Privacy Law Issues for Developers and Deployers of Generative Artificial Intelligence

Maintained

by *Nancy Libin* and *David Rice*, Davis Wright Tremaine, LLP

The development and deployment of generative artificial intelligence tools present unique challenges under U.S. and EU privacy laws and regulations. This practice note explores these obligations at a high level and identifies emerging legal issues and best practices for attorneys advising clients on these matters.

For more practical guidance resources on generative artificial intelligence, see *Generative Artificial Intelligence (AI) Resource Kit*.

To stay on top of artificial intelligence legislation, see *Artificial Intelligence Legislation Tracker (2025)* and *Artificial Intelligence State Law Survey*.

Generative Artificial Intelligence (GAI) Technology and Definitions

Artificial intelligence (AI) enables computers and systems to solve complex problems and perform tasks that typically require or involve human intelligence. Until recently, AI was limited to performing specific tasks, such as recognizing and identifying individuals by analyzing images of their faces and comparing them to facial templates stored in a database. This technology is known as predictive AI.

While predictive AI has been around for years, GAI is relatively new. Unlike predictive AI, which uses machine learning to analyze historical data and make predictions about future events, GAI analyzes data to generate entirely new content. GAI is made possible by the availability of both massive computing power and vast quantities of data necessary to train foundation models that can be used to produce new content (outputs) in response to queries (inputs). GAI has been defined in several ways, but as noted above, it generally means a type of AI that produces new content in the form of text, images, and so forth. See Executive Office of the President of the United States, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (Oct. 30, 2023) (explaining that GAI is "the class of AI models that emulate the structure and characteristics of input data in order to generate derived synthetic content" and "can include images, videos, audio, text, and other digital content"); and Congressional Research Service, Generative Artificial Intelligence and Data Privacy: A Primer (May 23, 2023) (defining GAI as "a type of AI that can generate new content—such as text, images, and videos—through learning patterns from data"). GAI applications include chatbots (e.g., OpenAI's ChatGPT), voice clones, and image generators (e.g., DALL-E).

For purposes of this practice note, entities that develop, produce, or substantially modify a GAI model are "Developers." *Colo. Rev. Stat. § 6-1-1701(7)*; *Cal. Civ. Code § 3110(b)*. Entities that use GAI models created by Developers are "Deployers." *Colo. Rev. Stat. § 6-1-1701(5)*, *(6)*. Here, the term "GAI model" describes the algorithm that Developers create by using training data, and the term "GAI system" means the underlying model plus the infrastructure that enables users to use the model—the user interface and the fine-tuning, for example. Developers and Deployers face different privacy issues, as highlighted below.

Overview of GAI Privacy Issues

Privacy laws are triggered by the processing of personal data, which is broadly defined under most privacy laws to mean any information that is linked or linkable, directly, or indirectly, to an identified or identifiable natural person. See Article 4(a), *Regulation (EU) 2016/ 679, GDPR (EU GDPR); Cal. Civ. Code § 1798.140(v)(1)*. Because Developers of GAI typically train their models on enormous amounts of data—some of which may be personal data—they must consider the potential impact of privacy laws on their operations.

As the National Institute of Standards and Technology (NIST) explained, "GAI systems raise several risks to privacy. GAI system training requires large volumes of data, which in some cases may include personal data. The use of personal data for GAI training raises risks to widely accepted privacy principles, including transparency, individual participation (including consent), and purpose specification. For example, most model developers do not disclose specific data sources on which models were trained, limiting user awareness of whether personally identifiably information (PII) was trained on and, if so, how it was collected." *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile,* National Institute of Standards and Technology, NIST AI 600-1, Section 2.4, page 7 (July 2024).

Similarly, Deployers of GAI often use data to fine-tune and tailor models to their specific needs. If Deployers use personal data for fine-tuning, then they, too, must ensure compliance with applicable privacy laws. Privacy laws also come into play when end users enter prompts (inputs) and obtain outputs that include personal data and when such inputs and outputs are used in connection with consequential decision-making about individuals. NIST AI 600-1, Section 2.4, page 7.

Developers and Deployers of GAI have different privacy considerations, although they overlap somewhat. Because personal data is often inevitably included in the massive data sets used to train GAI, Developers must ensure that they:

- Know the source of their training data
- Understand what type of data is used for training
- Have legal authority to collect such data –and–
- Comply with applicable laws regarding the collection, use, and disclosure of such data

To prevent Deployers from using a Developer's GAI services improperly, Developers must also establish acceptable use policies that require Deployers to use their GAI services only for their intended purpose.

Because Deployers use GAI models created by others, they lack important information about the data used to train the GAI model and cannot confirm the legal status of that data independently. Deployers should, however:

- Confirm that the Developer complied with applicable privacy laws and regulations when the Developer collected the training data and used such data to train the model (This can be handled through negotiated data warranties and indemnities along these lines in the service agreement.)
- Know the source and type of data they use for fine-tuning and any legal restrictions regarding the use and disclosure of personal data that they use or obtain through the model

Note that if Deployers substantially modify the GAI model through fine-tuning, they may be regulated as Developers. See, e.g., *Colo. Rev. Stat. § 6-1-1701(7)* (defining "developer" to mean a person who "intentionally and substantially modifies an artificial intelligence system").

In addition, both Developers and Deployers must develop, adopt, and implement policies and procedures to ensure that their own employees, contractors, and users comply with GAI service restrictions and applicable privacy laws.

Below we analyze the specific obligations that Developers and Deployers of GAI systems have under U.S. federal and state privacy laws and—at a high-level—EU privacy laws regarding the processing of personal data in the context of GAI, including issues related to data collection, use, and disclosure; transparency; data protection assessments; secondary use; consumer rights; automated decision-making and profiling; and security.

Data Collection

Data that Developers use to train GAI models and that Deployers use for fine-tuning may be collected directly from consumers, licensed from third parties, scraped from websites and other online platforms, or obtained from inputs provided by users of the GAI model, among other sources. Below, we consider the legal implications of these various sources of personal data.

First-Party Data

Personal data that Developers and Deployers collect directly from consumers is first-party data. Laws typically place more specific obligations on collectors of first-party data, because such collectors are in the best position to manage disclosures to consumers and respond to consumers' privacy rights requests. Although they have these obligations, first-party data collectors are also likely to have clearer rights to use data than do entities that obtain data from third-party sources.

Federal Laws

Section 5 of the Federal Trade Commission Act (FTC Act) prohibits companies from engaging in unfair or deceptive acts or practices, requires companies to adhere to their representations regarding the processing (including collection) of personal data, and prohibits companies from omitting material information about their personal data practices or otherwise engaging in unfair practices. *15 U.S.C. § 45(a)(1)*. The FTC Act also prohibits companies from making material changes to their privacy policies without prior notice and—if those changes are applied retroactively to previously collected personal data—prior consent from the consumer. Developers and Deployers should ensure that they:

- Disclose in their privacy policies accurate and complete information about how they will use consumers' personal data to train or fine-tune GAI models
- Update their privacy policies to provide notice If they make material changes to the purposes that they initially disclosed (e.g., if they later decide to train or fine-tune GAI models using personal data) (If they intend to use previously collected personal data for this purpose, they must obtain opt-in consent before doing so.)

Sector-specific laws—including the Gramm-Leach-Bliley Act (GLBA), *15 U.S.C. §§ 6801–6809*, the Children's Online Privacy Protection Act (COPPA), *15 U.S.C. §§ 6501–6506*, and the Fair Credit Reporting Act (FCRA), *15 U.S.C. § 1681 et seq.*,—govern the collection of nonpublic financial information, personal data collected online from users under 13 years of age, and certain consumer credit report information, respectively. Although these laws do not specifically reference GAI, they can directly impact its development because of the restrictions that they impose on data collection.

Developers and Deployers of GAI services should determine whether personal information that they collect to train or fine-tune GAI models is governed by a sector-specific law and ensure compliance with those laws.

State Privacy Laws

State privacy laws also impact the development and deployment of GAI services that involve the processing of personal data. These laws require companies to notify consumers regarding the categories of personal data that they collect and to describe the purposes for which they will use the data. Most states to date that have enacted comprehensive privacy laws also require companies to obtain opt-in consent before processing "sensitive" personal data. And at least one state thus far—Maryland—requires companies to limit their collection of personal data and processing of "sensitive" data to only what is "reasonably necessary and proportionate" to "provide or maintain a specific product or service requested by the consumer to whom the data pertains," regardless of whether the consumer consents. *Md. Code Ann., Com. Law § 14-4707(b)(1)(i)*.

Developers and Deployers must determine whether they meet the jurisdictional threshold for these state laws (i.e., whether they annually process the requisite amount of personal data or derive the required amount of annual revenue from sales of such data). If they do, they must:

- Provide an accurate description in their privacy policies regarding what personal data they collect from consumers and how they will use and disclose such data to develop or deploy GAI services
- Collect only personal data that is reasonably necessary and proportionate to train or fine-tune the GAI model, as applicable –and–
- Obtain opt-in consent for any "sensitive" personal data used to train or fine-tune the model

EU and U.K. Data Protection Law

Unlike U.S. laws, which allow companies, for the most part, to collect non-sensitive personal information so long as they are not prohibited from doing so and they provide proper notice, the EU GDPR and U.K. Data Protection Act require controllers (i.e., entities that control the means and purposes of data processing) to have one of the following six "legal bases" for collecting the personal data of individuals located in the EU or U.K. (known as "data subjects"):

- **Consent.** The data subject has consented to the processing of personal data for specific purposes.
- **Contract.** Data processing is necessary for the performance of a contract with the data subject.
- **Legal obligation.** Data processing is necessary for compliance with a legal obligation that applies to the controller under EU law.
- **Vital interests.** Data processing is necessary to protect the vital interests of the data subject or another natural person.
- **Public interest.** Data processing is necessary to perform a task in the public interest.
- **Legitimate interests.** Data processing is necessary for the legitimate interests of the controller or a third party so long as the legitimate interest is not outweighed by the fundamental rights and freedoms of the data subject.

Articles 6 and 47, EU GDPR.

Developers will find it easier to satisfy one of these criteria when they use first-party data that they have collected directly from a data subject than when they use publicly available information collected initially by others. The most plausible legal bases for controllers collecting first-party personal data to train or fine-tune GAI are consent, contract, or legitimate interest:

- **Consent.** Controllers that have a direct relationship with data subjects are well positioned to provide the required notice to, and obtain consent from, the data subject before collecting and using their personal data for training GAI models. Consent is therefore a plausible lawful basis for processing in this scenario.
- **Contract.** Deployers that provide GAI services to data subjects who specifically enter into agreements or create online accounts to use such services—and who must provide their own personal data as inputs or for fine-tuning—may be able to rely on this legal basis for processing. Developers that do not have a direct relationship with data subjects whose data they process to train models cannot rely on this legal basis, however.
- **Legitimate interests.** Developers and Deployers that act as controllers but cannot rely on consent or contract could make the case that they have a legitimate interest in processing data subjects' personal data to train or fine-tune GAI models, so long as they can demonstrate that the processing is necessary to further those interests and that the rights and freedoms of the data subjects do not override the controller's legitimate interests in training or fine-tuning the GAI models.

The European Data Protection Board (EDPB) recently issued guidance stating that training an AI model could be deemed an interest that is "legitimate," citing as examples such activities as "developing the service of a conversational agent to assist users" or deploying an AI model to improve "threat detection in an information system." EDPB, *Opinion 28/2024* at p. 3.

The EDPB made clear that the use of personal data for training the model also would have to be the least intrusive means to achieve the objective, however. Therefore, Developers using personal data collected from or about data subjects in the EU or U.K. should ensure that the amount of personal data they process is reasonable and proportionate, although it is unclear what this means in the context of training large language models (LLMs) with enormous amounts of data. EDPB, *Opinion 28/2024* at p. 3.

The EDPB also emphasized the importance of considering the data subjects' reasonable expectations when conducting the balancing test. Specifically, in determining whether the rights and freedoms of the data subject outweigh the controller's legitimate interest, Developers and Deployers must assess whether the data subjects would reasonably expect that their personal data would be used to train or fine-tune the model. EDPB, *Opinion 28/2024* at p. 3.

Licensed Personal Data

The legal obligations identified above also would apply to Developers and Deployers that license personal data to train GAI systems, although the analysis is different because Developers and Deployers in this context have no direct relationship with the consumer to whom the data pertains. Because Developers and Deployers that license personal data collected by others do not have a direct relationship with the underlying consumers, they must rely on the licensor for assurance that consumers have received the proper notice regarding the purposes for which their personal data was collected and the entities to whom it would be disclosed and for what purpose. This is challenging as a practical matter because the licensor may have obtained the data from a third party and may not be certain what consumers were told when their personal information was collected. Indeed, the personal information may have changed hands many times, and its origin and date of collection may be unknown. This situation presents risks similar to those associated with purchasing lead lists for marketing communications.

The license agreement must include a data protection addendum that describes the parties' roles and responsibilities under applicable privacy laws to provide notice, obtain any required consents, provide opt out mechanisms, conduct or assist in conducting data protection impact assessments, respond to consumers' requests to exercise their privacy rights, and so forth.

Publicly Available Data

Publicly available data is critical to GAI model development. As one large technology company explained, "[p]ublicly available information is at the core of how AI models are trained" and is "foundational to model quality and functionality." Sam Clark, *Generative AI training must involve personal data despite risks, Google urges,* MLex Market Insight (June 4, 2024). This presents privacy risks because such data is likely to include personal data: "Although there are techniques to reduce certain highly specific personal data collected and processed in the training phase, currently personal data is key to training models to understand language and cannot be removed easily or without potential quality implications." Id.

Just because personal data is publicly available, however, does not mean that companies may collect it to develop GAI models without triggering legal obligations. Indeed, several laws govern the collection of publicly available information. Which laws apply will depend on the jurisdiction in which the individual to whom the data pertains resides or—for purposes of the EU GDPR—is located.

Federal Laws

As noted previously, the FTC Act requires companies to ensure that their disclosures regarding the collection and use of personal data—including publicly available personal data—are accurate and to avoid surreptitious and retroactive privacy policy changes to support GAI training. Developers and Deployers that use publicly available information must therefore, ensure that they disclose in their privacy policies that they collect personal data from publicly available sources and use that data to train or fine-tune GAI models. The FTC's Office of Technology has observed that the "incentive to constantly ingest additional data can be at odds with a company's obligations to protect users' data, undermining peoples' privacy or resulting in the appropriation of a firm's competitively significant data." Staff in the Office of Techn., FTC, *AI Companies; Uphold Your Privacy and Confidentiality Commitments*, Fed. Trade Comm'n Techn. Blog, (Jan 9, 2024).

The federal Computer Fraud and Abuse Act (CFAA), *18 U.S.C. § 1030*, also regulates scraping of public information online in some circumstances through its prohibition on unauthorized access, or exceeding authorized access to computer systems. *18 U.S.C. § 1030(a)(3)(C)*. The Ninth Circuit in *hiQ Labs v. LinkedIn* outlined what types of scraping would violate the CFAA. *hiQ Labs, Inc. v. LinkedIn Corporation, 31 F.4th 1180 (9th Cir. 2022)*. In that case, the court rejected LinkedIn's argument that hiQ's scraping of publicly available personal information on LinkedIn's website violated the CFAA. hiQ, 31 F.4th at 1184. The court held that the "CFAA's prohibition on accessing a computer 'without authorization' is violated when a person circumvents a computer's generally applicable rules regarding access permissions, such as username and password requirements, to gain access to a computer." hiQ, 31 F.4th at 1201. In contrast, "[i]t is likely that when a computer network [like LinkedIn] generally permits public access to its data, a user's accessing that publicly available data will not constitute access without authorization under the CFAA." hiQ, 31 F.4th at 1201. The court then determined that, "[t]he data hiQ seeks to access is not owned by LinkedIn and has not been demarcated by LinkedIn as private using such an authorization system. HiQ has therefore raised serious questions about whether LinkedIn may invoke the CFAA to preempt hiQ's possibly meritorious tortious interference claim." hiQ, 31 F.4th at 1201. The court added that "entities that view themselves as victims of data scraping are not without resort, even if the CFAA does not apply: state law trespass to chattels claims may still be available. And other causes of action, such as copyright infringement, misappropriation, unjust enrichment, conversion, breach of contract, or breach of privacy, may also lie." hiQ, 31 F.4th at 1201.

The Ninth Circuit, citing the U.S. Supreme Court's decision in *Van Buren v. United States*, explained further that CFAA liability depends on a "gates-up-or-down" analysis: "if authorization is required [for computer system access] and has been given, the gates are up; if authorization is required and has *not* been given, the gates are down" hiQ, 31 F.4th at 1199. "[A]pplying the 'gates' analogy to a computer hosting publicly available webpages, that computer has erected no gates to lift or lower in the first place," meaning that the CFAA's prohibition on access "'without authorization' does not apply to public websites." hiQ, 31 F.4th at 1199. The data that hiQ was scraping "was available to anyone with a web browser" and was obtained from profiles "visible to the general public." hiQ, 31 F.4th at 1184 and 1199.

This case did not involve GAI but is relevant to the scraping of publicly available personal information for use in training that technology.

State Privacy Laws

States that have enacted their own comprehensive consumer privacy laws generally *exclude* "publicly available information" from the definition of "personal data." This is potentially helpful for Developers and Deployers that want to use personal data that was scraped from public online sources, because they would not need to comply with those privacy laws with respect to any such data that met the definition.

State laws define the term "publicly available information" differently, however, making it difficult to develop a uniform nationwide approach to compliance. For example, under privacy laws in California, Utah, and Virginia, "publicly available information" means information that:

- Is lawfully made available from government records
- A business has a reasonable basis to believe that it is lawfully made available to the general public by either the consumer or widely distributed media –or–

- Is made available by a person to whom the consumer has disclosed the data, so long as the consumer did not restrict the data to a specific audience

Cal. Civ. Code § 1798.135(v)(2); *Utah Code Ann. § 13-61-101(29)*; *Va. Code Ann. § 59.1-575*.

Under Connecticut's privacy law, however, "publicly available information" means information that both:

- Is lawfully made available through government records or widely distributed media –and–
- The controller has a reasonable basis to believe the consumer has lawfully made available to the general public

Conn. Gen. Stat. § 42-515(18). Developers and Deployers can handle the exception for "publicly available information" in one of two ways:

- They can add metadata to personal information (also known as data tagging) to identify the state law that should govern. This approach allows companies to leverage the broader exceptions that some states have adopted, but it is labor-intensive and can be difficult to implement and manage.
- Alternatively, companies can adopt the most limited definition of "publicly available information" and apply that to all personal information obtained online. Companies that take this approach will lose the opportunity to leverage the exemption in some states, but they will simplify compliance.

EU and U.K. Data Protection Law

As discussed above, Developers and Deployers that act as controllers must identify a lawful basis under EU and U.K. law before collecting and using publicly available personal data of persons in the U.K. or EU to train or fine-tune GAI models.

- **Consent.** Controllers that want to collect and use publicly available personal data to train or fine-tune GAI models will have a difficult time relying on consent as a legal basis given that they likely have no relationship with the consumers whose data they wish to use. Indeed, it would be impractical—if not impossible—to obtain from such data subjects consent that is "freely given, specific, informed and unambiguous" and is in the form of a statement or "clear affirmative action," as is required under the GDPR. Article 4(11), EU GDPR.
- **Performance of a contract.** Controllers that develop or fine-tune LLMs cannot rely on necessity to perform a contract as a legal basis for collecting and processing the publicly available data of data subjects in the EU and U.K. due to the absence of any contractual relationship between the parties.
- **Legitimate interests.** A controller's commercial and economic interests may be valid "legitimate interests." See EU GDPR Recital 47 (stating that "[t]he processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest"). However, in determining whether a legitimate interest overrides data subjects' rights and freedoms, controllers must conduct a "careful assessment, including whether a data subject can reasonably expect at the time and in the context of collection of the personal data that processing for that purpose may take place." EU GDPR Recital 47. At least one commentator has opined that "[i]t is reasonable to suggest that the typical internet user does not expect, nor intend, for their data to be utilized as training material for LLMs." Ruschemeier, Hannah, *Generative AI and Data Protection* (May 2, 2024). The EDPB Opinion on AI Models, described above, specifically notes that "web scraping practices" may require risk mitigation as part of a legitimate interest balancing test. EDPB, *Opinion 28/2024* at pp. 26 and 29.

In addition, Deployers who act as controllers must confirm that the personal data that Developers used to train was processed lawfully. EDPB, *Opinion 28/2024* at p. 33.

Type of Personal Data

In addition to knowing how data is collected (from publicly available sources, licensed, or directly from a consumer), Developers and Deployers must know the type of data used to train or fine-tune GAI so that they can determine whether the data is governed by sector-specific privacy laws (e.g., protected health information governed by the Health Insurance Portability and Accountability Act (HIPAA)), *42 U.S.C. § 1320d et seq.*, subject to heightened protections (e.g., personal data that reveals an identifiable person's race or ethnicity is "sensitive" and therefore subject to heightened protection under privacy laws), or exempt from protection (e.g., de-identified data).

Sector-specific state privacy laws, such as biometric privacy laws (Illinois's Biometric Information Privacy Act (BIPA), *740 Ill. Comp. Stat. 14/1 et seq.*; Texas's Capture or Use of Biometric Identifier Act (CUBI), *Tex. Bus. & Com. Code § 503.001*; Colorado's biometrics law, *Colo. Rev. Stat. § 6-1-1314* (effective July 1, 2025); and Washington's biometrics statute (HB 1493), *Wash. Rev. Code § 19.375.010 et seq.*) and health data privacy laws (e.g., Washington's My Health My Data Act (MHMD), *Wash. Rev. Code Ann. §§ 19.373.005–19.373.900*), may also be implicated depending on the type of personal data collected. Some of these laws impose unique requirements that companies must comply with, such as the obligation under Washington's MHMD to obtain prior, opt-in consent from consumers to process their "consumer health data" for a specified purpose, unless processing is necessary to provide a specific product or service that the consumer requested. *Wash. Rev. Code Ann. § 19.373.030*. This may be a difficult standard to meet for some types of GAI services that train on health-related data from uncertain or licensed sources.

Transparency

Policymakers are beginning to require Developers to provide information about the data they use to train their GAI systems. For instance, the recently enacted California Artificial Intelligence Training Data Transparency Act (CA AI Training Data Transparency Act), *Cal. Civ. Code §§ 3110 through 3111*, requires Developers that make a "generative artificial intelligence system or service" available to California consumers, or that substantially modify such a system, to post on their website a "high-level summary of the datasets used in the development of the generative artificial intelligence system or service." *Cal. Civ. Code § 3111(a)*. The CA AI Training Data Transparency Act specifically requires Developers to disclose whether the data they use to train GAI systems includes personal information and aggregate information, as those terms are defined in the California Consumer Privacy Act. *Cal. Civ. Code § 3111(a)(7)*, *(a)(8)*.

Data Protection Assessments

State consumer privacy laws generally require controllers to conduct data protection assessments when they process personal information that creates a "heightened risk of harm to a consumer." See, e.g., *Colo. Rev. Stat. § 6-1-1309(1)*, *(2)*. Developers and Deployers should consider whether their GAI systems create such a risk. State consumer privacy laws have slightly different definitions of both the kind of processing that poses a "heightened risk" and the elements of an acceptable data protection assessment. Developers and Deployers will likely need a harmonized, multi-jurisdictional approach to handle these assessments efficiently.

States typically categorize the following activities as creating a heightened risk of harm to consumers:

- Processing of personal information for the purposes of targeted advertising
- "Profiling" that creates the risk of unfair or deceptive treatment or an unlawful disparate impact on a consumer, financial or physical injury, or offensive intrusion into private concerns
- Selling personal data
- Processing sensitive data –and–
- Processing that causes "other substantial injury to consumers"

See, e.g., *Colo. Rev. Stat. § 6-1-1309(2)*. Examples of GAI that in certain circumstances may create a heightened risk of harm under this standard could include:

- Using agentic AI to guide consumers through a financial transaction
- Using a chatbot to answer consumer health questions based on the consumer's prompts that contain sensitive data
- Personalizing content and advertising based on sensitive data
- Summarizing legal content, such as tenants' rights –and–
- Using consumer inputs to train AI models, which could result in a data "sale" under the laws of some states

See, e.g., State of California Department of Technology Office of Information Security, "*Generative Artificial Intelligence Risk Assessment*," SIMM 5305-F (July 2024). Data protection assessments must identify and weigh the benefits of these services against the risks that they create for the consumer. See, e.g., *Colo. Rev. Stat. § 6-1-1309(3)*.

Secondary Use

Any use of personal information for purposes that are incompatible with the processing purposes disclosed to the consumer at or before the time of collection is a "secondary use." Under both the FTC Act and the EU GDPR, controllers must obtain consent from consumers before repurposing their personal data for a secondary use.

Secondary uses are a common challenge for controllers seeking to use previously collected personal data to train GAI models. While it can be tempting for a controller to modify a privacy policy after the fact to include AI training among the purposes for which it will use personal data, the FTC has advised that "quietly" changing terms of service to permit AI training without providing sufficient notice, and in some cases obtaining consent, may be unfair or deceptive. Staff in the Office of Technology and the Division of Privacy and Identity Protection, FTC, AI (and other) Companies: *Quietly Changing Your Terms of Service Could Be Unfair or Deceptive*. Fed. Trade. Comm'n Technology Blog (Feb. 13, 2024).

Practitioners evaluating the legal risks of using a previously collected data set for GAI training should consider the following:

- What was the consumer told regarding the purpose for which their personal data was collected? If initially collected for a purpose other than training GAI, subsequent use of their personal data for training GAI is a secondary use.
- Do exceptions apply? Is GAI model training reasonably aligned with consumer expectations and compatible with the context of the relationship between the business and the consumer?
- Does the data set include personal data from consumers who reside in a state that does not allow secondary use, even with consent (e.g., Maryland)?
- If the data set is licensed from a third party, what limits on data use were imposed when the data was collected? What were consumers told, if anything, at the time of data collection about the use of personal data to train or fine-tune GAI?

Consumer Rights

Nineteen states thus far have enacted comprehensive state privacy laws. At the time of this publication, omnibus state privacy laws in California, Colorado, Connecticut, Delaware, Iowa, Nebraska, New Hampshire, New Jersey, Utah, Virginia, Oregon, and Texas are in effect. Montana, Minnesota, Illinois, Kentucky, Tennessee, Rhode Island, and Maryland passed comprehensive privacy laws that will take effect between 2025 and 2026. Florida enacted a privacy law that applies only to a small number of large technology companies. Washington and Nevada have health privacy laws with broad scope and many elements in common with the general state privacy laws.

Each of these laws gives rights to consumers that Developers and Deployers must honor in the context of the development and deployment of GAI. These include the rights to confirm that a consumer's personal data is being processed, access such data, request that such data be deleted, correct inaccurate personal data, opt in to the processing of sensitive data (or in California, limit the processing of sensitive data), and opt out of sales and sharing of personal data as well as profiling in furtherance of certain types of decisions based on automated decision-making. See *Cal. Civ. Code §§ 1798.105–1798.125*. Developers that are subject to these laws must establish internal processes to fulfill consumers' requests to exercise these rights.

Some of these rights may be difficult to implement in the context of GAI, however. Indeed, it is not clear whether:

- Deleting personal data from the training data set so that the data is not used to further train or fine-tune the model would suffice to fulfill a request to delete personal data –or–
- A Developer is responsible for "correcting" personal data that a GAI system produces as an output or whether the user who provides the prompt that elicits the incorrect response would be responsible for using a particular input

These are not just theoretical problems. Privacy activist Max Schrems has brought an action against OpenAI for producing an inaccurate birthdate in response to a query. "*ChatGPT provides false information about people, and OpenAI can't correct it*," NYOB.EU News (April 29, 2024). Certain technical solutions may resolve these issues, but they have not yet been tested for legal adequacy. The *EDPB report* regarding ChatGPT stated that "technical impossibility cannot be invoked to justify non-compliance with these [GDPR] requirements." European Data Protection Board, Report of the work undertaken by the ChatGPT Task force, (May 23, 2024). For example, if a consumer requests deletion of personal information, the GAI system could block outputs that contain that information, or the Developer could mask personal data in a training data set or model or substitute synthetic data for personal data in the training data set. The EDPB identified as one possibility "measures to mask personal data or to substitute it with fake personal data in the training set (e.g., the replacement of names and email addresses with fake names and fake email addresses). This measure may be particularly appropriate when the actual substantive content of the data is not relevant to the overall processing (e.g., in LLM training)." EDPB *Opinion 28/2024* at pp. 28–29.

Consumer Rights Exceptions

Developers and Deployers may be able to leverage some of the exceptions to avoid having to grapple with these technical challenges. In addition to the "publicly available information" exception covered above, state privacy laws provide several exceptions, including:

- **Using personal data for internal research to improve services, where reasonably aligned with consumers' expectations.** It is not clear, however, whether the use of personal data to train GAI models would be considered "research," and if so, whether consumers would expect that their personal data would be used for this purpose. It also is not clear what kind of disclosure would suffice to provide notice. For example, a Deployer might wish to do so by enabling a GAI customer service chatbot to identify itself as such to users, prominently disclose that user prompts train the GAI model, and incorporate a link to the Deployer's privacy policy.
- **Personal data used to provide a good or service requested or reasonably anticipated by the consumer within the context of a business's ongoing business relationship with the consumer.** One example of personal data used to provide a good or service requested or reasonably anticipated by the consumer within the context of a business's ongoing business relationship with the consumer would be an application that uses GAI on photos submitted by users to alter the users' image, such as by showing the user as either older or younger. The Developer may be able to argue that a consumer should expect the use of

GAI technology, so long as its use and training are prominently disclosed and are necessary to provide the service.

- **Personal data used to enable solely internal operations that are reasonably aligned with a consumer's expectations based on the relationship with the business and compatible with context in which the consumer provided information.** As noted above, it is unclear whether the use of personal data to train GAI models would be "reasonably aligned" with consumers' expectations. Perhaps, as GAI technology becomes more common and over time is integrated into the vast majority of computing systems, a reasonable consumer might expect that any processing of their personal information will involve GAI.

- **Data that is pseudonymous data, which is exempt from consumer rights requests, or de-identified or anonymized data, which is not "personal data" under privacy laws.** Pseudonymous data generally is defined as "personal data that cannot be attributed to a specific natural person without the use of additional information, provided that such additional information is kept separately and is subject to appropriate technical and organizational measures to ensure that the personal data is not attributed to an identified or identifiable natural person." *Va. Code Ann. § 59.1-575*. Developers and Deployers would need to analyze the data to determine whether it met the definition.

  A Developer also could take the position that the information in the training data set is de-identified or anonymized, in which case the consumer does not have the right to request deletion. The EDPB Opinion on AI Models states that under certain circumstances, artificial intelligence training data and models may contain only anonymized data rather than personal data. EDPB *Opinion 28/2024* at p. 16.

  Another possibility is that data used to train the GAI model is personal data, but the GAI model itself is not, because such models represent "statistical relationships between 'tokens' or 'chunks' of text representing commonly occurring sequences of characters." Jordan Francis, et al., *Do LLMs Contain Personal Information? California AB 1008 Highlights Evolving, Complex Techno-Legal Debate*, Future of Privacy Forum Blog, Oct. 25, 2024.

Regulators are now exploring the viability of these approaches as well as risk mitigation measures. For example, the EDPB recently noted that "relevant [risk mitigation] measures may include . . . [p]seudonymisation measures," which "could, for example, include measures to prevent any combination of data based on individual identifiers." EDPB *Opinion 28/2024* at p. 28. The EDPB added that "[t]hese measures may not be appropriate where the [Supervisory Authority] considers that the controller demonstrated the reasonable need to gather different data about a particular individual for the development of the AI system or model in question." EDPB *Opinion 28/2024* at p. 28.

The applicability of these exceptions must be analyzed on a case-by-case basis.

Automated Decision-Making Technology and Profiling

Policymakers have sought to address how AI systems may perpetuate bias and algorithmic discrimination. See, e.g., *Colo. Rev. Stat. § 6-1-1703* (providing that Deployers have a duty to prevent algorithmic discrimination in high-risk AI systems). To that end, data protection laws in the EU and U.K. and privacy laws in the U.S. regulate the processing of personal data for "profiling," which under U.S. state privacy laws is generally defined as "any form of automated processing of personal data to evaluate, analyze, or predict personal aspects concerning an identified or identifiable individual's economic situation, health, personal preferences, interests, reliability, behavior, location, or movements." *Colo. Rev. Stat. § 6-1-1301*. Specifically, consumers "have the right to opt out of [p]rofiling . . . when the [p]rofiling is done in furtherance of a decision that results in the provision or denial of financial or lending services, housing, insurance, education enrollment or opportunity, criminal justice, employment opportunities, health-care services, or access to essential goods or services . . ." *4 CCR § 904-3*:9.02(B). Regulations regarding automated decision-making technology may ultimately have more impact on predictive AI technologies than GAI, but practitioners should be aware of this issue.

The California Privacy Protection Agency is conducting a rulemaking focused on "access and opt-out rights with respect to businesses' use of automated decision-making technology, including profiling and requiring businesses' response to access requests to include meaningful information about the logic involved in those decision-making processes, as well as a description of the likely outcome of the process with respect to the consumer." *Cal. Civ. Code § 1798.185(16)*. Colorado regulations implementing the Colorado Privacy Act impose similar obligations with respect to profiling. See *4 CCR § 904-3*-9.03 (2025). Practitioners also should be aware of the interplay between state AI laws, like the Colorado Artificial Intelligence Act, and state consumer privacy laws, as both types of laws will impact how Developers and Deployers process personal data in this context.

Security

The massive amounts of data used to train GAI models present challenging security risks. Indeed, these databases are targets for bad actors who could use the information to engage in identity theft and social engineering scams. Bad actors also could introduce erroneous data into the datasets to reduce the reliability of the GAI models, and insiders who have access to these databases could misuse them for personal gain. Describing reasonable security measures for these databases is beyond the scope of this practice note, but practitioners should be aware of these concerns.

Other Sources of Regulation

The Executive Branch under the Biden administration was very active in developing AI policy. A summary of some of its major initiatives is below:

- **Biden Executive Order**. The Biden White House issued the *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (Biden EO) on October 30, 2023. Certain aspects of the Biden EO related to companies in the private sector that provide products and services to governmental entities and proposed requirements for entities operating very large foundation models. President Trump rescinded the Biden EO on the first day of his second term in office.
- **Trump Executive Order.** After rescinding the Biden EO, President Trump issued his own AI executive order, stating a new U.S. policy on AI: to "sustain and enhance America's global AI dominance to promote human flourishing, economic competitiveness, and national security."
- **Office of Management and Budget (OMB).** The OMB issued a Request for Information that, among other things, sought information about how privacy impact assessments and vendor disclosures help to mitigate privacy risks, including those related to *AI. Request for Information: Responsible Procurement of Artificial Intelligence in Government, 89 Fed. Reg. 22196, 22197 (Mar. 29, 2024)*.
- **Department of Health and Human Services (HHS).** HHS released guidance regarding the use of AI to increase algorithm transparency and fairness for predictive AI and health records. HHS Finalizes Rule to Advance Health IT Interoperability and Algorithm Transparency, HHS (Dec 13, 2023).

EU AI Act

The European Union Artificial Intelligence Act (EU AI Act), Regulation (EU) 2024/1689, which became effective on August 1, 2024, is the EU's first comprehensive effort to regulate AI technology, including GAI. A full discussion of the EU AI Act is beyond the scope of this practice note, but practitioners should be aware of several high-level considerations:

- While the EU AI Act is a technology-focused AI regulation, it operates in tandem with and as a companion to the EU GDPR, which is a personal data protection regulation. See EU AI Act, Article 2(1). The threshold consideration for companies that process personal data to train GAI models is whether the proposed processing is consistent with data subjects' rights under the GDPR, including whether the Developer or

Deployer has a lawful basis to process the personal data for this purpose. Also relevant are the GDPR provisions on profiling, as the EU AI Act adopts the GDPR definition of profiling. See EU AI Act, Articles 3(52) and 6(3).

- The EU AI Act regulates AI technology according to the following risk levels:
  - o **Unacceptable risk.** AI systems are prohibited.
  - o **High-risk.** AI systems are subject to extensive requirements, including those regarding transparency.
  - o **Limited risk.** AI systems trigger only transparency requirements because they interact with people. –and–
  - o **Minimal risk.** AI systems do not trigger any obligations.

  The EU AI Act has separate requirements for general-purpose AI technology. Any GAI system that is high-risk under the EU AI Act likely requires a risk assessment under the GDPR if it processes or is trained on personal data.
- GAI Deployers and Developers must honor data subjects' rights throughout the AI system life cycle. "The right to privacy and to protection of personal data must be guaranteed throughout the entire lifecycle of the AI system." EU AI Act, Recital 69.
- The EU AI Act is generally applicable to Deployers and Developers that place AI technologies into market in the EU, in addition to other situations. See generally EU AI Act, Article 2 (Scope); See Article 113 (Entry into Force and Application).

Model Disgorgement

Regulators have imposed severe penalties on companies that have used personal data to train or fine-tune AI models in violation of privacy laws. For instance, the FTC has brought enforcement actions against several companies, alleging that they engaged in unfair or deceptive acts or practices when they used personal data to train AI models without making clear to consumers that they were doing so. These FTC investigations include the following:

- ***In re Cambridge Analytica,*** **LLC, FTC Docket No. 9383, Final Order (Dec. 6, 2019).**The FTC directed Cambridge Analytica to "[d]elete or destroy all Covered Information collected from consumers through GSRApp, and any information or work product, including any algorithms or equations, that originated, in whole or in part, from this Covered Information."
- ***In re Everalbum, Inc.,*** **FTC Docket No. C-4743, Final Order (Jan. 11, 2021).** The FTC ordered Everalbum to delete "models or algorithms developed in whole or in part using Biometric Information [that Everalbum] collected from Users of the "Ever" mobile application." –and–
- ***In re Rite Aid Corporation,*** **FTC Docket No. 2023190, Final Order (Mar. 8, 2024).** Rite Aid used facial recognition technologies to track shoppers who were deemed to present security risks without informing those customers. The FTC, after finding the activity to be "unfair," ordered Rite Aid to "[d]elete, and direct third parties to delete, any images or photos they collected because of RiteAid's facial recognition system as well as any algorithms or other products that were developed using those images and photos."

To settle these allegations, these companies agreed to consent orders that required them, among other things, to destroy both the personal data that they processed and the algorithms developed with such data. FTC Commissioner Rebecca Slaughter explained the rationale for model disgorgement: "When companies collect data illegally, they should not be able to profit from either the data or any algorithm developed using it," and she explained that the "authority to seek this type of remedy [model disgorgement] comes from the Commission's power to order relief reasonably tailored to the violation of the law." Rebecca Kelly Slaughter, *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade,* 23 Yale Journal of Law & Technology, Special Issue 1, p. 39 (August 2021). Commissioner Slaughter also stated that "[t]his innovative enforcement approach should send a clear message to companies engaging in illicit data collection in order to train AI models: Not worth it." Id.

Emerging Challenges Going Forward

Practitioners who advise clients regarding the development and deployment of GAI systems should analyze these issues in the context of applicable privacy and data protection laws and ensure they are aware of new laws and regulatory priorities. The challenges of complying with privacy laws will increase as agentic AI begins to emerge. Agentic AI, powered by GAI technology, acts as a personal assistant that can perform tasks requiring access to multiple platforms, customized to an individual's preferences, and will engage in reasoning, planning, and complex problem-solving with less direct human oversight. This will impact the travel, health, and financial industries as well as many others. And, it will yield valuable data about consumer purchasing patterns that could be further leveraged. We are only at the beginning of learning what GAI can do and managing the risks associated with it. Practitioners will need to keep current regarding these developments.

Current as of: ***03/05/2025***

End of Document