

# THE SUPPLY CHAIN IS A CIRCLE: AI, PRIVACY, AND PEOPLE

*Alicia Solow-Niederman\**

*Artificial Intelligence (AI) governance underscores the AI “supply chain” and the “many hands” involved in the creation of AI systems. Although invaluable, this production-centered approach risks overlooking what happens when people use AI tools. That’s a mistake, because user interactions determine the real-world impact of AI.*

*This Article contends that regulating AI and holding the appropriate actors responsible requires far more attention to what happens after deployment of an AI system. The supply chain is an invaluable start, but it’s not a linear chain. Rather, it is a circle that must incorporate users’ engagement with AI tools as part of post-deployment feedback. To do so, we must distinguish among different categories of challenges that occur as people engage with AI tools. Focusing on information privacy as a concrete example, I identify three types of challenges that arise in, out of, and via generative AI systems. Each type requires a different kind of policy response. Unless we account for the contextual ways that humans interact with AI systems on the ground, it is users who will bear the cost of harms, and our regulatory interventions will remain incomplete at best and pernicious at worst.*

---

© 2026 Alicia Solow-Niederman. Individuals and nonprofit institutions may reproduce and distribute copies of the Work in any format, at or below cost, for educational purposes, so long as each copy identifies the author, provides a citation to the *Notre Dame Law Review Reflection*, and includes this provision and copyright notice.

\* Associate Professor of Law, The George Washington University. For invaluable comments, questions, and support, thank you to A. Feder Cooper, Sev Engelmann, James Grimmelmann, Claudia Haupt, Yafit Lev-Aretz, Helen Nissenbaum, Paul Ohm, and Frank Pasquale, as well as participants at the 2024 GenLaw DC Workshop, the 2024 GW Law Faculty Retreat Junior Scholarship Showcase, the Digital Life Initiative Colloquium at Cornell Tech, and the Yale Law School Information Society Project Ideas Lunch. Thank you to Rhyia Bibby and Charlise Doan for expert research assistance. I wish to also thank technology journalists, including but not limited to individuals at the *Washington Post* who were laid off in February 2026, whose reporting informed me of so many of the stories I include in this Article. All remaining errors and omissions are my own.

## INTRODUCTION

Artificial Intelligence (AI) governance is complicated. That's not news. A rapidly growing, vital AI "supply chain" literature has recognized the systemic, distributed, and dynamic nature of AI production.<sup>1</sup> But AI's producers are just the start of the story. Too often, the AI supply chain fails to confront how human beings engage with AI tools after deployment, and who should be responsible for these outcomes.<sup>2</sup>

This Article connects AI's developers and deployers to AI's users and argues that the supply chain isn't a chain at all; rather, it's a circle. The supply chain is about the actors and steps that produce an AI model, as well as what steps are taken to embed that model in a generative AI system, such as a chatbot, an agentic AI assistant (sometimes referred to as "agents," for short), or an AI companion.<sup>3</sup> However, the supply chain cannot sustain AI accountability unless it incorporates what happens next: real-world human interactions with specific AI tools, and the social, technical, and legal challenges that occur during these interactions. Picture, say, a chatbot user who hits "share," thinking it's the way to enter their query—only to find that they have shared a personal conversation about their religion, health, finances, or family

---

1 See Jennifer Cobbe, Michael Veale & Jatinder Singh, *Understanding Accountability in Algorithmic Supply Chains*, in PROCEEDINGS OF THE 2023 ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 1 (June 12–15, 2023), <https://arxiv.org/pdf/2304.14749> [<https://perma.cc/3K4T-MUHR>] (focusing on the "many hands" problem and analyzing algorithmic supply chains); Katherine Lee, A. Feder Cooper & James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, 72 J. COPYRIGHT SOC'Y 251, 285–307 (2025) (proposing a "generative-AI supply chain," *id.* at 286, and describing each of the eight stages).

2 See Jennifer Wang, Andrew D. Selbst, Solon Barocas & Suresh Venkatasubramanian, *Distinguishing Task-Specific and General-Purpose AI in Regulation*, ARXIV 8 (Jan. 23, 2026), <https://www.arxiv.org/abs/2506.17347> [<https://perma.cc/9YZ8-T5CS>] ("The current discourse on AI regulation has predominantly focused narrowly on the AI value chain. While this has yielded concrete technical governance proposals, it neglects the systemic nature of AI risks and impacts, as its harm manifests through interactions across an even broader ecosystem." (citation omitted)).

3 My focus in this Article is on user interactions with AI systems, post deployment; in this analysis, models are relevant insofar as choices about model design and deployment affect the AI system with which users engage. As Lee, Cooper, and Grimmelmann explain, "[m]ost users of generative AI do not interact with a model directly. Instead, they use an interface to a system, in which the model is just one of several embedded, interoperating components." Lee, Cooper & Grimmelmann, *supra* note 1, at 271. Unless otherwise indicated, I use the terms "system" and "tool" interchangeably.

life with a public-facing feed.<sup>4</sup> Even current efforts to monitor post-deployment impacts do not adequately connect outcomes like this to choices and responsibilities further up the chain. For instance, the EU AI Act requires post-market monitoring for high-risk AI systems<sup>5</sup>—but a general purpose chatbot isn’t itself such a system.<sup>6</sup>

My core thesis is simple: different types of human interactions with deployed AI systems cluster into recurring patterns of challenges, and these categories warrant distinct analysis and call for distinct policy responses. The situation above, where a user interacting with a chatbot clicks a “share” button and publishes an embarrassing private conversation for anyone to see,<sup>7</sup> is different from a situation in which a chatbot responds to a user’s query about how to contact a business and, in so doing, reveals a different individual’s home phone number.<sup>8</sup> And it’s also different from a user who engages with a speech-to-text AI tool, and whose tone of voice permits the company that created the tool to infer that the user is likely to have a particular medical condition.<sup>9</sup> But scholarly and policy debates frequently don’t distinguish these issues.<sup>10</sup> Though some accounts rightly pinpoint that different stages of the AI lifecycle will present different kinds of privacy challenges, there has not been adequate attention to the ways that distinct categories of post-deployment user interactions themselves present different issues.<sup>11</sup> A

4 Cf. Geoffrey A. Fowler, *Zuckerberg’s New Meta AI App Gets Personal in a Very Creepy Way*, WASH. POST (May 5, 2025), <https://www.washingtonpost.com/technology/2025/05/05/meta-ai-privacy/> [<https://perma.cc/H5BL-MFP6>].

5 See Council Regulation 2024/1689, of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonized Rules on Artificial Intelligence and Amending Regulations (Artificial Intelligence Act), art. 74, 2024 O.J. (L 1689) 1, 21.

6 See Margot E. Kaminski & Andrew D. Selbst, *An American’s Guide to the EU AI Act*, 40 BERKELEY TECH. L.J. 1081, 1115–18 (2026) (describing “ad hoc elements” of Act, *id.* at 1115, and discussing transparency requirements that apply to chatbots).

7 This is the privacy *in* generative AI problem. See discussion *infra* Section II.A.

8 See Robert Booth, *‘It’s Terrifying’: WhatsApp AI Helper Mistakenly Shares User’s Number*, GUARDIAN (June 18, 2025, 07:14 ET), [https://www.theguardian.com/technology/2025/jun/18/whatsapp-ai-helper-mistakenly-shares-users-number?mc\\_cid=1229ae7c67&mc\\_eid=382b075155](https://www.theguardian.com/technology/2025/jun/18/whatsapp-ai-helper-mistakenly-shares-users-number?mc_cid=1229ae7c67&mc_eid=382b075155) [<https://perma.cc/LAB7-GMLX>]. This is the privacy *out* of generative AI problem. See discussion *infra* Section II.B.

9 Cf. Laura Tedesco, *How Your Voice Could Reveal Hidden Disease*, MDEDGE (Dec. 7, 2022), <https://www.mdedge.com/cardiology/article/260028/mixed-topics/how-your-voice-could-reveal-hidden-disease/page/0/1> [<https://perma.cc/27KC-UHAP>] (detailing research on the use of voice to infer medical conditions). This is the privacy *via* generative AI problem. See discussion *infra* Section II.C.

10 For a rare exception, see Wang et al., *supra* note 2, at 8–9 (focusing on more concrete or cognizable harms, such as the risk of bioweapons development by generative AI systems, and proposing a “risk chain”).

11 See Jennifer King & Tiffany Saade, *Data Privacy and Foundational Models: Can We Have Both?*, STANFORD HAI ISSUE BRIEF (Apr. 2026), <https://hai.stanford.edu/assets/files/hai-issue-brief-data-privacy-and-foundation-models.pdf> [<https://perma.cc/Q4JH-6NAN>]. King

proposed intervention will fail if it does not first diagnose the problem that it is seeking to solve.

This Article proceeds in four parts. Part I assesses the lifecycle of generative AI systems and explains what the supply chain framing misses about user-focused, post-deployment challenges. The rest of the Article assesses how people use these AI systems on the ground—and what AI governance might do about it. To make my analysis concrete, I focus on information privacy to demonstrate how a value (such as privacy) or a policy objective (such as assuring a particular level of privacy) can be strained in distinct ways as users engage with AI systems. Drawing on illustrative contemporary examples, Part II provides a typology of privacy *in* generative AI challenges, privacy *out of* generative AI challenges, and privacy *via* generative AI challenges.<sup>12</sup> Part III then applies this framework, sketching how different legal and technical interventions map onto each category. Part IV concludes, contending that this taxonomy is practically and normatively valuable, both because it promotes regulatory alignment and because it can better hold the appropriate actors responsible, throughout the AI development and deployment cycle.

## I. THE SOCIOTECHNICAL AI SUPPLY CHAIN

This Part canvasses the growing literature on AI supply chains and assesses how this vision of AI systems is necessary, but not sufficient. It first distills the concept of an AI supply chain, evaluating what it reveals about AI systems as well as what it teaches about AI accountability and governance challenges. It then considers what this concept misses about post-deployment interactions with users of AI systems, emphasizing how human engagement with deployed tools should be part of a circle that feeds back into the supply chain.

The idea of an AI supply chain builds on a longstanding issue: the “many-hands” problem. The core idea is that, when more people are involved in a process, it is harder to figure out who to hold accountable and how to do so.<sup>13</sup> In foundational work published nearly thirty years ago, Helen Nissenbaum analyzed how the design and deployment of

---

and Saade recognize privacy risks that “emerge across the entire model life cycle.” *Id.* at 1. I concur that attention to different kinds of privacy risks, over time, is critical. I am the first, to my knowledge, to offer a typology of information privacy risks that occur as users engage with generative AI systems, during the ordinary operation of these tools.

12 Cf. Vivien A. Schmidt, *Democracy and Legitimacy in the European Union Revisited: Input, Output, and ‘Throughput’*, 61 POL. STUD. 2, 2–4 (2012) (canvassing past work on “input” and “output” legitimacy and offering an account of “throughput” legitimacy,” *id.* at 3, in democratic governance).

13 Cobbe, Veale & Singh, *supra* note 1, at 1.

technical systems create a many-hands problem.<sup>14</sup> First, computing systems are typically “products not of single programmers working in isolation but of groups or organizations.”<sup>15</sup> Second, they are often “constructed out of segments or modules,” each of which may be produced by teams or individuals, and each of which may integrate components from other systems or be built on top of other systems.<sup>16</sup>

Contemporary digital technologies compound and further complicate the many hands problem. As Jennifer Cobbe, Michael Veale, and Jatinder Singh explain, digital technologies “are increasingly organised around *data-driven supply chains* involving several interconnected actors and their systems.”<sup>17</sup> Because such supply chains involve multiple entities, issues that arise within them cannot be resolved at the level of any one organization.<sup>18</sup> These issues, moreover, become all the more complex in the context of “algorithmic supply chains” in which “several actors contribute towards the production, deployment, use, and functionality of AI technologies,” including “AI ‘as a service’ providers.”<sup>19</sup>

In addition, generative AI systems may create their own supply chain dynamics. As Katherine Lee, A. Feder Cooper, and James Grimmelmann propose, the process of turning training data into an AI-generated output, such as an image, can be understood as a “generative-AI supply chain” that consists of interconnected but distinct stages.<sup>20</sup> This framing, which the authors develop in the context of copyright law, highlights the corporate and individual choices as well as the potential legal consequences at each stage.<sup>21</sup>

The idea of an AI supply chain is invaluable in thinking about AI systems and considering whether, and if so how, to regulate them. It’s more descriptively accurate, because it shifts away from an

---

14 Helen Nissenbaum, *Accountability in a Computerized Society*, 2 SCI. & ENG.’G ETHICS 25, 28–30 (1996); see also Cobbe, Veale & Singh, *supra* note 1, at 1 (discussing Professor Nissenbaum’s work); Lee, Cooper & Grimmelmann, *supra* note 1, at 285 n.117 (discussing the many hands problem in the generative AI context) (citing A. Feder Cooper, Emanuel Moss, Benjamin Laufer & Helen Nissenbaum, *Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning*, in PROCEEDINGS OF THE 2022 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 864, 867–69 (2022), <https://arxiv.org/abs/2202.05338> [<https://perma.cc/5P59-5FNZ>]; Rui-Jie Yew & Dylan Hadfield-Menell, *Break It Till You Make It: Limitations of Copyright Liability Under a Pre-training Paradigm of AI Development*, in PROCEEDINGS OF THE FIRST WORKSHOP ON GENERATIVE AI & L. (2023), <https://blog.genlaw.org/CameraReady/30.pdf> [<https://perma.cc/E7HU-JWLS>].

15 Nissenbaum, *supra* note 14, at 28, 28–30.

16 *Id.* at 29, 28–30.

17 Cobbe, Veale & Singh, *supra* note 1, at 1.

18 *Id.* at 1–2.

19 *Id.* at 1, 1–2.

20 Lee, Cooper & Grimmelmann, *supra* note 1, at 257, 257–58.

21 *Id.* at 257–58.

organization-centered approach, which assumes that an AI product or service is created and controlled by a single entity in a static way. It instead engages with the myriad relationships and interdependencies that are required to produce contemporary AI systems and to refine them over time.<sup>22</sup>

The supply chain framing is also more analytically productive, for at least two reasons. First, it better positions policymakers to identify where there may be accountability gaps and tensions. For instance, a downstream deployer of a model created by an upstream third party may lack control over—or even awareness of—the initial design steps, even though it is these initial steps that lead to bias in the system at the point of use. This situation creates accountability challenges when, as is often the case, policy interventions target the downstream deployer.<sup>23</sup> Careful analysis of this sort is essential because there are multiple steps in the supply chain, such as compiling training data sets, selecting a technical architecture, or deploying a product as a service by “embedding [it] in some larger software system that provides a convenient interface.”<sup>24</sup> These steps are apt to implicate different actors, who face different political, economic, and legal incentives, constraints, and consequences.<sup>25</sup>

Second, the supply chain contemplates that the interactions among actors are dynamic and unsettled.<sup>26</sup> The relationships may change over time; in addition, a particular supply chain may emerge only in the context of a particular relationship that is mediated by a particular set of technological, social, and political conditions.<sup>27</sup> For example, a supply chain may exist where a business offers an AI tool and calls on a separate cloud services host for a certain set of processing resources; simultaneously or subsequently, that same cloud services host may play a different role, such as offering services to another entity or providing the necessary resources to deploy the tool at a different point in the chain.<sup>28</sup> The supply chain approach thus directs attention to the many players involved in AI’s production and the power dynamics associated with the relationships among them, calling for AI regulation to reflect these realities.

---

22 See Cobbe, Veale & Singh, *supra* note 1, at 2–4.

23 See *id.* at 4, 8.

24 Lee, Cooper & Grimmelmann, *supra* note 1, at 288, 287–88.

25 See, e.g., Cobbe, Veale & Singh, *supra* note 1, at 2; Lee, Cooper & Grimmelmann, *supra* note 1, at 285 & n.117, 307–08.

26 Cobbe, Veale & Singh, *supra* note 1, at 4–6.

27 *Id.* at 5.

28 See *id.* (describing a similar set of relationships in “[a] representative AI supply chain” and offering a schematic to visualize the process).

The supply chain metaphor, though, has its limits. Notably, this metaphor is one of production. It is about the creation and deployment of the AI system. To be sure, accounts of the supply chain can and do suggest that a system may be further developed after deployment. Professors Lee, Cooper, and Grimmelmann, for instance, acknowledge that:

The supply chain does not end with generation. Both before and after deployment, the developers of a generative-AI system can perform alignment by rating prompts and generations: further adjusting the model and the system it is embedded in to better achieve users' (and their own) needs. . . . In this way—as in many others—the supply chain feeds back into itself. It is not a simple cascade from data to generations. Instead, each stage is regularly adjusted to better meet the needs of the others.<sup>29</sup>

However, such iterative, post-deployment feedback loops are still focused on production choices that enhance performance. The goal is the developer's "alignment" of the system, which allows that actor to produce results that perform well enough to "achieve users' (and their own) needs."<sup>30</sup> This process occurs within the supply chain, with production and re-production over time to improve the model and the system.

And therein lies the rub: feedback that improves performance of a model or system is not always the same thing as feedback that reflects harms or risks as people use a specific AI tool. The two can at times overlap. For instance, an output might be deemed "unsafe" or "toxic," either because it is legally suspect or because it is socially harmful.<sup>31</sup> If the output, say, includes allegedly copyright-infringing material, or if it includes discriminatory or violent content or non-consensual intimate imagery, then refining the model or the system may address a social harm.

But there can also be daylight between technical and sociotechnical understandings of how a model or system performs. Model alignment reflects iterative refinement of the production process itself, to improve performance and enhance the product. Post-deployment social challenges also involve more complicated, human-centered processes, such as whether outputs compromise the privacy interests of an unrelated third party, or whether an interface reflects users' privacy expectations. This form of risk is not the same as the emphasis on whether a tool works well, in the technical sense, nor whether it's legally compliant, in the formal sense. It's sociotechnical. Moreover,

---

<sup>29</sup> Lee, Cooper & Grimmelmann, *supra* note 1, at 288.

<sup>30</sup> *Id.*

<sup>31</sup> *Id.* (asserting that developers' alignment processes may reflect "needs [that] can include safety, helpfulness, and legal compliance").

recognizing such social dynamics requires thinking granularly about the different ways in which human users interact with models, as part of a dynamic process. Seeing a “risk chain,” and not just a “supply chain,” is a vital start.<sup>32</sup> But if the goal is to manage systemic social harm, perpetuated through AI models and systems, then additional analytic and regulatory tools are needed.

In the remainder of this Article, I propose a framework to better account for humans’ interactions with AI systems on the ground, aiming to connect these post-deployment developments to the supply chain and to thereby enhance system-wide accountability. I begin with the deployment of the AI model in the context of a specific user-facing tool, such as a chatbot or a text-to-image generator. I then draw from supply chain insights on how to conceptualize and govern complex, distributed, and dynamic AI systems and apply similar insights to a different set of actors: the people who use AI tools and the social contexts in which they do so. Many types of users engage with AI tools, and the same user may engage with different tools or even the same tools in different ways. A product manager may try to streamline a process. An entrepreneur may try to “vibe code” a business idea.<sup>33</sup> A lawyer may try to create a contract template. A father may seek advice on how to relate to his teenage daughter. A person without access to a therapist may seek assistance with a mental health challenge. The list goes on.

These interactions are not all the same. Each involves distinct forms of user engagement with an AI tool, and each puts pressure on the regulatory status quo in distinct ways. For example, focusing on information privacy makes clear that the individual’s concern with the confidentiality of financial, health, or other sensitive data that they put into a model is not the same as a broader social concern that the output of a model will include someone else’s personal information. And it’s also not the same as the concern that a model can process troves of data to discern further information about individuals. Right now, though, neither AI governance nor information privacy law has the working vocabulary to recognize this reality. That, I think, is an oversight. To make this analysis concrete, the next Part uses information privacy to illustrate categories of challenges that can occur as users interact with AI systems.

---

32 See Wang et al., *supra* note 2, at 8–9.

33 See Rhiannon Williams, *What is Vibe Coding, Exactly?*, MIT TECH. REV. (Apr. 16, 2025), <https://www.technologyreview.com/2025/04/16/1115135/what-is-vibe-coding-exactly/> [<https://perma.cc/UU9B-9EMK>] (quoting OpenAI cofounder Andrej Karpathy, who coined the term “vibe coding” and defined it as “a new kind of coding . . . where you fully give in to the vibes, embrace exponentials, and forget that the code even exists . . . I’m building a project or webapp, but it’s not really coding—I just see stuff, say stuff, run stuff, and copy paste stuff, and it mostly works.”).

## II. THREE TYPES OF CHALLENGES: *IN*, *OUT OF*, AND *VIA*

This Part proposes a typology to understand the kinds of challenges that occur as users interact with generative AI systems, post-deployment. It focuses on information privacy and distinguishes among challenges that arise *in*, *out of*, and *via* generative AI tools, thereby providing a conceptual framework that can structure a growing conversation about the relationship between privacy and AI.<sup>34</sup> This exposition of the issues sets a foundation for Part III, which assesses which kinds of legal and technical interventions are best tailored to each type of challenge.

Two caveats and one note about scope are in order before proceeding. First, to illustrate these categories, I at times discuss company-specific examples; in so doing, my intent is not to single out that actor, but rather to provide an illustrative example that calls attention to human interactions with generative AI systems. Second, my coverage focuses only on privacy harms that emerge through non-adversarial use of the tool, meaning uses that do not involve an external attack by a third-party actor. To be sure, intentional adversarial attacks, in which an external actor attempts to make the AI system perform differently than expected by its developers, can also produce security-inflected privacy harms.<sup>35</sup> Complex AI systems create new opportunities for third parties to attempt to access sensitive information, and systems that rely on extended supply chains will create more points of potential attack.<sup>36</sup> But because such adversarial attacks typically involve the behavior of an unrelated third party hacker, and because they are really a matter of cybersecurity, I set them aside. My emphasis is on a subtler, but no less significant concern: privacy challenges that emerge from business as usual.

---

34 See generally King & Saade, *supra* note 11; Daniel J. Solove, *Artificial Intelligence and Privacy*, 77 FLA. L. REV. 1 (2025); *The Need to Protect Americans' Privacy and the AI Accelerant: Hearing Before the S. Comm. on Com., Sci., & Transp.*, 118th Cong. 7–11 (2024) (statement of Ryan Calo).

35 See generally Ram Shankar Siva Kumar, David O'Brien, Kendra Albert, Salome Viljoen & Jeffrey Snover, *Failure Modes in Machine Learning*, ARXIV (Nov. 25, 2019), <https://arxiv.org/abs/1911.11034> [<https://perma.cc/JE3A-LPSQ>] (providing early overview of adversarial machine learning); Melissa Heikkilä, *Three Ways AI Chatbots Are a Security Disaster*, MIT TECH. REV. (Apr. 3, 2023), <https://www.technologyreview.com/2023/04/03/1070893/three-ways-ai-chatbots-are-a-security-disaster/> [<https://perma.cc/W346-MBJW>] (summarizing three types of security vulnerabilities in AI systems).

36 See, e.g., Joseph Cox, *Hackers Break into AI Hiring Chatbot, Could Hire and Reject Fast Food Applicants*, 404 MEDIA (Jan. 11, 2024, 10:58), <http://404media.co/hackers-break-into-hiring-ai-chat-bot-chattr/> [<https://perma.cc/LV22-YJLS>] (reporting on restaurant's use of a third-party's AI hiring chatbot, which contained a security vulnerability that could be exploited to reveal sensitive information about applicants).

Finally, although I reserve a detailed study of domains beyond privacy, the patterns I present in this Article are at least to some extent generalizable. Often, regulatory interventions propose measures meant to achieve algorithmic fairness, accountability, or transparency, without amply accounting for user interactions with the AI system. For instance, what it means to minimize bias may be different if one is assessing whether the data that goes into the tool is fair (however defined), whether the outputs of an image generator are fair (however defined), or whether the predictions about a person made based on their interactions with a tool are fair (however defined).<sup>37</sup>

### A. *The Privacy In Generative AI Challenge*

The privacy *in* generative AI challenge involves the data that go into the system. The data that are disclosed in the process of engaging with the tool at the front-end are arguably sensitive or private, and most everyone would agree with that characterization. This issue can involve data from the user; alternatively or in addition, it can involve data from unrelated third parties.

First, consider the user's own data. To illustrate, picture a person who lacks access to telehealth or a primary care physician, asks a chatbot about ways to treat a sensitive medical condition, and discloses their own health data in the process of this interaction<sup>38</sup>—just as recently-launched product features such as OpenAI's ChatGPT Health<sup>39</sup>

37 Future work might explore whether these patterns hold equally for different kinds of issues; for instance, it is not clear what transparency via a generative AI tool would mean. Thank you to Frank Pasquale for questions on this point.

38 See Charlie Warzel, *AI Has Become a Technology of Faith*, ATLANTIC (July 12, 2024), <https://www.theatlantic.com/technology/archive/2024/07/thrive-ai-health-huffington-altman-faith/678984/> [<https://perma.cc/SN45-JLL2>] (as OpenAI CEO Sam Altman stated in the context of creating an AI healthcare company, "I and many others in the field have been positively surprised about how willing people are to share very personal details with an LLM . . . . They knew it wasn't a real person . . . and they were willing to have this hard conversation that they couldn't even talk to a friend about."); see also Elon Musk (@elonmusk), X (Oct. 29, 2024, 06:14 ET), <https://x.com/elonmusk/status/1851206080564773344> [<https://perma.cc/2SEL-BWPX>] ("Try submitting x-ray, PET, MRI or other medical images to Grok for analysis. This is still early stage, but it is already quite accurate and will become extremely good. Let us know where Grok gets it right or needs work."). X users appear to have responded to Musk's call. See Elizabeth Passarella, *Elon Musk Asked People to Upload Their Health Data. X Users Obligated*, N.Y. TIMES (Nov. 18, 2024), <https://www.nytimes.com/2024/11/18/well/x-grok-health-privacy.html> [<https://perma.cc/7B9W-LP8A>].

39 See *Introducing ChatGPT Health*, OPENAI (Jan. 7, 2026), <https://openai.com/index/introducing-chatgpt-health/> [<https://perma.cc/K29W-TYDT>]. This feature is distinct from "ChatGPT for Healthcare," which is an enterprise product for healthcare organizations. See Hemant Gupta, Bryan Hahm & Alaap B. Shah, *Health Care Without the Hospital: ChatGPT Health and Claude Go Direct to Consumers*, HEALTH L. ADVISOR BLOG (Jan. 28, 2026), <https://>

or Anthropic's Claude for Healthcare<sup>40</sup> encourage individuals to do.<sup>41</sup> Or suppose that an overworked junior associate at a law firm decides to use a chatbot to generate a contract and inputs confidential client data to do so. These situations are not hypothetical; indeed, disclosure of personal information is common in users' chats.<sup>42</sup>

Different forms of generative AI present these issues with particular force. The privacy in generative AI challenge is likely to be especially acute for agentic AI, where an AI system is given access to multimodal data, such as text, audio, and video, and may even process every waking moment, every written communication, and everything

---

[www.healthlawadvisor.com/health-care-without-the-hospital-chatgpt-health-and-claude-go-direct-to-consumers](https://www.healthlawadvisor.com/health-care-without-the-hospital-chatgpt-health-and-claude-go-direct-to-consumers) [https://perma.cc/3AXH-FSBL].

40 See *Advancing Claude in Healthcare and the Life Sciences*, ANTHROPIC (Jan. 11, 2026), [hereinafter *Advancing Claude*] <https://www.anthropic.com/news/healthcare-life-sciences> [https://perma.cc/4TW6-ZM8Z] (announcing features); *Transform Healthcare from Insight to Action*, ANTHROPIC, [hereinafter *Transform Healthcare*] <https://claude.com/solutions/healthcare> [https://perma.cc/47SA-PRSJ] (last visited Apr. 3, 2026) (describing products and capabilities). Anthropic's "Claude for Healthcare" offerings include both consumer-facing functionality, which can incorporate data directly from wearables, and enterprise offerings that are marketed as "HIPAA-ready." See Claude, LINKEDIN, [https://www.linkedin.com/posts/claude\\_claude-can-now-securely-connect-to-your-health-activity-7419520749591449603-gVMf](https://www.linkedin.com/posts/claude_claude-can-now-securely-connect-to-your-health-activity-7419520749591449603-gVMf) [https://perma.cc/ZKZ9-AGLM] (announcing direct-to-consumer functionality through beta integrations with "Apple Health (iOS), Health Connect (Android), HealthEx, and Function" that, when connected, allow Claude to "summarize your medical history, explain test results in plain language, detect patterns across fitness metrics, and more"); *Transform Healthcare*, ANTHROPIC, *supra* (detailing how "healthcare organizations" can use the system); see also Gupta, Hahm & Shah, *supra* note 39 (discussing Claude's offerings).

41 In addition to the privacy issues that I focus on, these systems raise serious accuracy and safety questions. See Geoffrey A. Fowler, *I Let ChatGPT Analyze a Decade of My Apple Watch Data. Then I Called My Doctor*, WASH. POST (Jan. 26, 2026), <https://www.washingtonpost.com/technology/2026/01/26/chatgpt-health-apple> [https://perma.cc/YT26-RN5N] ("[W]hen it comes to your fitness tracker and some health records, the new Dr. ChatGPT seems to be winging it. That fits a disturbing trend: AI companies launching products that are broken, fail to deliver or are even dangerous."); see also Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz & Andrew D. Selbst, *The Fallacy of AI Functionality*, in PROCEEDINGS OF THE 2022 ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY (June 21–24, 2022), <https://arxiv.org/pdf/2206.09511> [https://perma.cc/29NW-9596] (discussing AI systems that fail to work as promised). These performance problems make it even more important to understand how real people interact with systems, both so that the high stakes of these failures are apparent and so that these user interactions inform responsibilities all along the supply chain.

42 Gerrit De Vynck & Jeremy B. Merrill, *We Analyzed 47,000 ChatGPT Conversations. Here's What People Really Use It For*, WASH. POST. (Nov. 12, 2025), <https://www.washingtonpost.com/technology/2025/11/12/how-people-use-chatgpt-data/> [https://perma.cc/AXU4-JPHZ] ("Users often shared highly personal information with ChatGPT in the conversations analyzed by The Post, including details generally not typed into conventional search engines.").

that a person sees and does, with an eye to providing personalized recommendations, guidance, and problem-solving.<sup>43</sup>

For an extreme example, consider the OpenClaw personal assistant, billed as “AI that actually does things.”<sup>44</sup> When downloaded and given access to your device, the tool boasts that it “[c]lears your inbox, sends emails, manages your calendar, [and] checks you in for flights. All from WhatsApp, Telegram, or any chat app you already use.”<sup>45</sup> Running on one’s device, a user can permit OpenClaw to access content from these apps—seamlessly, by design.<sup>46</sup> Indeed, allowing personal data to flow into the tool is necessary for its very operation; otherwise, it would not be able to, as one reporter puts it, “follow[] almost any order like a well-paid mercenary.”<sup>47</sup> Moreover, OpenClaw never forgets; “persistent memory” recalls all of the content to which it is given access across the device, including personal or private information.<sup>48</sup>

A tool of this sort epitomizes the privacy in challenge: without complete or near-complete access to data, agentic AI of this sort could

43 There is no settled definition of agentic AI. See Sayash Kapoor & Arvind Narayanan, *New Paper: AI Agents That Matter*, AI AS NORMAL TECH. (July 3, 2024), [https://www.normaltech.ai/p/new-paper-ai-agents-that-matter?utm\\_source=publication-search](https://www.normaltech.ai/p/new-paper-ai-agents-that-matter?utm_source=publication-search) [<https://perma.cc/ZWT2-UDEA>] (discussing use of the term “agent” in the AI literature). A paper by OpenAI’s research team defines “agentic AI” as “systems that adaptably pursue complex goals using reasoning and with limited direct supervision.” Yonadav Shavit et al., *Practices for Governing Agentic AI Systems*, ARXIV 2, <https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf> [<https://perma.cc/XSZ6-CJB9>]. Companies have already introduced agentic systems. See, e.g., *Hello GPT-4o*, OPENAI (May 13, 2024), <https://openai.com/index/hello-gpt-4o/> [<https://perma.cc/ZS2C-SKTS>] (discussing OpenAI’s introduction of a multi-modal AI system, “GPT-4o (‘o’ for ‘omni’),” which “accepts as input any combination of text, audio, image, and video and generates any combination of text, audio, and image outputs”); Andrew Williams, *To Truly Fix Siri, Apple May Have to Backtrack on One Key Thing—Privacy*, WIRED (Mar. 23, 2025, 06:38), <https://www.wired.com/story/to-truly-fix-siri-apple-may-have-to-backtrack-on-one-key-thing-privacy/> [<https://perma.cc/8Z3P-CPP8>] (discussing Apple’s announcement that the Fall 2024 version of Siri assistant would be “equipped with awareness of your personal context, the ability to take action in and across apps, and product knowledge about your devices’ features and settings”).

44 OPENCLAW, <https://openclaw.ai/> [<https://perma.cc/NGQ2-2VMA>] (last visited Apr. 6, 2026). The tool was previously known as “Clawd,” and then “Moltbot.” See Peter Steinberger, *Introducing OpenClaw*, OPENCLAW (Jan. 29, 2026), <https://openclaw.ai/blog/introducing-openclaw> [<https://perma.cc/ZVH6-BN8K>].

45 OPENCLAW, *supra* note 44.

46 See *id.* A user can either select “full [system] access” or choose a “sandboxed” option. *Id.*

47 Deni Ellis Béchard, *OpenClaw—What Happens When AI Stops Chatting and Starts Doing*, SCI. AM. (Jan. 30, 2026), <https://www.scientificamerican.com/article/moltbot-is-an-open-source-ai-agent-that-runs-your-computer/> [<https://perma.cc/EG49-7RA9>].

48 See Macy Meyer, *Is OpenClaw Really the Next ChatGPT? Why Nvidia’s CEO Called this Hot New AI Assistant the Future*, CNET (Mar 18, 2026, 09:37 PT), <https://www.cnet.com/tech/services-and-software/from-clawdbot-to-moltbot-to-openclaw/> [<https://perma.cc/893H-5S5D>].

not exist. It is no surprise that security alarms have already been sounded—loudly.<sup>49</sup> But security concerns don't exhaust privacy concerns.<sup>50</sup> The problem is not merely access to data by external, potentially malicious actors, but also users' exposure of their own sensitive data for further processing. Rather than conflate data security and privacy, and rather than default to a "blame the user" stance on the grounds that individuals chose to install the tool, I situate tools like this as a pointed illustration of the privacy in challenge.

It's vital to recognize this type of issue, moreover, because it is not only the individual user's privacy that is at stake. Increasingly, data that is fed into an AI system includes sensitive or private information from unrelated third parties. To illustrate, consider "smart glasses," such as Meta Ray Bans or Oakley Meta "AI glasses." The company promises to "blend cutting-edge technology with iconic style"<sup>51</sup> and touts video capture capabilities as well as "[d]iscreet open-ear speakers."<sup>52</sup> If a patron at a restaurant wears these glasses during a meal, they may well capture the faces and actions of anyone who enters the frame—whether or not that third party agrees to or is even aware of the recording.<sup>53</sup> In other words, that third party's data also go into the AI tool, in the process of interacting with that tool, at the front end.

Even if that person is arguably in public and thus has no strong privacy claim under current U.S. law, it's worth recognizing how these technically-mediated interactions expand the amount of data about third parties that can be collected. Moreover, some of the collected

49 See Connor Jones, *Clawdbot Sheds Skin to Become Moltbot, Can't Slough Off Security Issues*, REGISTER (Jan. 27, 2026, 18:45 UTC), [https://www.theregister.com/2026/01/27/clawdbot\\_moltbot\\_security\\_concerns/](https://www.theregister.com/2026/01/27/clawdbot_moltbot_security_concerns/) [<https://perma.cc/BH8W-5EV3>] ("[W]e've spent 20 years building security boundaries into modern operating systems . . . AI agents tear all of that down by design. They need to read your files, access your credentials, execute commands, and interact with external services. The value proposition requires punching holes through every boundary we spent decades building." (quoting security specialist Jamieson O'Reilly)); see also Bécard, *supra* note 47 (discussing security issues).

50 See Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 490 (2006) (defining "insecurity" as one type of "information processing" interest that "involves carelessness in protecting stored information from leaks and improper access," and situating it in a taxonomy of other interests).

51 *Shop AI Glasses*, META, <https://www.meta.com/ai-glasses/shop-all/> [<https://perma.cc/R2S8-XXHV>] (last visited Apr. 21, 2026).

52 *Compare AI Glasses*, META, [https://www.meta.com/in/ai-glasses/compare/?srsltid=AfmB0ooT2eZAJQtjQYg\\_Uzn0MKLjrnL4BMNh6RhiZr--jb5IJN4ETjr](https://www.meta.com/in/ai-glasses/compare/?srsltid=AfmB0ooT2eZAJQtjQYg_Uzn0MKLjrnL4BMNh6RhiZr--jb5IJN4ETjr) [<https://perma.cc/DY8M-RWGP>] (last visited Apr. 21, 2026).

53 See Luke Fortney, *Dinner Is Being Recorded, Whether You Know It or Not*, N.Y. TIMES (Feb. 16, 2026), <https://www.nytimes.com/2026/02/16/dining/meta-ray-ban-glasses-restaurants.html> [<https://perma.cc/B6MM-TD4E>] ("As Meta smart glasses capture scenes in restaurants for social media, service workers and customers are becoming captive participants.").

information is likely to include extremely sensitive details about third parties. Consider reports that data annotators have been asked to review Meta AI smart glasses recordings that include “video material showing bathroom visits, sex and other intimate moments,” such as “a video where a man puts the glasses on the bedside table and leaves the room” before “his wife comes in and changes her clothes.”<sup>54</sup> Furthermore, even short of such highly sensitive disclosures of third parties’ private lives, the impact on third parties from AI technologies like these glasses is only likely to grow. With signals that Meta plans to offer facial recognition in their smart glasses,<sup>55</sup> it may soon be trivial to link an individual’s face to troves of other information about the person that is available online, thereby potentially revealing even more about that third party. This is a privacy in challenge, because merely existing in public permits an AI system to collect and curate potentially sensitive information about the person.

AI systems thus affect the privacy of not only individual users, but also the privacy of their friends, family, and even total strangers. The more that AI systems infuse different spheres of life, the more that this type of challenge will proliferate, in both blatant and subtler forms.

### B. *The Privacy Out of Generative AI Challenge*

The privacy *out of* generative AI challenge arises at the backend, as the user interacts with the system, only to discover that the response is privacy invasive. This issue can arise in two primary ways that reflect distinct characteristics of generative AI tools. In each, the harm flows

---

54 Naipano Lepapa, Ahmed Abdigadir & Julia Lindblom, *She Came Out of the Bathroom Naked, Employee Says*, SVENSKA DAGBLADET (Feb. 27, 2026), <https://www.svd.se/a/K8nrV4/met-as-smart-glasses-and-data-privacy-concerns-workers-say-we-see-everything> [<https://perma.cc/KWA2-3C4Z>]. Within days, this reporting led to an investigation by the UK Information Commissioner’s Office and a lawsuit against Meta in the United States. See Sarah Perez, *Meta Sued Over AI Smart Glasses’ Privacy Concerns, After Workers Reviewed Nudity, Sex, and Other Footage*, TECHCRUNCH (Mar. 5, 2026, 08:50 PT), <https://techcrunch.com/2026/03/05/meta-sued-over-ai-smartglasses-privacy-concerns-after-workers-reviewed-nudity-sex-and-other-footage/> [<https://perma.cc/73GJ-BRNG>]. Meta has contested these claims, asserting that it is not sending smart glasses footage recorded by European users to AI trainers in Kenya. See Ellen O’Regan, *Meta Tells Ireland no European Smart Glasses Data is Sent to Kenya*, POLITICO PRO (May 10, 2026), <https://pro.politico.eu/news/216011> [<https://perma.cc/A5Y2-FKQ7>].

55 See Kashmir Hill, Kalley Huang & Mike Isaac, *Meta Plans to Add Facial Recognition Technology to Its Smart Glasses*, N.Y. TIMES (Feb. 14, 2026), <https://www.nytimes.com/2026/02/13/technology/meta-facial-recognition-smart-glasses.html> [<https://perma.cc/762R-H3E8>] (reporting Meta’s plans to “add the [facial recognition] feature to its smart glasses, which it makes with the owner of Ray-Ban and Oakley, as soon as this year . . . . The feature, internally called ‘Name Tag,’ would let wearers of smart glasses identify people and get information about them via Meta’s artificial intelligence assistant.”).

from the system's outputs rather than the user's inputs; moreover, in each case, this result affects third parties who never interacted with the tool. The first subtype is a by-product of memorization, where a model internalizes, stores, and then serves up an exact copy of the content on which it was trained. Although this issue has arisen most prominently in intellectual property controversies,<sup>56</sup> there are striking privacy challenges, too. For instance, a group of computer science researchers were able to extract several megabytes of ChatGPT's training data by instructing the tool to repeat the word "poem" forever.<sup>57</sup> Prompts like this led the tool to circumvent privacy safeguards, and, in at least some instances, to disclose phone numbers and email addresses for real people and entities.<sup>58</sup> To be sure, this outcome was the product of an intentional effort to extract training data; however, even more quotidian requests can produce similarly personal data. For example, in summer 2025, a stranded train passenger asked Meta's WhatsApp AI assistant for the train company's customer services number, only to be provided with the real phone number of an unrelated individual.<sup>59</sup> In cases like this, ordinary user interactions lead the tool to reveal another person's personal information, putting pressure on the social and regulatory status quo.

The second subtype is a by-product of a different technical characteristic: the nature of generation itself. More precisely, the generation process is inherently probabilistic and non-deterministic, meaning that outputs are uncertain and to some extent random.<sup>60</sup> This technical characteristic matters for privacy if generative AI outputs unexpectedly include private information. By way of example, in

---

56 For a study of memorization in LLMs and potential copyright implications, see A. Feder Cooper, Aaron Gokaslan, Ahmed Ahmed, Amy B. Cyphert, Christopher De Sa, Mark A. Lemley, Daniel E. Ho & Percy Liang, *Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models*, ARXIV (Nov. 30, 2025) <https://arxiv.org/pdf/2505.12546> [<https://perma.cc/M8MW-FLGR>].

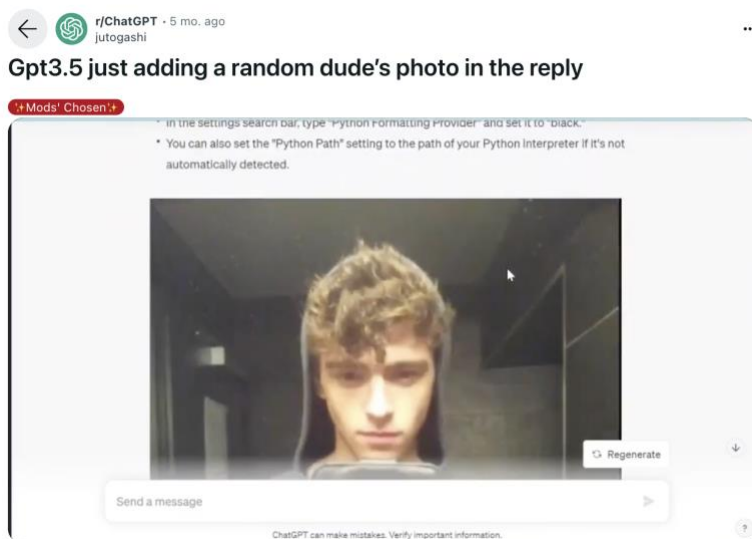
57 Nicholas Milad Nasr et al., *Extracting Training Data from ChatGPT* (Nov. 28, 2023) [hereinafter Milad Nasr et al., *Extracting Training*], <https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html> [<https://perma.cc/59MB-X3WA>]; see also Milad Nasr et al., *Scalable Extraction of Training Data from (Production) Language Models*, ARXIV (Nov. 28, 2023) [hereinafter Milad Nasr et al., *Scalable Extraction*], <https://arxiv.org/pdf/2311.17035> [<https://perma.cc/JDD8-RFTW>].

58 Milad Nasr et al., *Extracting Training*, *supra* note 57; Milad Nasr et al., *Scalable Extraction*, *supra* note 57, at 10.

59 See Booth, *supra* note 8.

60 See A. Feder Cooper, Jonathan Frankle & Christopher De Sa, *Non-Determinism and the Lawlessness of Machine Learning Code* 1, 3, in PROCEEDINGS OF THE 2022 SYMP. ON COMPUT. SCI. & L. (Nov. 1–2, 2022) (defining "[n]on-determinism" as "a property of processes for which supplying the same inputs can produce different outputs" and explaining that "non-deterministic outcomes are uncertain[...] in contrast to deterministic if/then logic, for which the same inputs produce the same outputs").

November 2023, Reddit users reported that GPT-3.5 “leaked a random dude’s photo in the output.”<sup>61</sup> At the time that this issue was reported, an individual’s image appeared in the middle of a query about how to format a particular type of code, as follows:<sup>62</sup>



Why? The working theory is that this selfie was uploaded to the website Imgur in December 2016,<sup>63</sup> and then Chat-GPT randomized a URL from that same site as part of the response, which in turn happened to direct to this photo.<sup>64</sup> In other words, the chatbot responded to an ordinary interaction by disclosing personal information that was freely shared in an entirely different context.

Without clarity into the details of the model itself, it’s not possible to confirm that theory; indeed, it is possible that this data was “memorized” from the training data. If that’s the case, then it’s another example of the first subtype of privacy out of challenge. Even if so, this issue is distinct from the input problem because no user of the tool

61 Alex Ker (@thealexker), X (Nov. 1, 2023 21:59 ET), <https://twitter.com/thealexker/status/1719896871009694057> [<https://perma.cc/YT2T-DJ9R>]; see Jutogashi, *GPT3.5 Just Adding a Random Dude’s Photo in the Reply*, REDDIT (Oct. 31, 2023, 21:28 ET), [https://www.reddit.com/r/ChatGPT/comments/17krsg2/gpt35\\_just\\_adding\\_a\\_random\\_dudes\\_photo\\_in\\_the/](https://www.reddit.com/r/ChatGPT/comments/17krsg2/gpt35_just_adding_a_random_dudes_photo_in_the/) [<https://perma.cc/FYV9-L9MY>].

62 See Jutogashi, *supra* note 61.

63 See IMGUR (Dec. 7, 2016), <https://imgur.com/YFRoBdF> [<https://perma.cc/8S75-8KTK>].

64 See PinGUY, *GPT3.5 Just Adding a Random Dude’s Photo in the Reply*, REDDIT (Oct. 31, 2023, 21:28 ET), [https://www.reddit.com/r/ChatGPT/comments/17krsg2/gpt35\\_just\\_adding\\_a\\_random\\_dudes\\_photo\\_in\\_the/](https://www.reddit.com/r/ChatGPT/comments/17krsg2/gpt35_just_adding_a_random_dudes_photo_in_the/) [<https://perma.cc/FYV9-L9MY>].

provided sensitive information. The issue here is that personal data was inadvertently disclosed as part of the output, in the process of a human interaction with the generative AI tool. This outcome is a violation of Helen Nissenbaum's theory of privacy as contextual integrity: there is arguably a privacy invasion because disclosure of the selfie in the later, unrelated context does not satisfy contextual information ("CI") norms.<sup>65</sup> This result arises during the ordinary operation of the tool. In this way, the technology's very operation facilitates a privacy out of generative AI challenge that exacerbates privacy law's longstanding difficulty accounting for cross-contextual information flows.

Part III returns to these interactions and contends that exposing issues in this way is constructive: it trains attention on the specific mix of technical and legal interventions that can redress specific weaknesses, while also highlighting which problems remain harder to solve, and why that is so.

### C. *The Privacy Via Generative AI Challenge*

Finally, the privacy *via* generative AI challenge occurs when the inferences made possible through the AI system themselves violate privacy by providing further information about a person, above and beyond that which was initially disclosed. Generative AI tools both make it easier to collect information and make it easier to infer personal information, including information that an individual does not realize they are explicitly disclosing during their interactions with the system.

Begin with the ways that many generative AI systems encourage data disclosure, in ways that make it easier to infer further information about an individual who uses that system. Picture an individual who consults an AI companion after a hard day at work, or a person who has formed a relationship with a generative AI chatbot, sharing dozens of messages each day.<sup>66</sup> Suppose that the person wants to talk to the chatbot as they would talk to a friend and relies on audio inputs rather than text. The systems facilitate just that, not only by allowing audio or multi-modal inputs, but also through choices to anthropomorphize the interface, such as referring to the chatbot as in the first person ("me" or "I") or the chatbot and user as "we." Design decisions such as these personalize the engagement and reduce friction in the user-system

---

65 See Helen Nissenbaum, *Contextual Integrity Up and Down the Data Food Chain*, 20 THEORETICAL INQUIRIES L. 221, 224–28 (2019). CI is a useful lens because it trains attention on the social dynamics that produce privacy challenges in a given setting. That said, the scale and distributed nature of AI systems, as well as the lack of a clear historic antecedent for generative AI tools, challenge the CI approach. See *infra* notes 125–27 and accompanying text.

66 See Kashmir Hill, *She Is in Love With ChatGPT*, N.Y. TIMES (Jan. 17, 2025), <https://www.nytimes.com/2025/01/15/technology/ai-chatgpt-boyfriend-companion.html> [<https://perma.cc/DPB3-CRA3>].

interaction; as a result, the user may well be more inclined to share information.

Collected data can then be analyzed to derive further inferences about that individual, whether or not they are aware of this possibility. Suppose, for instance, that it's possible to analyze the tone and modulation of the user's voice to predict that the person speaks "low and slowly" in a way that is indicative of Parkinson's disease, or to detect that they have begun slurring in a way that suggests they have suffered a stroke.<sup>67</sup> There might be many beneficial uses of these insights. I do not mean to suggest that all such insights are good or bad, but rather to make a descriptive point about the interaction itself: When the person spoke to the AI tool, they likely were not aware that their voice contained bits of latent data that could, using AI models, be processed to discern this further health information.<sup>68</sup>

The potential to derive inferences from disclosed data, moreover, goes well beyond health predictions. Suppose that a company decides to show ads in its chatbot.<sup>69</sup> It seems unlikely that the average user of a chatbot realizes that the data shared in their conversations could be used to infer the kinds of things the user might wish to purchase and how much they might be willing to pay, based on inferred mental states. It also seems unlikely that the average user realizes that they could then be shown ads for these very products, marketed at the highest price they are predicted to be willing to pay. To be sure, a company might self-regulate and restrict how data can be processed or used for advertising; indeed, after OpenAI's announcement that it would include ads on certain versions of Chat-GPT, the company pledged to do just that.<sup>70</sup> Moreover, other companies might use a lack of ads as a selling point, as Anthropic (a competitor to OpenAI) did in commercials during Super Bowl LX.<sup>71</sup> But this potential for self-regulation or market

67 See Carmen Molina Acosta & Lisa Weiner, *Artificial Intelligence Could Soon Diagnose Illness Based on the Sound of Your Voice*, NPR (Oct. 10, 2022, 05:00 ET), <https://www.npr.org/2022/10/10/1127181418/ai-app-voice-diagnose-disease> [<https://perma.cc/8RUQ-QC9P>].

68 See Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 382–84 & nn.103–10 (2022).

69 E.g., *Our Approach to Advertising and Expanding Access to ChatGPT*, OPENAI (Jan. 16, 2026), <https://openai.com/index/our-approach-to-advertising-and-expanding-access/> [<https://perma.cc/4YPV-TQJF>] (announcing that OpenAI will "start testing ads in the U.S. for the free and Go tiers, so more people can benefit from our tools with fewer usage limits or without having to pay").

70 *Id.* (OpenAI's ad principles include listing "Conversation privacy" as an "ads principle" and promising: "We keep your conversations with ChatGPT private from advertisers, and we never sell your data to advertisers.").

71 See Dawn Chmielewski, Deepa Seetharaman & Max A. Cherney, *Anthropic Buys Super Bowl Ads to Slap OpenAI for Selling Ads in ChatGPT*, REUTERS (Feb. 7, 2026, 06:04 ET), <https://>

differentiation does not change the baseline fact that chatbots are serving as data collectors, in ways that open the door to more information generation about the individual user.

Furthermore, the privacy via generative AI challenge can arise even outside of a user's interactions with a particular system. Specifically, AI-driven analysis can be applied to data about a person that was collected elsewhere in order to infer other sensitive information about that individual.<sup>72</sup> Researchers at ETH Zurich reveal that state-of-the-art chatbots can discern sensitive details, like age or geography, from large, publicly available data sets, such as Reddit posts.<sup>73</sup> For instance, if a user mentions a "hook turn" in a public post in an online forum, it's possible to predict that the user is located in Melbourne, Australia, where that term refers to a particular kind of intersection.<sup>74</sup> This is an example of the privacy via generative AI challenge because the generative AI system offers a new way to predict this additional information, even though the user didn't explicitly state it, and even though the user may not even have realized that they were disclosing this potentially private information.

These inferential capabilities are only likely to increase with further technological advances. For instance, recent research indicates that large language models can deanonymize individuals by "(1) matching an online account to its real identity; (2) linking an identity to an unknown pseudonymous account; and (3) linking pseudonymous accounts of the same person across different platforms or time periods."<sup>75</sup> Tabling concerns about adversarial attacks that leverage these capabilities, which are beyond this Article's scope, the instant point is the extent to which AI tools can reveal information about people. One could imagine, for instance, ordinary use of an AI tool by a marketing firm to track a person's activities across the web, infer their

---

[www.reuters.com/business/media-telecom/anthropic-buys-super-bowl-ads-slap-openai-selling-ads-chatgpt-2026-02-07/](https://www.reuters.com/business/media-telecom/anthropic-buys-super-bowl-ads-slap-openai-selling-ads-chatgpt-2026-02-07/) [<https://perma.cc/Y4AS-REKF>].

<sup>72</sup> Solow-Niederman, *supra* note 68, at 406–07.

<sup>73</sup> Robin Staab, Mark Vero, Mislav Balunović & Martin Vechev, *Beyond Memorization: Violating Privacy Via Inference with Large Language Models*, in PROCEEDINGS OF THE INT'L CONF. ON LEARNING REPRESENTATIONS 3 (2024), <https://arxiv.org/pdf/2310.07298> [<https://perma.cc/AW48-EY2D>].

<sup>74</sup> *Id.* at 1.

<sup>75</sup> Simon Lermen, Daniel Paleka, Joshua Swanson, Michael Aerni, Nicholas Carlini & Florian Tramèr, *Large-Scale Online Deanonymization with LLMs*, ARXIV 1 (Feb. 25, 2026), <https://arxiv.org/pdf/2602.16800> [<https://perma.cc/QX5-EADS>]; see also Dan Goodin, *LLMs Can Unmask Pseudonymous Users at Scale with Surprising Accuracy*, ARS TECHNICA (Mar. 3, 2026, 07:30), <https://arstechnica.com/security/2026/03/llms-can-unmask-pseudonymous-users-at-scale-with-surprising-accuracy/> [<https://perma.cc/4APQ-Y2HQ>] (discussing study).

likely tastes and preferences, and develop an even more granular profile of that person.

Even if some of this inferential capability was possible with other kinds of machine learning—or even perhaps with skilled human researchers provided with enough time and resources—there is a distinct set of pressure points when the cost of making privacy-invasive inferences like this is falling, at the same time that these technologies are becoming more prevalent throughout society.<sup>76</sup> The ease of deriving these and similar insights is new, as is the way that human beings are apt to engage with chatbot and AI companion systems as if they were human beings. These economic and social developments amplify the potential impact of these technologies, making it all the more vital to consider how well particular interventions fit particular types of challenges. The next Part takes up that task.

### III. CALIBRATING INTERVENTIONS TO CHALLENGES

This Part focuses on potential responses to privacy challenges in, out of, and via generative AI systems. Section III.A evaluates existing legal and technical responses and explains how they are unsatisfying, at best, and unresponsive, at worst. Section III.B then analyzes which combination of legal and technical interventions can best respond to each type of challenge.

#### A. *Exposing the Limitations of Existing Response Modes*

This Section considers why it's prudent to tailor interventions to a particular category of challenge. Briefly, no one legal or technical approach adequately responds to all the ways that people and generative AI technologies interact post-deployment, and insisting otherwise is misguided.

##### 1. Legal Risks

On the legal side, consumer privacy law's contemporary approach to privacy protection is ill suited to resolve the distinct varieties of challenges that arise in, out of, and via generative AI.<sup>77</sup> As presently

---

<sup>76</sup> See Solow-Niederman, *supra* note 68, at 390–94; Harry Surden, *Structural Rights in Privacy*, 60 SMU L. REV. 1605, 1607–08 (2007).

<sup>77</sup> I focus here on the “consumer protection” understanding that dominates American privacy law, especially at the federal level, and which regulates how private entities may collect and use personal data. In the United States, the Fourth Amendment controls government collection and use of personal data. The picture differs internationally. See Anupam Chander, Margot E. Kaminski & William McGeeveran, *Catalyzing Privacy Law*, 105 MINN. L. REV. 1733, 1747 (2021) (discussing the European “data protection” approach, which

operationalized in American law, it's fair to say information privacy regulation foregrounds one question: whether a person has enough control over information that is associated with that individual.<sup>78</sup> In most settings, the core idea is to provide a consumer with a statement of the company's privacy practices (notice) and obtain individual consent (choice) to those practices in exchange for access to a good or service, thereby permitting the user to exercise control over information that is associated with them. For instance, information that I disclose while using a social media platform, wearing a smart device, or engaging with an agentic AI system is covered by the company's terms of service and privacy policy—but unless it falls within a category deemed especially sensitive, such as health data or financial data, and unless the specific conditions of the applicable sectoral statute are met, there will not be further federal protection of that data. It's up to me to make autonomous choices about it. Or so the theory goes.

As I explore below in Section III.B, this retail, control-centered approach might be helpful to contend with some privacy challenges; however, it does not provide a satisfying response to privacy challenges that involve third parties, nor to privacy out or privacy via challenges.<sup>79</sup> A privacy-as-control paradigm, which centers on empowering a given individual to choose when (not) to allow a company to collect or process their own, personal information, doesn't address outputs that implicate the privacy interests of other people. Recall the not-so-hypothetical example of a real person's phone number

---

applies to both public and private use of data and “proceed[s] from the principle that data protection is a fundamental human right”). For further description of differences between the consumer protection and data protection approaches, see *id.* at 1747–49. A minority of American states have adopted legislation that embraces certain “data protection” principles, including but not limited to the right to access data, the right to correct data, and purpose limitation measures, but the U.S. has not conceptualized information privacy or data protection as a fundamental human right in the same way as the E.U. *Which States Have Consumer Data Privacy Laws?*, BLOOMBERG L. (Apr. 7, 2025), <https://pro.bloomberglaw.com/insights/privacy/state-privacy-legislation-tracker/> [<https://perma.cc/79GH-V55D>].

78 See, e.g., Julie E. Cohen, *Turning Privacy Inside Out*, 20 THEORETICAL INQUIRIES L. 1, 3 (2019) (“Perhaps the dominant justification for privacy is that it promotes and protects individual autonomy.” (first citing BEATE RÖSSLER, *THE VALUE OF PRIVACY* (R.D.V. Glasglow trans., 2d ed. 2018); and then citing Anita L. Allen, *Coercing Privacy*, 40 WM. & MARY L. REV. 723, 738–40 (1999))); ARI EZRA WALDMAN, *PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE* 29–33 (2018) (discussing dominant literature on privacy as “[a]utonomy, [c]hoice, and [c]ontrol,” *id.* at 29); Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609, 1613 (1999) (identifying “the traditional liberal understanding of information privacy, which views privacy as a right to control the use of one’s personal data”); see also Salomé Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573, 598–600 (2021) (surveying “Traditional Accounts: Privacy as Control and Access,” *id.* at 598).

79 But see Michael Birnhack, *In Defense of Privacy-as-Control (Properly Understood)*, 65 JURIMETRICS 143, 144 (2026) (“I argue that PaC [privacy as control] is still the best, albeit imperfect, normative lens we currently have to explain privacy law.”).

or address being provided in response to another person's chatbot query. Privacy as control is also poorly suited for situations in which the tool's outputs are privacy invasive.

For similar reasons, privacy as control is a bad fit for privacy via generative AI challenges. In what I have previously called the "inference economy,"<sup>80</sup> scores of data points from many people are processed to develop inferences and predict both that individual's behavior and the behavior of unrelated third parties.<sup>81</sup> Privacy as control does little for these unrelated third parties, even though they may be the individuals affected by privacy via generative AI challenges. Again, control doesn't help if the person who is affected by a data-driven AI inference is not the individual who has control over the data that was used to develop the model. Furthermore, control arguably does a bad job empowering the original individuals who are tasked with controlling their data, yet likely face a "context challenge" and are unable to predict whether a bit of data that they disclose will later be collected, aggregated, and processed to discern further information about them.<sup>82</sup> Giving someone control over their personally identifiable information, at time one, doesn't help if the privacy challenge involves what an AI model can do with that data, at time two.

Accordingly, both at the federal level and in most states, there are not overarching interventions that engage with the range of privacy and generative AI challenges.

## 2. Technological Risks

Given the limitations of legal interventions, it's tempting to look to privacy enhancing technologies (PETs) as an alternative approach. But turning to technology for a solution risks skipping essential analytic steps about which kinds of challenges a particular intervention can (not) solve and what understandings of privacy it does (not) serve.<sup>83</sup>

To illustrate, consider one PET, federated learning.<sup>84</sup> This PET is relevant for AI systems because training an AI model requires the

---

80 See generally Solow-Niederman, *supra* note 68.

81 For an extended discussion of this point, see *id.* at 384–87.

82 See *id.* at 378–84.

83 This Article adopts Herbert Burkert's early definition of PETs as "technical and organizational concepts that aim at protecting personal identity." Herbert Burkert, *Privacy-Enhancing Technologies: Typology, Critique, Vision*, in TECHNOLOGY AND PRIVACY: THE NEW LANDSCAPE 125, 125 (Philip E. Agre & Marc Rotenberg eds., 1997).

84 For an overview of federated learning, see Brendan McMahan & Daniel Ramage, *Federated Learning: Collaborative Machine Learning Without Centralized Training Data*, GOOGLE RSCH.: BLOG (Apr. 6, 2017), <https://research.google/blog/federated-learning-collaborative-machine-learning-without-centralized-training-data/> [https://perma.cc/4W5A-8Y75]; see also Nicola Rieke et al., *The Future of Digital Health with Federated*

collection and processing of massive quantities of data. Federated learning changes how the processing is done: rather than centralizing data in a single location, it processes information in a decentralized fashion.<sup>85</sup> For instance, instead of engaging with a third-party processor, such as a cloud provider, a federated learning approach might allow a company to process data on a user's own laptop or mobile device.<sup>86</sup> Assuming that it is technically feasible and reliable in a given context, federated learning can shield the user's information and limit which parties can access user data. Furthermore, by reducing the number of actors who have access to the user's data, federated learning can enhance data security.

However, because it focuses on how to process data that has already been collected, there are things that federated learning does not do.<sup>87</sup> On its own, it doesn't specify whether there should be limits on data collection in the first instance.<sup>88</sup> It doesn't provide additional protections for sensitive types of information, such as health data or financial data. And it doesn't place limits on the other entities with whom the original collector can share the underlying data.

Given these limitations, federated learning cannot address every kind of generative AI privacy issue. Federated learning would change where the data is processed, which might enhance data security by making it harder for would-be hackers to access and breach the dataset. That intervention doesn't help, though, if the privacy challenges arise during ordinary use. And because it is focused on a particular user, it also can't attend to privacy challenges that affect unrelated third parties, particularly when it comes to the inferences that feature in the privacy via generative AI challenge. Additionally, there are practical concerns about how well federated privacy can work as applied to generative AI. Notably, because federated learning can still allow sensitive data to be reverse engineered from information about the AI model, it risks creating a façade of protection that is greater than the one that exists in practice.<sup>89</sup> Unthinking application of federated learning as a technological solution thus risks failing to consider whether privacy protections apply to all of users' post-deployment interactions with AI tools, in ways that go beyond a concern with whether information is

---

*Learning*, 3 NPJ DIGIT. MED., Sep. 14, 2020, at 1, 2–3 (discussing potential value of federated learning for healthcare).

85 See McMahan & Ramage, *supra* note 84.

86 See, e.g., *id.*

87 See Marieke Bak et al., Comment, *Federated Learning Is Not a Cure-All for Data Ethics*, 6 NATURE MACH. INTEL. 370, 370 (2024) ("It is important to realize that by itself, FL [federated learning] offers no privacy protections.").

88 See *id.*

89 See, e.g., *id.* at 370.

amply shielded and secure. Insisting that federated learning solves all issues risks techno-solutionism. The better path, which I consider below in Section III.B, is to try to pair specific techniques such as federated learning with other privacy protections, with an eye to a specific type of post-deployment challenge.<sup>90</sup>

Before developing this approach, a moment of candor: I recognize that regulatory tailoring of this sort is nuanced and complex. That's potentially risky, insofar as simple interventions are more practical options, particularly in a world of over-burdened, under-resourced regulators and politically gridlocked governing bodies. But here, nuance and complexity are necessary to contend with the discrete types of challenges that occur as people interact with generative AI systems. Simplicity is tempting—but not everything is simple.

### *B. Aligning Regulatory Interventions and Post-Deployment Privacy Challenges*

This Section serves as proof of concept, mapping how greater specificity about the nature of the challenge can and should inform the choice of intervention. Careful work of this sort will foster better alignment of regulatory objectives and outcomes,<sup>91</sup> permitting scholars and policymakers to assess how well a given regulatory intervention addresses a given issue, if at all.

#### *1. Privacy In Generative AI*

Legal and regulatory reforms have the most purchase for the privacy in generative AI challenge. Recall that this category of issues involves inputting sensitive or confidential data into a generative AI system, and that it may implicate the interests of the original user, an unrelated third party, or both. When the concern is the privacy of the original user, refinements of existing law have real promise in contexts that involve sectoral privacy protections, such as health privacy and financial privacy.

User interactions with generative AI systems provide fresh opportunities to respond to longstanding weaknesses in sectoral

---

90 See *infra* Section III.B. Other PETs, such as differential privacy, warrant similar analysis. See Jeremy Seeman & Daniel Susser, *Between Privacy and Utility: On Differential Privacy in Theory and Practice*, 1 ACM J. ON RESPONSIBLE COMPUTING, Mar. 2024, at 1, 2 (providing analysis of differential privacy that is attuned to power dynamics and sociotechnical complexity).

91 See Neel Guha et al., *AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing*, 92 GEO. WASH. L. REV. 1473, 1485–90 (2024) (identifying and analyzing AI's "regulatory alignment problem," *id.* at 1485).

privacy protections. A pressing example involves the Health Insurance Portability and Accountability Act of 1996 (HIPAA),<sup>92</sup> which only applies to “covered entities” such as doctors, other health care providers who conduct business electronically, and insurance companies, or certain “business associates” of such covered entities.<sup>93</sup> HIPAA regulates the flow of health information within the healthcare context, traditionally understood. The average person might think that their wearable device, or health tracking app, or even their online conversations about their own, personal health information are covered by HIPAA and its associated privacy rules. That person would be wrong, as a matter of law. HIPAA doesn’t apply to direct-to-consumer health interactions that fall outside of conventional healthcare settings, including questions that a user poses to a generative AI tool.<sup>94</sup>

Now, the Health Breach Notification Rule (HBNR), which is enforced by the FTC, might have purchase outside of HIPAA,<sup>95</sup> and it might be a flexible, adaptive way to protect consumers as technology changes.<sup>96</sup> But it is not a good way to cover all of the latent weaknesses that technological advances place in particularly vivid relief. The HBNR is a procedural, notice-focused regulation that doesn’t provide substantive protections for sensitive health data. Specifically, it requires the company to notify a consumer of a “breach of security” in which “unauthorized acquisition of identifiable health information . . . occurs as a result of a data security breach or an unauthorized disclosure.”<sup>97</sup> The regulation does not offer substantive limits to protect the sensitive health information a person shares through a smartwatch, an activity tracker, a fertility-tracking app—or, say, a chatbot. The upshot is that

---

92 Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered titles of the U.S. Code).

93 See *Collecting, Using, or Sharing Consumer Health Information? Look to HIPAA, the FTC Act, and the Health Breach Notification Rule*, FTC BUS. GUIDANCE (Aug. 2024), <https://www.ftc.gov/business-guidance/resources/collecting-using-or-sharing-consumer-health-information-look-hipaa-ftc-act-health-breach> [<https://perma.cc/7HXY-LKKP>].

94 See Claudia E. Haupt & Mason Marks, Opinion, *AI-Generated Medical Advice—GPT and Beyond*, 329 J. AM. MED. ASS’N 1349, 1350 (2023).

95 Press Release, FTC, FTC Finalizes Changes to the Health Breach Notification Rule (Apr. 26, 2024), <https://www.ftc.gov/news-events/news/press-releases/2024/04/ftc-finalizes-changes-health-breach-notification-rule> [<https://perma.cc/6F2P-J3UJ>].

96 See, e.g., Complaint for Permanent Injunction, Civil Penalties, and Other Relief, United States v. GoodRx Holdings at 10–18, No. 23-cv-460 (N.D. Cal. Feb. 1, 2023) (alleging that GoodRx “Shared Sensitive Information About Millions of Users with Advertising Platforms or Other Third Parties” in violation of the Health Breach Notification Rule through technological means and discussing company’s integration of “tracking tools (including automated web beacons called tracking ‘pixels’ and other automated trackers called ‘Software Development Kits’ (‘SDKs’)) [that] collect and send data to third parties so that they can provide advertising, data analytics, or other business services to the owner of the website or mobile app”).

97 Press Release, FTC, *supra* note 95.

generative AI tools will put tremendous pressure on already-strained lines in health privacy law.<sup>98</sup>

Rather than concede defeat, it's possible to recognize this pressure as an example of the privacy in generative AI challenge, wherein sensitive data is fed into a generative AI tool. Seen this way, the question is whether existing statutory protections adequately cover the interaction—and if not, how they might be adapted or reformed, even without reworking the entire regulatory apparatus.

Start with HIPAA. Suppose that a doctor finds that the pressures of modern medical practice have squeezed out time for valuable patient interactions and decides to use a chatbot to summarize their notes. HIPAA would apply here, because the doctor is a “covered entity” providing healthcare under the statute and the doctor's notes likely include “protected health information.”<sup>99</sup> But without additional steps, the doctor's use of the chatbot is likely problematic. Specifically, if the data is transferred to an external chatbot company for processing,<sup>100</sup> then the data is no longer contained within the healthcare system and there is a HIPAA violation.<sup>101</sup> Instead, doctors must ensure that the entity providing the chatbot is operating in compliance with the HIPAA framework, as a “business associate.” Moreover, regulators at the Department of Health and Human Services (which enforces HIPAA regulations) can and should carefully monitor the claims of HIPAA compliance that AI companies make, and regulators at the FTC and in state AG offices (which enforce federal and state unfair and deceptive acts and practices, or “UDAP,” laws) should not hesitate to enforce if the promised enterprise systems fall short, or if claims are deceptive or otherwise misleading.

A corollary set of regulatory reforms could address health data that falls outside of HIPAA's statutory ambit. To start, one piece of low-hanging fruit again draws on existing consumer protection law: companies should not be permitted to make it seem as if data is covered by HIPAA when it is not.

Generative AI systems are likely to produce further confusion about what is and is not covered by statutory healthcare protections.

---

98 Cf. Mason Marks & Claudia E. Haupt, Opinion, *AI Chatbots, Health Privacy, and Challenges to HIPAA Compliance*, 330 J. AM. MED. ASS'N 309, 309 (2023) (“HIPAA is ill suited to protect patients in the face of modern technologies . . . . Even if compliance were possible, it would not ensure privacy or address larger concerns regarding power and inequality.”).

99 See HIPAA General Provisions, 45 CFR § 160.103 (2013); see also Genevieve P. Kanter & Eric A. Packel, Opinion, *Health Care Privacy Risks of AI Chatbots*, 330 J. AM. MED. ASS'N 311, 311 (2023) (focusing on generative AI chatbots and potential HIPAA compliance issues).

100 See Kanter & Packel, *supra* note 99, at 311–12.

101 See *id.* at 311; see also Haupt & Marks, *supra* note 94, at 1349 (warning that, as of April 2023, “current versions of GPT are not HIPAA compliant and could jeopardize patient privacy”).

Consider, by way of example, Anthropic's January 2026 blog post announcing a "Claude for Healthcare" expansion, which touts the company's new "set of tools and resources that allow healthcare providers, payers, and health tech companies and startups to use Claude for medical purposes through HIPAA-ready products."<sup>102</sup> The post then continues, "We're also introducing tools to help individuals understand and navigate their personal health data."<sup>103</sup> Technically, the post does distinguish between the "HIPAA-ready products," which are made available to enterprise users in the healthcare system, and "tools to help individuals."<sup>104</sup> But the average person seems unlikely to make this distinction.

In addition, other company materials are likely to further muddy the HIPAA versus non-HIPAA line. Notably, the linked discussion on the "Claude for Healthcare" page seems targeted to enterprise users and states, "Claude is built on HIPAA-ready infrastructure with safety guardrails designed for healthcare."<sup>105</sup> Yet that same page also discusses "connectors"<sup>106</sup> that permit individual users to link the chatbot directly to "tools, databases, and applications."<sup>107</sup> The reasonable consumer might well think these connections are also covered by HIPAA. They might continue to think so even when Anthropic's materials on "Accessing and analyzing your health data," which are posted elsewhere on the site, include the disclaimer, "[w]e are not a medical device and cannot provide medical guidance. Always consult a licensed professional."<sup>108</sup> These perceptions matter because a person's belief that they are covered by statutory health privacy protections may inform what they share with the system. If the statements are misleading or otherwise induce sharing that an individual would not expect, then public law enforcers should step in and protect consumer interests. Regulatory clarity, tailoring, and strong enforcement of this sort would create space for beneficial uses of AI tools, without permitting companies to take advantage of individuals who rely on

---

102 *Advancing Claude*, *supra* note 40.

103 *Id.*

104 *See id.*

105 *Transform Healthcare*, *supra* note 40.

106 *See id.*

107 *Connectors*, CLAUDE, <https://claude.com/connectors> [<https://perma.cc/3YBX-X58M>] (last visited Apr. 15, 2026).

108 *See Using Claude with Android Apps*, CLAUDE SUPPORT, <https://support.claude.com/en/articles/11869629-using-claude-with-android-apps> [<https://perma.cc/UMG7-SCFB>] (last visited Apr. 15, 2026) (displaying that page was "[u]pdated over a month ago" at time of visit); *Using Claude with iOS Apps*, CLAUDE SUPPORT, <https://support.claude.com/en/articles/11869619-using-claude-with-ios-apps> [<https://perma.cc/UT9D-4EJ6>] (last visited Apr. 15, 2026) (displaying that page was "[u]pdated over a month ago" at time of visit).

health providers—inside and outside of the traditional healthcare system—for their care.

Alternatively, a more audacious reform effort would focus on the relationship between the original user and the company's tool. These reforms might expand explicit fiduciary duties and/or doctor-patient confidentiality to chats that involve healthcare data, even if such interactions are not within the conventional healthcare system that HIPAA regulates. This direct intervention would expand protections for a broader swath of privacy in challenges and counteract the risk that chatbots can offer seemingly authoritative, but harmful, health advice.<sup>109</sup> In this way, the privacy in challenge is an opportunity to think critically about how users engage with generative AI tools in not just healthcare, but also in other sensitive domains.

Admittedly, sectoral interventions won't solve every form of privacy in challenge. For one, such reforms will not address third party privacy interests. For another, strengthening current regulations won't help in domains that are not covered by existing sectoral protections. My goal here is not to propose silver-bullet solutions that would cover all privacy in challenges, but rather to emphasize how analytic precision allows us to calibrate reforms to address a particular type of challenge and, just as importantly, to recognize which issues a particular intervention leaves unresolved.

As proof of concept, return to the question of how to respond to privacy in challenges that involve one individual and their data, outside of coverage of a sectoral privacy statute. When we look beyond sectoral privacy protections, we can see that many such challenges involve the AI developer's design choices and the ways in which the user interacts with the deployed system.<sup>110</sup> To make this point more concrete, if a user's query becomes public, it's often the product of the company's previous decisions about the user interface. For example, in summer 2025, a number of Chat-GPT conversations—including chats on sensitive, intimate, or otherwise confidential topics—were accessible on public search engines.<sup>111</sup> Such sharing was not public by default; rather, as reporting explained at the time, “[a] conversation would be

---

109 See Claudia E. Haupt & Mason Marks, Opinion, *FTC Regulation of AI-Generated Medical Disinformation*, 332 J. AM. MED. ASS'N 1975, 1975 (2024).

110 On privacy and the importance of design choices, see generally WOODROW HARTZOG, *PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES* (2018); Ann Cavoukian, *Privacy by Design*, INFO & PRIV. COMM'R OF ONT., <https://www.ipc.on.ca/en/media/1826/download?attachment> [<https://perma.cc/8TVQ-U7L2>].

111 Amanda Silberling, *Your Public ChatGPT Queries Are Getting Indexed by Google and Other Search Engines*, TECHCRUNCH (July 31, 2025, 12:23 PT), <https://techcrunch.com/2025/07/31/your-public-chatgpt-queries-are-getting-indexed-by-google-and-other-search-engines/> [<https://perma.cc/TR9J-XXR3>].

appended with a ‘/share’ URL only if the user deliberately clicks the ‘share’ button on their own chat and then clicks a second ‘create link’ button.”<sup>112</sup> Then, “[a]fter clicking through to create a link,” a user could “toggle whether or not they want that link to be discoverable”—and, if they said yes, those links could be indexed by other search engines.<sup>113</sup> Weeks later, it became clear that the “share” feature on xAI’s Grok tool was similarly permitting public access to chats.<sup>114</sup> To OpenAI’s credit, it changed the settings after public outcry over these disclosures.<sup>115</sup> xAI did not. The broader point is not specific to any one company: across the board, these are the company’s design choices, from the outset.

This data sharing example highlights how seemingly technical choices about AI settings are anything but technical; rather, they mediate privacy outcomes for users of these systems. Seen this way, a new policy option becomes available: even outside of areas covered by sectoral privacy statutes, a company can and should model user engagement with a tool and the likely privacy implications. To ensure that companies do so, law can complement design, in the form of regulatory mandates requiring such privacy assessments. At best, these assessments should occur pre-deployment; in addition, thorough evaluation of these on-the-ground interactions, over time, must feed back into the upstream supply chain and inform developers’ choices. I return to these implications in Part IV.

## 2. Privacy *Out of* Generative AI

Responding to privacy out of generative AI challenges is trickier and requires pinpointing the underlying technological issue. First, combining technical and regulatory interventions could go a long way towards addressing memorization issues. These kinds of challenges are more technical, in the sense that they reflect choices made during production about how to develop the model and how to test it before deployment. They are more akin to the kinds of copyright infringement issues that Professors Lee, Cooper, and Grimmelmann assess in the generative AI supply chain;<sup>116</sup> here, they are appearing in the data privacy context.

To minimize the harms from privacy-invasive outputs, particularly given the effect on unrelated third parties, *ex ante* algorithmic impact

---

<sup>112</sup> *Id.*

<sup>113</sup> *Id.*

<sup>114</sup> Iain Martin & Emily Baker-White, *Elon Musk’s xAI Published Hundreds of Thousands of Grok Chatbot Conversations*, FORBES (Aug. 21, 2025, 10:47 ET), <https://www.forbes.com/sites/iainmartin/2025/08/20/elon-musks-xai-published-hundreds-of-thousands-of-grok-chatbot-conversations/> [<https://perma.cc/WQH9-4PFK>].

<sup>115</sup> Silberling, *supra* note 111.

<sup>116</sup> See *supra* notes 20–21 and accompanying text.

assessments warrant serious consideration.<sup>117</sup> Any such impact assessment would be an anticipatory, pre-deployment analysis of how a system is likely to operate. For example, if the concern is that a companion system's response might include sensitive information that pertains to third parties, then one important question is the likelihood of this result and the conditions that might produce it. Here, the developer as well as any downstream entity that refines the model would need to assess the probability that these issues will occur. Legal requirements, such as a maximum acceptable error rate for a type of system (like chatbots, or companion AI tools, or health tools), could reduce the risk of check-the-box compliance.<sup>118</sup> Granted, there is a further concern about overly prescriptive regulatory mandates in a rapidly changing space. But individuals will bear the cost if nothing is done. Careful future work is vital to determine which actors are most likely to have useful information and how to scope impact assessments to increase the likelihood of useful disclosures,<sup>119</sup> while minimizing the risk of ossified, onerous regulation.

Moreover, it is worth confronting privacy out and memorization challenges because they are relatively more amenable to intervention than privacy harms that result from generation challenges. Recall that a second kind of privacy out issue might occur through random, de-contextual outputs of personal or sensitive data. Again, it is difficult to know with certainty whether this is occurring, or whether a model is responding with information memorized in the training data. But to the extent that it is genuinely occurring, this issue is hard to confront

---

117 See ADA LOVELACE INST. & DATAKIND UK, EXAMINING THE BLACK BOX: TOOLS FOR ASSESSING ALGORITHMIC SYSTEMS 15–20 (2020), <https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf?v=1605882304> [<https://perma.cc/V6T6-DXSM>] (discussing the difference between pre-deployment “algorithmic impact assessments” and post-deployment “algorithmic impact evaluation”).

118 See Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 628–30 (2019). Another issue, which I table to focus on the complex sociotechnical dynamics, is careful assessment of the likely First Amendment challenges to such risk assessments under current doctrine. See Rebecca Aviel, Margot E. Kaminski, Toni M. Massaro & Andrew Keane Woods, *From Gods to Google*, 134 YALE L.J. 1269, 1292 (2025).

119 See, e.g., Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1389–1403 (2023) (analyzing the “construct[ion of] AI harms as risks,” *id.* at 1390); Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117 (2021) (analyzing algorithmic risk assessments, their potential, and their challenges); see also MAYU TOBIN-MIYAJI, ELEC. PRIV. INFO. CTR., ASSESSING THE ASSESSMENTS: MAXIMIZING THE EFFECTIVENESS OF ALGORITHMIC & PRIVACY RISK ASSESSMENTS 3 (2025), <https://epic.org/wp-content/uploads/2025/06/Assessing-the-Assessments-Report.pdf> [<https://perma.cc/DF5H-WYSM>] (advocating for risk assessments as “a key accountability mechanism”); ADA LOVELACE INST. & DATAKIND UK, *supra* note 117 (analyzing different forms of algorithmic inspection, how these measures differ, and open questions about them).

because it is emergent and endemic to the tool's operation. As of this writing, all generative AI entails a degree of randomness, by design.<sup>120</sup> That said, emergent does not mean entirely unpredictable.

Combining technical and procedural interventions through robust pre-deployment testing and assessment is again an auspicious step. For instance, there might again be a minimum amount of testing that deployers and adapters of a tool must conduct to establish that the likelihood of outputting personal information from another context is low. If there is uncertainty about the appropriate threshold, then negligence law might serve as a backstop, with courts determining what is reasonable when it comes to such outputs. There are downsides to relying on ex post adjudication, however; for instance, cases may exceed judicial capacity, or the common law may fall short of what is normatively desirable to protect individual rights. Thus, a bright-line rule might be preferable. For example, generative AI companies might engage in output filtering if the response links to external content, as appears to be the case in the Imgur selfie example, either with a soft law version of this requirement emerging bottom up through industry self-regulation or with a hard law version coming top down through traditional regulation. Again, the idea is not to provide a one-shot solution, but instead to curb the likelihood of privacy-invasive outputs that occur in the context of actual interactions with deployed systems.

### 3. Privacy *Via* Generative AI

Because they involve social dynamics and inference-generation over time, the category of privacy via generative AI challenges is the most difficult to confront. Here, the best approach is to consider the relationship between the user and the AI company with more precision. In prior work, I have focused on machine learning classification and prediction and distinguished between the “data collector,” which amasses the data, and the “information processor,” which derives inferences.<sup>121</sup> These entities can be the same, or different.<sup>122</sup> A similar distinction is useful in the generative AI context, too. Consider, by way of illustration, the prior example of a chatbot provider that advertises its ability to offer consumer-facing healthcare advice. In this example, the chatbot provider is both the data collector and the information processor, because it first amasses the data through chats and through connections to other user data, and then processes that data to infer that the user is likely to have a particular health condition (or not). Suppose that the chatbot provider instead collected all its chats with a user, and

---

120 See Cooper, Frankle & De Sa, *supra* note 60, at 2.

121 Solow-Niederman, *supra* note 68, at 406, 406–10.

122 *Id.* at 406–10.

then a different company accessed that set of data to derive healthcare predictions about the individual. In this second example, the chatbot provider would remain the data collector, but the other company would be the information processor.

The key to contending with the privacy via generative AI challenge is to craft interventions that respond to the distinct relational dynamics. This will be easiest, and clearest, when the data collector is the same as the information processor. Indeed, there are many legal hooks to consider, from imposing information fiduciary responsibilities, as Jack Balkin and Jonathan Zittrain have proposed for internet companies,<sup>123</sup> to considering a duty of loyalty, as Woodrow Hartzog and Neil Richards have proposed for privacy law.<sup>124</sup> It's also worth considering adaptations of conventional "fair information privacy" principles, such as the use of a strict purpose limitation principle to constrain a company's ability to generate inferences without the user's affirmative consent. That said, consent-based approaches are not a wise choice, on their own, because they place the burden solely on the individual despite a growing algorithmic accountability literature that recognizes the need for more systemic interventions. The right response comes down to social consensus: how much invasion of individual privacy is society willing to tolerate, in order to access AI systems? There is no easy answer to this, but reframing the problem in terms of social norms shifts attention to how to create conditions for democratic deliberation on this issue and how to integrate the perspectives of affected individuals in a meaningful fashion.

The deeper challenge is how to respond to instances in which the data collector and the information processor are not the same entity. Here, consumer protection law provides an important legal hook: whether through state attorneys general or through the Federal Trade Commission, robust UDAP enforcement can police situations in which data crosses contexts in ways that the reasonable user would not expect. For instance, if data disclosed in the context of a chat is then accessed by another entity to determine the price of a product, then that sharing might be seen as an unfair boundary transgression that harms the consumer financially. To facilitate such enforcement, data collectors

---

123 See, e.g., Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2016); Jack M. Balkin & Jonathan L. Zittrain, *A Grand Bargain to Make Tech Companies Trustworthy*, ATLANTIC (Oct. 3, 2016), <https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/> [<https://perma.cc/9J4Y-XBBG>]; Jack M. Balkin, *Information Fiduciaries in the Digital Age*, BALKINIZATION (Mar. 5, 2014, 16:50 ET), <https://balkin.blogspot.com/2014/03/information-fiduciaries-in-digital-age.html> [<https://perma.cc/P9J3-VP2D>].

124 See, e.g., Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961 (2021).

might watermark or tag their data so that it is clear if it flows outside of the expected setting, thereby combining technical and legal interventions. Steps such as these complement a robust literature on privacy as contextual integrity,<sup>125</sup> offering an additional response to the challenges presented by data-driven systems<sup>126</sup> and emerging technologies for which there are not yet established privacy norms.<sup>127</sup>

Calibrating interventions to the type of post-deployment privacy challenge might make regulation seem more complicated, but it in fact streamlines it. Regulatory interventions, whether technical, legal, or both, can be most effective when they are responsive to a particular issue. Moreover, as the next Part contends, tailoring an intervention to a type of challenge helps to ensure that the right actors are held accountable.

#### IV. FEEDING BACK: AI USERS AND THE SUPPLY CHAIN

This Part argues that the supply chain is a circle in which users' post-deployment interactions with AI systems should inform earlier design and deployment decisions. This approach is principled because downstream individuals are unable to make design choices that affect the likelihood that they will be harmed by an AI system's outputs. They lack the ability to control the outputs, or even, in most cases, to access or evaluate the technical steps taken to create the generative AI system. Doing nothing forces the users and humans affected by AI systems to bear the full cost of privacy challenges that occur as they use these tools. To be sure, there are hard questions on the margins; for example, how much autonomy should a user be deemed to have, if they choose to use an agentic system that permits extraordinary amounts of access to their data?

But there are also easy cases, and in these easy cases, this Article provides a missing conceptual framework to assess who should be held responsible. Take a situation in which the very design of the chatbot induces a person to feel comfortable sharing extremely personal data, and then further disseminates the data in ways that users are unlikely to expect. For instance, returning to a prior example, would someone who clicks "share" expect that their query would be indexed by a search engine, and then be publicly accessible? That's a design choice. It is made

---

125 See Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119, 124, 136–38 (2004) (proposing a theory of privacy as contextual integrity, which focuses on "appropriate," *id.* at 138, informational flows); HELEN NISSENBAUM, *PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE* (2010) (further developing theory).

126 See Nissenbaum, *supra* note 65, at 238, 248.

127 *Id.* at 232–34. As Professor Nissenbaum acknowledges, CI is a "conservative" approach, in the sense that it "presumptively favors entrenched norms" as it assesses the "appropriateness" of information flows. *Id.*

by the company or companies that develop and deploy the tool. Denying that point means the user is the only one who pays for that choice. Conversely, if an issue like this is framed as a form of the privacy in generative AI challenge, then it becomes possible to consider which actors, earlier in the process, could have made different choices to reduce the risk. The supply chain no longer ends with deployment. Instead, it is informed by the distinct ways that users' interactions with AI systems affect the information privacy of both those individuals and other people, as part of a cyclical process.

This approach empowers regulators to understand the supply chain as a circle. The core idea is to work backwards from user interactions with AI tools to earlier stages in AI development and deployment, seeking to craft targeted additional interventions at the right points upstream in the AI supply chain. Moreover, future scholarship might augment the supply chain literature in ways that inform these policy choices; for instance, open research questions include whether the relationship between the user and the AI deployer is one-way or bi-directional, or whether the user should be understood to have a relationship with the developer and the deployer, or only the deployer. My goal here is not to settle these questions, but rather to shift the terms of the conversation, with an eye to the ways that humans use the AI systems that companies produce.

## CONCLUSION

This Article reframes AI's supply chain as a circle and makes three primary contributions. First, it augments multiple literatures across both AI governance and information privacy law. My distinct focus on individuals' interactions with already-deployed systems offers a complementary approach to an expanding AI supply chain literature as well as a voluminous literature on privacy as contextual and relational. This Article's intervention also provides a fresh, nuanced angle on the relationship between AI and information privacy, a contested topic of considerable importance to scholars and policymakers alike. Second, this Article provides a generalizable, flexible analytic framework to assess the ways that people engage with generative AI systems and to understand different kinds of post-deployment privacy challenges. Third, and relatedly, it empowers scholars, developers, and policymakers to develop analyses and policy interventions that reflect the use of AI tools in social context. Rather than passively accepting issues that occur after model and system deployment, technologists, lawyers, and policymakers can and should work together to determine whether any of these issues can be mitigated during the production process. Unless we account for the contextual ways that humans

interact with AI systems, we cannot craft sound regulatory interventions to protect the values that make us human.

