in *Journal of AI Law and Regulation*, 2, 2025, 37 ff.

[123] M.E. Kaminski – G. Malgieri, *The Right to Explanation in the AI Act*, in *University of Colorado Law Legal Studies Research Paper No. 25-9,* 2025.

[124] Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, 2017.

meaningful human involvement[125]. As the CJEU's recent Dun & Bradstreet judgment clarified for GDPR contexts that meaningful information about algorithmic decision-making must enable affected persons to effectively exercise their right to express their point of view and contest decisions[126], art. 86 extends this principle to collaborative human-AI systems, recognizing that semi-automated decisions, which are the predominant model in today's judicial practice, equally require explanation rights to ensure effective contestation.

The requirement that explanations be «clear and meaningful» merits additional interpretation. Drawing from the AI Act's broader transparency framework, particularly provisions on disclosures to deployers under art. 13 and human oversight under art. 14, meaningful explanations must be understandable for their recipients[127]. Therefore, explanations should go beyond strictly technical specifications and must describe the role of the AI system relative to the human involvement, disclosing the extent to which human decision-makers exercised discretion to disregard, override, or reverse the system's output, along with any measures taken to ensure the correct interpretation of outputs against automation bias[128].

Therefore, the substantive content of explanations under art. 86 must address the «main elements of the decision taken», requiring individualized *ex post* explanations of specific decisions rather than abstract descriptions of system logic. This partially contrasts with earlier interpretations of GDPR's art. 15, para. 1(h), though the Dun & Bradstreet judgment clarified that even GDPR language should be understood to require disclosure of «procedures and principles actually applied» to individual decisions[129]. Read in light of Recital 171's requirement that explanations «provide a basis on which affected persons are able to exercise their rights»[130], the question arises whether the information that providers must disclose to deployers under art. 13, para. 3(b), namely the intended purpose of use, accuracy levels, known circumstances posing fundamental rights risks, performance regarding specific persons or groups, and input data characteristics, could provide a substantive baseline also for art. 86. In fact, deployers might be able to adapt this existing information for affected individuals, barring legitimate trade secret concerns,[131] though this still remains an open question.

Since art. 86 is understood as operating "on demand" rather than by de-

[125] Art. 86, para. 1, referring to decisions made «on the basis of the output» rather than «solely automated»; see M.E. Kaminski – G. Malgieri, *The Right to Explanation in the AI Act*, cit.

[126] Case C-203/22, Dun & Bradstreet (2024), § 57.

[127] Art. 13 concerns the information that must be provided by the AI system's provider to the AI system's deployer, while art. 14 concerns the downstream relationship between the AI system's deployer and the natural person to whom oversight is assigned.

[128] I. Carnat, *Human, All Too Human: Accounting for Automation Bias in Generative Large Language Models*, cit.

[129] Case C-203/22, Dun & Bradstreet (2024), § 58.

[130] G. De Gregorio – S. Demková, *The Constitutional Right to an Effective Remedy in the Digital Age*, cit.

[131] M.E. Kaminski – G. Malgieri, *The Right to Explanation in the AI Act*, cit., 19.

fault[132], is it crucial to couple it with the notification requirement under art. 26, para. 11 to ensure its effectiveness. Based on this latter provision, deployers must inform individuals not just that they are subject to AI-based decision-making, but also that it might have significant effects on their health, safety, and – most importantly for the case at hand – fundamental rights. This interpretation aligns and complements the GDPR's art. 22. Therefore, without effective notification, art. 86 appears devoid of practical remedial power, particularly for vulnerable individuals[133].

4.5. Human oversight and the fundamental requirement of AI literacy

Finally, the requirements of human oversight and the general imperative of AI literacy close the framework of the most constitutionally relevant provisions in the AI Act. Together, the provisions in arts. 14 and 4 seek to operationalize the principle that judicial decision-making power cannot be delegated to algorithmic systems, even when such systems materially influence judicial reasoning, in respect of judicial independence.

Art. 14 mandates that high-risk AI systems must be designed to enable effective oversight by natural persons, requiring both appropriate human-machine interface tools and measures commensurate to the risks, level of autonomy, and context of use. This regulatory requirement directly responds to the constitutional concern that judges must retain ultimate decision-making authority rather than becoming validators of algorithmic recommendations. The provision explicitly acknowledges automation bias as a widely known phenomenon, requiring that human overseers remain aware of the tendency to automatically rely on algorithmic outputs[134].

However, the effectiveness of human oversight mechanisms depends fundamentally on whether judicial operators possess sufficient understanding of AI systems' capabilities and limitations to exercise meaningful control, which brings art. 4's AI literacy requirements into play. Art. 4 establishes obligations for providers, deployers, and affected persons to ensure a sufficient level of AI literacy, defined in art. 3(56) as the skills, knowledge, and understanding necessary for informed deployment and awareness of AI opportunities, risks, and potential harms. This provision recognizes that constitutional safeguards cannot function effectively if judicial actors lack the technical competence to critically evaluate algorithmic outputs. In fact, operators require training that encompasses not merely the system's functioning but also exposure to cases of system failures, operational procedures, available tools and interfaces for human intervention, and guidance criteria for assessing the AI system's recommendation quality

[132] *Ibid.,* 20.

[133] M.-L. Rebrean – G. Malgieri, *Vulnerability in the EU AI Act: Building an Interpretation*, Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, 2025.

[134] European Data Protection Supervisor, *TechDispatch: Human Oversight of Automated Decision Making*, 2025.

in contexts requiring human discretion[135]. In this sense, the requirement for AI literacy is a form of democratic control, since judges cannot fulfill their duty to state reasons if they were not properly trained to understand how AI systems function and how their recommendations can impact the affected individual's fundamental rights[136].

This integrated framework of human oversight and AI literacy operationalizes the constitutional constraint on judicial independence even in the prospect of GenAI systems that influence the judicial decision-making process. Art. 14 ensures structural mechanisms, while art. 4 ensures the cognitive capacity, to exercise meaningful human control[137]. Interpreted together, they translate abstract constitutional principles into concrete obligations: providers must design systems that enable oversight, deployers must ensure operators receive appropriate training, and judicial authorities must maintain sufficient AI literacy to fulfill their constitutional role as independent decision-makers. Nonetheless, faced with the risk of positioning human operators as scapegoats for algorithmic harms when errors typically stem from factors beyond their control[138], the effectiveness of this operationalization ultimately depends on whether these provisions are implemented acknowledging that they serve not merely technical compliance functions but fundamental constitutional imperatives safeguarding the rule of law in an age of algorithmic governance, while ensuring accountability is appropriately distributed among all actors in the AI value chain rather than concentrated on judicial overseers alone[139].

5. Concluding remarks

This article examined how the deployment of AI systems – most recently generative large language models that bring about new potential for automation of cognitive tasks – in judicial decision-making may constitute a *de facto* delegation of decision-making power that challenge existing accountability mechanisms and thus demand rigorous constitutional scrutiny.

The identified constitutional constraints, grounded in the rule of law and the fundamental right to an effective judicial remedy, establish non-negotiable boundaries for AI deployment in judicial contexts. These constraints

[135] S. Sterz – S. Aliman – E. Krakowski – P. Leon – V. Dignum, *On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives*, The 2024 ACM Conference on Fairness, Accountability, and Transparency, 2024.

[136] See Recitals 20 and 91 AI Act.

[137] S. Nyholm, *Responsibility Gaps, Value Alignment, and Meaningful Human Control over Artificial Intelligence*, cit.; L. Methnani – A. Esquivel – M. Alzghoul – F. Bacchus – M. Hüllermeier, *Let Me Take Over: Variable Autonomy for Meaningful Human Control*, in *Frontiers in Artificial Intelligence*, 4, 2021; F. Santoni de Sio – J. van den Hoven, *Meaningful Human Control over Autonomous Systems: A Philosophical Account*, in *Frontiers in Robotics and AI*, 5, 2018.

[138] B. Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, in *Computer Law & Security Review*, 45, 2022.

[139] Wagner, *Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems*, in *Policy & Internet*, 11, 2019, 104 ff. R. Crootof – M.E. Kaminski – W.N. Price II, *Humans in the Loop,* in *Vanderbilt Law Review,* 76, 2023, 429.

require that any delegation of decision-making power must preserve judicial independence, ensure transparent and reasoned decision-making, and maintain effective accountability mechanisms. This demands a democratic governance framework that addresses the entire AI value chain rather than technical solutions alone.

Against this background, the AI Act represents a significant legislative response to this challenge that bears the potential to operationalize said constitutional constraints through its risk-based regulatory framework. To this purpose, the most prominent and promising provisions were analyzed, including the fundamental rights impact assessment, the right to an explanation, human oversight and AI literacy obligations, to showcase how abstract constitutional principles may be translated in actionable accountability mechanisms along the entire AI value chain. Nevertheless, some critical interpretative and regulatory gaps remain, especially concerning the regulation of general-purpose AI systems and the novel derogation provision to the presumption of high-risk of AI systems that allegedly do not materially influence the decision-making process.

Yet, the proposed algorithmic accountability framework synthesizes constitutional principles with the AI Act's regulatory requirements, establishing multilayered governance that distributes responsibility appropriately among providers, deployers, and surveillance authorities. This framework emphasizes continuous monitoring, feedback loops between actors in the AI value chain, and iterative risk assessment throughout the system's lifecycle.

Looking forward to possible developments of this framework, the central role of public procurement practices emerges as another democratic governance dimension[140], whereby the constitutional constraints identified here are further operationalized through contractual agreements whenever AI systems are provided by private economic actors[141]. This becomes particularly critical given the risk of transferring normative decision-making authority to profit-driven entities that may prioritize efficiency over constitutional safeguards. The proposed accountability framework therefore serves a dual purpose: providing concrete grounds for trustworthiness verification in public procurement practices while ensuring that procurement officials possess sufficient capacity to conduct meaningful due diligence, auditing, and negotiate contractual terms for the procurement of AI systems that by design preserve judicial accountability and enable effective oversight throughout the AI system's lifecycle.

On a final note, this article acknowledges the importance of interdisciplinary socio-technical expertise in navigating these challenges. AI-driven automation, enforced by the efficiency argument, need not be the inevitable fate of judicial decision-making. The landmark COMPAS case was regarded as the first wake-up call[142], yet the recent instances of judges and

---

[140] C. Coglianese, *Procurement and Artificial Intelligence*, in *Handbook on Public Policy and AI*, 2024.

[141] M. Hickok, *Public Procurement of Artificial Intelligence Systems: New Risks and Future Proofing*, in *AI & Society*, 39, 2024, 1213 ff.; A. Sanchez-Graells, *Digital Technologies and Public Procurement: Gatekeeping and Experimentation in Digital Public Governance*, Oxford, 2024.

[142] K. Martin, *Ethical Implications and Accountability of Algorithms*, cit.

even lawyers[143] (over)relying on ChatGPT represent a much stronger cry for a more effective constitutional guidance. The framework proposed here demonstrates that addressing these challenges requires constitutional principles to descend from the ivory tower of abstract doctrine and engage substantively with the pragmatic, risk-based regulatory approach embodied in the AI Act: translating the principles of judicial independence, transparency, and accountability into concrete technical requirements, contractual obligations, and risk management measures that can guide AI system development, procurement, and deployment. Only through this integration of constitutional imperatives with product safety regulation can democratic societies ensure that the efficiency gains promised by AI do not come at the expense of the fundamental right to an effective judicial remedy and the constitutional safeguards that underpin the rule of law.

[143] For notable cases of AI hallucinations in legal briefs, see *AI Hallucination Cases Database*, in *damiencharlotin.com*, 2025.

Abstract

The deployment of generative AI systems in judicial decision-making constitutes a *de facto* delegation of power that threatens constitutional principles governing the administration of justice, particularly the fundamental right to an effective judicial remedy under art. 47 of the Charter of Fundamental Rights. This article operationalizes constitutional constraints within the EU AI Act's risk-based regulatory framework by analyzing relevant provisions on risk classification, fundamental rights impact assessments, human oversight, and the right to explanation. The proposed algorithmic accountability framework distributes responsibility among providers, deployers, and surveillance authorities, demonstrating how constitutional principles must engage substantively with pragmatic product safety regulation to preserve judicial independence and accountability in an age of algorithmic governance.

Keywords

automated decision-making – algorithmic accountability – fundamental rights protection – human oversight – Generative AI