Artificial Intelligence Tools: Privacy Pitfalls

*Contributed by **William Ridgway**, **Ken Kumayama** and **Brian O'Connor**, Skadden, Arps, Slate, Meagher & Flom LLP*

August 2023

The rapid advance of generative artificial intelligence has captured the attention of regulators and new AI legislation is anticipated around the globe. Yet businesses that seek to use AI tools should also be mindful of the many ways that AI tools can implicate laws that are already on the books, especially when it comes to data privacy.

Use of Personal Information

Businesses that use AI should keep applicable privacy regulations in mind when doing so. AI tools can generate new content—such as text, images, and videos—through learning patterns from data, which is often referred to as "training" and "fine-tuning." Often, an AI provider "trains" the AI tool it provides, and users of that tool do not. But users of an AI tool may sometimes engage in "fine-tuning" of the tool. Users can also input information into a tool for analysis and other purposes. If the data an AI tool relies on for any of these activities includes personal information—whether as part of its training or fine-tuning or as a prompt or other input from the user—that may trigger various U.S. privacy law requirements.

The **Federal Trade Commission Act** regulates both "unfair" and "deceptive" acts, and the potential application of those standards to AI is quickly evolving. See **Chatbots, deepfakes, and voice clones: AI deception for sale**. State laws such as the **California Consumer Privacy Act** (CCPA) also require that businesses disclose how they use personal information and regulate when companies may take personal information collected for one purpose and reuse it for another. In some cases, businesses must give consumers the right to opt-out from the use of their personal information for use with AI or other automated decision-making systems.

Companies should carefully analyze these requirements before using AI tools on personal information collected from employees, vendors, or customers, especially if doing so in ways that may deviate from the purpose of the initial collection. That may prove challenging with AI tools—especially generative AI—because the same data could be used to train different algorithms for various purposes, and a tool's algorithm can use data in unpredictable—and even unknowable—ways.

Other privacy issues arise when AI providers use data from AI users—which may include personal information—to further train their AI tools. Typically, AI providers require users to secure consents from customers and satisfy any other legal requirements to allow the vendor to use the information for such training purposes. Businesses using AI tools should be aware that in certain states, the "sale" or "sharing" of personal information to third parties—which might include allowing service providers to use personal information for further training of AI models—may also create notice and opt-out requirements and may, in certain instances, be subject to opt-in requirements or be prohibited by applicable law.

One predictor for litigation over the *use* of AI tools is found in recent suits brought against *creators* of these tools. In **In the Matter of Everalbum and Paravision**, the Federal Trade Commission (FTC) took action against a company that drew on its users' personal information in violation of its privacy policies to train an AI facial recognition model. Notably, in its **order** the FTC required the company to delete AI models that were trained using personal information they did not have permission to use, and to destroy all such data—a dramatic remedy known as "algorithmic disgorgement" that can result in significant costs.

Private plaintiffs have also attempted to bring similar claims, although general state privacy laws such as the CCPA typically do not allow private rights of action—or if they do, they are limited to narrow circumstances. Plaintiffs have claimed, for example, that AI tools violate the CCPA's use, notice and opt-out provisions, such as in ***Doe 1 v. GitHub, Inc.***, where plaintiffs sued over GitHub's AI program that assists programmers by "suggest[ing] code," which was trained on "'billions of lines' of publicly available code, including code from public GitHub repositories." **2023 BL 169580**, *2 (N.D. Cal. May 11, 2023). Plaintiffs claimed that GitHub "improperly used Plaintiffs' 'sensitive personal data' by incorporating the data into [the program] and therefore selling and exposing it to" third-party users of the program in violation of the CCPA. The court dismissed the privacy claims for failing to specify the sensitive data at issue, leaving uncertain how courts will evaluate claims where specific private information is involved.

Although early AI litigation has targeted creators of the tools, businesses that use AI tools may face potential liability as well, especially given the representations, warranties and indemnification rights typically found in the terms of use for these tools, which often place responsibility on the user of the tool for complying with applicable laws. See, e.g., **GitHub Docs' terms of service**. In cases where vendors use their customers' data to train and improve the vendors' AI models, vendors will typically require the customers to secure all requisite permissions for the vendor to use the customers' data for such purposes, and then require that the customers indemnify the vendor if the vendor is sued as a result of the vendor's use of such data. See, e.g., **Poe's terms of service**. This risk is even more pronounced where the AI tools are included as a component or feature in a company's product or service offerings.

Consumer Rights to Delete, Correct, or Access

Several state privacy laws also give consumers the right to delete or correct their personal information. Laws creating such rights include the recently enacted **Virginia Consumer Data Protection Act**, **Colorado Privacy Act**, and **Utah Consumer Privacy Act**, as well as the CCPA. Complying with these rules can be difficult or impossible for AI programs, which often cannot remove or "unlearn" individual pieces of data. Instead, the program would likely need to be "retrained" without that data point—a process that would typically be impracticable from a cost and efficiency perspective. As with other data privacy requirements, it is uncertain how courts will analyze these provisions in the AI context. The *GitHub* case, for example, also involved a claim that its AI program failed to offer a right to alter or delete the personal information that its tool trained on—a claim that the court dismissed before reaching the merits.

Businesses should bear in mind that any given data set may persist as a part of an algorithm's training source, resulting in unpredictable forms of reproduction. As a result, they should exercise caution before inputting any personal information into AI tools. Given the potential remedy of algorithmic disgorgement, which has now been imposed by the Federal Trade Commission on three separate occasions, companies investing heavily in AI systems should take special care when including personal information in training data sets. In certain cases, it may be possible to eliminate personal information from a data set through anonymization techniques or use of synthetic data for training purposes. Companies that do intend to use personal information for purposes of training AI models should consider whether privacy harms may be mitigated through the use of privacy-enhancing technologies such as differential privacy or pseudonymization.

Collection of Personal Information

AI tools that collect or use sensitive or private information may also risk constitutional and common law privacy claims. California, for example, **allows claims** for "unwanted access to data by electronic or other covert means, in violation of the law or social norms." ***In re Google Inc. Cookie Placement Consumer Privacy Litig.***, **806 F.3d 125**, 151 (3d Cir. 2015). In recent years, plaintiffs have brought privacy claims based on the use of un-permissioned personal information to train AI models. In one **suit** against Clearview AI, the plaintiff claimed that the collection of "sensitive and confidential" biometric information scraped from publicly available photographs violated the California Constitution. *In re Clearview AI, Inc., Consumer Priv. Litig.*, **585 F. Supp. 3d 1111**, 1129–30 (N.D. Ill. 2022).

AI tools may also enable novel and expansive data collection techniques that could risk claims about subverting consumers' expectations of privacy. For example, courts applying California law have **explained** that whether an expectation of privacy is reasonable turns on context including applicable "customs" and the opportunity for notice and consent. *Google Cookie Placement*, **806 F.3d at 151**. Critical to the analysis is whether businesses abide by their representations about data collection and use. For example, in one case, the court reasoned that because allegations that Google tracked users by "overriding the plaintiffs' cookie blockers, while concurrently announcing in its Privacy Policy that internet users could reset your browser to refuse all cookies" were "[c]haracterized by deceit and disregard," they raised "different issues than tracking or disclosure alone." *Id.* at 150. See also *Calhoun v. Google LLC*, **526 F. Supp. 3d 605**, 631 (N.D. Cal. 2021), rejecting argument that data collection "served a legitimate commercial purpose," because the commercial purpose was one factor among many and Google's "surreptitious" practice that contravened its representations outweighed that purpose.

Businesses should thus carefully consider their public disclosures and strive for transparency about their use of AI tools, especially if consumers are unlikely to expect those uses. Similarly, businesses should consider whether any public disclosures are required regarding the use of consumers' personal information to develop or train AI systems, whether by businesses or their vendors.

Businesses should also consider whether implementing AI systems may run afoul of any wiretap laws—particularly for tools that are public-facing such as on a website. A wave of recent California wiretap litigation has centered on third-party "live chat" tools on customer-facing websites that allegedly capture and analyze communications between a website and its users. See, e.g., *Licea v. Old Navy, LLC*, No. 5:22-cv-01413-SSS-SPx, **2023 BL 132425** (C.D. Cal. Apr. 19, 2023); ***Byars v. Goodyear Tire & Rubber Co.***, No. 5:22-cv-01358-SSS-KKx, **2023 BL 42222** (C.D. Cal. Feb. 3, 2023).

Biometric Privacy

Certain AI programs may rely on the collection or processing of biometric data or enable the identification of individuals based on a given prompt. These programs can implicate biometric privacy laws such as the **Illinois Biometric Information Privacy Act** (BIPA) that regulate the collection and use of such data.

Unlike the general state privacy laws such as the CCPA, BIPA has a broad private right of action that has been bolstered by several favorable decisions by Illinois courts. For example, in *McDonald v. Symphony Bronzeville Park, LLC*, **2022 IL 126511** (Feb. 3, 2022), the Illinois Supreme Court ruled that BIPA violations are not preempted by the Illinois Workers Compensation Act, and in *Rosenbach v. Six Flags Entertainment Corp*, **2019 IL 123186** (Jan. 25, 2019), it held that anyone whose rights under BIPA were violated qualifies as "aggrieved." Accordingly, it has generated a high volume of civil litigation in recent years. Businesses should pay careful attention to any use of AI tools that implicates biometric privacy.

AI tools have featured in biometric privacy cases across different contexts. In one case, a plaintiff alleged that McDonald's collected biometric "voiceprints" through its automated ordering system, using AI "to identify unique customers regardless of which location they visit and present them certain menu items based on their past visits." Carpenter v. McDonald's Corp., **580 F. Supp. 3d 512**, 517 (N.D. Ill. 2022). The court denied a motion to dismiss, in part because "the technology [could] effectively interpret and understand customer orders," which showed "that it detects and analyzes human speech in a way that a mere recording device does not." In other recent cases, operators of mobile applications that allegedly use AI "to extract a person's face from a photo" and generate new types of images have been sued for collecting biometric data in the form of "facial geometry" from their users. *Gutierrez v. Wemagine.AI LLP*, No. 21 C 5702, **2022 BL 28418**, at *1 (N.D. Ill. Jan. 26, 2022); see also ***Flora et al v. Prisma Labs, Inc.***, Docket No. 23-cv-00680 (N.D. Cal. Feb. 15, 2023).

Vendor Contracts & Terms of Service

A business decision about using AI should also align with any applicable privacy-related contractual obligations. Many businesses have vendor contracts, terms of service, and other obligations that guarantee that personal information will be kept confidential. Again, when personal information is entered into an AI tool, the tool's algorithm may use that data in unpredictable ways that could exceed what the business has permission to do under applicable terms. Plaintiffs have already brought claims against AI creators for violating contractual obligations in the form of privacy policies and terms of service. See ***GitHub***; ***Andersen v. Stability AI Ltd.***, Docket No. 3:23-cv-00201 (N.D. Cal. Jan. 13, 2023). Users of these tools face similar risks with respect to privacy obligations in their own contracts and policies.

Moreover, as noted above, AI provider contracts typically require the providers' customers to obtain all necessary rights and consents to provide the personal information to the provider for the provider's use—which may include further improvement of the provider's own AI systems. Companies considering the use of AI tools should therefore update their vendor due diligence processes to identify such risks. Where companies intend to rely on third-party AI models in any material aspect of their businesses, such companies should consider negotiating and entering into an enterprise license, as opposed to relying on the standard vendor terms, which may be updated by the vendor with little notice.

Takeaways

Even though new AI-specific laws are expected in the years ahead, one should not ignore the existing laws and regulations these tools may implicate. Indeed, federal agencies including the Consumer Financial Protection Bureau, Department of Justice, Equal Employment Opportunity Commission, and Federal Trade Commission recently issued a **joint statement** noting that each undersigned agency has enforcement authority that applies to automated systems and pledging "to vigorously use our collective authorities to protect individuals' rights regardless of whether legal violations occur through traditional means or advanced technologies."

Businesses looking to use AI should scrutinize how such tools will collect and use data, including any protections for personal information, and obtain any appropriate consents required to use that data. Businesses contracting with AI providers should also consider requirements for appropriate privacy protections and indemnification for potential violations of privacy laws. And businesses should develop and implement a rigorous vendor diligence process for AI providers before creating any dependencies on their products. Ultimately, as the possible uses of AI tools rapidly expand, there will be a corresponding need for careful legal analysis to mitigate unnecessary risks.