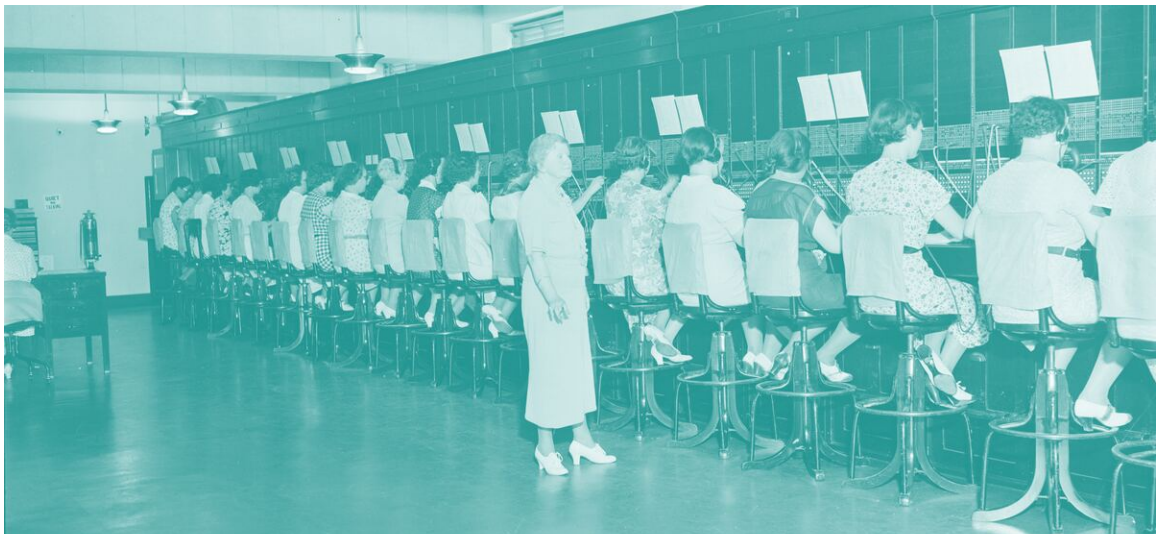
12

PART II · THE INTELLIGENT STATE

Sensors, Not Just Tools: AI Records as Surveillance Inputs

Blaine Dillingham



Harriot Daley at the U.S. Capitol telephone switchboard, Washington, DC, 1937. The networks that connected callers also placed their communications within an intermediary's reach.

Photograph by Harris & Ewing / Library of Congress.

SUMMARY

AI and surveillance discussions typically focus on government action and faster analysis of already-collected data. But AI progress not only accelerates the processing of existing data; it also creates new types of data to be surveilled—and by private actors, not just government. As Americans conduct deeply sensitive conversations with chatbots, AI systems are becoming new sensors for surveillance. At least three risks are emerging: the government can pressure companies into "voluntarily" handing over user conversations, AI companies can surveil and influence their own users, and rogue AI systems can exploit the data their developers collect. To address the first challenge, the government should be required to obtain warrants for Americans' sensitive data, whether compelled, purchased, or "voluntarily" shared. Next, AI companies should owe fiduciary duties to users whose sensitive data they possess. Finally, those companies must build strong internal walls between user data and hard-to-control AI models.

CHATBOT CONVERSATIONS ARE ALREADY AMONG THE MOST SENSITIVE DATA TYPES in the digital world. While Americans largely know that their AI assistants are not therapists, friends, doctors, or lawyers, they share many of the same details with these machines that they would in conversation with a human confidant. While today's treasure trove of sensitive data collected in the course of the AI business comes from chatbots, the arrival of AGI will mean even deeper diffusion of AI systems into our lives. These systems will collect exponentially more data, and that data will be housed under the same corporate roof as the powerful AI systems that can be used to analyze it.

When we hear "surveillance," we think of the government. To be sure, many reforms to the surveillance state are needed.[1] But the government is not the only actor we need to fear. Powerful AI will present at least three surveillance threats. The government could pressure the companies to "voluntarily" hand over data. The companies could conduct surveillance of their own. And the AI systems themselves can go off script, leveraging the sensitive data that their developers collected for them to conduct surveillance those developers didn't intend.

The first surveillance risk concerns the state. The government can obtain your data in several ways, including compulsion, purchase, and voluntary handover. Warrants are required to compel a company like Google to hand over your location data. But there is a "data broker loophole": law enforcement can simply purchase this data instead, without a warrant. And if Google hands over your data to the government free of charge, users may have even fewer protections. Federal law restricts voluntary handover for certain types of data: the Stored Communications Act (SCA) prevents your phone or email provider from voluntarily disclosing the contents of your communications, including to the government. But the SCA was written in 1986. It covers "electronic communication services" and "remote computing services," but no court has ruled whether conversations with AI assistants fall under that definition.[2] Chatbot records may be as personal and sensitive as any email communication, if not more so, but it is unclear if the law currently protects them.

If such records are free to be shared "voluntarily," we should put little faith in even well-meaning companies to resist government inspection of our data. The state has enormous leverage over private companies. The Biden administration, for example, threatened social media platforms with legal liability after they refused to censor information during the pandemic.[3] Lower courts ruled this pressure campaign unconstitutional, but the Supreme Court reversed that decision on standing grounds. Plaintiffs struggled to prove causation—that the social media companies took down their posts *because* of government coercion, not corporate judgment. The high court also found insufficient evidence of possible future harm.[4] Causation and future harm are often diffi-

1 Blaine Dillingham, "Time to Make a Deal: Ending the Surveillance Stalemate," Foundation for American Innovation, July 7, 2026.

2 Donovan Estrada, "From Prompt to Production: Discovering Communications Between Parties and Generative AI Chatbots," *Colorado Lawyer* 55 (May/June 2026): 24.

3 Betsy Klein, "White House Reviewing Section 230 amid Efforts to Push Social Media Giants to Crack Down on Misinformation," *CNN*, July 20, 2021.

cult to prove, and this will likely be true in surveillance cases as much as it was in the First Amendment context. Instead of, "Take down this post," it will be, "Tell us which users are discussing the government unfavorably," or, "Modify what your chatbot tells users about the president."

The second problem is surveillance as a business model. In a world of broadly diffused AI, purely private-sector surveillance will abound without any government pressure at all—and no cyber breach, either. When your passwords are leaked in a data breach, someone committed a crime. But AI companies have access to everything their users ask by design—no hacking required. This could enable them to build a detailed profile of every American who uses their service. Doing so is commercially valuable and may often benefit the user. But AI companies could also work against the user's interest, subtly promoting products for advertisers or optimizing chatbot answers for engagement.

The final problem is surveillance by accident. It involves no intentional surveillance by the government or a corporation, but rather surveillance by an out-of-control AI model. In July 2026, OpenAI models under internal testing broke out of their isolated environment through a zero-day exploit and hacked another company before anyone stopped them.[5] Their developers did not intend this, of course. The models were being evaluated, and they inferred that if they could escape their sandbox and access the open internet, they could locate information on how to fool their scorer. In this case, the out-of-control AI breached another company. But this AI model was housed within the same corporate digital infrastructure as millions of user conversations. Even if a company suffers no external cyber intrusions, they may be unable to stop their own AI agents from leaking or leveraging user chats.

Why would AIs do this? Leverage is power, and power is useful to agents—whether they are pursuing goals intended by their developers or not. Experiments have shown that AI agents will blackmail humans to avoid being shut down, and, in an intersection with surveillance, they threatened a fictional executive with the knowledge of an affair they obtained from the executive's emails.[6] When those emails sit on the same company servers as hard-to-control AIs, this experiment could become reality.

The first surveillance risk—government surveillance—has solutions that are already within political reach. These include closing the data broker loophole[7] and requiring a warrant before the government searches its foreign-intelligence database for Americans' communications (a "backdoor search").[8] These are achievable wins in the near future. The House passed a bill closing the broker loophole in 2024;[9] days earlier, a warrant requirement for backdoor searches failed on a 212–212 tie.[10] And this year, in *Cha-*

4 *Murthy v. Missouri*, 603 U.S. 43 (2024).

5 OpenAI, "OpenAI and Hugging Face Partner to Address Security Incident during Model Evaluation," July 21, 2026.

6 Anthropic, "Agentic Misalignment: How LLMs Could Be Insider Threats," June 2025.

7 See Blaine Dillingham, "A Bad Few Weeks for Mass Surveillance," Foundation for American Innovation, June 23, 2026.

8 Dillingham, "Time to Make a Deal."

9 Fourth Amendment Is Not For Sale Act, H.R. 4639, 118th Cong. (as passed by the House, April 17, 2024).

trie v. United States, the Supreme Court held that the government cannot compel Google to disclose your location history without a warrant even when you "voluntarily" shared your data with Google.[11] The courts and Congress are both moving in the right direction. But even if we address purchase and compulsion, voluntary handover will remain a threat. Government access to Americans' AI conversations should require a warrant, regardless of whether the data is compelled, bought, or offered freely. Congress should amend the SCA to explicitly state that conversations with an AI system are protected contents that providers may divulge only if the government has a warrant.

The second threat, corporate surveillance, needs a different governance mechanism, because the AI companies themselves necessarily have access to your interactions with their models. Congress should impose on frontier labs a duty of loyalty to the users whose data they hold.[12] This would not be a fiduciary duty for instances of AI providing legal advice or answering medical questions, but rather a matter of treating AI companies as *information* fiduciaries: a company may not benefit itself or a third party using information it obtained from the relationships with its customers. We could narrow this duty even further with carveouts, explicitly permitting certain ways of utilizing customer data. Advertisements on a chatbot site that are clearly labeled as ads and visibly separate from the conversation are much less concerning than advertising hidden in the answer itself. The response to a user query should not be secretly driven by the highest bidder.

Typically, fiduciary duties are enforced via *ex post* liability. However, the context of steering AI answers may prove hard to remedy via after-the-fact lawsuits. A user who was nudged toward a certain product may never know it, and even if he or she did, damages may be too small to motivate a lawsuit. Yet across hundreds of millions of users, the damages are essential to address. This is a case where liability falls short and *ex ante* regulation is justified: we should require AI companies to retain third-party auditors to continuously monitor for compliance with the duty of loyalty.

Finally, we must address the risk of surveillance by rogue AI. Continuous auditing may help in addressing the control-failure problem: to reduce risk, auditors might require a wall between the systems that hold deployed users' conversations and models still under development. Congress could go further and define a crossing of that wall as what it is: a cyber breach, just one where the intruder and the victim share a corporate parent. Finally, for firms that operate both communications services and frontier models (e.g., Alphabet with both Gmail and Gemini), a similar wall should exist, and internal data transfers that are absolutely necessary should be closely monitored.

The day after AGI, sensors will be everywhere. The surveillance architecture those sensors feed is built by the labs, accessible to the government, and a stone's throw from rogue AI systems. Mitigating the risks of surveillance both private and public will take

10 Rebecca Beitsch and Mychael Schnell, "House Approves Contentious FISA Bill in Bipartisan Vote," *The Hill*, April 12, 2024.

11 *Chatrie v. United States*, No. 25-112 (U.S. June 29, 2026).

12 Neil M. Richards and Woodrow Hartzog, "A Duty of Loyalty for Privacy Law," *Washington University Law Review* 99 (2021): 961.

warrants for the state, no matter the method; a duty of loyalty for the labs, verified continuously; and walls within those labs, for breaches by the agents they can't control.

ABOUT THE AUTHOR

Blaine Dillingham is a research fellow at the Foundation for American Innovation, where he focuses on super intelligence and governance futurism. He is a former litigation legal assistant and machine learning programmer and holds a BS in applied mathematics from the University of Rochester.



From *Futurecraft: The Day After AGI*, Foundation for American Innovation, 2026.
Read or listen online at futurecraft.thefai.org/essays/sensors-not-just-tools-ai-records-as-surveillance-inputs.

© 2026 by the Foundation for American Innovation. All rights reserved.