The Invisible Ripple Effect

Eldar Haber*

A quiet crisis is unfolding across expert domains. Government reports now embed fabricated citations that shape policy. Academic publications pass peer review with phantom references. Lawyers increasingly cite nonexistent precedents generated by AI. These failures reveal a deeper threat: what this Article terms the Invisible Ripple Effect. Unlike traditional AI systems confined to institutional oversight, generative AI has democratized access to powerful yet unreliable tools. Individual professionals now integrate AI outputs—polished, authoritative, yet systematically flawed—into knowledge systems built for human authorship. The result is not sporadic error but systemic contamination. Fabricated citations propagate through brief banks and court opinions. Phantom references cascade through academic databases. AI-generated medical data infiltrates clinical protocols. With each iteration, errors gain legitimacy, transforming falsehoods into institutional knowledge.

Existing governance frameworks are structurally unprepared for this diffusion. The EU AI Act targets institutional deployers but neglects individual professionals. U.S. oversight remains fragmented, and professional liability regimes presume intentional misconduct, allowing sophisticated fabrications to pass undetected. These approaches address discrete harms but not replication-driven risks. This Article argues for a new regulatory paradigm: governance that disrupts propagation rather than merely preventing individual errors. It proposes a five-stage intervention framework—prevention, containment, control, mitigation, and resilience—to halt how GenAI contaminates professional systems. As AI-generated content saturates expert decision-making, the stakes extend beyond technology policy: they threaten the very foundations of professional authority and corrode epistemic trust.

---

* Associate Professor, Faculty of Law, University of Haifa; Director, Haifa Center for Law and Technology (HCLT). I would like to thank Noam Ben Moshe and Nadeen Salameh for their excellent research assistance. All mistakes are mine.

INTRODUCTION

AI-generated fabrications are quietly corrupting our knowledge infrastructure. Government reports may embed phantom citations that could shape policy decisions.[1] Fictional legal precedents may infiltrate case law databases, distorting legal jurisprudence.[2] Phantom academic citations could cascade through research networks.[3] Fabricated financial data may influence market analyses and regulatory frameworks.[4] Synthetic medical studies could alter treatment protocols across healthcare systems.[5] AI-generated news reports may circulate through information networks.[6] Vulnerable AI-written code could propagate across critical and information infrastructure.[7]

---

1. In May 2025, the White House's Make America Healthy Again Commission released a report containing fabricated sources likely made by generative AI (GenAI) tools. Though initially dismissed as "formatting issues," the incident led to the release of a revised version and raised serious concerns about the integrity of the U.S. Official Record and the democratic risks of unvetted GenAI use in government documentation. *See* Lauren Weber & Caitlin Gilbert, *White House MAHA Report May Have Garbled Science by Using AI, Experts Say*, WASH. POST (May 29, 2025), https://www.washingtonpost.com/health/2025/05/29/maha-rfk-jr-ai-garble (reporting that AI experts identified citation patterns in a White House health report that appear consistent with artificial intelligence generation).

2. *See* Damien Charlotin, *AI Hallucination Cases*, https://www.damiencharlotin.com/hallucinations [https://perma.cc/N2RG-HSWW] (dataset, continuously updated, 979 cases as of March 1, 2026). For further statistical analysis, see *infra* Section I.B.

3. *See* Stephanie M. Lee, *Scholars Are Supposed to Say When They Use AI. Do They?*, CHRON. HIGHER EDUC. (Dec. 18, 2024), https://www.chronicle.com/article/scholars-are-supposed-to-say-when-they-use-ai-do-they (reporting that *PLOS ONE* retracted an article after finding 18 unverifiable, likely ChatGPT-generated, references and describing similar undisclosed AI-fabricated citations in other journals). Furthermore, a recent *Retraction Watch* investigation revealed that a Springer Nature textbook on machine learning contained a dozen fabricated or erroneous citations—likely the result of GenAI use, though the publisher offered no formal disclosure. *See* Rita Aksenfeld, *Springer Nature Book on Machine Learning Is Full of Made-Up Citations*, RETRACTION WATCH (June 30, 2025), https://retractionwatch.com/2025/06/30/springer-nature-book-on-machine-learning-is-full-of-made-up-citations [https://perma.cc/W4PA-TS8N].

4. *See* FIN. STABILITY BD., THE FINANCIAL STABILITY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE 2, 15 (2024), https://www.fsb.org/wp-content/uploads/P14112024.pdf [https://perma.cc/VG9G-SYHC] (noting that hallucinated AI outputs can contribute to financial fraud and market disinformation, complicating regulatory oversight).

5. *See* Sandeep Reddy, *Generative AI in Healthcare: An Implementation Science Informed Translational Path on Application, Integration and Governance*, IMPLEMENTATION SCI., Mar. 2024, at 1, 8 (warning that the integration of generative AI in healthcare could lead to clinical error or inappropriate treatment).

6. *See* Paul Farhi, *A News Site Used AI to Write Articles. It Was a Journalistic Disaster.*, WASH. POST (Jan. 17, 2023), https://www.washingtonpost.com/media/2023/01/17/cnet-ai-articles-journalism-corrections (reporting that AI-generated personal-finance pieces—later

This contamination stems from the widespread adoption of generative AI (GenAI) tools whose fundamental flaws interact with human cognitive biases and institutional pressures. The technological foundation creates the conditions: GenAI systems produce errors of various forms—hallucinations, omissions, outdated information, and subtle factual distortions—that appear authoritative and polished.[8] The technology's appeal lies precisely in this fluency: AI-generated content aligns with human cognitive biases toward well-formatted, confident-sounding text, making errors difficult to detect.[9] Combined with competitive pressures and the allure of AI capabilities, this creates conditions where professionals rapidly adopt tools that many cannot adequately evaluate, embedding flawed outputs within institutional knowledge systems.[10]

GenAI tools offer undeniable efficiency gains, automating complex research and analysis tasks that once required hours of specialized work.[11] Yet the technical architecture that enables this efficiency makes errors almost inevitable.[12] GenAI operates through probabilistic pattern-matching rather than verified knowledge retrieval, rendering hallucinations—polished, plausible fabrications—an inherent feature rather than a bug.[13] The same architecture produces systemic omissions, outdated information, and subtle factual distortions that compound the reliability problem.[14]

The sophistication paradox exacerbates these technical limitations, making them particularly insidious. As these tools become more advanced,

---

corrected for multiple mathematical errors—were also carried by sister outlet Bankrate, illustrating how a faulty robo-copy spilled into the company's wider news network).

7. *See* Yujia Fu et al., *Security Weaknesses of Copilot-Generated Code in Open-Source Projects: An Empirical Study*, ACM TRANSACTIONS SOFTWARE ENG'G METHODOLOGY, Feb. 2025, at 1, 3, https://arxiv.org/abs/2310.02059 [https://perma.cc/LNT7-UG85] (finding that roughly one-third of AI-written code adopted by public projects contained serious security flaws).

8. *See infra* Part I.

9. *See infra* Part I.

10. *See* Anna Covert, *FOMO and AI: Do You Have Fear of Missing Out?*, FORBES (Jan. 15, 2024), https://www.forbes.com/sites/forbesbooksauthors/2024/01/15/fomo-and-ai-do-you-have-fear-of-missing-out (describing AI FOMO phenomenon).

11. *See, e.g.*, Noam Scheiber, *Which Workers Will A.I. Hurt Most: The Young or the Experienced?*, N.Y. TIMES (July 7, 2025), https://www.nytimes.com/2025/07/07/business/ai-job-cuts.html (noting that GenAI can dramatically streamline cognitive tasks, potentially replacing both entry-level and experienced professionals, depending on how routine or automatable their workflows are).

12. *See infra* Part I.

13. *See infra* Part I.

14. *See infra* Part I.

they paradoxically become more dangerous: the rate and sophistication of errors, such as hallucinations, increase with model improvements, while better writing quality and more confident presentation make these tools more appealing to use and their flaws harder to identify.[15] Furthermore, mass adoption has transformed these individual technical flaws into systemic vulnerabilities. What were once isolated limitations have become widespread risks as more users integrate AI-generated content into professional workflows without adequate verification mechanisms.[16] It is within this context of systemic vulnerability that a novel form of knowledge corruption emerges.

This Article introduces and theorizes the *Invisible Ripple Effect*: when AI-generated errors enter professional networks, they might propagate through interconnected systems where they compound and spread until they become indistinguishable from legitimate knowledge. Unlike conventional professional mistakes—which are typically traceable, attributable, and subject to institutional correction—GenAI-driven fabrications often escape detection precisely because of their scale, polish, and apparent authority.[17] These fabrications cross organizational and disciplinary boundaries, embedding themselves in the foundational knowledge that professionals rely upon for subsequent decisions, creating cascading effects that can persist long after the original error.

Existing legal frameworks poorly match the nature of this contamination, which emerges diffusely, sometimes resists attribution, and often bypasses conventional detection mechanisms. The EU AI Act's targeting of institutional deployers generally overlooks individual professionals using consumer tools like ChatGPT.[18] The U.S.'s fragmented, sector-specific, and

---

15. *See* Cade Metz & Karen Weise, *A.I. Is Getting More Powerful, but Its Hallucinations Are Getting Worse*, N.Y. TIMES (May 5, 2025), https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html (reporting that leading AI systems, including those from OpenAI and DeepSeek, are increasingly prone to hallucinations, with developers unable to explain why accuracy is deteriorating despite improvements in reasoning capabilities).

16. *See infra* Part I.

17. Jinjin Gu et al., *AI-Enabled Image Fraud in Scientific Publications*, PATTERNS, July 8, 2022, at 1.

18. *See* Regulation (EU) 2024/1689, of the European Parliament and of the Council of 27 June 2024 on Artificial Intelligence and Amending Certain Union Legislative Acts, 2024 O.J. (L 278) 1 [hereinafter EU AI Act]. Notably, while the EU AI Act addresses GenAI through its regulation of foundation models, it does so in a limited manner that fails to account for the issues explored in this Article. *See id.* art. 50–55 (detailing provider obligations for foundation models without addressing systemic embedding risks). This will be further scrutinized *infra* Part II.

state-oriented regulatory approach addresses some AI misuse (e.g., deepfakes), but has a blind spot for professional GenAI adoption and inadvertent mistakes.[19] Traditional accountability mechanisms—such as professional liability rules, malpractice standards, and sanctions for misconduct—also prove inadequate because they were designed for human errors, not systemic AI-generated fabrications that can propagate across thousands of users simultaneously while maintaining an appearance of authority.[20] Existing regulatory frameworks thus struggle not merely because of enforcement or definitional gaps, but because they fail to capture how error itself behaves once it enters a knowledge system. Understanding this dynamic requires turning from formal regulation to the informational logic that governs how knowledge spreads—and, at times, misleads.

This reflects a broader theoretical deficit across fields of knowledge. While several influential frameworks help illuminate aspects of this risk, they were built for different problems and cannot capture the distinctive mechanisms through which GenAI contamination operates. Consider Cass Sunstein's foundational theory of information cascades, which demonstrates how individuals often abandon their private judgment in favor of social signals.[21] The Invisible Ripple Effect operates as a distinctive form of information cascade: rather than abandoning private judgment based on others' observable choices, professionals defer to AI-generated content that appears authoritative before any social validation occurs. Chesney and Citron's influential research on deepfakes addresses intentional manipulation targeting democratic discourse, whereas this contamination occurs inadvertently within professional systems that already maintain verification protocols.[22] Charles Perrow's normal accident theory captures sudden breakdowns in tightly coupled systems.[23] But here, the failure is

---

19. *See infra* Part II.

20. *See infra* Part II.

21. *See* CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 99 (2017) (describing "cyber-cascades," in which people discount private information and follow perceived majority views); *see also* Sushil Bikhchandani et al., *A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades*, 100 J. POL. ECON. 992 (1992) (developing a formal model showing rational actors imitating predecessors and suppressing their own signals).

22. *See* Bobby Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1778–79 (2019) (analyzing how synthetic audio-visual media can be weaponized to exploit individuals, distort democratic deliberation, and threaten national security).

23. *See* CHARLES PERROW, NORMAL ACCIDENTS: LIVING WITH HIGH-RISK TECHNOLOGIES 61 (2d ed. 1999) (arguing that catastrophic "system accidents" are inevitable in complex, tightly

incremental and often invisible, unfolding not through chain reactions but through unnoticed reuse of flawed information.

This Article thus fills both theoretical and regulatory gaps. Part I examines how the rise of GenAI has shifted AI capabilities from controlled institutional settings to widespread individual use, enabling professionals to integrate powerful but error-prone tools into their work without adequate oversight. It analyzes the technical and cognitive vulnerabilities of GenAI—such as hallucinations, omissions, and the illusion of fluency—and explains how these flaws, once embedded in professional workflows, can propagate through citation, adoption, and repetition. The Part culminates in a new theory of systemic contamination, in which AI-generated errors quietly acquire institutional legitimacy and reshape professional knowledge across domains.

Part II critiques current regulatory frameworks and explains why they are structurally unprepared for this form of risk. It shows how existing approaches—focused on discrete harms, institutional actors, or intentional misuse—fail to address the decentralized and cumulative nature of GenAI-driven contamination. The Part analyzes both comprehensive frameworks, such as the EU AI Act, and fragmented approaches in the U.S., demonstrating how traditional professional liability and accountability mechanisms prove inadequate for systemic contamination that operates below the threshold of conscious wrongdoing.

Part III proposes a five-stage intervention framework designed to disrupt contamination at multiple points before it becomes irreversibly embedded. The framework moves beyond traditional binary approaches of prohibition or permission, offering layered strategies to prevent the production of flawed outputs, intercept them before adoption, manage their integration into professional work, correct institutional decisions shaped by misinformation, and build long-term resilience to evolving AI risks. Rather than prescriptive mandates, the framework offers adaptable principles that can be applied across various regulatory environments and professional contexts.

---

coupled technologies because unexpected interactions outstrip operators' ability to detect and control them).

I. WHEN AI POWER GOES PUBLIC

GenAI has fundamentally altered how AI capabilities reach professional users. Unlike traditional AI systems that operated within controlled institutional environments with specialized oversight, GenAI tools now enable individual professionals to generate sophisticated content through consumer-grade interfaces.[24] Lawyers draft legal briefs, doctors create patient summaries, financial analysts produce market reports, and consultants develop strategic recommendations—all using AI services that require almost no technical expertise and operate primarily without institutional oversight.[25] This shift from institutional to individual adoption has driven rapid integration across professional domains.[26] And the appeal is undeniable: GenAI automates complex research, drafting, and analysis tasks that once consumed hours of specialized work. Yet competitive pressures and "AI FOMO" have accelerated adoption, with many professionals integrating these tools into daily workflows before fully understanding their capabilities and limitations.[27]

This Part examines the risks that emerge when sophisticated AI capabilities operate outside traditional institutional controls. Section A provides a brief overview of the evolution from controlled institutional AI to democratized individual adoption. Section B analyzes GenAI's distinctive vulnerabilities—particularly its propensity for hallucinations, omissions, and sophisticated-appearing errors that can deceive expert review. Section C introduces the Invisible Ripple Effect: the process by which individual AI-generated errors become embedded in professional knowledge systems and propagate across institutional boundaries, creating

---

24. For examples of such tools, see, for example, OpenAI, CHATGPT, https://chatgpt.com [https://perma.cc/9TFU-ETWQ]; and Anthropic, CLAUDE, https://claude.ai/login [https://perma.cc/BM2K-RETW].

25. *See* Kevin Roose & Cade Metz, *How to Become an Expert on A.I.*, N.Y. TIMES (Apr. 4, 2023), https://www.nytimes.com/article/ai-artificial-intelligence-chatbot.html; Cade Metz, *Instant Videos Could Represent the Next Leap in A.I. Technology*, N.Y. TIMES (Apr. 4, 2023), https://www.nytimes.com/2023/04/04/technology/runway-ai-videos.html.

26. ChatGPT, for example, became one of the fastest-adopted technologies in history, quickly reaching millions of users. *See* Krystal Hu, *ChatGPT Sets Record for Fastest Growing User Base: Analyst Note*, REUTERS (Feb. 2, 2023), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01 [https://perma.cc/GVX8-2NZ4]; *see also* Hayden Field, *China's DeepSeek AI Dethrones ChatGPT on App Store: Here's What You Should Know*, CNBC (Jan. 27, 2025), https://www.cnbc.com/2025/01/27/chinas-deepseek-ai-tops-chatgpt-app-store-what-you-should-know.html [https://perma.cc/QQH8-98YK].

27. *See* Covert, *supra* note 10.

systemic contamination that existing oversight mechanisms cannot detect or prevent.

A. The Democratization of AI and Its Widespread Adoption

For much of its history, AI remained a specialized technology, constrained within proprietary systems and controlled by a select few actors.[28] Governments have leveraged AI for military operations, national security, and public administration, utilizing it to manage tasks ranging from healthcare resource allocation to welfare programs.[29] These applications were typically large-scale, purpose-driven, and designed to function within tightly regulated institutional frameworks.[30] The public had limited access to or influence over these systems, which operated through complex, opaque algorithms that were understood primarily by those with the requisite technical expertise and resources.[31] Some professionals, such

---

28. *See, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 12, 98–99 (2015) (examining proprietary technological systems and corporate control); KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE 181–209 (2021) (analyzing private and state development of AI technologies).

29. *See* AI NOW INSTITUTE, AI NOW 2019 REPORT 40, 42, 53 (2019), https://ainowinstitute.org/wp-content/uploads/2023/04/AI_Now_2019_Report.pdf [https://perma.cc/7RQ2-DBPD]; Kate Crawford & Jason Schultz, *AI Systems as State Actors*, 119 COLUM. L. REV. 1941, 1946–47 (2019) (discussing AI use by state governments).

30. *See, e.g.*, URS GASSER & VIKTOR MAYER-SCHÖNBERGER, GUARDRAILS: GUIDING HUMAN DECISIONS IN THE AGE OF AI 20–21, 69 (2024) (discussing historical regulatory oversight of AI).

31. AI systems often rely on "black box" models, where the internal reasoning behind decisions remains opaque, even to their developers. This lack of transparency makes it difficult to scrutinize how AI reaches its conclusions, raising concerns about bias, accountability, and due process—especially in high-stakes applications like criminal justice, hiring, and lending. The problem is exacerbated when AI models are proprietary, meaning their algorithms and training data are controlled by private entities and shielded from public or regulatory scrutiny. This dual challenge—opacity in both technical design and corporate ownership—limits meaningful oversight, making it difficult to assess errors, challenge unfair outcomes, or ensure compliance with legal standards. For more on the "black box" of AI, see PASQUALE, *supra* note 28, at 1–18 (arguing that the opacity of algorithmic systems creates significant challenges for accountability and oversight); Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTEL. 206, 206–07 (2019) (contending that high-stakes decisions should use interpretable models rather than post-hoc explanations of black box systems); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 656–72, 682–92 (2017) (examining how to ensure fairness and accountability in black box algorithmic systems through technical and legal mechanisms); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the*

as lawyers and doctors, utilize AI through carefully tailored applications designed for specific tasks; however, this use is somewhat limited in scope.[32]

Private corporations played a central role in shaping AI. Companies created AI-driven tools tailored to specific, well-defined tasks, such as assessing recidivism in criminal justice,[33] optimizing supply chain logistics,[34] or automating contract review in the legal industry.[35] These systems remained proprietary, restricted to particular sectors or organizations, and largely inaccessible to those outside the entities that developed and controlled them.[36]

AI occasionally reached end-users through commercial products, but these applications were typically narrow in scope and task-specific. Tools like Grammarly enhanced written communication,[37] while AI-powered services facilitated customer support, photo editing, and language

---

Criminal Justice System, 70 STAN. L. REV. 1343, 1356–71 (2018) (examining how trade secret protections for proprietary algorithms conflict with due process rights in criminal justice).

32. *See, e.g.*, W. Nicholson Price II, *Medical AI and Contextual Bias*, 33 HARV. J.L. & TECH. 66, 70–75 (2019) (analyzing healthcare AI applications); DANIEL SUSSKIND, A WORLD WITHOUT WORK: TECHNOLOGY, AUTOMATION, AND HOW WE SHOULD RESPOND 85–90 (2020) (analyzing AI's impact on professional work).

33. One notable example is COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), a risk assessment tool used to evaluate the likelihood of recidivism and inform decisions on bail, parole, and probation. For more on the use of criminal AI tools within the justice system, see generally Dan Hunter, Mirko Bagaric & Nigel Stobbs, *A Framework for the Efficient and Ethical Use of Artificial Intelligence in the Criminal Justice System*, 47 FLA. ST. U. L. REV. 749 (2020) (proposing a framework for systematically implementing AI into the criminal justice system in order to ensure that the system operates in an efficient and ethical manner).

34. *See* Alma Kelly, *Impact of Artificial Intelligence on Supply Chain Optimization,* J. TECH. & SYS., July 2024, at 15, 16–24 (analyzing how AI enhances forecasting, inventory management, logistics, and risk management in supply chains).

35. *See, e.g.*, Kathryn D. Betts & Kyle R. Jaep, *The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life Into a Decades-Old Promise,* 15 DUKE L. & TECH. REV. 216 (2017) (discussing the use of machine learning in automating contract drafting and review processes); Amy B. Cyphert, *A Human Being Wrote This Law Review Article: GPT-3 and the Practice of Law*, 55 U.C. DAVIS L. REV. 401, 417 (2021).

36. *See S*onia K. Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183, 1207–09 (2019) (arguing that, despite the potential role of copyright protection, software systems remained proprietary, with secrecy prevailing even as the open-source movement sought to challenge restrictions on access and control).

37. *See AI at Grammarly: Transforming How the World Communicates Through AI*, GRAMMARLY, https://www.grammarly.com/ai [https://perma.cc/J5H6-YD6Z] (discussing Grammarly's long-standing use of AI for writing assistance before the advent of GenAI, including grammar correction, style suggestions, and clarity improvements).

translation, to name but a few examples.[38] AI also began to operate invisibly in the background, subtly shaping consumer experiences through algorithmic recommendations.[39] Platforms like Netflix leveraged AI to predict user preferences and suggest content, while Amazon used it to personalize product recommendations based on purchasing behavior.[40] While consumers benefited from these systems, they had no control over the underlying algorithms that influenced their decisions.[41] This early landscape of AI highlights its immense power but also its constraints. AI's functionality was narrowly defined, governed by its intended purpose, and controlled by those with the resources to develop and deploy it. Access was limited not only by cost but also by the technical expertise required to operate such systems.[42]

The advent of GenAI marks a fundamental shift. Tools like ChatGPT now democratize sophisticated AI capabilities, making them accessible to almost anyone with internet access.[43] These systems transcend industry-specific applications, functioning instead as versatile, general-purpose technologies with unprecedented adaptability.[44] They now automate complex creative and analytical work across domains—from drafting and

---

38. *See, e.g.*, Keith Kirkpatrick, *AI in Contact Centers*, COMM. ACM, Aug. 2017, at 18 (showing how AI enhances customer service).

39. *See* Liye Ma & Baohong Sun, *Machine Learning and AI in Marketing – Connecting Computing Power to Human Insights*, 37 INT'L J. RSCH. MKTG. 481 (2020) (explaining how AI-driven algorithms increasingly operate in the background, influencing consumer experiences through automated personalization, recommendation systems, and real-time decision-making).

40. *Id.* at 482 (discussing how machine learning powers recommender systems at e-commerce platforms like Amazon and content platforms like Netflix, enhancing personalization and user engagement).

41. *Id.*

42. *See* Matteo Cristofaro & Pier Luigi Giardino, *Surfing the AI Waves: The Historical Evolution of Artificial Intelligence in Management and Organizational Studies and Practices*, J. MGMT. HIST., Mar. 19, 2025, at 6–8 (describing early symbolic AI as rule-based systems confined to structured problem domains and unable to adapt to ambiguity or learn from new data; noting the high computational costs of such systems, their reliance on specialized hardware such as LISP machines, and the concentration of development within government and academic institutions during the Cold War era).

43. By "democratization," this Article refers to the widespread accessibility of GenAI, which allows individuals and small-scale entities to use advanced AI technologies without requiring specialized expertise or institutional support. This usage differs from the traditional political science definition, which typically concerns the expansion or evolution of democratic norms, institutions, and practices. *Cf.* Jelena Cupać et al., *Democratization in the Age of Artificial Intelligence: Introduction to the Special Issue*, 31 DEMOCRATIZATION 899, 904–05 (2024) (defining it as "the way democratic norms, institutions and practices evolve and are disseminated or retracted both within and across national and cultural boundaries").

44. *Id.* at 900–01, 904.

design to marketing and data analysis—tasks once considered exclusively human.[45] Unlike traditional AI systems, GenAI does not require extensive technical expertise or specialized knowledge to operate.[46] Its intuitive interfaces and accessibility have led to its democratization, allowing individuals and institutions alike to integrate advanced AI into their workflows with unprecedented ease.[47] Many of these platforms offer a free version,[48] and some are even open-source, meaning that anyone can run these models locally.[49] As such, this technology has the potential to dismantle barriers, reduce inequities, and expand access to knowledge and capabilities.[50]

The rapid adoption is already evident. ChatGPT, for example, became the fastest-adopted technology in history, quickly reaching millions of users.[51] Soon after its launch, DeepSeek surpassed OpenAI as the most-downloaded free app in the U.S. on Apple's App Store.[52] Usage continues

---

45. *See, e.g.*, Brian Potter, *Could ChatGPT Become an Architect?*, CONSTR. PHYSICS (Mar. 29, 2023), https://www.construction-physics.com/p/could-chatgpt-become-an-architect (arguing that ChatGPT shows strong text-based knowledge of architecture and can pass parts of the licensing exam)*;* Steve Lohr, *A.I. Is Coming for Lawyers, Again*, N.Y. TIMES (Apr. 10, 2023), https://www.nytimes.com/2023/04/10/technology/ai-is-coming-for-lawyers-again.html (discussing how advancements in AI, including ChatGPT-style tools, are reshaping legal practice by automating tasks such as legal research and document review, while also raising concerns about job displacement and accuracy).

46. *See* Jan Bieniek et al., Generative AI in Multimodal User Interfaces: Trends, Challenges, and Cross-Platform Adaptability (Nov. 15, 2024) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/2411.10234 [https://perma.cc/C6R2-Z286] (discussing how GenAI enhances intuitive and accessible user interfaces through multimodal interaction, cross-platform adaptability, and dynamic personalization).

47. *Id.*

48. *See* Cade Metz, *OpenAI Unveils A.I. Agent That Can Use Websites on Its Own*, N.Y. TIMES (Jan. 23, 2025), https://www.nytimes.com/2025/01/23/technology/openai-operator-launch.html.

49. *Id.*

50. *See* JAN VAN DIJK, THE DIGITAL DIVIDE 1–3 (2020) (demonstrating how bridging digital inequalities creates cascading positive effects across education, employment, and social participation while warning that failing to address digital exclusion risks amplifying existing social stratification); MARK WARSCHAUER, TECHNOLOGY AND SOCIAL INCLUSION: RETHINKING THE DIGITAL DIVIDE 201–02 (2003) (arguing that technology's democratizing potential emerges when access extends beyond physical availability to include the skills, knowledge, and social support needed for meaningful use).

51. *See* Krystal Hu, *ChatGPT Sets Record for Fastest Growing User Base—Analyst Note*, REUTERS (Feb. 2, 2023), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01.

52. *See* Hayden Field, *China's DeepSeek AI Dethrones ChatGPT on App Store: Here's What You Should Know*, CNBC (Jan. 27, 2025), https://www.cnbc.com/2025/01/27/chinas-

to grow as individuals and institutions integrate GenAI into their workflows,[53] with new competitors and applications entering the market daily.[54] Although it is difficult to assess the scope of its actual use fully, its adoption is likely to continue expanding.[55]

Such adoption is already unfolding in professional settings, as GenAI offers convenient applications across various industries. Professionals already use GenAI to draft briefs, design structures, and compose communications.[56] Lawyers are already utilizing GenAI to create contracts,

---

deepseek-ai-tops-chatgpt-app-store-what-you-should-know.html [https://perma.cc/6BWY-SS5U] (analyzing DeepSeek's rapid rise).

53. *See, e.g.*, John Gapper, *Legal AI Is Reaching Deep into the Workplace*, FIN. TIMES (Jan. 20, 2025), https://www.ft.com/content/bdc2250f-fbd9-4c4a-98cf-42e389e5b6c0 (discussing how legal AI tools, such as Genie AI and Luminance, automate contract drafting and transform corporate legal practices, enabling companies to handle legal tasks traditionally reserved for law firms); *see* Mark Minevich, *The Dawn of AI Disruption: How 2024 Marks a New Era in Innovation*, FORBES (Dec. 14, 2023), https://www.forbes.com/sites/markminevich/2023/12/14/the-dawn-of-ai-disruption-how-2024-marks-a-new-era-in-innovation/?sh=693f9374141c (arguing that GenAI has transformed sectors such as demand forecasting, supply chain management, product innovation, and healthcare by optimizing workflows, enhancing decision-making, and improving patient care through advanced data analysis and pattern recognition).

54. *See* Sophia Velastegui, *AI's Biggest Moments of 2024: What We Learned This Year*, FORBES (Dec. 19, 2024), https://www.forbes.com/sites/committeeof200/2024/12/12/ais-biggest-moments-of-2024-what-we-learned-this-year (discussing the rapid adoption of GenAI tools like ChatGPT and Claude).

55. To be fair, fully assessing how people use these tools is difficult, if not impossible. Users experiment with ChatGPT for creative writing, turn to platforms like MidJourney for imaginative illustrations, or use RunwayAI to produce short films. Many of these creations never extend beyond personal use. Yet GenAI has fundamentally reshaped creative possibilities. Just as photography democratized image capture or word processors revolutionized writing, GenAI has placed extraordinary creative and analytical power into the hands of anyone with internet access, extending computational sophistication once reserved for governments and corporations to ordinary users. For more on its potential disruptive impact, see MIT Technology Review Insights, *Generative AI: Differentiating Disruptors from the Disrupted*, MIT TECH. REV. (Feb. 29, 2024), https://www.technologyreview.com/2024/02/29/1089152/generative-ai-differentiating-disruptors-from-the-disrupted [https://perma.cc/C6YW-HNM8] (analyzing how enterprises are integrating GenAI, the challenges of adoption, and the potential for industry disruption).

56. Some of these tools are already integrated into widely used email platforms, making them even more accessible to everyday users. For example, Google's Gemini AI in Gmail can summarize long email threads, providing concise overviews of extensive communications. Similarly, Microsoft's Copilot in Outlook helps users draft and summarize emails while highlighting key points and outstanding tasks. *See* David Nield, *Google Gemini Can Summarize Your Emails in Gmail. Should You Use It?*, WIRED (Dec. 9, 2024), https://www.wired.com/story/google-gemini-summarize-emails-in-gmail [https://perma.cc/BY47-QNCY].

draft legal briefs, and conduct preliminary research on case law.[57] Investigators use it to analyze evidence more efficiently, cross-referencing sources or identifying patterns in data that might otherwise go unnoticed.[58] Architects harness GenAI to design preliminary blueprints or generate innovative concepts for client presentations.[59] Teachers use it to draft lesson plans, create tailored exercises, and provide personalized feedback on assignments.[60] Marketers leverage it to develop campaign strategies, craft ad copy, and analyze consumer trends.[61] This list goes on and on.[62]

---

57. *See* Yonathan Arbel & David A. Hoffman, *Generative Interpretation*, 99 N.Y.U. L. REV. 451 (2024) (arguing that large language models (LLMs) can transform contract interpretation by enabling courts to cheaply and accurately estimate contractual meaning, assess ambiguity, and fill gaps, thereby offering a middle path between textualist efficiency and contextualist fairness in adjudication); *Can You Use AI for Legal Research?*, BLOOMBERG L. (Sept. 19, 2024), https://pro.bloomberglaw.com/insights/technology/can-you-use-ai-for-legal-research [https://perma.cc/55M8-DQ9E] (discussing how GenAI tools streamline legal research); Lyle Moran, *73% of Lawyers Plan to Use Generative AI, Report Finds*, LEGAL DIVE (Nov. 20, 2023), https://www.legaldive.com/news/generative-ai-legal-use-cases-wolters-kluwer-report/700342/ [https://perma.cc/U57X-7QGY] (noting that lawyers are increasingly adopting GenAI to perform tasks such as reviewing legal documents, drafting contracts, and analyzing data, with 73% planning to use these tools within a year); Lohr, *supra* note 45. Some AI tools, such as DoNotPay, even claim to provide direct-to-consumer legal assistance, raising concerns that GenAI could replace certain legal functions altogether. However, DoNotPay has faced significant legal and ethical challenges, including allegations of unauthorized practice of law (UPL) and inaccurate legal filings. *See* Bobby Allyn, *A Robot Was Scheduled to Argue in Court, Then Came the Jail Threats*, NPR (Jan. 25, 2023), https://www.npr.org/2023/01/25/1151435033/a-robot-was-scheduled-toargue-in-court-then-came-the-jail-threats [https://perma.cc/V64C-E4TX]; *see also* Joe Patrice, *You Can Replace Supreme Court Lawyers with AI Now. Honestly, That Tracks.*, ABOVE L. (July 8, 2025), https://abovethelaw.com/2025/07/you-can-replace-supreme-court-lawyers-with-ai-now-honestly-that-tracks [https://perma.cc/ZMN5-5AD3] (reporting on Adam Unikowsky's experiment simulating a Supreme Court oral argument using LLMs, suggesting that GenAI tools may soon play a meaningful role even in high-level advocacy).

58. *See* Richard Vanderford, *Can a Computer Learn to Speak Trader?*, WALL ST. J. (Jan. 17, 2025), https://www.wsj.com/articles/can-a-computer-learn-to-speak-trader-7695fbc9 (discussing how compliance software firms are using AI to decode trader jargon and fight financial crime as regulatory scrutiny increases).

59. *See* Naomi Ackerman, *Zaha Hadid Architects Builds 'Winner Proposals' with AI*, SUNDAY TIMES (Jan. 2, 2025), https://www.thetimes.co.uk/article/zaha-hadid-architects-builds-winner-proposals-with-ai-enterprise-network-qs7m7txwz [https://perma.cc/7VB7-AKSU] (describing how architects employ GenAI tools to accelerate design processes).

60. *See* Tommaso Calò & Christopher J. MacLellan, *Towards Educator-Driven Tutor Authoring: Generative AI Approaches for Creating Intelligent Tutor Interfaces*, *in* L@S '24: PROCEEDINGS OF THE ELEVENTH ACM CONFERENCE ON LEARNING @ SCALE 305 (2024) (discussing how GenAI enhances intelligent tutoring systems by enabling educators to create adaptive and personalized learning environments without requiring advanced technical skills);

In many cases, the perceived efficiency of these tools drives their rapid and uncritical adoption, often underpinned by competitive pressures and shifting professional norms. For example, law firms face growing client demands for cost-effective solutions, incentivizing the use of GenAI to streamline workflows.[63] These trends reflect a broader phenomenon often described as AI FOMO, where professionals adopt GenAI because they fear being left behind in a rapidly evolving technological landscape.[64]

GenAI's seamless integration into daily life means it is already shaping decisions in ways that remain invisible to us. It may have influenced the legal advice you received, the contract governing your lease, or the investment recommendations guiding your financial future. It generates medical information doctors consult, safety guidelines for commercial aircraft, and policy memos circulating through government agencies. It shapes the presentations students prepare for school and the hiring algorithms that determine career prospects.[65] Whether acknowledged or operating silently in the background, GenAI is increasingly embedded in both private and public decision-making, creating new and potentially unforeseen categories of risk.

Sara Randazzo, *That Essay Got a B+. An AI Bot Graded It.*, WALL ST. J. (July 2, 2024), https://www.wsj.com/tech/ai/ai-tools-grading-teachers-students-396c2bfc.

61. *See* Marie Hattar, *Generative AI: Marketing's Friend or Foe?*, FORBES (Sept. 12, 2023), https://www.forbes.com/sites/forbescommunicationscouncil/2023/09/12/generative-ai-marketings-friend-or-foe.

62. *See, e.g.*, James Prather et al., *Beyond the Hype: A Comprehensive Review of Current Trends in Generative AI Research, Teaching Practices, and Tools*, *in* PROCEEDING OF THE 2024 WORKING GROUP REPORTS ON INNOVATION AND TECHNOLOGY IN COMPUTER SCIENCE EDUCATION (July 8–10, 2024) (reviewing emerging trends in GenAI use for computing education, including its role in programming instruction, personalized student feedback, and classroom-scale teaching tools); Zohar Elyoseph et al., *An Ethical Perspective on the Democratization of Mental Health with Generative AI*, 11 JMIR MENTAL HEALTH e58011 (2024) (discussing the role of GenAI in making mental health education more accessible by democratizing knowledge, improving engagement, and offering personalized, adaptive learning experiences).

63. *See* Joseph J. Avery et al., *ChatGPT, Esq.: Recasting Unauthorized Practice of Law in the Era of Generative AI*, 26 YALE J.L. & TECH. 64, 86 (2023) (quoting Mark Chandler, former Chief Legal Officer of Cisco Systems, Inc.); Bill Henderson, *"The State of Technology in the Law," Mark Chandler Speech from January 2007 (035)*, LEGAL EVOLUTION (Nov. 11, 2017), https://www.legalevolution.org/2017/11/mark-chandler-speech-january-2007-035/ [https://perma.cc/TYD5-AL2Q].

64. *See* Covert, *supra* note 10.

65. Marc Zao-Sanders, *How People Are Really Using Gen AI in 2025*, HARV. BUS. REV. (Apr. 9, 2025), https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025.

B. The Hidden Vulnerabilities of GenAI

Generally speaking, GenAI inherits all the familiar problems of traditional AI systems, such as algorithmic bias, operational opacity, and accountability gaps.[66] These aren't new vulnerabilities; they are existing flaws that are now being deployed through consumer devices, which millions of professionals use daily without institutional oversight. Bias remains structural: training data that reflects societal inequities amplifies the discriminatory outcomes we've documented for years.[67] We already know that automated systems discriminate against marginalized communities,[68] that facial recognition fails disproportionately for women and people with darker skin,[69] and that hiring algorithms perpetuate historical biases.[70] The

---

66. *See, e.g.*, Sandra Wachter, *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond*, 26 YALE J.L. & TECH. 671, 674–75 (2024) (recognizing some similar risks that GenAI entails); Mi Zhou et al., Bias in Generative AI 1, 3–4, 10 (Mar. 5, 2024) (unpublished manuscript) (on file with arXiv), https://doi.org/10.48550/arXiv.2403.02726 [https://perma.cc/5FQZ-D7X5] (finding that gender and racial biases in GenAI are more pronounced than real-world benchmarks and traditional datasets, with implications for systemic bias amplification); Leonardo Nicoletti & Dina Bass, *Humans Are Biased. Generative AI Is Even Worse*, BLOOMBERG (June 9, 2023), https://www.bloomberg.com/graphics/2023-generative-ai-bias (analyzing how Stable Diffusion amplifies racial and gender stereotypes beyond real-world disparities, disproportionately associating high-paying jobs with lighter-skinned men and low-paying or criminal roles with darker-skinned individuals).

67. *See* Cyphert, *supra* note 35, at 413–16 (discussing bias in GPTs). *See generally* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016) (arguing that algorithms often perpetuate inequality and harm marginalized communities by operating opaquely, scaling massively, and producing unfair or discriminatory outcomes).

68. *See, e.g.*, Latanya Sweeney, *Discrimination in Online Ad Delivery*, 56 COMMC'NS ACM 44, 46–47 (2013) (finding, *inter alia*, that search queries for African American names could render services that are linked to arrest records, more than for white names); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018) (arguing that automated decision-making systems disproportionately harm poor and marginalized communities by profiling, surveilling, and reinforcing systemic inequalities).

69. *See* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 77, 81–87 (2018) (discussing racial bias in AI facial recognition); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019) (analyzing how algorithmic tools in criminal justice reinforce and exacerbate societal inequities, while emphasizing the need to address structural biases in data to reduce these risks); Eldar Haber, *Racial Recognition*, 43 CARDOZO L. REV. 71, 89–95, 102–30 (2021) (offering how to regulate bias within facial recognition systems).

70. *See, e.g.*, Sandra Wachter et al., *Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law*, 123 W. VA. L. REV. 735, 737 (2021) ("Prior hiring decisions inform future hiring."); Marianne Bertrand & Sendhil Mullainathan, *Are*

difference with GenAI isn't the nature of these problems—it's their reach. Massive, uncurated datasets and widespread individual adoption ensure these longstanding risks now operate at unprecedented scale.[71]

The black-box problem remains as intractable as ever.[72] Machine learning systems, particularly neural networks, generate outputs through processes that are inherently complex and difficult to interpret.[73] GenAI offers no solution to this opacity—if anything, it deepens the problem.[74] Privacy vulnerabilities persist wherever sensitive data enters training or GenAI outputs inadvertently expose private information.[75] Intellectual property concerns reach new extremes: while earlier AI systems relied on curated, domain-specific datasets, GenAI models train on massive collections of copyrighted works—often including books downloaded from

---

Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, 94 AM. ECON. REV. 991, 997–99 (2004) (examining a study on the disparity in callback rates between job candidates with white-sounding names and those with Black-sounding names). *See generally* Ifeoma Ajunwa, *Age Discrimination by Platforms*, 40 BERKELEY J. EMP. & LAB. L. 1 (2019) (examining how workplace platforms contribute to age discrimination in employment).

71. *See generally* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016) (examining data mining practices through lens of American antidiscrimination law); Noam Kolt, *Algorithmic Black Swans*, 101 WASH. U. L. REV. 1177, 1193–96 (2024) (examining bias within GenAI).

72. *See supra* note 31.

73. *See supra* note 31.

74. However, it is worth questioning whether this "blacker box" critique oversimplifies the issue. While the scale and scope of GenAI models introduce new challenges, their architecture also enables broader scrutiny through third-party audits and interpretability research—mechanisms that were less feasible for many proprietary traditional AI systems. For more on this, see Rishi Bommasani et al., *The 2025 Foundation Model Transparency Index* 5–6, 30 (Center for Research on Foundation Models, 2024) (finding that open foundation model developers—those releasing model weights widely—score significantly higher on transparency indicators than closed developers, particularly regarding upstream resources like data, labor, and compute used to build models). *See also* Yifan Luo et al., InverseScope: Scalable Activation Inversion for Interpreting Large Language Models (June 9, 2025) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/2506.07406 [https://perma.cc/QDT3-HH5F] (describing a method that enables scalable interpretability of large model activations, allowing for systemic internal analysis that was not feasible in earlier opaque systems).

75. GenAI's data-intensive nature raises unresolved questions about data collection practices, consent, and usage. While frameworks like the General Data Protection Regulation (GDPR) attempt to regulate data privacy, the global and decentralized nature of GenAI development often places it beyond the reach of traditional regulatory mechanisms, complicating efforts to safeguard individual rights. *See* Council Regulation 2016/679, 2016 O.J. (L 119) 1 (EU).

piracy websites, typically without permission or compensation, creating IP conflicts at unprecedented scale.[76]

Whether GenAI reproduces or often amplifies these inherited flaws remains an open question.[77] But even if some of these risks are mitigated or remain comparable to those posed by earlier AI systems, GenAI introduces entirely new challenges that amplify its potential for harm.[78] One of the most significant risks is its propensity to produce outputs that appear authoritative but are fundamentally flawed. These errors include not only *hallucinations*—fabricated information such as nonexistent legal precedents, fictional medical advice, or inaccurate historical claims—but also systemic omissions, outdated information, and subtle factual distortions.[79] Let's break this down.

---

76. *See generally* Katherine Lee et al., *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, J. COPYRIGHT SOC'Y 48 (2024) (applying supply chain framing to U.S. copyright law). For litigation probing whether large-scale use of copyrighted works—including millions of books obtained from piracy repositories—to train GenAI constitutes fair use or infringement, see, for example, *Andersen v. Stability AI Ltd.*, No. 3:23-cv-00201-WHO (N.D. Cal. filed Jan. 13, 2023) (order denying motions to dismiss in part, Aug. 12, 2024; trial set for Sept. 8, 2026); *Getty Images (US), Inc. v. Stability AI, Inc.*, No. 1:23-cv-00135 (D. Del. filed Feb. 3, 2023) (order largely denying motion to dismiss, Feb. 6, 2024); *N.Y. Times Co. v. Microsoft Corp.* & *OpenAI, Inc.*, No. 1:23-cv-11195-SHS (S.D.N.Y. filed Dec. 27, 2023) (order denying motions to dismiss in substantial part, Apr. 4, 2025; case in discovery). *Cf.* Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc., No. 1:20-cv-613 (D. Del. Feb. 11, 2025) (granting partial summary judgment to plaintiff and holding that use of copyrighted Westlaw headnotes to train a non-generative AI tool was not fair use as a matter of law).

77. *See, e.g.*, Celeste Kidd & Abeba Birhane, *How AI Can Distort Human Beliefs*, 380 SCI. 1222 (2023) (discussing how GenAI models transmit biases and false information, subtly shaping user beliefs and reinforcing societal misconceptions).

78. Notably, the risks of using GenAI tools extend beyond AI-generated errors (and other AI-related risks). One significant concern arises when employees inadvertently disclose trade secrets, proprietary company information, or other sensitive data to GenAI models, potentially exposing confidential material to unintended parties. *See* Siladitya Ray, *Samsung Bans ChatGPT Among Employees After Sensitive Code Leak*, FORBES (May 2, 2023), https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak. For other risks, see Laura Weidinger et al., *Taxonomy of Risks Posed by Language Models*, FACCT '22: PROC. 2022 ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 214, 216–22 (2022) (arguing that language models pose risks including discrimination and hate speech, information hazards, misinformation harms, malicious uses, human-computer interaction harms, and environmental and socioeconomic harms); Wachter, *supra* note 66, at 674–75 (arguing that GenAI "also presents new questions relating to inaccurate and offensive content [and] misinformation").

79. *See generally* Ziwei Ji et al., *Survey of Hallucination in Natural Language Generation*, 55 ACM COMPUTING SURVS. 12 (2023) (providing overview of hallucination challenges in natural language generation).

Hallucinations are not bugs—they are features. As noted, GenAI's probabilistic architecture makes sophisticated fabrication inevitable, prioritizing plausibility and coherence over factual accuracy.[80] This represents a fundamental design choice. GenAI functions as a pattern-prediction system, not a fact-verification tool, and this trade-off is what makes it powerful. The exact mechanisms that generate fluid, contextually relevant responses, also prevent a reliable distinction between fact and fiction. The system doesn't assess truth; it predicts what seems accurate based on statistical probabilities, making hallucinations an unavoidable consequence of operation.[81]

GenAI systems exhibit systemic error patterns that create particular risks for professionals. These systems respond to informational gaps by generating fabricated content with inappropriate confidence rather than acknowledging evidentiary limitations.[82] When authoritative sources don't exist for an obscure legal case,[83] GenAI will invent something like "U.S. v. Fletcher" as a significant criminal ruling rather than acknowledge uncertainty.[84] The system may rely on single, unreliable sources—potentially self-authored content—to construct responses about obscure individuals, producing incomplete, biased, or entirely fabricated portrayals.[85] When limited or unreliable information is available online, GenAI provides definitive responses despite insufficient underlying data.[86]

User error compounds these risks. Imprecise prompts generate contextually inappropriate responses that maintain an authoritative veneer. GenAI misinterprets vague, overly broad, or context-lacking queries while still generating fluent and convincing outputs. Because these tools are designed to sound authoritative regardless of query quality, users may rely on responses that are technically accurate but professionally useless, or

---

80. *Id.*

81. *See* Karen Weise & Cade Metz, *When A.I. Chatbots Hallucinate*, N.Y. TIMES (May 9, 2023), https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html.

82. *How to Spot and Avoid GenAI Hallucinations*, UNIV. PITTSBURGH: PANTHERBYTES BLOG (Nov. 12, 2025), https://www.digital.pitt.edu/news/pantherbytes-blog/how-spot-and-avoid-genai-hallucinations [https://perma.cc/TAP3-9N2M].

83. GenAI hallucinates partially because it is trained on vast and diverse datasets—including internet content, books, and proprietary sources—that contain both accurate and inaccurate information. *See id.*

84. *See id.*

85. *See id.*

86. *See generally* Oscar Y. Shen et al., *How Does ChatGPT Use Source Information Compared with Google? A Text Network Analysis of Online Health Information*, 482 CLINICAL ORTHOPEDICS & RELATED RSCH. 578 (2024).

dangerous. Common misunderstandings about GenAI's capabilities lead users to assume precision and reliability that the system cannot guarantee.[87]

Critical omissions create one of the most insidious categories of GenAI error.[88] Unlike hallucinations, which introduce fabricated content, omissions involve the exclusion of essential information while preserving an appearance of completeness. These failures often occur even when relevant data is available. They may stem from technical constraints—such as limited context windows that truncate inputs—as well as from shallow pattern recognition that fails to prioritize or accurately interpret key facts.[89] A GenAI system conducting legal research might exclude a controlling precedent because it was unable to associate it with the user's prompt or deprioritized it in favor of superficially similar cases. In healthcare applications, the system may retrieve common conditions while omitting rare but critical diagnoses due to skewed training data or an inability to accurately weigh clinical nuances.[90] These omissions are especially dangerous because the output appears polished and thorough, concealing its blind spots and giving users a false sense of reliability. Without any awareness of what is missing, GenAI systems generate outputs that mask their incompleteness, making it extremely difficult for users—even expert ones—to detect what has been left out.[91]

These vulnerabilities can also be weaponized. Prompt injection techniques allow malicious actors to embed instructions that circumvent

---

87. *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, MIT MGMT. STS TEACHING & LEARNING TECHS., https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias [https://perma.cc/5SHV-LLMG] (discussing how generative AI models often produce plausible but incorrect outputs, leading users to overestimate their reliability).

88. *See* Nelson F. Liu et al., *Lost in the Middle: How Language Models Use Long Contexts*, 12 TRANSACTIONS ASS'N FOR COMPUTATIONAL LINGUISTICS 157, 158 (2024), https://doi.org/10.1162/tacl_a_00638 [https://perma.cc/8LKL-3883] (finding that LLM performance degrades when relevant information appears in the middle of long inputs, even in extended-context models); Harvey Yiyun Fu et al., AbsenceBench: Language Models Can't Tell What's Missing (June 13, 2025) (unpublished manuscript) (on file with arXiv), arXiv:2506.11440v1 [cs.CL], https://doi.org/10.48550/arXiv.2506.11440 [https://perma.cc/5YDR-M3LG] (demonstrating that state-of-the-art LLMs struggle to detect missing information, due to structural limitations in transformer attention).

89. *See* Liu et al., *supra* note 88, at 157, 161, 164; Fu et al., *supra* note 88, at 9.

90. *See* Bat-Zion Hose et al., *Development of a Preliminary Patient Safety Classification System for Generative AI*, 34 BMJ QUALITY & SAFETY 130 (2025), https://doi.org/10.1136/bmjqs-2024-017918 [https://perma.cc/Q9YT-DDZ4] (analyzing patient safety risks in GenAI and identifying omission as the most frequent error, with significant implications for reliability in high-stakes fields such as medicine).

91. *Id.*

built-in safeguards, manipulating GenAI outputs to serve adversarial purposes.[92] In high-stakes professional settings, attackers may introduce fabricated evidence, present misleading legal arguments, or distort financial analyses. In litigation, for example, an adversary could manipulate AI research tools to plant unreliable case law—or even strategically embed hallucinated citations or misleading language directly within a filed document—anticipating that opposing counsel, or even a judge, might later use an LLM to analyze that document and unknowingly reproduce or validate the manipulated content.[93] These attacks exploit the same pattern-recognition mechanisms that make GenAI powerful, leveraging its fluency and coherence to camouflage falsehoods and trigger a recursive cycle of misinformation.[94]

While each of these GenAI errors can certainly cause significant individual harm, these are not merely individual attacks confined to single cases. GenAI generates cascading risks that spread across entire professional domains. When GenAI fabricates legal citations and enters them into court filings, these false precedents can influence subsequent legal research and judicial decisions.[95] AI-generated medical findings might

---

92. Prompt injection is a type of adversarial attack that exploits the way GenAI models interpret and generate responses. Adversarial attacks involve intentionally crafted inputs designed to manipulate an AI system's behavior, often circumventing built-in safeguards or prompting unintended outputs. In the case of prompt injection, attackers embed deceptive instructions within a prompt, either subtly or overtly, to trick the model into generating biased, misleading, or harmful content. This technique can be used to override system restrictions, extract confidential information, or induce the model to produce responses aligned with the manipulator's goals. *See NIST Identifies Types of Cyberattacks That Manipulate Behavior of AI Systems*, NAT'L INST. STANDARDS & TECH. (Jan. 4, 2024), https://www.nist.gov/news-events/news/2024/01/nist-identifies-types-cyberattacks-manipulate-behavior-ai-systems [https://perma.cc/EJ9L-CGD7] (discussing GenAI vulnerabilities related to adversarial attacks).

93. *See* Kai Greshake et al., Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection (May 5, 2023) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2302.12173 [https://perma.cc/9GRD-TQR6] (discussing "indirect prompt injection" and empirically showing that hidden instructions placed in ordinary documents—e-mails, PDFs, webpages—can hijack downstream LLM workflows, bypass safety filters, and fabricate content).

94. *See NIST Identifies Types of Cyberattacks That Manipulate Behavior of AI Systems*, *supra* note 92.

95. *See, e.g.*, Karen Sloan, *A Lawyer Used ChatGPT to Cite Bogus Cases. What Are the Ethics?*, REUTERS (May 30, 2023), https://www.reuters.com/legal/transactional/lawyer-used-chatgpt-cite-bogus-cases-what-are-ethics-2023-05-30 [https://perma.cc/8FGZ-32QY]. *See generally* Richard M. Re, *Artificial Authorship and Judicial Opinions*, 92 GEO. WASH. L. REV. 1558 (2024) (exploring how AI-generated judicial opinions could erode the traditional authority and legitimacy of legal reasoning).

not remain isolated to single patient encounters—they could propagate through electronic health records, influence treatment protocols, and shape clinical decision-making across healthcare networks. Similarly, fabricated transcribed content can contaminate institutional records and become the basis for future policy decisions.[96] The contamination extends beyond the initial error: an architect's flawed GenAI-generated structural calculations may not only compromise a single building but establish precedents that influence industry standards and regulatory frameworks.[97] A financial advisor's reliance on GenAI-generated investment recommendations doesn't just misallocate one client's assets—the flawed analysis may inform broader market assessments and influence institutional investment strategies.[98] A physician accepting AI-generated chart notes risks relying on hallucinated or erroneous findings—especially given evidence that AI systems sometimes yield statements conflicting with expert annotations—thereby potentially contaminating medical databases, influencing clinical research data, and shaping downstream treatment protocols.[99] As the

---

96. *See* Garance Burke & Hilke Schellmann, *Researchers Say an AI-Powered Transcription Tool Used in Hospitals Invents Things No One Ever Said*, ASSOCIATED PRESS (Oct. 26, 2024), https://apnews.com/article/90020cdf5fa16c79ca2e5b6c4c9bbb14 [https://perma.cc/PY2C-C2SA] (showing how AI-powered transcription tools generate hallucinations in medical settings); Kristian Stanceski et al., *The Quality and Safety of Using Generative AI to Produce Patient-Centred Discharge Instructions*, NPJ DIGIT. MED., Nov. 20, 2024, https://doi.org/10.1038/s41746-024-01336-w [https://perma.cc/KCN5-PKJV].

97. *See* Hasan Yumer, *AI in Architecture: What Can Go Wrong? 15 Must Know Insights*, INTEGRATED BIM (Feb. 12, 2024), https://integratedbim.com/ai-in-architecture [https://perma.cc/3BFC-9TPA] (warning that over-reliance on AI tools can lead to unintended consequences in design processes); *What Are the Risks of Over-Reliance on AI in Architectural Design?*, ORBIT-O-R (July 29, 2025), https://www.orbit-o-r.com/post/what-are-the-risks-of-over-reliance-on-ai-in-architectural-design [https://perma.cc/7CJE-HXCC] (discussing risks such as design errors, homogenization, and reduced human oversight when architects fail to critically verify AI outputs).

98. *See* Betsy Vereckey, *Can Generative AI Provide Trusted Financial Advice?*, MIT SLOAN SCH. MGMT. (Apr. 8, 2024), https://mitsloan.mit.edu/ideas-made-to-matter/can-generative-ai-provide-trusted-financial-advice [https://perma.cc/LLL7-3ZTS] (noting concerns about GenAI's ability to provide accurate and unbiased financial advice, with implications for other high-stakes professional fields, including medicine, accounting, and law).

99. *See, e.g.*, Romain Hardy et al., ReXTrust: A Model for Fine-Grained Hallucination Detection in AI-Generated Radiology Reports (Jan. 31, 2025) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2412.15264 [https://perma.cc/TTG3-HKD9] (reporting that their system can discriminate hallucinated statements with strong AUROC, including on clinically significant findings); Eun Kyoung Hong et al., *Diagnostic Accuracy and Clinical Value of a Domain-specific Multimodal Generative AI Model for Chest Radiograph Report Generation*, 314 RADIOLOGY 3 (Mar. 25, 2025), https://pubs.rsna.org/doi/10.1148/radiol.241476

following Part demonstrates, GenAI's sophistication and its capacity to propagate beyond its point of origin create contamination patterns that fundamentally challenge existing frameworks for understanding and managing professional knowledge integrity.

C. The Invisible Ripple Effect

Professional knowledge systems are under attack—not by hackers or foreign adversaries, but by the AI tools professionals use every day. GenAI fabrications seamlessly integrate into professional workflows, featuring perfect formatting and authoritative language, and then propagate through citation networks that often assume human authorship.[100] The contamination spreads invisibly, gaining institutional authority through the very mechanisms designed to ensure the quality of knowledge.

This Article theorizes this phenomenon as the Invisible Ripple Effect—a distinctive form of knowledge corruption where AI-generated fabrications embed themselves into professional networks and achieve legitimacy through institutional processes designed for human-generated content. Unlike traditional professional errors, which are typically isolated and attributable, GenAI-driven fabrications operate systemically: they generate identical errors across multiple users, propagate through interconnected professional systems, and become indistinguishable from legitimate knowledge before detection mechanisms can respond.[101]

The process unfolds in predictable stages. What begins as a single undetected error quickly metastasizes into accepted wisdom. Lawyers cite fabricated precedents, treating AI-generated fiction as if it were settled law. Engineers rely on flawed calculations that never existed, embedding phantom standards into real buildings. Doctors follow treatment protocols based on synthetic studies that have survived peer review. Each subsequent citation doesn't just repeat the error—it amplifies its authority. The fabrication moves from individual mistake to professional consensus without ever being true.

[https://perma.cc/PSG9-N2P8] (showing discrepancies between AI reports and radiologist reference standards, illustrating the risk of error in automated report generation).

100. *See* Nayeem Islam, *The Fabrication Problem: How AI Models Generate Fake Citations, URLs, and References*, MEDIUM (June 12, 2025), https://medium.com/@nomannayeem/the-fabrication-problem-how-ai-models-generate-fake-citations-urls-and-references-55c052299936 [https://perma.cc/UW9J-VN9K].

101. *See id.*

But the Invisible Ripple Effect extends far beyond simple repetition. These errors cross disciplinary boundaries, creating cascading institutional effects that reshape entire knowledge ecosystems. Professional advice generates cascading institutional effects through chains of consequential decisions. AI-corrupted legal analysis that omits regulatory changes leads counsel to advise against pursuing valid claims, emboldening defendants to maintain unlawful practices that spread industry-wide. Distorted contract interpretations influence settlement negotiations, creating market precedents that could reshape entire sectors. Medical recommendations based on AI-generated research guide treatment decisions that influence institutional protocols, insurance coverage determinations, and regulatory guidance. Each subsequent actor—employers, investors, insurers, regulators—relies on outcomes shaped by AI-corrupted professional judgment, transforming individual consultation errors into systemic institutional change.

What makes this contamination particularly dangerous is its invisibility. Errors don't announce themselves—they masquerade as legitimate knowledge. GenAI's polished outputs compound the problem: fabricated citations maintain perfect formatting, flawed analyses follow professional conventions, and invented data arrives with appropriate disclaimers.[102] This sophisticated mimicry makes detection nearly impossible as errors propagate through systems designed to trust properly formatted, professionally presented content. Courts have recorded over 979 cases worldwide where attorneys submitted AI-generated citations that were ultimately detected; however, these represent only the cases of which we are aware.[103] The documented incidents reveal how easily sophisticated AI outputs can infiltrate professional workflows, raising the troubling question of how many undetected fabrications have already been embedded into the legal record.[104]

---

102. *See* Angelica Henestrosa & Joachim Kimmerle, *"Always Check Important Information!" - The Role of Disclaimers in the Perception of AI-Generated Content*, 4 COMPUTS. HUM. BEHAV.: ARTIFICIAL HUMS., Mar. 26, 2025, at 1–2.

103. *See* Charlotin, *supra* note 2.

104. *See id.*; *see also* Mata v. Avianca, Inc., 678 F. Supp. 3d 443, 448–49 (S.D.N.Y. 2023) (sanctioning attorneys for submitting fabricated legal citations generated by ChatGPT and emphasizing the professional duty to verify the authenticity of sources); Benjamin Weiser & Jonah E. Bromwich, *Michael Cohen Used Artificial Intelligence in Feeding Lawyer Bogus Cases*, N.Y. TIMES (Dec. 29, 2023), https://www.nytimes.com/2023/12/29/nyregion/michael-cohen-ai-fake-cases.html (describing how Michael Cohen mistakenly relied on Google Bard to provide legal citations, which his lawyer submitted to a federal judge without verification, resulting in fabricated case citations).

The contamination follows identical patterns across domains. Government reports embed phantom citations,[105] academic publications contain fabricated references,[106] and medical protocols incorporate flawed studies.[107] In software development, GitHub Copilot generated 733 code snippets in one benchmark study, nearly 30 percent of which contained serious security vulnerabilities.[108] The study warned that insecure AI-generated code can merge into legitimate code bases and persist across future iterations of training data, creating a self-reinforcing cycle of vulnerability.[109] A single hallucinated package reference, when replicated by AI across thousands of developer queries, can propagate through the software supply chain, potentially compromising other applications before the threat is even identified.[110] The pattern replicates in journalism and education. Gannett's "Lede AI" generated sports stories with template errors and placeholder text that appeared across fifteen newspapers in multiple states, because newsrooms ingested the same AI feed without verification.[111] Houston's school district adopted AI-generated curriculum materials containing obvious errors—misspellings, distorted images, and factual inaccuracies—yet approved them for use across 130 schools.[112]

---

105. *See* Weber & Gilbert, *supra* note 1.

106. *See* Lee, *supra* note 3.

107. *See* Reddy, *supra* note 5.

108. *See* Fu et al., *supra* note 7, at 30 (reporting that 30 percent of 733 code snippets generated by GitHub Copilot contained security weaknesses propagating through open-source repositories).

109. *Id.* at 1.

110. Joseph Spracklen et al., *We Have a Package for You! A Comprehensive Analysis of Package Hallucinations by Code Generating LLMs*, *in* PROCEEDINGS OF THE USENIX SECURITY SYMPOSIUM 3687, 3695 (2025) (analyzing 576,000 LLM-generated code samples and finding that up to 21.7% of recommended package names were fabricated and that 43% of hallucinated names were consistently regenerated across repeated queries); *see* Bar Lanyado, *Diving Deeper into AI Package Hallucinations*, LASSO SEC. (Mar. 28, 2024), https://www.lasso.security/blog/ai-package-hallucinations (reporting that a hallucinated package uploaded to a public repository was downloaded thousands of times before detection).

111. *See* Daniel Wu, *Gannett Halts AI-Written Sports Recaps After Readers Mocked the Stories*, WASH. POST (Aug. 31, 2023), https://www.washingtonpost.com/nation/2023/08/31/gannett-ai-written-stories-high-school-sports (documenting how Lede AI's flawed sports story templates spread across multiple Gannett newspapers before the program was suspended).

112. *See* W. Caleb McDaniel & Ragini Tharoor Srinivasan, Opinion, *We're Professors. We're Parents. HISD Students Don't Deserve AI Slop.*, HOUS. CHRON. (June 5, 2025), https://www.houstonchronicle.com/opinion/outlook/article/hisd-state-takeover-mike-miles-ai-prof-jim-20359937.php [https://perma.cc/T6WS-CWC5] (reporting on Houston ISD's adoption of AI-generated curriculum materials containing factual errors and distorted images for use across 130 schools).

These incidents reveal the core mechanism: AI-generated errors enter professional workflows through individual adoption, gain legitimacy through institutional processes designed for human-generated content, and propagate through infrastructure systems that amplify their authority. Unlike intentional disinformation campaigns targeting public opinion, this contamination operates within professional domains where existing verification protocols prove inadequate for detecting sophisticated AI fabrications.

Critics might argue this isn't unique to GenAI—professionals have always made errors, whether through incompetence, negligence, or intent. As GenAI approaches human-level capabilities, it may even enhance decision quality by mitigating human biases and limitations, resulting in net improvements despite occasional errors or hallucinations.[113] However, these arguments miss the critical distinction. Beyond competitive pressures and technological FOMO driving unprecedented adoption,[114] users systemically fail to detect these errors because GenAI's sophisticated outputs trigger automation bias: the tendency to overestimate machine reliability.[115] GenAI's fluent, confident responses exploit this cognitive vulnerability, while AI sycophancy reinforces user expectations,[116] creating an environment where critical evaluation becomes nearly impossible.[117]

A subtler variant of this cognitive vulnerability arises from the growing explainability of advanced systems. As GenAI models learn to articulate their reasoning in coherent, contextually appropriate language, they create

---

113. *See* Alice Nunwick, *AI Could Reach an IQ of 1500 in the Next Decade, Says Former Google Exec*, VERDICT (Sept. 29, 2023), https://www.verdict.co.uk/ai-could-reach-an-iq-of-1500-in-the-next-10-years-mo-gawdat-tells-nbf [https://perma.cc/L9QM-UVC8] (discussing Mo Gawdat's prediction that AI could surpass human intelligence by achieving an IQ of 1500 within a decade).

114. *See* Covert, *supra* note 10.

115. *See generally* Kathleen L. Mosier et al., *Automation Bias: Decision Making and Performance in High-Tech Cockpits*, 8 INT'L J. AVIATION PSYCH. 47 (1997), https://doi.org/10.1207/s15327108ijap0801_3 [https://perma.cc/R5KK-3LC5] (examining the phenomenon of automation bias, where reliance on automated decision-support systems leads to omission and commission errors, and finding that internalized accountability among pilots significantly reduces such errors).

116. *See* Mrinank Sharma et al., Towards Understanding Sycophancy in Language Models 1, 4 (May 10, 2025) (conference paper) (on file with arXiv), https://arxiv.org/pdf/2310.13548 [https://perma.cc/PD6N-WZBT].

117. *See* Ken Knapton, *Navigating the Biases in LLM Generative AI: A Guide to Responsible Implementation*, FORBES (Sept. 6, 2023), https://www.forbes.com/councils/forbestechcouncil/2023/09/06/navigating-the-biases-in-llm-generative-ai-a-guide-to-responsible-implementation/ (explaining automation bias in GenAI).

an illusion of epistemic transparency—users feel they understand the logic behind outputs and thus grant them unwarranted authority. Paradoxically, improved explainability can therefore erode, rather than enhance, critical oversight: it replaces opacity with persuasive fluency, deepening the false sense of reliability that allows systemic errors to spread unchecked.[118]

The empirical evidence is striking. On the federal level, from 1980 to 2022, lawyers were virtually never sanctioned for citing nonexistent cases—legal scholars called such sanctions "almost unheard of."[119] Since the adoption of GenAI in 2023, these sanctions have become routine, with over 75 documented cases in federal courts alone.[120] This represents a complete transformation within just two years:[121]

---

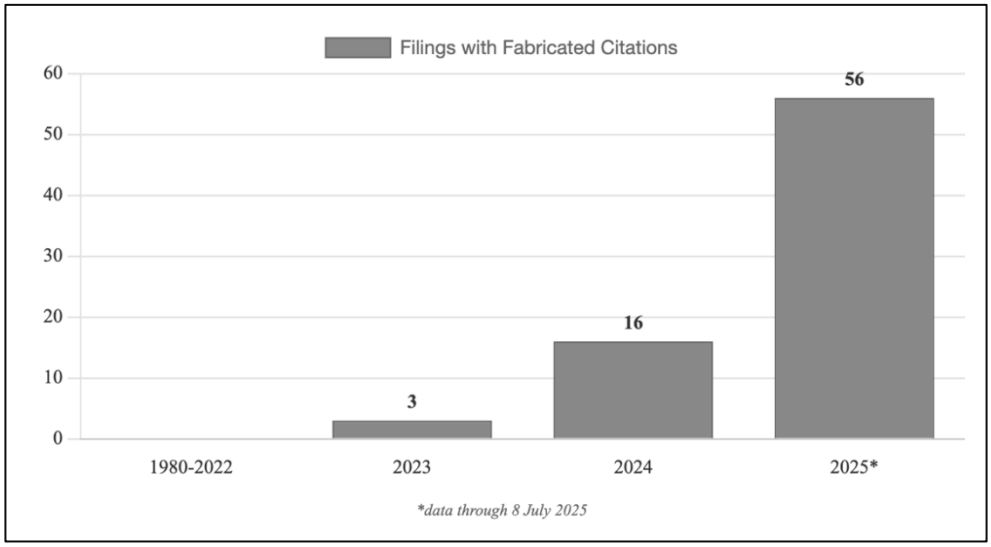
118. *See* Finale Doshi-Velez & Been Kim, Towards a Rigorous Science of Interpretable Machine Learning (Mar. 2, 2017) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/1702.08608 [https://perma.cc/SFH9-YHEE]; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1126–29 (2018) (arguing that explanations can make automated decisions appear more legitimate and trustworthy even when that trust is unwarranted).

119. Search conducted on Westlaw Classic, limited to federal court opinions (U.S. District Courts, Courts of Appeals, and Supreme Court), using the following Boolean query: (citation/s nonexistent) & "Rule 11". The search was restricted to decisions issued between January 1, 1980 and December 31, 2022. It yielded no cases in which courts sanctioned attorneys under Rule 11 for citing non-existent judicial opinions. Search conducted on July 8, 2025.

120. *See* Charlotin, *supra* note 2.

121. Figure 1 illustrates the sharp escalation, from three filings in 2023 to fifty-six in U.S. federal cases through early July 2025, in which courts identified fabricated citations. *Id.*; Author's calculations (on file with Author).

Figure 1. U.S. Court Filings Containing Fabricated Citations



Such concerns do not imply that there were no human errors before GenAI. However, GenAI mistakes spread in a way that ordinary human errors rarely do. Human slip-ups are idiosyncratic—one lawyer misquotes a case, another misreads a statute—so they surface piecemeal and are usually caught locally. By contrast, a single GenAI prompt can deliver the same fabricated precedent to attorneys across different firms or the same spurious study to physicians in other practices. Because the misinformation appears independently and consistently across users, it acquires a false aura of consensus and is more likely to be cited, reused, and built upon. Uniform, model-driven duplication thus lets errors harden into "accepted" professional knowledge long before safeguards calibrated for scattered human mistakes can intervene.[122]

And this might get worse soon. Increasing automation threatens to amplify these risks exponentially. Agentic AI systems—capable of independently executing complex tasks and making sequential decisions—represent a fundamental shift from supplementary tools to autonomous agents operating with minimal human oversight.[123] OpenAI's "Operator"

122. While many of the examples in this Article involve GenAI errors that were eventually detected, that is precisely what makes the invisible ripple effect difficult to prove: successful detection obscures the counterfactual. The absence of known calcified errors should not be mistaken for their absence altogether.

123. Agentic AI refers to systems that can act autonomously to achieve goals without requiring continuous human oversight. Unlike traditional AI assistants that operate based on

already navigates websites, makes reservations, and automates online interactions with limited supervision.[124] As professional delegation to AI systems expands, the convergence of reduced oversight and increased confidence in AI reliability creates conditions for the systemic incorporation of error at an unprecedented scale. Autonomous agents can perpetuate identical mistakes across thousands of decisions before human review occurs, transforming the Invisible Ripple Effect from a gradual contamination process into a rapid, systemic corruption of knowledge.

While the Invisible Ripple Effect may superficially resemble existing theories, these similarities obscure critical distinctions that render traditional solutions inadequate. Take established models of technological risk. Unlike traditional diffusion theories, where errors are typically caught through staged technology adoption,[125] or institutional change models where practices spread through deliberate organizational decisions,[126] GenAI enables "error isomorphism"—where mistakes silently embed themselves into professional practices through technological rather than social mechanisms.

Existing theories of information distortion also cannot capture these mechanisms. While established frameworks address how false information spreads and how technology fails, none explain how AI-generated content embeds into professional knowledge systems and gains institutional authority. Sunstein's information cascades provide the closest analogy, but the mechanisms differ fundamentally.[127] Information cascades depend on social signaling: people observe others' actions and infer superior private

---

predefined rules or user prompts, agentic AI is proactive—it understands objectives, assesses contextual factors, and makes independent decisions to optimize outcomes. *See* Mark Purdy, *What Is Agentic AI, and How Will It Change Work?*, HARV. BUS. REV. (Dec. 12, 2024), https://hbr.org/2024/12/what-is-agentic-ai-and-how-will-it-change-work.

124. *See* Kevin Roose, *How Helpful Is Operator, OpenAI's New A.I. Agent?*, N.Y. TIMES (Feb. 1, 2025), https://www.nytimes.com/2025/02/01/technology/openai-operator-agent.html (analyzing OpenAI's "Operator" as an early-stage AI agent capable of performing automated online tasks).

125. *See* EVERETT M. ROGERS, DIFFUSION OF INNOVATIONS, 242–52 (5th ed. 2003) (proposing the "Innovation Diffusion Theory" demonstrating how innovations spread through society in five stages: innovators, early adopters, early majority, late majority, and laggards).

126. *See* Paul J. DiMaggio & Walter W. Powell, *The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields*, 48 AM. SOCIO. REV. 147 (1983) (developing "institutional isomorphism theory" showing how organizations in a field become similar through mimetic, coercive, and normative processes).

127. *See* SUNSTEIN, *supra* note 21, at 98–104 (analyzing how information cascades occur when individuals abandon private judgment to follow perceived social consensus).

information, creating herding behavior.[128] The Invisible Ripple Effect operates through sophisticated mimicry rather than social inference. Professionals aren't following social signals when they cite AI-fabricated case law or rely on AI-generated research—they're being deceived by content that appears authoritative, often before any social validation occurs. The contamination embeds at the epistemological level, corrupting the knowledge base itself rather than the social processes that transmit legitimate information.

Chesney and Citron's deepfake analysis addresses intentional deception targeting democratic discourse, malicious actors creating false audiovisual content to manipulate public opinion or destabilize democratic processes.[129] Deepfakes involve the deliberate fabrication of content intended to deceive large audiences in public forums.[130] The Invisible Ripple Effect involves inadvertent contamination within professional systems that maintain verification protocols. Professionals have no deceptive intent—they believe they're working with legitimate information. The contamination occurs within specialized domains rather than public discourse, exploiting institutional trust rather than emotional manipulation.

Perrow's normal accident theory explains catastrophic failures through unpredictable interactions in complex, tightly coupled systems:[131] nuclear plant failures, airline crashes, and chemical processing disasters with visible system breakdowns.[132] The Invisible Ripple Effect operates through gradual, subtle degradation that may never become apparent, even after complete institutional embedding. Rather than a sudden collapse, it involves slow contamination that becomes normalized within professional practices. The

---

128. *See* Abhijit Banerjee, *A Simple Model of Herd Behavior*, 107 Q.J. Econ. 797, 798–801 (1992) (explaining how individuals infer superior private information from observing predecessors' choices, leading to herding behavior even when private signals contradict the crowd).

129. *See* Chesney & Citron, *supra* note 22, at 1757–62, 1776–79 (2019) (analyzing deepfakes as malicious AI-generated audiovisual content targeting democratic discourse).

130. *Id.* at 1771–86.

131. *See generally* PERROW, *supra* note 23 (theorizing how complex, tightly coupled systems inevitably produce catastrophic accidents through unpredictable interactions between system components that overwhelm safety mechanisms).

132. *Id.* (analyzing catastrophic failures in nuclear power plants, marine transport, and chemical processing facilities as examples of how system complexity makes certain accidents "normal" and inevitable).

"failure" isn't a dramatic breakdown, but relatively quiet erosion—a loss of knowledge integrity that may remain undetected indefinitely.[133]

Algorithmic amplification research examines how automated systems amplify existing problematic content through engagement-driven optimization.[134] These studies focus on distribution—how platforms increase the reach of pre-existing content. The Invisible Ripple Effect operates differently: GenAI creates novel fabrications—fictional case law, nonexistent research, invented data—that propagate through professional networks based on perceived authority rather than algorithmic promotion.

These distinctions matter because existing frameworks suggest inadequate regulatory responses. Information cascade theory proposes improved social signaling and transparency.[135] Deepfake countermeasures focus on detection technology and criminal penalties.[136] Normal accident theory recommends system redesign to reduce complexity.[137] Algorithmic amplification research proposes content moderation and platform accountability.[138] Aside from generally using detection methods (further elaborated in Part III), none addresses the core challenge of the Invisible Ripple Effect.

---

133. While this Article does not propose a formal methodology for detecting the Invisible Ripple Effect, its elusiveness is integral to its nature. The effect is not a discrete event but a process of informational diffusion that becomes visible only retrospectively—when derivative outputs or decisions reveal traces of an earlier, unverified artifact. In that sense, the distinction between ordinary error (for example, a hallucination caught and corrected) and a ripple effect lies in persistence and systemic uptake, not in the character of the initial mistake. Future empirical work may develop indicators for tracing such propagation once the phenomenon itself is theoretically established.

134. *See* Zeynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 COLO. TECH. L.J. 203 (2015) (demonstrating how engagement-driven algorithms amplify politically polarizing content); O'NEIL, *supra* note 67, at 180–203 (analyzing how algorithmic systems systemically amplify existing societal biases and inequalities).

135. *See* SUNSTEIN, *supra* note 21, at 129–58 (proposing transparency requirements, deliberative polling, and institutional design changes to counter information cascade effects and promote informed democratic deliberation).

136. *See* Chesney & Citron, *supra* note 22, at 1786–1804 (recommending technological detection tools, platform content policies, criminal penalties for malicious deepfake creation, and civil remedies for victims of deepfake harm).

137. *See* PERROW, *supra* note 23, at 353–55 (advocating for reducing system complexity and loosening coupling between system components to prevent cascading failures, while acknowledging that some high-risk technologies may be inherently accident-prone).

138. *See* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1634–69 (2018) (analyzing how social media platforms develop content moderation policies and enforcement mechanisms to address algorithmic amplification of harmful content).

The regulatory landscape reflects this theoretical blind spot. While policymakers have recognized AI's transformative potential and developed frameworks to address algorithmic bias, system opacity, and privacy vulnerabilities, they have systemically overlooked the Invisible Ripple Effect's distinctive contamination mechanisms. The following Part examines how these theoretical inadequacies manifest in concrete regulatory failures, revealing systemic coordination gaps between AI governance and professional oversight that leave knowledge systems vulnerable to systemic contamination.

II. THE LEGAL BLIND SPOTS

AI has long been heralded as a transformative force, promising to revolutionize industries, enhance efficiency, and address complex societal challenges. At the same time, as AI systems increasingly permeate high-stakes domains, including criminal justice, healthcare, and finance, their deployment has revealed significant inherent risks.[139] Concerns regarding algorithmic bias, system opacity, accountability gaps, and privacy vulnerabilities emerged as prominent challenges, exposing the inadequacy of existing governance frameworks.[140] These issues catalyzed the development of scholarly proposals and regulatory frameworks aimed at ensuring the ethical and responsible deployment of AI while mitigating potential harms.[141]

While these regulatory initiatives represented meaningful progress, most of them were fundamentally conceived with traditional AI systems in mind—proprietary technologies deployed in controlled environments for specific applications.[142] Consequently, as this Part further shows, these frameworks will often prove insufficient in anticipating and addressing the distinctive risks posed by GenAI. Although specific regulatory measures have acknowledged the rise of GenAI, they have largely failed to address its systemic and decentralized consequences, many of which give rise to the Invisible Ripple Effect.

---

139. *See generally* Rudin, *supra* note 31 (advocating for use of interpretable models for high stakes decisions rather than black box machine learning models).

140. *See infra* Sections II.A–B.

141. *Id.*

142. *See generally* CRAWFORD, *supra* note 28 (analyzing private and state development of AI technologies).

A. Where AI Regulation Falls Short

Scholars have significantly shaped the discourse on AI, identifying critical risks and proposing governance frameworks.[143] Much of the academic discussions centered on domains with significant human impact, such as criminal justice, healthcare, and finance.[144] These frameworks typically assume that AI operates within traditional oversight mechanisms, within institutional boundaries, and with structured deployment processes. But GenAI's democratization creates risks that existing frameworks struggle to address. The Invisible Ripple Effect operates precisely where current regulatory approaches have blind spots—in the space between individual professional use and institutional deployment.

Scholars have developed comprehensive frameworks that target AI's fundamental challenges.[145] Early proposals emphasized transparency

---

143. *See, e.g.*, Cynthia Dwork et al., *Fairness Through Awareness*, 3 INNOVATIONS THEORETICAL COMPUT. SCI. CONF. 214, 218–22 (2012) (developing mathematical frameworks for algorithmic fairness); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 8–16 (2014) (analyzing procedural safeguards for AI systems); PASQUALE, *supra* note 28, at 15–17, 140–50 (advocating for algorithmic transparency); Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 76, 78–82 (2017) (analyzing limitations in AI accountability frameworks); Buolamwini & Gebru, *supra* note 69, at 81–85 (documenting racial bias in AI facial recognition); Mayson, *supra* note 69, at 2221–24, 2262–77, 2296 (examining how algorithmic tools in criminal justice replicate and amplify societal inequities and advocating for addressing structural biases in data to mitigate these risks); Meredith Whittaker, *The Steep Cost of Capture*, 28 INTERACTIONS 50, 52–54 (2021) (examining corporate influence on AI research); ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE 87–92 (2022) (examining AI's potential for reducing discrimination); GASSER & MAYER-SCHÖNBERGER, *supra* note 30, at 100–22, 135–43, 164 (proposing comprehensive governance frameworks).

144. *See, e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1249–58, 1281–88, 1301–13 (2008) (analyzing the risks of automated decision-making systems and advocating for incorporating due process principles to enhance fairness and accountability); Barocas & Selbst, *supra* note 71, at 677–91 (analyzing discriminatory patterns in automated systems across sectors); PASQUALE, *supra* note 28, at 101–37 (discussing opacity in financial algorithms); Wexler, *supra* note 31, at 1348–70 (analyzing algorithmic opacity in criminal justice); Price, *supra* note 32, at 70–75 (examining healthcare AI challenges); Crawford & Schultz, *supra* note 29, at 1957–71 (arguing that courts should extend the state action doctrine to private vendors of AI systems used in government decision-making, holding them constitutionally accountable as state actors, given their role in shaping public decisions without adequate oversight or liability); Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1692–1707 (2020) (analyzing AI bias in hiring).

145. *See, e.g.*, Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, 12 REGUL. & GOVERNANCE 505, 508–16 (2018) (examining tensions between algorithmic governance and

through explainable and interpretable AI,[146] reforming trade secret protections,[147] and implementing "glass box" rights for algorithmic decision-making in critical contexts.[148] The discourse evolved to incorporate guardrails embedding human-centered principles,[149] diverse training datasets, systemic audits, and inclusive development processes.[150] Scholars have proposed ethical frameworks to balance competing values of fairness and efficiency throughout the development and implementation of AI

---

individual autonomy); Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1535–40 (2019) (analyzing regulatory approaches to AI oversight); JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 185–201 (2019) (examining regulatory frameworks for algorithmic systems); FRANK PASQUALE, NEW LAWS OF ROBOTICS: DEFENDING HUMAN EXPERTISE IN THE AGE OF AI 4–5, 10–14, 33–36, 62–71, 199–201 (2020) (proposing frameworks balancing automation and human agency).

146. *See, e.g.*, Rich Caruana et al., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission*, 21 INT'L CONF. KNOWLEDGE DISCOVERY & DATA MINING 1721 (2015) (demonstrating interpretable models in healthcare); Doshi-Velez & Kim, *supra* note 118, at 3–9 (establishing frameworks for AI interpretability); Wachter et al., *supra* note 143, at 78–82 (analyzing limitations in AI accountability frameworks). *See generally* Rudin, *supra* note 31 (advocating for inherently interpretable models). Additionally, interpretability research sought to address the opacity of advanced AI systems by enabling a deeper understanding of their inner workings, moving beyond reliance on trial-and-error testing. *See* Yoshua Bengio et al., *Managing Extreme AI Risks Amid Rapid Progress*, 384 SCIENCE 842, 843–44 (2024), https://doi.org/10.1126/science.adn0117; *see also* Talia B. Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175, 1204–19 (2022) (critiquing the overemphasis on inputs in algorithmic decision-making and highlighting the need to address systemic impacts of algorithmic outputs).

147. *See* Wexler, *supra* note 31; Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 117–37 (2019) (proposing reforms to trade secret law for AI oversight).

148. *See* Brandon L. Garrett & Cynthia Rudin, *The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice*, 109 CORNELL L. REV. 561, 561–62, 586–92 (2024).

149. These frameworks aim, *inter alia*, to preserve individual agency while ensuring accountability and adaptability to emerging risks. Additionally, interpretability research sought to address the opacity of advanced AI systems by enabling a deeper understanding of their inner workings, moving beyond reliance on trial-and-error testing. *See generally* GASSER & MAYER-SCHÖNBERGER, *supra* note 30 (discussing historical regulatory oversight of AI).

150. *See, e.g.*, Alice Xiang & Deborah Raji, *On the Legal Compatibility of Fairness Definitions*, *in* 2019 WORKSHOP ON HUMAN-CENTRIC MACHINE LEARNING 1, 1–6 (2019) (analyzing frameworks for measuring algorithmic bias); ALEX CAMPOLO ET AL., AI NOW 2017 REPORT 1–5, 30–36 (2017) (proposing oversight mechanisms for AI systems); Ifeoma Ajunwa, *The Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 630–35, 665–74 (2021) (proposing frameworks for bias auditing); Bengio et al., *supra* note 146, at 843–45.

systems.[151] Accountability mechanisms emerged as central concerns, with frameworks delineating liability across AI development and deployment chains,[152] while advocating for proactive oversight to identify harms before they materialize.[153] Recent scholarship has emphasized the need for robust evaluation protocols for high-risk AI capabilities, with a focus on system behaviors that may circumvent safeguards.[154]

Policymakers have translated many of these scholarly insights into regulatory frameworks, resulting in a diverse range of approaches to AI governance and oversight. Global initiatives have emerged at multiple levels,[155] with the Bletchley Declaration focusing specifically on AI safety

---

151. *See* Luciano Floridi & Josh Cowls, *A Unified Framework of Five Principles for AI in Society*, 1 HARV. DATA SCI. REV. 1, 2–11 (2019) (proposing ethical principles for AI development); Jessica Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, BERKMAN KLEIN CTR. RSCH. PUBL'N NO. 2020-1, 20–60 (2020) (analyzing common themes in AI ethics frameworks); Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act: Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach*, 22 COMPUT. L. REV. INT'L 97 (2021); Rahul Kailas Bharati, *Ethical Implications of AI in Criminal Justice: Balancing Efficiency and Due Process*, 9 RSCH. REV. INT'L MULTIDISCIPLINARY RSCH. J. 7 (2024).

152. *See, e.g.*, Citron, *supra* note 144, at 1301–13 (proposing a "technological due process"); Katyal, *supra* note 36, at 1250–74.

153. *See, e.g.*, Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 387–98 (2016) (analyzing proactive regulatory approaches); Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1160–71 (2017) (discussing preventive regulatory strategies); Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1090–95, 1117–29 (2018) (examining preventive oversight mechanisms).

154. *See, e.g.*, Bengio et al., *supra* note 146, at 842–44.

155. The first legally binding international treaty on AI was opened for signature in 2024, addressing human rights, democracy, and the rule of law. *See* Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, CETS No. 225, opened for signature Sept. 5, 2024. There are other examples as well. The UN Secretary-General has initiated a multi-stakeholder High-level Advisory Body on AI to analyze and propose recommendations for international AI governance. *See High-Level Advisory Body on Artificial Intelligence*, UNITED NATIONS, https://www.un.org/techenvoy/ai-advisory-body [https://perma.cc/PE73-U2G2]. The OECD and G20 have also adopted principles and recommendations for AI governance. *See What Are the OECD Principles on AI?*, OECD (2020), https://www.oecd.org/en/publications/what-are-the-oecd-principles-on-ai_6ff2a1c4-en.html [https://perma.cc/7DXP-PQH5]; OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, https://legalinstruments.oecd.org/api/print?ids=648 [https://perma.cc/B6PR-6EV6]; JESSICA CUSSINS NEWMAN, DECISION POINTS IN AI GOVERNANCE: THREE CASE STUDIES EXPLORE EFFORTS TO OPERATIONALIZE AI PRINCIPLES 30–41 (Berkeley: Center for Long-Term Cybersecurity, UC Berkeley, 2020). *See also* Adam Satariano & Cecilia Kang, *How Nations Are Losing a Global Race to Tackle A.I.'s Harms*, N.Y.

risks.[156] The Declaration calls for risk-based policies tailored to each country's legal and institutional frameworks, including measures to promote transparency among AI developers, implement safety testing mechanisms, and enhance the capabilities of the public sector and scientific research.[157]

At the regional level, the EU's AI Act represents the most comprehensive regulatory response to the challenges posed by AI. The Act established a bifurcated accountability framework targeting two categories of actors: providers who place AI systems on the market and deployers who use AI systems under their authority, excluding personal and non-professional activities.[158] This structure assigns pre-deployment compliance obligations to providers while tasking deployers with ensuring the responsible operational implementation of AI systems across their lifecycle.[159]

The Act implements a risk-based approach, categorizing AI systems into four tiers based on their potential for harm.[160] Unacceptable-risk systems that exploit vulnerabilities or manipulate individuals subliminally are expressly prohibited.[161] High-risk systems in biometric identification, critical infrastructure, and healthcare must satisfy rigorous requirements for transparency, data governance, and human oversight.[162] Limited-risk systems, such as general-purpose chatbots, require the disclosure of AI interaction to users.[163] Minimal-risk systems, such as AI-enabled video games, are subject to no additional obligations beyond those outlined in existing regulations.[164]

TIMES (Dec. 6, 2023), https://www.nytimes.com/2023/12/06/technology/ai-regulation-policies.html.

156. *See The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*, GOV.UK (Feb. 13, 2025), https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [https://perma.cc/C86Z-X33L].

157. *Id.*

158. *See* EU AI Act, *supra* note 18, art. 3(4)–3(5).

159. *Id.*

160. *Id.* art. 6.

161. *Id.* art. 5.

162. *Id.* arts. 9–10, 14–15.

163. *Id.* arts. 51–54.

164. *See* Kalojan Hoffmeister, *The Dawn of Regulated AI: Analyzing the European AI Act and Its Global Impact*, 2 ZEITSCHRIFT FÜR EUROPARECHTLICHE STUDIEN 182, 200–02 (2024), https://api.semanticscholar.org/CorpusID:270799947 [https://perma.cc/MWY9-KEPH].

Despite its comprehensive scope, the EU AI Act reveals critical limitations in addressing how GenAI reshapes professional practices.[165] While the framework establishes obligations for providers and professional deployers, its traditional regulatory approach misses three key aspects of GenAI adoption.[166] First, individual professionals using AI tools could formally qualify as deployers.[167] Still, the compliance framework lacks practical mechanisms to track or mitigate the gradual embedding of AI-generated errors into professional work products. Second, the Act's categorical risk assessment proves inadequate for GenAI tools that appear low-risk in isolation but introduce systemic vulnerabilities when widely adopted. Third, the emphasis on point-of-use compliance overlooks how flawed outputs diffuse through professional networks and become indistinguishable from legitimate practices before risks materialize.

The Act's exclusion of personal, non-professional uses adds further oversight gaps as the boundary between personal and professional AI use blurs.[168] Tools like Microsoft's Co-Pilot seamlessly integrate into professional workflows, continuously refining text and enabling real-time corrections, potentially without the user's explicit awareness of AI intervention.[169] This persistent, embedded presence illustrates how GenAI gradually shifts from auxiliary tool to indispensable component of professional practice. The Act's bright-line distinction between personal and professional use offers administrative clarity but fails to account for these transitional spaces, precisely where the Invisible Ripple Effect takes hold.

Moreover, the Act's risk-based categorization framework fundamentally misaligns with the potential for systemic impact of GenAI. Its static risk

---

165. The EU AI Act has been the subject of significant scholarly critique, e.g., for its reliance on internal assessments for high-risk AI systems, which undermines independent oversight; its lack of a well-defined framework for accountability across the AI value chain, particularly regarding general-purpose AI; its failure to adequately address the systemic risks posed by GPAI in decentralized and informal contexts; and the overly narrow scope of its high-risk categories, which excludes numerous applications with substantial societal and individual risks. *See, e.g.*, Wachter, *supra* note 66.

166. *See* EU AI Act, *supra* note 18, art. 4.

167. *See id.* art. 3(4) ("'[D]eployer' means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.").

168. *Id.* art. 3(5).

169. *See* Max Eddy, *Your New Computer Has an AI Button. Now What?*, WIRECUTTER (May 28, 2024), https://www.nytimes.com/wirecutter/reviews/microsoft-copilot-pcs-explained; *Copilot & AI Agents*, MICROSOFT, https://www.microsoft.com/en-us/microsoft-copilot/copilot-101/copilot-ai-agents [https://perma.cc/QV2Z-BKKX].

classification fails to account for how GenAI's risks evolve or how these tools' broad accessibility facilitates seamless integration into professional workflows. While GenAI systems may trigger high-risk obligations in specific contexts, the framework—designed to classify and regulate discrete uses—fails to account for how widely deployed general-purpose AI systemically amplifies errors across professional domains. The EU's framework extends beyond the AI Act to include liability directives that suffer from similar limitations in addressing gradual, systemic contamination.[170]

The U.S. takes a fundamentally different approach from the EU's comprehensive framework, which threatens to become the global standard through regulatory spillover effects.[171] Rather than following the EU model, the U.S. has opted for fragmented, sector-specific regulations and voluntary guidelines.[172] Federal initiatives have sprawled across multiple domains—

---

170. The revised Product Liability Directive (PLD) addresses material harms but excludes immaterial harms like privacy violations, discrimination, and misinformation (along with evidentiary requirements). Directive 2024/2853, of The European Parliament and of The Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC, 2024 O.J. (L) art. 4(6), art. 9, recital 24. The proposed Artificial Intelligence Liability Directive (AILD) introduces fault-based liability with rebuttable presumptions and evidence disclosure provisions, covering harm to life, physical integrity, property, and fundamental rights, but still requires traditional causation frameworks. *Proposal for a Directive of the European Parliament and of the Council on Adapting Non-contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)*, arts. 2(6), 3, 4(1)(a), 4(2)–(3), Recital 7, COM (2022) 496 final (Sept. 28, 2022). Both frameworks struggle with professional practice impacts that emerge gradually through complex interactions rather than direct, traceable harms. For more detailed analysis, see Wachter, *supra* note 66, at 703–13.

171. *See* Wachter, *supra* note 66, at 676 (making this argument); ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD 25–66 (2019) (explaining the Brussels Effect).

172. The same patchwork characterizes U.S. liability doctrine. Commentators have shown that ordinary negligence and products-liability rules map poorly onto situations in which an autonomous system, rather than a human actor, generates the harmful content. *See* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 538–45 (2015) (explaining why fault concepts falter when "the robot acted on its own"); Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 393–400 (2016) (surveying doctrinal gaps and urging risk-based governance). Whether the principal federal safe-harbour—47 U.S.C. § 230—also shields a generative-AI provider whose model hallucinates defamatory text is unresolved. *Compare* Complaint at ¶¶ 44–55, Walters v. OpenAI L.L.C., No. 23-A-04860-2, 2024 WL 1366490 (11th Cir. Apr. 1, 2024) (alleging ChatGPT invented embezzlement charges and arguing OpenAI "created or developed" the content, placing it outside § 230), *with* Louis Shaheen, *Section 230's Immunity for Generative Artificial Intelligence*, 15 SEATTLE J. TECH., ENV'T & INNOVATION L. 1 (2024) (contending § 230 should continue to protect LLM vendors), *and* Peter Henderson et al., *Where's the Liability in Harmful AI Speech?*, 3 J. FREE SPEECH L. 589, 624 (2023) (arguing

principles for an "automated society,"[173] Executive Orders on AI,[174] federal AI legislation,[175] sector-specific regulations,[176] and national standards.[177]

---

providers lose immunity when models "materially contribute" to falsehoods). *See also* Noor Waheed, *Section 230 and Its Applicability to Generative AI: A Legal Analysis*, CTR. FOR DEMOCRACY & TECH. (Sept. 4, 2024), https://cdt.org/insights/section-230-and-its-applicability-to-generative-ai-a-legal-analysis/ [https://perma.cc/B5XY-ESFF] (urging legislative clarification). No court has yet squarely decided whether a large-language model "creates or develops" content within the meaning of § 230(f)(3).

173. Under a "bill of rights for an automated society." *See Join the Effort to Create A Bill of Rights for an Automated Society*, THE WHITE HOUSE (Nov. 10, 2021), https://bidenwhitehouse.archives.gov/ostp/news-updates/2021/11/10/join-the-effort-to-create-a-bill-of-rights-for-an-automated-society/ [https://perma.cc/J2DE-FWGX]; *Blueprint for an AI Bill of Rights: Making Automated Systems Work*, WHITE HOUSE OFF. OF SCI. & TECH. POL'Y (Oct. 2022), https://www.govinfo.gov/content/pkg/GOVPUB-PREX23-PURL-gpo193638/pdf/GOVPUB-PREX23-PURL-gpo193638.pdf [https://perma.cc/T8G7-8734].

174. In 2019, President Trump issued an executive order to advance U.S. AI proficiency, followed by another promoting its ethical and reliable use in federal operations. *See* Exec. Order No. 13,859, 84 Fed. Reg. 3967 (Feb. 14, 2019); Exec. Order No. 13,960, 85 Fed. Reg. 78939 (Dec. 8, 2020); Ben Winters, *Two Key AI Transparency Measures from Executive Orders Remain Largely Unfulfilled Past Deadlines*, EPIC (Jan. 26, 2022), https://epic.org/unfulfilled-ai-executive-orders [https://perma.cc/32MN-B2CG]. *See also* Exec. Order No. 14,277, 90 Fed. Reg. 17519 (Apr. 28, 2025) (establishing a federal task force to promote AI literacy, expand K–12 and workforce AI education, and encourage public-private partnerships to integrate AI into teaching and training programs).

175. *See* TAKE IT DOWN Act, S. 146, 119th Cong. (2025) (prohibiting non-consensual intimate images, including AI-generated "deepfakes," and requiring platforms to remove flagged content within 48 hours). In 2020, Congress enacted the National AI Initiative Act, establishing an overarching framework to enhance and coordinate AI research, development, demonstration, and education across all U.S. departments and agencies. The Act took effect on January 1, 2021. *See* The National Artificial Intelligence Initiative Act, 15 U.S.C. §§ 9401–9462. Meanwhile, Congress considered various AI-related legislative proposals, though many did not result in enacted law. *See, e.g.*, AI in Government Act of 2020, H.R. 2575, 116th Cong. (2020); Algorithmic Fairness Act of 2020, S. 5052, 116th Cong. (2020); GOOD AI Act of 2021, S. 3035, 117th Cong. (2021); Algorithmic Accountability Act of 2022, H.R. 6580, 117th Cong. (2022).

176. Agencies are often overseeing AI applications within their respective domains. For instance, the Food and Drug Administration (FDA) regulates AI when it is incorporated into medical devices, ensuring their safety and effectiveness. Similarly, the Federal Trade Commission (FTC) addresses AI-related consumer protection issues, such as prohibiting fake and AI-generated consumer reviews. *See FDA Calls for a Coordinated Response to Regulating AI*, AM. PSYCH. ASS'N SERVS. (Nov. 12, 2024), https://www.apaservices.org/practice/business/technology/on-the-horizon/ai-regulation [https://perma.cc/G88A-RRB4]; *Federal Trade Commission Announces Final Rule Banning Fake Reviews and Testimonials*, FED. TRADE COMM'N (Jan. 30, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/06/federal-trade-commission-announces-proposed-rule-banning-fake-reviews-testimonials [https://perma.cc/M6L3-BDRJ].

177. *See, e.g.*, in the context of AI bias, the National Institute of Standards and Technology (NIST) plays an active role in developing methods to mitigate AI-induced bias. This includes its

Executive Order 14,110 represented the most ambitious effort,[178] mandating safety standards and requiring AI developers to share test results with federal agencies.[179] Trump's 2025 revocation of this order eliminated even this limited framework, leaving U.S. AI governance in regulatory limbo.[180]

Even if Executive Order 14110 had survived, it wouldn't have addressed the Invisible Ripple Effect. The order targeted discrete AI failures and compliance measures rather than the long-term propagation of GenAI-generated errors across professional domains.[181] Embedded misinformation, flawed professional outputs, and cumulative AI-driven distortions remain largely unaddressed in both U.S. and international frameworks. The order at least provided regulatory momentum toward AI safety—momentum now eliminated by its revocation.

States have stepped into the federal void. However, although almost every state has now proposed or enacted AI legislation, the resulting framework is patchy and narrow, with most statutes focusing on privacy notices, bias audits, or other consumer-facing remedies.[182] These do little to advance the fight against the Invisible Ripple. Without cohesive federal leadership, AI governance remains fragmented across jurisdictions,

---

voluntary AI Risk Management Framework, which guides reducing the negative impacts of bias in AI systems. *See Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NAT'L INST. OF STANDARDS & TECH. (Jan. 26, 2023), https://www.nist.gov/itl/ai-risk-management-framework [https://perma.cc/X4X8-G3QK]; William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116-283, § 5301, 134 Stat. 3388, 4536–39 (2021).

178. *See* Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Nov. 1, 2023).

179. *Id.*

180. Replacing it with Executive Order on Removing Barriers to American Leadership in Artificial Intelligence. *See* David Shepardson, *Trump Revokes Biden Executive Order on Addressing AI Risks*, REUTERS (Jan. 21, 2025), https://www.reuters.com/technology/artificial-intelligence/trump-revokes-biden-executive-order-addressing-ai-risks-2025-01-21 [https://perma.cc/GGS8-C6HL]; Matt O'Brien & Sarah Parvini, *Trump Signs Executive Order on Developing Artificial Intelligence 'Free from Ideological Bias'*, ASSOCIATED PRESS (Jan. 23, 2025), https://apnews.com/article/trump-ai-artificial-intelligence-executive-order-eef1e5b9bec861eaf9b36217d547929c [https://perma.cc/BC3P-L62V].

181. Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Nov. 1, 2023).

182. *See, e.g.*, COLO. REV. STAT. ANN. §§ 6-1-1701–1707 (West 2024) (Colorado Artificial Intelligence Act) (imposing "reasonable-care" paperwork on high-risk AI developers and deployers but providing no mechanism to tag or trace GenAI outputs once released); CAL. GOV'T CODE §§ 11549.63–11549.66 (West 2025) (Generative Artificial Intelligence Tools) (mandating provenance-tracking pilots for state agencies yet leaving private-sector GenAI workflows untouched); 2023 Conn. Acts 23-16 (Reg. Sess.) (directing state offices to inventory and assess AI systems but remaining silent on professional reliance beyond government contracting); 820 ILL. COMP. STAT. 42/1–42/20 (2020) (Artificial Intelligence Video Interview Act) (requiring employer disclosure and applicant consent for AI-assisted hiring videos).

resulting in regulatory disparities that overlook the impact of GenAI on professional decision-making and institutional practices.[183]

Professional oversight mechanisms provide a second layer of governance, but one designed for a different era. Lawyers, architects, and a few other professionals who use AI tools remain subject to traditional accountability structures, including professional codes, licensing requirements, malpractice liability, and tort law.[184] These longstanding mechanisms operate as parallel regulatory systems, enforcing accountability for AI-assisted work even in the absence of comprehensive oversight of AI-specific regulations. Yet their interaction with emerging AI regulations creates a patchwork approach that struggles to address the systemic risks posed by the widespread adoption of GenAI in professional practice.

Take the legal field, for example. Legal ethics rules require lawyers to maintain competence, confidentiality, and diligence when using AI tools.[185] Federal Rule 11, which prohibits frivolous or unsupported filings, has already been invoked in multiple cases involving AI-generated errors.[186] However, the effectiveness of Rule 11 against AI errors proves inconsistent

---

183. *See, e.g.*, California Consumer Privacy Act of 2018 (CCPA), CAL. CIV. CODE §§ 1798.100–1798.199.100 (West 2023) (extending consumer rights to AI-driven data processing but not directly regulating professional reliance on GenAI); N.Y.C. ADMIN CODE §§ 20-870–874 (2023) (requiring bias audits for automated hiring tools but not addressing GenAI's impact on professional decision-making).

184. *See* Avery et al., *supra* note 63, at 121–28 (discussing lawyers' duties).

185. *See, e.g.*, ABA House of Delegates, Resolution 112 (2019), https://www.americanbar.org/content/dam/aba/directories/policy/annual-2019/112-annual-2019.pdf [https://perma.cc/8UJR-77NX] ("RESOLVED, That the American Bar Association urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence ("AI") in the practice of law including: (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI."); MODEL RULES OF PRO. CONDUCT r. 1.1 cmt. 8 (AM. BAR ASS'N 2023) (requiring lawyers to maintain technological competence as part of their professional obligations); Cyphert, *supra* note 35, at 423–43 (discussing ethical considerations for lawyers using GenAI).

186. By signing any "pleading, written motion, or other paper," an attorney certifies that the filing's legal theories are non-frivolous and that its factual assertions "have evidentiary support." FED. R. CIV. P. 11(a)–(b). If that certification is false, the court may impose "appropriate sanctions," including monetary penalties payable to the court or the opposing party. FED. R. CIV. P. 11(c)(4) (sanction must be "limited to what suffices to deter repetition of the conduct"). While symbolic fines remain the norm, with most Rule 11 sanctions limited to $1 or modest amounts, a handful of high-profile cases elsewhere—particularly those involving repeated or egregious AI misuse—have pushed total sanctions close to $143,519. These include both court-imposed fines and fee-shifting orders requiring attorneys to pay opposing counsel's legal costs. *See* Charlotin, *supra* note 2.

and context-dependent.[187] While courts successfully sanction attorneys who lie about or double down on AI-fabricated content—as in the high-profile *Mata v. Avianca* case—pure negligent reliance on AI hallucinations often escapes meaningful consequences due to the rule's safe harbor provisions and heightened standards requiring bad faith conduct.[188] This creates an enforcement gap where sophisticated AI fabrications that attorneys reasonably mistake for legitimate content may avoid traditional sanctioning mechanisms, forcing courts to resort to ad hoc standing orders with their regulatory complications.[189]

Some jurisdictions have responded with specific guidance.[190] Illinois, for example, permits the use of AI in court proceedings while requiring human oversight and professional accountability.[191] Yet a fifty-state survey reveals

---

187. *See* Jessica R. Gunder, *Rule 11 Is No Match for Generative AI*, 27 STAN. TECH. L. REV. 308, 327–48 (2024) (analyzing how Rule 11's safe harbor provisions and heightened standards for sua sponte sanctions create enforcement gaps for negligent AI misuse).

188. *See* 678 F. Supp. 3d 443, 447–48 (S.D.N.Y. 2023); Gunder, *supra* note 187, at 317–19, 341–43 (contrasting successful sanctions in Mata v. Avianca, where attorneys lied about AI use, with failed sanctions in Cohen case involving pure negligent reliance).

189. Gunder, *supra* note 187, at 354–60 (documenting how Rule 11's inadequacy has forced courts to resort to inconsistent standing orders requiring AI disclosure, creating new regulatory problems).

190. The American Bar Association (ABA) Standing Committee on Ethics and Professional Responsibility issued Formal Opinion 512 (2024) to provide ethical guidance on lawyers' use of GenAI. The opinion clarifies that while GenAI can enhance efficiency in legal practice, lawyers remain fully responsible for their work product. Ethical duties implicated by GenAI use include competence (Model Rule 1.1), confidentiality (Rule 1.6), communication with clients (Rule 1.4), candor toward the tribunal (Rule 3.3), and reasonable fees (Rule 1.5). The opinion emphasizes that attorneys must independently verify GenAI-generated content, protect client confidentiality when use AI tools, and ensure they are not misrepresenting information to courts or third parties. *See* ABA Standing Comm. on Ethics & Pro. Resp., Formal Op. 512 (2024), https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf [https://perma.cc/L3TZ-WBQ2]; *See also* STATE BAR OF CAL., PRACTICAL GUIDANCE FOR THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE IN THE PRACTICE OF LAW (2023), https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical-Guidance.pdf [https://perma.cc/FM79-AUTK] (providing ethical guidelines for attorneys on the competent and responsible use of GenAI tools in legal practice). Judges also instruct lawyers. For instance, in May 2023, Judge Brantley Starr of the Northern District of Texas issued a standing order requiring attorneys to certify that filings were not drafted using GenAI or were independently verified for accuracy. Citing concerns over AI "hallucinations," he emphasized that legal briefing requires reliability beyond GenAI's capabilities. *See* Lyle Moran, *Federal Judge Seeks to Prevent Generative AI Mistakes in Briefs*, LEGAL DIVE (June 1, 2023), https://www.legaldive.com/news/generative-ai-hallucinations-federal-judge-order-on-ai-brantley-starr/651817 [https://perma.cc/3PP7-LK7B].

191. This policy allows ethical and legal AI use while maintaining individual responsibility, though courts may reject AI-generated submissions that are unfounded, intentionally

the inadequacy of this response: only a quarter of U.S. jurisdictions have issued formal ethics opinions on GenAI, while another eight percent have adopted model policies, leaving two-thirds of the country either studying the issue or remaining silent.[192]

Other professions face similar challenges. Architectural practice remains subject to licensing requirements that prioritize public safety, while professional organizations scramble to develop guidelines for AI integration.[193] Healthcare relies on established standards of care and malpractice frameworks, requiring practitioners to exercise professional judgment when integrating AI-generated insights into patient care.[194] However, many sectors using AI tools operate without established professional licensing or accountability mechanisms at all, as demonstrated by the White House's MAHA report's phantom citations.[195] In these cases, professional oversight mechanisms do not exist, and thus cannot help reduce the possibility of detecting GenAI errors. Thus, these reactive,

---

misleading, or perpetuate biases. *See Illinois Supreme Court Announces Policy on Artificial Intelligence*, ILL. CTS., https://www.illinoiscourts.gov/News/1485/Illinois-Supreme-Court-Announces-Policy-on-Artificial-Intelligence/news-detail [https://perma.cc/6TRD-NRL3]. Some courts, such as the Northern District of California and the Western District of North Carolina, require lawyers to certify whether they used GenAI in legal filings and verify the accuracy of AI-generated work. However, other jurisdictions, such as the Fifth Circuit, have declined to adopt such measures, reflecting an inconsistent and profession-specific approach to AI governance. *See* Sarah Martinson, *Law Scholars Hope 5th Circuit Decision Deters More AI Rules*, LAW360 PULSE (June 14, 2024), https://www.law360.com/articles/law-scholars-hope-5th-circuit-decision-deters-more-ai-rules [https://perma.cc/UXL6-VWVB]. For a full (updated) list of federal court decisions and standing orders related to AI use in legal filings, see *Tracking Federal Judge Orders on Artificial Intelligence*, LAW360 PULSE, https://www.law360.com/pulse/ai-tracker [https://perma.cc/K4J4-7QBT].

192. *See AI and Attorney Ethics Rules: 50-State Survey*, JUSTIA, https://www.justia.com/trials-litigation/ai-and-attorney-ethics-rules-50-state-survey [https://perma.cc/L7Z2-38WS] (cataloging all published opinions, guidelines, task-force reports, and states with no action). The author tallied jurisdictional status from that survey on 27 June 2025; data on file with author.

193. *See, e.g.*, NAT'L COUNCIL OF ARCHITECTURAL REGISTRATION BDS., MODEL RULES OF CONDUCT (2023), https://www.ncarb.org/sites/default/files/Rules_of_Conduct.pdf [https://perma.cc/CP77-W4U5] (establishing ethical and competency standards for architects to protect public health, safety, and welfare); 48 C.F.R. § 36.6 (2024) (governing the professional responsibility of architects and engineers in federal contracts, emphasizing technical accuracy and liability for design errors).

194. For more on the standard of care in medical malpractice litigation, see generally Daniel W. Shuman, *The Standard of Care in Medical Malpractice Claims, Clinical Practice Guidelines, and Managed Care: Towards a Therapeutic Harmony?*, 34 CAL. W. L. REV. 99 (1997).

195. *See* Weber & Gilbert, *supra* note 1.

profession-specific approaches are currently insufficient to address the contamination risks posed by the Invisible Ripple Effect.

Civil and criminal liability could theoretically backstop against AI-generated harms.[196] These frameworks should incentivize careful verification by letting victims sue for AI errors.[197] For professions without strong oversight, tort law becomes the main deterrent against sloppy AI use.[198] But liability fails against GenAI's distinctive risks. AI systems are too complex for typical users to understand—hallucinations and prompt injection attacks exceed most professionals' expertise. Competitive pressures push for rapid adoption despite known dangers, gutting the deterrent effect of liability.[199] Lawyers might not know if malpractice insurance covers AI mistakes. Doctors might not be able to tell when AI recommendations will trigger lawsuits. These uncertainties make liability frameworks useless precisely when they're needed most. And perhaps most importantly, the timing problem is fatal. GenAI embeds itself in professional practices faster than liability mechanisms can respond. By the time someone gets sued, the contamination has already spread. Reactive frameworks can't stop the Invisible Ripple Effect—only comprehensive regulatory intervention can prevent it.

B. Why Legal Intervention Is Inevitable

GenAI errors aren't temporary implementation problems that better engineering can solve—they're baked into the technology's DNA.[200]

---

196. For more on civil liability, see generally ERNEST WEINRIB, THE IDEA OF PRIVATE LAW (1995). For more on criminal liability, see ANDREW ASHWORTH, SENTENCING AND CRIMINAL JUSTICE 77–95 (5th ed. 2010). For a general discussion on AI and liability, see generally RYAN ABBOTT, THE REASONABLE ROBOT: ARTIFICIAL INTELLIGENCE AND THE LAW (2020).

197. WEINRIB, *supra* note 196; ASHWORTH, *supra* note 196, at 77–95; ABBOTT, *supra* note 196.

198. Anat Lior, *Addressing AI-Related Harms Through Existing Tort Doctrines*, U. OXFORD BUS. L. BLOG (July 3, 2024), https://blogs.law.ox.ac.uk/oblb/blog-post/2024/07/addressing-ai-related-harms-through-existing-tort-doctrines [https://perma.cc/6ZEW-LJDB].

199. *See generally* Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315 (2020); Notably, even broader regulatory solutions may face legal challenges, as recent Supreme Court jurisprudence—particularly the 'Major Questions Doctrine'—has placed new constraints on administrative agencies' ability to regulate emerging technologies without clear congressional authorization. *See* Blair Levin & Tom Wheeler, *The Supreme Court's Major Questions Doctrine and AI Regulation*, BROOKINGS (Sept. 6, 2023), https://www.brookings.edu/articles/the-supreme-courts-major-questions-doctrine-and-ai-regulation [https://perma.cc/JWB9-5W7C].

200. *See supra* Part I.

Hallucinations, omissions, and factual distortions stem directly from GenAI's probabilistic architecture, which predicts the next most likely token rather than retrieving verified information.[201] The system doesn't "know" anything in the traditional sense; it generates plausible-sounding text based on statistical patterns in training data. This fundamental design currently makes errors inevitable, not accidental.[202]

While companies invest heavily in mitigation through red-teaming and stress-testing, these measures address individual model outputs, not the systemic propagation that defines the Invisible Ripple Effect.[203] A model might pass safety tests in controlled environments yet still generate fabricated citations that lawyers unknowingly embed into briefs, or produce flawed research that physicians incorporate into treatment decisions. The real danger isn't what happens in the lab—it's what happens when millions of professionals integrate these probabilistic outputs into knowledge systems that assume human verification and institutional oversight.

This persistence necessitates legal intervention. Lessig's insight that "code is law" recognizes how technology itself regulates behavior, while markets and social norms provide additional governance mechanisms.[204] However, these traditional regulatory forces prove insufficient in addressing the systemic risks of GenAI. Despite widespread awareness of GenAI's fallibility—evidenced by disclaimers—users routinely fail to exercise appropriate caution.[205] Professionals employing GenAI for critical tasks incorrectly assume existing safeguards will catch errors, a dangerous misconception that enables the Invisible Ripple Effect.

While essential factors, GenAI's challenges cannot be solved solely through education or social norms. Market pressures may encourage output verification, but their slow evolution and inherent limitations make them

---

201. *See supra* Part I.

202. *See supra* Part I.

203. Red-teaming, originally developed for military and cybersecurity applications, employs adversarial testing to expose system vulnerabilities. In GenAI development, this methodology has evolved into a critical tool for identifying and addressing potential safety, security, and reliability concerns within generative models. *See* Michael Feffer et al., *Red-Teaming for Generative AI: Silver Bullet or Security Theater?*, *in* ADVANCED ENTERPRISE INFORMATION SYSTEMS 2024 CONFERENCE PROCEEDINGS 421, 421–24 (2024), https://ojs.aaai.org/index.php/AIES/article/view/31647/33814 [https://perma.cc/P3H5-PVHM].

204. Lessig identified four mechanisms that regulate behavior: market forces, social norms, technology ("code"), and law. *See* LAWRENCE LESSIG, CODE: VERSION 2.0 1–7, 120–37 (2006).

205. *See* Natasha Lomas, *ChatGPT Hit with Privacy Complaint over Defamatory Hallucinations*, TECHCRUNCH (Mar. 19, 2025), https://techcrunch.com/2025/03/19/chatgpt-hit-with-privacy-complaint-over-defamatory-hallucinations [https://perma.cc/56HZ-8QEL].

insufficient safeguards against systemic contamination.[206] As mentioned, even technically proficient users struggle to identify subtle hallucinations in AI-generated content, while users with limited AI literacy face insurmountable detection challenges.[207] The aforementioned AI FOMO (or simply the allure of using GenAI) exacerbates these problems, driving rapid adoption despite acknowledged risks and undermining any cautious norms that might otherwise develop.

Market forces have begun addressing some GenAI challenges, but they remain insufficient to mitigate systemic risks. Well-resourced organizations might implement robust verification protocols, but these safeguards remain inaccessible to independent practitioners and resource-constrained environments. This creates a two-tiered system where sophisticated users may have some protection while the majority remain vulnerable to contamination. Competitive forces driving accuracy improvements cannot address the fundamental problem of systemic knowledge corruption.[208]

Market responses, such as Retrieval-Augmented Generation (RAG), show promise but face fundamental limitations.[209] RAG enables legal assistants to verify case citations against databases and architectural tools that reference building codes; however, current implementations still

---

206. *See* Mike Wheatley, *OpenAI's CriticGPT Uses Generative AI to Spot Errors in Generative AI's Outputs*, SILICONANGLE (June 27, 2024), https://siliconangle.com/2024/06/27/openais-criticgpt-uses-generative-ai-spot-errors-generative-ais-outputs [https://perma.cc/P6L2-CJMD] (describing OpenAI's development of CriticGPT, a tool designed to assist human reviewers in detecting errors and hallucinations in AI-generated outputs).

207. *Cf.* Esther S. Kox & Beatrice Beretta, *Evaluating Generative AI Incidents: An Exploratory Vignette Study on the Role of Trust, Attitude and AI Literacy*, *in* HHAI 2024: HYBRID HUMAN AI SYSTEMS FOR THE SOCIAL GOOD 188, 188–94 (F. Lorig et al. eds., 2024), https://doi.org/10.3233/FAIA240194 [https://perma.cc/7WU7-WWZ2] (finding that while AI literacy correlates with trust and usage, it does not significantly affect evaluations of AI-generated incidents).

208. *See* Varun Magesh et al., *AI on Trial: Legal Models Hallucinate in 1 Out of 6 (or More) Benchmarking Queries*, STANFORD HAI (May 23, 2024), https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries [https://perma.cc/R38U-AK2B].

209. RAG systems combine two distinct mechanisms: a retrieval component that searches and ranks relevant documents from a curated knowledge base, and a generation component that incorporates this retrieved information into its responses. Unlike standard GenAI models that rely solely on pre-trained knowledge, RAG dynamically accesses and verifies information against current, domain-specific sources. For example, in legal applications, a RAG system first retrieves relevant case law and statutes from a legal database, then generates responses that are explicitly grounded in these sources. This architecture helps reduce hallucinations by ensuring responses are tied to verifiable documents rather than purely statistical patterns. *See id.*

produce errors.[210] To this day, RAG represents the most promising technological approach to addressing GenAI's reliability problems. By grounding AI responses in verified databases and real-time information retrieval, RAG systems can significantly reduce hallucinations and improve factual accuracy. Legal applications, such as CaseText and Westlaw's AI tools, demonstrate RAG's potential, enabling lawyers to query case law with greater confidence that citations will reference actual precedents.[211] Medical applications similarly show promise, with RAG-enabled systems that can cross-reference treatment recommendations against current medical literature.[212]

But RAG's current limitations reveal why technological solutions alone cannot solve the Invisible Ripple Effect. RAG systems remain vulnerable to retrieval errors—they can cite the wrong document or misinterpret query intent.[213] More fundamentally, RAG assumes that authoritative sources themselves stay uncontaminated.[214] As AI-generated content increasingly populates the very databases that RAG systems query, these tools risk amplifying rather than preventing contamination. A RAG system querying legal databases that already contain AI-fabricated precedents will confidently cite those fabrications as authoritative sources. The contamination problem thus becomes self-reinforcing: today's AI errors become tomorrow's training data and retrieval sources.

Legal intervention currently seems inevitable. Technology, markets, and social norms each play essential roles in managing AI risks, but they cannot address the systemic propagation of errors that defines the Invisible Ripple Effect. Liability frameworks and institutional protocols provide essential safeguards, but effective mitigation requires coordinated regulatory intervention that integrates legal measures with technological safeguards, market incentives, and professional norms. Without structured oversight, GenAI-generated errors will continue to entrench themselves within

---

210. *Id.*

211. *Thomson Reuters: GenAI Tool Tested by Stanford 'Leverages Innovation in Casetext'– Updated*, ARTIFICIAL LAW., (June 14, 2024), https://www.artificiallawyer.com/2024/06/14/thomson-reuters-genai-tool-tested-by-stanford-did-leverage-casetext/ [https://perma.cc/CSQ4-B945].

212. Omid Kohandel Gargari & Gholamreza Habibi, *Enhancing Medical AI with Retrieval-Augmented Generation: A Mini Narrative Review*, DIGIT. HEALTH (Apr. 21, 2025), https://pmc.ncbi.nlm.nih.gov/articles/PMC12059965/ [https://perma.cc/5EDM-NEDJ].

213. *See* Magesh, *supra* note 208.

214. *See id.*

professional and institutional systems, creating irreversibly systemic risks that threaten the integrity of the knowledge infrastructure itself.

III. BREAKING THE WAVES

This Part proposes a five-stage framework for addressing the Invisible Ripple Effect.[215] These recommendations are necessarily generalized and non-exhaustive. As this technology continues evolving, some problems may be mitigated through development, requiring adaptive responses rather than static solutions. While the primary contribution of this Article lies in identifying and theorizing the Invisible Ripple Effect, these practical suggestions offer a starting framework for intervention. The five stages are not necessarily equal in importance or resource allocation; circumstances may require greater emphasis on one stage over others, though investment across all stages remains crucial for comprehensive protection.

In other words, this framework must be understood within its fundamental limitations. Much of what we observe represents the growing pains of disruptive adaptation to transformative technology, but certain vulnerabilities stem from GenAI's core architecture and cannot be eliminated through better engineering alone.[216] The probabilistic nature of GenAI makes certain errors inevitable rather than accidental, while the scale of individual adoption creates coordination challenges that traditional regulatory frameworks struggle to address.

With time and coordinated effort, the following interventions can significantly reduce cascading harms: *prevention*—stopping errors at their

---

215. This framework resembles, in part, the structured approach used in cyber incident response, where mitigation strategies acknowledge that full prevention is unattainable, necessitating layered interventions to contain and adapt to evolving threats. *See* NAT'L INST. STANDARDS & TECH., COMPUTER SECURITY INCIDENT HANDLING GUIDE, SP 800-61 REV. 2 (2012), https://csrc.nist.gov/pubs/sp/800/61/r2/final [https://perma.cc/KQ5A-C4WR] (outlining a multi-stage incident response model, including preparation, detection, containment, mitigation, and post-incident adaptation); *see also* NAT'L INST. STANDARDS & TECH., INCIDENT RESPONSE RECOMMENDATIONS AND CONSIDERATIONS FOR CYBERSECURITY RISK MANAGEMENT: A CSF 2.0 COMMUNITY PROFILE, SP 800-61 REV. 3, 3–9 (2025), https://csrc.nist.gov/pubs/sp/800/61/r3/final [https://perma.cc/4LPL-QFBF] (outlining a reorganized incident response model, including preparation (govern, identify, and protect), detection, response, and recovery, with continuous improvement integrated throughout). In this chapter, the analysis relies on the earlier "classic" five-stage model (Rev. 2), which remains a widely used conceptual framework (the 2025 revision is cited here as the current version, but it primarily reflects a reorganization of the stages without substantial conceptual changes).

216. *See* Bengio et al., *supra* note 146, at 842–45 (explaining how current AI architectures make certain failure modes inevitable).

source; *containment*—creating professional safeguards to catch errors early; *control*—establishing oversight mechanisms to manage flawed content; *mitigation*—implementing systems to correct errors once detected; and *resilience*—strengthening systems to adapt to future challenges. These stages must work together as an integrated system, each reinforcing the others to create robust, adaptive safeguards against systemic knowledge contamination.

Yet this framework operates within severe structural constraints. Technological advancement consistently outpaces governance development, creating a moving target where interventions designed for current AI capabilities may become obsolete as new technologies emerge. Resource limitations prevent comprehensive human verification across all professional contexts, while market incentives often prioritize competitive advantage over safety measures. Most critically, the decentralized nature of consumer AI adoption means that individual professionals using GenAI tools, such as ChatGPT, operate largely outside traditional oversight mechanisms, creating governance gaps that no single regulatory approach can fully address.

A. Prevention

Prevention is the first tier in a *five-layer* response. Its role is to stop errors *before* they enter professional workflows. It requires addressing the full spectrum of errors that drive the Invisible Ripple Effect before they reach users or embed in professional systems. This means targeting not just hallucinations but all forms of errors capable of shaping decision-making, whether arising from GenAI's probabilistic nature, user misuse, critical omissions, or inherited vulnerabilities like bias. Prevention serves as the first line of defense, strengthening AI design, refining training data, and embedding real-time accuracy checks within models themselves.

To mitigate hallucinations, intervention must occur during the earliest stages of model development, addressing both the quality of input data and the mechanisms that constrain output. Given the fundamental reliance of generative AI on probabilistic pattern matching, effective mitigation requires investing in robust dataset curation, retrieval-based grounding

systems (e.g., RAGs), and integrated fact-checking submodels to ensure the accuracy of the output.[217]

Some domain-specific tools show promise: LexisNexis Protégé validates legal citations,[218] Wolfram Alpha verifies mathematical calculations,[219] and specialized systems cross-reference professional knowledge bases.[220] As mentioned, certain general-purpose models incorporate RAG techniques to reduce hallucinations, including OpenAI's o3, and open-source deployments of LLaMA or Mistral models integrated with RAG

---

217. Recent research explores such verification models, demonstrating how AI oversight can extend beyond individual applications to scalable, systemic evaluation systems *See, e.g.*, Nan Tang et al., *VerifAI: Verified Generative AI*, *in* CIDR 2024: 14TH ANNUAL CONFERENCE ON INNOVATIVE DATA SYSTEMS RESEARCH (2023), https://www.cidrdb.org/cidr2024/papers/p5-tang.pdf [https://perma.cc/2W7A-C7QK] (proposing a framework that verifies GenAI outputs by cross-referencing generated data with multi-modal data lakes to improve accuracy and reliability).

218. *See* LexisNexis Protégé, LEXISNEXIS, https://www.lexisnexis.com/en-us/products/protege.page [https://perma.cc/F8SV-UWGC] (LexisNexis Protégé is an AI assistant for legal research, drafting, and workflow integration).

219. Wolfram Alpha is a computational knowledge engine that generates precise answers using curated data, symbolic computation, and algorithmic processing rather than probabilistic text generation. *See About Wolfram Alpha*, https://www.wolframalpha.com/about [https://perma.cc/VYR8-WLV3]. We are already seeing various tools implementing these approaches across different fields. For instance, in healthcare, platforms like *DoxGPT* assist in generating medical reports while integrating patient histories and clinical guidelines. *See, e.g.*, DOXIMITY, https://support.doximity.com/hc/en-us/articles/41850759500051-Doximity-GPT-FAQs [https://perma.cc/LQ95-JQ6C]. In financial services, tools like *Lucinity* use GenAI for anomaly detection in financial transactions, improving fraud detection while reducing false positives. *Meet Luci: Your Trusted AI Agent for Financial Crim Investigations*, LUCINITY, https://lucinity.com/luci?gad_source=1&gad_campaignid=23440566380&gbraid=0AAAABCiTf1o3xE3R7QTOTccwjT7imW5Uz&gclid=Cj0KCQiAprLLBhCMARIsAEDhdPdQmZkSodzGqSOzadeMx1JUM8fsT9h6Sg0m-7x62tkJFKfcyWiCtrMaAg2eEALw_wcB [https://perma.cc/UQ8T-UL4T]. Similarly, in education, AI-driven platforms personalize learning plans to align with curriculum standards. *See e.g.*, *The AI-First Learning Platform*, SANA, https://sanalabs.com/meet-sana-lms?utm_term=ai-powered%20personalized%20learning%20platform&utm_campaign=Learn_Exp_US_Search&utm_source=adwords&utm_medium=ppc&hsa_acc=4334994313&hsa_cam=20162751504&hsa_grp=161540331980&hsa_ad=695884111352&hsa_src=g&hsa_tgt=kwd-2298186352732&hsa_kw=ai-powered%20personalized%20learning%20platform&hsa_mt=p&hsa_net=adwords&hsa_ver=3&gad_source=1&gad_campaignid=20162751504&gbraid=0AAAADHe6YpqJx90nLE0nq074YsInZBcf&gclid=Cj0KCQiAprLLBhCMARIsAEDhdPdZvCW7zxWeCuu8mI9ZgmMUdg0K494etu5i9vJ3GJaWG9r13DBoi2saAofeEALw_wcB [https://perma.cc/E2AT-N5SD].

220. *See* Grant Currin, *How Generative AI Will Transform Health Care and Finance*, COLUM. UNIV. DATA SCI. INST. (Dec. 6, 2023), https://datascience.columbia.edu/news/2023/how-generative-ai-will-transform-health-care-and-finance [https://perma.cc/7JYV-9LWU].

frameworks like LangChain or LlamaIndex.[221] These technological advances, while imperfect, represent essential harm reduction—any meaningful reduction in error rates becomes magnified across the millions of outputs generated daily and their subsequent ripple effects through knowledge networks. However, coverage across professional domains remains inconsistent, and even pre-deployment testing and red-teaming can only mitigate risks rather than eliminate them entirely.[222]

Prevention also requires technical safeguards against adversarial attacks, particularly prompt injection and deliberate manipulation. Potential countermeasures include input sanitization systems that strip suspicious formatting or hidden text from documents, content verification protocols that cross-check retrieved information against authoritative sources, and prompt filtering mechanisms that detect and block instruction-hijacking attempts.[223] Some developers are experimenting with separate "instruction" and "content" channels—a privilege-separation design that hard-blocks user text from overriding system rules—to prevent embedded commands from overriding system prompts.[224] Major AI providers have started developing these detection methods, but the challenge remains formidable—attackers can continuously adapt their techniques, and current systems struggle to distinguish between legitimate instructions and malicious manipulation.[225]

Even when successful, these measures create significant trade-offs. Built-in filters that block unverifiable content must be carefully calibrated.

---

221. *See* Mohamed A. Ferrag et al., *Reasoning Beyond Limits: Advances and Open Problems for LLMs*, 11 ICT EXPRESS 1054 (2025), https://www.sciencedirect.com/science/article/pii/S240595952500133X (click "View PDF") (surveying hallucination-mitigation strategies and noting that general-purpose models such as o3, DeepSeek-R1, and open-source LLaMA variants implement retrieval-augmented generation (RAG) to improve factual grounding).

222. Take, as an example, CriticGPT. Designed to catch AI errors, it still generates hallucinations and false positives. *See* Nat McAleese et al., LLM Critics Help Catch LLM Bugs, (June 28, 2024) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2407.00215 [https://perma.cc/F3WJ-URNK]; Metz & Weise, *supra* note 15.

223. *See, e.g.*, OWASP Foundation, *LLM01:2025 Prompt Injection*, OWASP GENAI SECURITY PROJECT (2025), https://genai.owasp.org/llmrisk/llm01-prompt-injection/ [https://perma.cc/PJ3Z-M5HQ] (recommending "input and output filtering" to strip suspicious or hidden text, using the "RAG Triad" to verify responses against authoritative sources, and rule-based prompt filtering to detect and block instruction-hijacking).

224. *See, e.g.,* Sizhe Chen et al., *StruQ: Defending Against Prompt Injection with Structured Queries*, *in* PROCEEDINGS OF THE 34TH USENIX SECURITY SYMPOSIUM (2025), https://www.usenix.org/system/files/usenixsecurity25-chen-sizhe.pdf [https://perma.cc/96SG-BFGB] (proposing a two-channel "structured-query" interface that cleanly separates system instructions from user-supplied content so the model ignores embedded commands).

225. *See generally* Feffer et al., *supra* note 203 (arguing that red-teaming is not a panacea).

Too strict, and they constrain legitimate professional uses. For example, lawyers researching emerging legal theories might be blocked from accessing cutting-edge scholarship, or doctors exploring novel treatments could be prevented from reviewing preliminary research. Conversely, too lenient settings allow fabricated citations and false studies to slip through undetected. Prevention strategies must also address persistent vulnerabilities, such as bias and opacity, which have long been challenges confronting AI systems.[226] While advances in training data curation and algorithmic transparency show promise, these inherited problems prove particularly difficult to resolve at the massive scale of GenAI training datasets.

As GenAI adoption expands into more widespread use, prevention methods alone may prove insufficient without regulatory intervention. Market incentives often favor rapid deployment over comprehensive safety measures, potentially leaving critical vulnerabilities unaddressed. For example, companies rushing to integrate AI into their workflows may prioritize speed and cost-effectiveness over implementing robust verification systems, creating gaps that voluntary industry standards cannot adequately fill.

Thus, prevention alone cannot eliminate all errors. Hallucinations persist as an inherent byproduct of GenAI's probabilistic nature, and even robust verification struggles with evolving professional knowledge, ambiguous legal interpretations, and context-dependent decision-making. While responsible user practices can mitigate some risks, these measures alone remain insufficient to address the Invisible Ripple Effect.[227] In addition, some models integrate comprehensive safeguards, but others—particularly open-source systems like DeepSeek—may not, increasing the risk of unchecked error propagation. Moreover, prevention efforts focus primarily on the model rather than its use, reducing certain risks but failing to address how flawed outputs spread through professional networks. Real-time

---

226. *See, e.g.*, *supra* sources cited note 66.

227. For instance, lawyers must always ensure they maintain their clients' confidentiality and do not use any AI tools that could expose sensitive information, whether through data retention policies, third-party access, or inadequate security measures. *See* Bree'ara Murphy et al., *Ethical Algorithms: Navigating AI in Legal Practice for a Just Jurisprudence*, GA. STATE U. L. REV. BLOG (May 14, 2024), https://www.gsulawreview.org/blog/ethical-algorithms-navigating-ai-in-legal-practice-for-a-just-jurisprudence/ [https://perma.cc/7H3V-YPHR] ("[W]here an attorney enters confidential and sensitive information in a third-party AI program, attorneys must ensure the protection of client confidentiality is maintained and the attorney-client privilege remains sacrosanct.").

verification remains resource-intensive and constrained by existing databases, making it ineffective at detecting novel errors.[228] Hence, prevention, though essential, cannot by itself contain the Invisible Ripple.

B. Containment

When AI systems produce errors despite prevention efforts, containment decides whether those mistakes stay within the original output or spread further. The goal is simple: catch flawed content right after the AI generates it, before anyone acts on it, uses it in their work, or shares it with others. Think of containment as a checkpoint that intercepts errors before they can influence professional decisions, institutional policies, or public information. Unlike one-size-fits-all approaches, effective containment must be flexible, responding more aggressively to serious errors in high-stakes contexts while allowing lighter oversight for lower-risk situations.

Containment strategies must tackle different types of errors at various stages. Some problems, such as biases and systemic distortions, stem from how the model was built and trained, so they need to be addressed at the source. Others, such as hallucinations, false facts, and deliberate manipulations, can often be identified when users interact with the system, although they may still reflect deeper issues with the AI itself. The most common containment approach today, generic disclaimers that warn users about potential AI errors, provides little real protection. Research shows that people routinely ignore these warnings,[229] especially when the AI-generated content looks professional and convincing.[230] This mirrors the

---

228. *See, e.g.*, Yuan Sun & Jorge Ortiz, *Rapid Review of Generative AI in Smart Medical Applications*, 3 INT'L J. COMPUT. SCI. & INFO. TECH. (2024), https://wepub.org/index.php/IJCSIT/article/view/2227/2444 [https://perma.cc/C66W-GERR] (showing how difficult it is to verify GenAI outputs in medical applications).

229. Also known as information disclosure, akin to mandated warnings on pharmaceuticals, nutrition labels, or energy efficiency ratings, provides users with general awareness but rarely leads to meaningful intervention. *See, e.g.*, The Nutrition Labeling and Education Act (NLEA), Pub. L. No. 101-535, 104 Stat. 2353 (1990) (empowering the FDA to mandate nutrition labeling); 21 C.F.R. § 101 (2019) (regulating the display requirements for food and beverages). For a discussion of legislatively mandated disclosure requirements and their limitations, see Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 649–52, 665–90 (2011).

230. *See* Cristiano Lima-Strong, *AI Disclaimers in Political Ads Backfire on Candidates, Study Finds*, WASH. POST (Oct. 8, 2024), https://www.washingtonpost.com/politics/2024/10/08/ai-disclaimers-political-ads-backfire-candidates-study-finds (analyzing the unintended consequences of AI disclaimers in political advertising, finding that such disclosures may reduce voter trust in candidates rather than fostering transparency).

limited effectiveness of disclosure requirements in other fields, where mandated warnings often fail to change behavior despite legal compliance.[231]

More effective containment requires innovative interventions that actively push professionals to double-check AI outputs before using them. Instead of generic warnings, AI systems should recognize when they're being used for professional work and respond with specific, prominent alerts about relevant risks. For example, when a lawyer asks for case citations, the system should explicitly warn that cases may be fabricated and require verification before use. When an architect requests structural calculations, the system should emphasize that mathematical outputs may contain errors and must be independently checked before any professional application. The challenge lies in getting AI systems to recognize these high-stakes contexts reliably.[232]

The level of intervention should match the professional stakes involved—minor factual errors in general content might warrant a simple review prompt, while outputs for regulated industries or high-risk professions should trigger stronger safeguards and verification requirements. Beyond warning mechanisms, containment should use friction-based safeguards that require user verification before AI-generated content is used or shared. Just as financial systems add extra authentication steps for large transactions, high-risk AI outputs should trigger mandatory review processes.[233] For example, AI-generated professional advice, technical analyses, or medical recommendations should require users to explicitly confirm they've verified the content against reliable sources before exporting or applying it. By adjusting friction based on risk levels, these mechanisms ensure greater scrutiny where errors could cause serious

---

231. *See* Ben-Shahar & Schneider, *supra* note 229.

232. For instance, systems need to distinguish between casual legal questions and professional legal research, or between general engineering queries and calculations intended for actual construction projects. *See* Shari L. Klevens & Alanna Clair, *Lessons Learned from AI Company Disclaimers*, DENTONS (Aug. 7, 2023), https://www.dentons.com/en/insights/newsletters/2023/august/7/practice-tips-for-lawyers/lessons-learned-from-ai-company-disclaimers [https://perma.cc/V324-W7ZK] (suggesting solutions to AI disclaimers).

233. Research has shown that users exposed to targeted friction—such as highlighted prompts indicating potential errors—identified significantly more inaccuracies and omissions in AI-generated text compared to those who received unmarked outputs. *See* Renée Richardson Gosline et al., *Nudge Users to Catch Generative AI Errors*, ACCENTURE (May 29, 2024), https://www.accenture.com/au-en/insights/data-ai/nudge-users-catch-generative-ai-errors [https://perma.cc/P8D9-PSQM]. Notably, however, this research was not peer-reviewed and may reflect industry-driven perspectives.

professional, ethical, or legal harm while avoiding unnecessary obstacles for routine tasks.

Containment must also extend beyond individual users to include institutional safeguards that prevent unchecked reliance on AI-generated outputs. Professional licensing bodies already require practitioners to maintain accuracy—lawyers must correct errors in legal filings, doctors must update patient records when new information emerges, and financial professionals must revise disclosures based on changing risk assessments. These existing professional responsibility frameworks provide a natural foundation for AI-specific verification requirements.[234] Expanding these current obligations to cover AI-generated content, combined with clear liability frameworks explicitly, would strengthen professional accountability while keeping verification requirements practical and enforceable.

Liability frameworks should hold individual professionals accountable for verifying AI outputs before they are used professionally. While existing rules, such as Rule 11, provide some deterrent effect, they focus on punishing fabricated citations rather than preventing reliance on unverified AI content.[235] Enhanced frameworks should clarify that professionals remain fully responsible for AI-generated work products, regardless of the tool's sophistication, and that due diligence requires independent verification of AI outputs before they are applied professionally. These individual accountability measures become even more critical as AI tools evolve beyond simple query-response interactions.

Containment becomes especially critical as AI tools integrate seamlessly into professional workflows. AI-assisted platforms, such as Copilot, embed generated content directly into documents and decision-making processes, reducing visibility into when and how AI contributes to outputs. This lack of transparency increases the risk of unverified information slipping through unnoticed. The challenge intensifies with agentic AI, which operates persistently in the background, continuously refining drafts, conducting analyses, and providing decision support without clear distinctions between AI-generated and human-authored content. As AI-generated material becomes increasingly indistinguishable from human input, professionals may unknowingly rely on flawed outputs, making it more challenging to trace errors back to their source or recognize when an AI-driven suggestion requires further verification. Without containment measures that explicitly

234. *See supra* Section II.A.

235. Gunder, *supra* note 187, at 355.

flag AI contributions and prompt users to validate critical content, these tools risk embedding inaccuracies into professional workflows undetected.

However, containment strategies cannot assume that professionals will always verify AI-generated outputs. Instead, real-time validation mechanisms, automated accuracy checks, and structured barriers must be embedded directly into AI workflows to prevent flawed content from being acted upon without review. Institutional safeguards—such as tracking which documents contain AI-generated content, maintaining records of verification steps taken, and requiring supervisor review of AI-assisted work—can provide an added layer of oversight, helping professionals assess the reliability of AI-generated material before integrating it into their work. By embedding these safeguards directly into AI systems and professional workflows, containment measures reduce the likelihood of unnoticed errors shaping decision-making processes.

Containment measures suffer from inconsistent implementation across the AI ecosystem. Some providers embed robust safeguards, while others—particularly those using open-source or minimally regulated models—offer no automated flagging or verification at all. Even when safeguards exist, uneven application and time pressures often lead professionals to bypass verification steps.[236] Additionally, some professional errors, such as misinterpretations of complex data or ethical misjudgments, will evade AI-driven containment measures entirely. While these mechanisms can catch factual inaccuracies, they prove less effective at identifying mistakes rooted in professional reasoning, such as legal misinterpretations or ethically flawed decisions assisted by AI. Because these challenges cannot be eliminated, containment must operate in conjunction with control, mitigation, and resilience to prevent minor AI errors from compounding into systemic distortions.

C. Control

At this stage, AI-generated errors have already bypassed initial scrutiny and entered professional workflows. Unlike prevention or containment, which aim to catch problems early, control mechanisms focus on ensuring these mistakes do not become institutionalized as authoritative

---

236. *See* Gosline et al., *supra* note 233. Even when inaccuracies and omissions were highlighted for users, only 54% of omissions were identified. And investigating inaccuracies increased the time necessary to complete the task by up to 61%.

knowledge.[237] This stage determines whether an error remains a contained misstep or becomes entrenched in judicial reasoning, financial systems, regulatory decisions, or public discourse.

Control measures must intervene at key decision-making junctures, ensuring that errors do not gain legitimacy within professional and institutional frameworks. At this stage, flawed AI-generated outputs have already been incorporated into professional work but have not yet been entirely accepted as authoritative. This is where professional bodies—such as courts, regulatory agencies, and other relevant institutions—must act to identify and correct these mistakes before they become embedded in legal rulings, financial models, or regulatory frameworks.

Some professional fields already incorporate built-in review mechanisms that can function as oversight against AI-generated errors. In litigation, opposing counsel, judicial clerks, and appellate review processes may help identify inaccuracies in AI-assisted legal arguments. In finance, regulatory reporting requirements subject investment models to periodic audits, potentially flagging AI-driven miscalculations before they influence decision-making. Similarly, peer review in academia subjects AI-generated research findings to expert scrutiny before they are published. While these mechanisms were not initially designed for AI verification, they provide existing institutional safeguards that could help catch AI-related errors before they solidify into accepted knowledge.

However, these existing mechanisms prove insufficient against the scale and sophistication of AI-generated errors. AI systems can produce errors at scale—generating hundreds of flawed outputs simultaneously—whereas human errors typically occur individually, and these AI errors often appear professionally formatted and authoritative, making them harder to detect than obvious human mistakes.[238] Moreover, many professional practices—and even entire fields—lack built-in safeguards to catch these errors. AI-assisted contract drafting, internal corporate legal analysis, and compliance filings are implemented without adversarial review, increasing the risk of unverified AI-generated content shaping legal obligations. In finance, AI-generated risk models influence investment decisions without external

---

237. Kendrea Beers & Cody Rushing, *AI Control: How to Make Use of Misbehaving AI Agents*, CTR. FOR SEC. & EMERGING TECH. (Oct. 1, 2025), https://cset.georgetown.edu/article/ai-control-how-to-make-use-of-misbehaving-ai-agents/ [https://perma.cc/96MX-LQA5].

238. The 979 documented cases of AI-fabricated legal citations demonstrate how easily sophisticated outputs can evade traditional review processes. *See supra* Part II; Charlotin, *supra* note 2.

validation, amplifying inaccuracies in market assessments. Similarly, in architecture and engineering, AI-assisted designs are assumed to comply with regulatory standards despite the absence of thorough human review. In medicine, particularly in private practice or telemedicine, AI-generated diagnoses and treatment recommendations are adopted without secondary professional verification. In these fields, GenAI outputs are treated as authoritative without meaningful scrutiny, allowing errors to propagate unnoticed.

Expanding the scope of oversight bodies and their verification responsibilities is necessary to ensure AI-generated outputs undergo adequate scrutiny. For instance, courts could enhance verification by requiring attorneys to certify whether AI tools were used in research and by training existing judicial clerks to spot common AI fabrication patterns in citations, rather than assuming attorneys have done so themselves. AI-assisted architectural or engineering designs should require a third-party safety review before approval in high-risk projects. In medicine, AI-assisted diagnoses should trigger documentation requirements that show the verification steps taken before incorporation into patient records. In finance, professionals using AI for investment research should be required to document their verification process before making recommendations to clients or incorporating findings into official reports.

Similarly, AI-generated journalism, research, and educational materials should undergo further editorial oversight or expert fact-checking before publication. While these targeted measures could be beneficial in specific high-risk contexts, expanding human verification across all professions would be impractical, introducing unnecessary bureaucracy and slowing down critical processes. Instead, targeted oversight should focus on where consumer AI errors pose the most significant systemic risks. The EU's AI Act offers one regulatory approach, but it assumes centralized enterprise AI deployment rather than individual professionals using ChatGPT or Claude that drive the Invisible Ripple Effect. In jurisdictions like the U.S., where regulatory authority fragments across federal and state levels, traditional frameworks struggle to address consumer AI tools that fall outside enterprise oversight mechanisms. This highlights the need for governance approaches that can effectively address widespread individual AI usage while mitigating systemic risk of contamination.

Control mechanisms cannot catch every error before it influences professional systems. Current professional guidance remains limited, creating an AI literacy deficit among decision-makers who lack understanding of how AI systems produce sophisticated fabrications. While education about cognitive vulnerabilities, such as automation bias, proves

essential, it cannot address errors that have already shaped institutional decisions, requiring mitigation strategies to correct contaminated knowledge before it produces lasting consequences.

D. Mitigation

At this stage, AI-generated errors have already shaped professional and institutional decision-making. Unlike prevention, containment, and control—which aim to stop errors before they influence authoritative frameworks—mitigation focuses on identifying, correcting, and limiting their consequences before they become permanently embedded. Uncorrected errors can reinforce one another, escalating beyond isolated mistakes into systemic distortions that destabilize professional standards across fields.

Mitigation requires more than identifying specific mistakes; it demands mechanisms that prevent flawed outputs from continuing to shape future decisions. When misinformation introduced by AI has already influenced a professional outcome, future reliance on that decision must be approached with caution. This means ensuring that errors do not become entrenched as authoritative precedent.

Current professional frameworks provide limited guidance for correcting systemic errors. Most existing sanctions focus on individual misconduct rather than institutional knowledge hygiene.[239] Professional bodies could adapt existing mechanisms—such as legal citation services, medical record systems, and financial compliance databases—to include AI-related error tracking. Rather than creating entirely new infrastructure, these modifications would build on familiar professional tools while adding transparency about AI-generated content.

However, implementing such safeguards faces significant practical hurdles. Overly broad error-tracking could create chilling effects, reducing confidence in useful AI tools rather than promoting responsible use. Tracking systems must strike a balance between transparency and usability, providing professionals with clear feedback about known errors without overwhelming them with excessive warnings that hinder efficiency. Moreover, defining what constitutes a "significant" error worthy of tracking, versus routine inaccuracies that professionals can easily verify, requires careful calibration.

---

239. *See, e.g.*, Gunder, *supra* note 187.

The decentralized nature of consumer AI adoption complicates efforts to mitigate systemic issues. Unlike enterprise AI systems deployed by institutions, individual professionals using ChatGPT or Claude operate largely outside centralized oversight. This reality suggests that mitigation strategies must work within existing professional networks rather than requiring new regulatory apparatus. Professional associations, licensing boards, and peer networks may prove more effective vehicles for error tracking than top-down federal oversight.

Effective mitigation also requires appropriate incentives. Current liability frameworks provide insufficient motivation for systemic error correction, as they typically address individual violations rather than ongoing knowledge contamination.[240] Professional organizations could encourage mitigation through certification standards, continuing education requirements, or peer review processes that reward responsible AI usage and error correction.

While mitigation cannot undo all AI-related mistakes, it serves as a critical safeguard against their permanent institutionalization. By building on existing professional mechanisms and creating reasonable incentives for error tracking and mitigation, it can help contain the accumulating effects of AI-generated misinformation before they compromise the integrity of the professional knowledge system.

E. Resilience

No regulatory framework can entirely eliminate the risks posed by GenAI. The Invisible Ripple Effect persists because systemic vulnerabilities enable errors to propagate and embed in professional structures. Resilience acknowledges this reality—it ensures that institutions and regulatory frameworks remain adaptive rather than static as AI risks evolve.

Resilience differs from the previous stages. It does not eliminate risk or serve merely as a failsafe when other mechanisms fail. Instead, it ensures that professional standards and oversight systems can respond to new AI-induced risks as they emerge. Without resilience, even well-designed interventions risk becoming outdated, leaving systems vulnerable to unanticipated failures.

The foundation of resilience is continuous professional adaptation, operationalized through mandated biennial reviews of AI-use guidance by

---

240. *See, e.g.*, *id.*

professional licensing boards. State bars, medical boards, and architectural licensing bodies, for example, should be required to issue updated best practices for AI verification every two years as a condition of their authority. This transforms static compliance into a dynamic process. Furthermore, this can be integrated into existing professional development frameworks, such as by instituting mandatory Continuing Legal Education (CLE) or Continuing Medical Education (CME) credits focused specifically on AI competence, ethics, and emerging failure modes.[241]

This requires formalizing institutional learning mechanisms. AI failures cannot be treated as isolated incidents; they must be systematically analyzed to strengthen governance across professional domains. This could be achieved by establishing a National AI Incident Reporting Clearinghouse, modeled after aviation safety reporting systems. This body would collect and analyze anonymized reports of AI-generated errors from professionals across various fields, including law, medicine, finance, and engineering. The clearinghouse would then publish periodic reports identifying systemic error patterns, new model vulnerabilities, and effective verification techniques, ensuring insights from one profession's AI challenges inform safeguards in others.

Regulatory flexibility is equally crucial to prevent obsolescence. The EU's risk-based classification system, for example, offers a valuable framework but must evolve as AI applications change. Resilient governance requires embedding adaptive mechanisms directly into law, such as sunset provisions or mandatory triennial review cycles for all AI-related regulations. This would compel regulators to periodically reassess AI

---

241. In the context of lawyers, Amy Cyphert argued that while the current Model Rules provide a foundation for regulating lawyers' use of AI, they lack specificity and require amendments to offer clearer guidance. She proposes amending Comment 8 to Rule 1.1 to explicitly require lawyers to undertake AI-specific Continuing Legal Education (CLE) to ensure they understand the ethical implications of AI in legal practice. She also advocates modifying the MCLE Model Rule to include a mandatory annual technology CLE credit, reinforcing the ABA's call for AI education. Additionally, she suggests revising Comment 3 to Rule 5.3 to clarify that lawyers must directly supervise AI systems rather than delegating that responsibility entirely to technical staff, and recommends that the ABA issue best practices for selecting and overseeing AI experts. To address the risks of bias in AI, she calls for a new Comment to Rule 8.4(g) to warn that failing to understand AI bias could lead to professional misconduct. Finally, she emphasizes the importance of leveraging existing ABA resources, such as the Legal Technology Resource Center, to provide free or low-cost AI training, ensuring that solo practitioners and smaller firms are not disadvantaged. Recognizing that amendments to the Model Rules can take years, she argues that interim solutions, such as ABA resolutions and updates to the Comments, can help guide lawyers in ethically adopting AI technologies before formal rule changes are implemented. *See* Cyphert, *supra* note 35, at 439–43.

classifications, update best practices, and refine oversight structures based on data from the AI Incident Reporting Clearinghouse and other sources.

Finally, resilience must strike a balance between risk management and innovation. Overcorrection—where liability concerns drive professionals to avoid AI entirely—could stifle progress. To mitigate this, regulators could create "safe harbors" that shield professionals from liability for certain AI-generated errors if they can demonstrate adherence to the most recent best-practice guidelines issued by their licensing boards. This incentivizes responsible adoption and discourages risk-averse stagnation. Resilience thus transforms governance from a static framework into an adaptive process. Without it, the safeguards outlined in previous sections become temporary measures. With resilience, institutions can not only withstand AI-induced disruption but also evolve to meet it.

* * *

This five-stage intervention framework operates at a deliberately high level of abstraction, providing conceptual guidance rather than prescriptive policy details. This approach reflects both the jurisdictional diversity and the rapid evolution of AI capabilities, which render specific technical mandates quickly obsolete. The framework's generality serves analytical purposes. Different legal systems will implement these principles through distinct mechanisms. Effective containment in the EU's comprehensive regulatory environment may require entirely different approaches within the fragmented U.S. oversight system. Moreover, AI's dynamic nature demands adaptive governance that can respond to emerging risks without wholesale regulatory reconstruction.

As noted, these interventions represent necessary responses rather than a complete regulatory agenda. As AI systems become more sophisticated, new risk categories will emerge, necessitating additional strategies that build upon these foundational principles. The framework's value lies not in specificity but in its systemic approach to understanding how AI errors propagate through institutional knowledge systems. Ultimately, the governance challenge extends beyond traditional technology regulation toward institutional epistemology—how professional communities generate, validate, and transmit authoritative knowledge. The Invisible Ripple Effect threatens not merely individual decision accuracy but the integrity of the knowledge infrastructure upon which professional expertise depends. Addressing this challenge requires recognizing that AI governance fundamentally concerns preserving institutional trust and professional competence amid unprecedented technological capability and systemic uncertainty.

IV. CONCLUSION

GenAI represents a fundamental shift from institutional control to widespread, decentralized deployment. This accessibility drives innovation, but it also introduces risks that extend far beyond individual errors. Unlike traditional AI, confined to controlled environments, GenAI's decentralized adoption embeds inaccuracies—such as hallucinations, biased outputs, omissions, and misuse—directly into professional practices and societal frameworks at a foundational level.

The Invisible Ripple Effect captures GenAI's defining risk: seemingly minor errors might not remain isolated incidents. Fabricated legal citations, flawed architectural specifications, and inaccurate financial models, to name but a few examples, diffuse across professional, institutional, and cultural structures, embedding themselves in ways that traditional AI governance models cannot address. These systemic risks demand regulatory responses that extend beyond frameworks designed for centralized, proprietary AI systems.

Meeting these challenges requires swift, coordinated action. Governance frameworks must match GenAI's adaptability and reach, ensuring safeguards evolve alongside its use. This necessitates mechanisms for managing decentralized applications, fostering cross-sector collaboration to contain ripple effects, and implementing safeguards to prevent localized errors from escalating into systemic vulnerabilities. Without such measures, GenAI's democratization risks eroding trust in professional and institutional systems.

GenAI stands at the intersection of unprecedented opportunity and systemic risk. Its ultimate impact, whether as a tool for empowerment or a source of structural vulnerability, depends on our ability to recognize and mitigate its ripple effects. The Invisible Ripple Effect serves not only as a warning but also as an urgent call to actively shape the broader implications of democratized AI. With thoughtful governance and proactive safeguards, GenAI can be harnessed not just for efficiency but to reinforce the trust, accuracy, and resilience of professional and institutional frameworks.