



**Congressional
Research Service**

Informing the legislative debate since 1914

Generative Artificial Intelligence and Data Privacy: A Primer

May 23, 2023

Congressional Research Service

<https://crsreports.congress.gov>

R47569

Contents

Overview	1
What Is Generative AI?	2
How Do Generative AI Models Use Data?	3
Where Does the Data Come From?	4
What Happens to Data Shared with Generative AI Models?	4
Policy Considerations for Congress	5
Existing Data Privacy and Related Laws	5
Proposed Privacy Legislation.....	6
Existing Agency Authorities	6
Regulation of Data-Scraping.....	7
Research and Development for Alternative Technical Approaches	7

Figures

Figure 1. Examples of Generative AI Models	2
--	---

Contacts

Author Information.....	8
-------------------------	---

Overview

Since the public release of Open AI’s ChatGPT, Google’s Bard, and other similar systems, some Members of Congress have expressed interest in the risks associated with “generative artificial intelligence (AI)”. Although exact definitions vary, generative AI is a type of AI that can generate new content—such as text, images, and videos—through learning patterns from pre-existing data. It is a broad term that may include various technologies and techniques from AI and machine learning (ML).¹

Generative AI models have received significant attention and scrutiny due to their potential harms, such as risks involving privacy, misinformation, copyright, and non-consensual sexual imagery. This report focuses on privacy issues and relevant policy considerations for Congress. Some policymakers and stakeholders have raised privacy concerns about how individual data may be used to develop and deploy generative models. These concerns are not new or unique to generative AI, but the scale, scope, and capacity of such technologies may present new privacy challenges for Congress.

Generative AI at a Glance

Major Developers and Selected Products:²

- **OpenAI** (with partnerships and funding from Microsoft)—“ChatGPT” chatbot, “DALL-E” image generator
- **Google**—“Bard” chatbot
- **Meta**—“LLaMA” research tool, “Make-A-Video” video generator
- **Anthropic** (founded by former employees of OpenAI)—“Claude” chatbot
- **Stability AI**—“Stable Diffusion” image generator
- **Hugging Face**—BLOOM language model
- **NVIDIA**—“NeMo” chatbot, “Picasso” visual content generator

Types of Applications:

- **Chatbots**—systems that simulate human conversation, often in question-and-answer format
- **Image generators**—systems that generate images based on an input or “prompt”
- **Video generators**—systems that generate videos based on an input or “prompt,” sometimes called deepfakes
- **Voice clones**—systems that generate speech and voice sounds, sometimes called audio deepfakes

¹ There are various definitions of AI in statute and agency guidance. For example, the National Artificial Intelligence Initiative Act of 2020 (P.L. 116-283) defines AI as “a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. Artificial intelligence systems use machine and human-based inputs to—(A) perceive real and virtual environments; (B) abstract such perceptions into models through analysis in an automated manner; and (C) use model inference to formulate options for information or action.” Artificial intelligence and machine learning (ML) are often used interchangeably, but ML is a subfield of AI that focuses on systems that can “learn” and improve through experience and data. For more information on this distinction, see Columbia Engineering “Artificial Intelligence (AI) vs. Machine Learning,” <https://ai.engineering.columbia.edu/ai-vs-machine-learning/>.


² Six of the seven listed companies were identified based on participation in White House initiatives to develop “public assessments” of existing generative AI systems. Meta was not included in the White House announcement. “Fact Sheet: Biden-Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans’ Rights and Safety,” White House, May 4, 2023, <https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/>.

What Is Generative AI?

Generative AI can generate new content—such as text, images, and videos—through learning patterns from data.³ There are many types of generative AI models (see **Figure 1**), which can produce content based on different inputs or “prompts.” For example, some models can produce images from text prompts (e.g., MidJourney, Stable Diffusion, DALL-E), while others create videos (e.g., Gen2 or Meta’s Make-A-Video).

Some scholars and policymakers have recently coined the term “**general-purpose models**” (GPAI) to describe applications like ChatGPT that can complete various functions.⁴ These GPAI models may have a wide range of down-stream applications compared to single-purpose models designed for a specific task. Many general-purpose AI applications are built on top of **large language models** (LLMs) that can recognize, predict, translate, summarize, and generate language.⁵ LLMs are a subset of generative AI and are characterized as “large” due, in part, to the massive amount of data necessary for training the model to learn the rules of language.

Figure 1. Examples of Generative AI Models

Image Generator	Chatbot
<p>Prompt: “The Library of Congress building in the style of Van Gogh.”</p>	<p>Prompt: “Write a 4 sentence poem about the Library of Congress.”</p>
<p>Response:</p>	<p>Response:</p>
	<p>“In the heart of D.C., there stands a place A temple of knowledge, a palace of grace The Library of Congress, a marvel to see A monument to learning, for you and me</p> <p>Rows upon rows of books, towering high A treasury of history, never to die Where stories and knowledge, together reside A beacon of wisdom, that cannot be denied”</p>

Source: Stable Diffusion and ChatGPT, via CRS. The image was generated by Stable Diffusion and the text response was generated by ChatGPT.

³ Generative AI models may use different technical approaches and techniques, such as generative adversarial networks (GANs) or generative pre-trained transformers (GPTs). The colloquial term “deep fakes,” which refers to realistic machine-generated images, videos, and audio, may also fall under the umbrella term of “generative AI.” Deepfakes are typically generated through GANs.

⁴ European Parliament, “General-Purpose Artificial Intelligence,” [https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA\(2023\)745708_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf).

⁵ Bender, Gebru, et al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” March 2023, ACM Conference on Fairness, Accountability, and Transparency, <https://doi.org/10.1145/3442188.3445922>; Samuel Bowman, “Eight Things to Know About Large Language Models,” April 2023, <https://arxiv.org/pdf/2304.00612.pdf>.

How Do Generative AI Models Use Data?

Data is essential to train and fine-tune AI models. Generative AI models require especially large datasets for training and fine-tuning.

Definitions

- **Training a model** refers to providing a model with data to learn from, often called a training dataset. After a model is trained to recognize patterns from one dataset, some models can be provided with new data and still recognize patterns or predict results.
- **Fine-tuning a model** refers to training a previously trained model on new data, or otherwise adjusting an existing model.⁶

Generative AI models, particularly LLMs, require massive amounts of data. For example, OpenAI’s ChatGPT was built on a LLM that trained, in part, on over 45 terabytes of text data obtained (or “scraped”) from the internet. The LLM was also trained on entries from Wikipedia and corpora of digitized books.⁷ Open AI’s GPT-3 models were trained on approximately 300 billion “tokens” (or pieces of words) scraped from the web and had over 175 billion parameters, which are variables that influence properties of the training and resulting model.⁸

Critics contend that such models rely on privacy-invasive methods for mass data collection, typically without the consent or compensation of the original user, creator, or owner.⁹ Additionally, some models may be trained on sensitive data and reveal personal information to users. In a company blog post, Google AI researchers noted, “Because these datasets can be large (hundreds of gigabytes) and pull from a range of sources, they can sometimes contain sensitive data, including personally identifiable information (PII)—names, phone numbers, addresses, etc., even if trained on public data.”¹⁰ Academic and industry research has found that some existing LLMs may reveal sensitive data or personal information from their training datasets.¹¹

Some models are used for commercial purposes or embedded in other downstream applications. For example, companies may purchase subscription versions of ChatGPT to embed in their various services or products. Khan Academy, Duolingo, Snapchat, and other companies have partnered with OpenAI to deploy ChatGPT in their services.¹² However, individuals may not know their data was used to train models that are monetized and deployed across such applications.

⁶ For example, fine-tuning could include adjusting a pre-existing model’s parameters.

⁷ Brown, et al., “Language Models Are Few-Shot Learners,” July 22, 2020, <https://arxiv.org/abs/2005.14165>.

⁸ According to OpenAI, their models were trained on some datasets with a total of 300 billion tokens. A token is a piece of a word. One token is around ¾ of a word. Brown, et al., “Language Models Are Few-Shot Learners,” July 22, 2020, <https://arxiv.org/abs/2005.14165>.

⁹ Matt Burgess, “ChatGPT Has a Big Privacy Problem,” *Wired*, April 4, 2023, <https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/>.

¹⁰ Nicholas Carlini, “Privacy Considerations in Large Language Models,” December 15, 2020, Google Research Blog, <https://ai.googleblog.com/2020/12/privacy-considerations-in-large.html>.

¹¹ Nicholas Carlini et al., “Extracting Training Data from Large Language Models,” Jun 15, 2021, <https://arxiv.org/abs/2012.07805>.

¹² Alex Heath, “Snapchat Is Releasing Its Own AI Chatbot Powered by ChatGPT,” *The Verge*, February 27, 2023, <https://www.theverge.com/2023/2/27/23614959/snapchat-my-ai-chatbot-chatgpt-openai-plus-subscription>; “Introducing Duolingo Max, a Learning Experience Powered by GPT-4,” March 14, 2023, Duolingo Blog, <https://blog.duolingo.com/duolingo-max/>; Sal Khan, “Harnessing GPT-4 So That All Students Benefit. A Nonprofit Approach for Equal Access,” Khan Academy, March 14, 2023, <https://blog.khanacademy.org/harnessing-ai-so-that-all-students-benefit-a-nonprofit-approach-for-equal-access/>.

Some countries have taken action against AI developers for improper use of personal information. For example, the Italian Data Protection Authority issued a temporary ban preventing OpenAI from using Italian users' data.¹³ After agreeing to certain changes—such as allowing users to submit removal requests for personal data under the EU's General Data Protection Regulation (GDPR)—OpenAI restored access to its service for users in Italy.¹⁴

Where Does the Data Come From?

Many AI developers do not disclose the exact details of their training datasets. For generative AI, most training data is “scraped” from publicly available webpages before it is repackaged and sold, or in some cases, made freely available to AI developers.

Some AI developers rely on popular large datasets such as “Colossal Clean Crawled Corpus” (C4) and “Common Crawl,” which are amassed through web crawling (i.e., software that systematically browses public internet sites and collects information from each available webpage). Similarly, AI image generators are commonly trained on a dataset called LAION, which contains billions of images scraped from internet sites and their text descriptions.¹⁵ Some companies might also use proprietary datasets for training.

Generative AI datasets can include information posted on publicly available internet sites, including PII and sensitive and copyrighted content. They may also include publicly available content that is erroneous, pornographic, or potentially harmful. Since data may be scraped without the creator's consent, some artists, content creators, and others have begun to use new tools such as “HaveIBeenTrained” to identify and report their own content in such databases. In a 2023 investigation, the *Washington Post* and Allen Institute for AI analyzed the websites scraped for the C4 dataset, which is used by AI developers including Google, Facebook, and OpenAI.¹⁶ The investigation found that the C4 dataset included websites with copyrighted content as well as potentially sensitive information, such as state voter registration records.

These forms of data collection may also raise questions about copyright ownership and fair use. For a discussion of copyright issues and generative AI, see CRS Legal Sidebar LSB10922, *Generative Artificial Intelligence and Copyright Law*, by Christopher T. Zirpoli.

What Happens to Data Shared with Generative AI Models?

Some critics have also raised concerns that user data shared with a generative AI application—such as a chatbot—may be misused or abused without the user's knowledge. For example, a user may reveal sensitive health information while conversing with a healthcare chatbot without realizing their information could be stored and used to re-train the models or for other

¹³ Italian Data Protection Authority (Garante per la protezione dei dati personali or GPDP), March 31, 2023, <https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english>.

¹⁴ Adi Robertson, “ChatGPT Returns to Italy After Ban,” April 28, 2023, *The Verge*, <https://www.theverge.com/2023/4/28/23702883/chatgpt-italy-ban-lifted-gpdp-data-protection-age-verification>

¹⁵ Marissa Newman and Aggi Cantrill, “The Future of AI Relies on a High School Teacher's Free Database,” *Bloomberg*, April 23, 2023, <https://www.bloomberg.com/news/features/2023-04-24/a-high-school-teacher-s-free-image-database-powers-ai-unicorns#xj4y7vzkg>.

¹⁶ Kevin Schaul, Szu Yu Chen, and Nitasha Tiku, “Inside the Secret List of Websites That Make AI Like ChatGPT Sound Smart,” April 19, 2023, *Washington Post*, <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>.

commercial purposes. Many existing chatbots have terms of service that allow the company to re-use user data to “develop and improve their services.”

These concerns may be particularly pertinent for generative models used in interactions or services that commonly result in the disclosure of sensitive information such as advising, therapy healthcare, legal, or financial services. In response, some critics have argued that chatbots and other generative AI models should require affirmative consent from users or provide clear disclosure of how user data is collected, used, and stored.

Policy Considerations for Congress

Existing Data Privacy and Related Laws

The United States does not currently have a comprehensive data privacy law. Congress has enacted a number of laws that create data requirements for certain industries and subcategories of data, but these statutory protections are not comprehensive. For example, the Gramm-Leach-Bliley Act (P.L. 106-102) regulates financial institutions’ use of nonpublic personal information, while the Health Insurance Portability and Accountability Act (HIPAA, P.L. 104-191) requires covered entities to protect certain health information. Under current U.S. law, generative AI may implicate certain privacy laws depending on the context, developer, type of data, and purpose of the model. For example, if a company offers a chatbot in a videogame or other online service directed at children, the company could be required to meet certain requirements under the Children’s Online Privacy Protection Act (COPPA, P.L. 105-277).

Additionally, certain state laws on privacy, biometrics, and AI may have implications for generative AI applications. In many cases, the collection of personal information typically implicates certain state privacy laws that provide an individual a “right to know” what a business collects about them; how data is used and shared; the “right to access and delete” their data; or the “right to opt-out” of data transfers and sales.¹⁷ However, some of these laws include exemptions for the collection of public data, which may raise questions about how and whether they apply to generative AI tools that use information scraped from the internet.

In the absence of a comprehensive federal data privacy law, some individuals and groups have turned to other legal frameworks (e.g., copyright, defamation, right of publicity) to address potential privacy violations from generative AI and other AI tools. For example, some companies have faced class action lawsuits for possible violations of right of publicity state laws, which protect against unauthorized use of an individual’s likeness for commercial purposes.¹⁸

Congress may consider enacting comprehensive federal privacy legislation that specifically addresses generative AI tools and related concerns. In doing so, Congress may consider and evaluate similar state and international efforts. For example, the European Union’s (EU) proposed AI Act includes various articles on data regulation, disclosures, and documentation, among other requirements. The EU AI Act recently added a category for general purpose AI systems and

¹⁷ For example, the California Consumer Privacy Act (CCPA) and California Consumer Privacy Regulation (CPR) provide certain privacy rights to California residents. California Office of the Attorney General, “California Consumer Privacy Act (CCPA),” May 10, 2023, <https://oag.ca.gov/privacy/ccpa>. For more information on the CCPA or data privacy laws, see CRS Legal Sidebar LSB10213, *California Dreamin’ of Privacy Regulation: The California Consumer Privacy Act and Congress*, coordinated by Eric N. Holmes; and CRS Report R45631, *Data Protection Law: An Overview*, by Stephen P. Mulligan and Chris D. Linebaugh.

¹⁸ Isaiah Poritz, “AI Celebrity ‘Deepfakes’ Clash with Web of State Publicity Laws,” *Bloomberg Law*, April 14, 2023, <https://news.bloomberglaw.com/ip-law/ai-celebrity-deepfakes-clash-with-web-of-state-publicity-laws>.

foundation models, another term used for AI models that train on large amounts of data and can be adapted for various tasks.¹⁹

Proposed Privacy Legislation

Some Members of Congress have proposed various comprehensive or targeted privacy bills with requirements that could impact generative AI applications. These are three common mechanisms included in various privacy bills:

Notice and Disclosure Requirements. Currently, most generative AI applications do not provide notice or acquire consent from individuals to collect and use their data for training purposes. Congress may consider requiring companies developing or deploying generative AI systems to (1) acquire consent from individuals before collecting or using their data, or (2) notify individuals that their data will be collected and used for certain purposes, such as training models. Some scholars dispute the efficacy of notice and consent requirements.²⁰

Opt-out Requirements. Congress may consider requiring companies to provide users an option to opt-out of data collection. Of note, opt-out systems may not necessarily protect data that is publicly-scraped from the web, and may be cumbersome for individuals to exercise.

Deletion and Minimization Requirements. Congress may also consider requiring companies to provide mechanisms for users to delete their data from existing datasets or require maximum retention periods for personal data. Currently, most leading chatbots and other AI models do not provide options for users to delete their personal information.

In considering such proposals, Congress may also wish to consider practical challenges users may face exercising specific privacy rights as well as potential challenges for companies in complying with certain types of legal requirements and user requests.

Existing Agency Authorities

Various federal agencies may enforce laws relevant to AI and data privacy. The Federal Trade Commission (FTC) has been active in addressing data privacy issues and has taken various actions involving AI. The FTC has applied its broad authorities over “unfair or deceptive acts or practices in commerce” to cases related to data privacy and data security. In recent months, the Commission reaffirmed that its authorities also apply to new AI tools.²¹ Chairman Khan stated, “there is no AI exemption to the laws on the books, and the FTC will vigorously enforce the law to combat unfair or deceptive practices or unfair methods of competition.”²²

¹⁹ European Parliament, “AI Act: A Step Closer to the First Rules on Artificial Intelligence,” press release, May 11, 2023, <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>.

²⁰ Claire Park, “How ‘Notice and Consent’ Fails to Protect Our Privacy,” New America Open Technology Institute, March 23, 2020, <https://www.newamerica.org/oti/blog/how-notice-and-consent-fails-to-protect-our-privacy/>.

²¹ The Federal Trade Commission (FTC) released multiple blog posts to caution companies against using AI to deceive or mislead consumers. The FTC’s recent blog post states that the FTC’s “unfair or deceptive acts or practices” (UDAP) authorities could apply to companies that develop, sell, or use an AI system that is “effectively designed” to deceive consumers, “even if not the system’s original purpose.” Michael Atleson, “Keep Your AI Claims in Check,” FTC Business Blog, February 27, 2023, <https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check>; Michael Atleson, “Chatbots, Deepfakes, and Voice Clones: AI Deception for Sale,” FTC Business Blog, March 20, 2023, <https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale>.

²² Federal Trade Commission, “FTC Chair Khan and Officials from DOJ, CFPB and EEOC Release Joint Statement on (continued...)”

The data collection practices of AI companies may also raise competition concerns. At the 2023 Annual Antitrust Enforcers Summit, Chair Khan stated, “As you have machine learning that depends on huge amounts of data and also depends on huge amounts of storage, we need to be very vigilant to make sure that this is not just another site for the big companies becoming bigger and really squelching rivals.”²³ The development of AI models may also require significant computational and financial resources, which may preclude new competitors and entrench incumbents.²⁴

In evaluating existing agency authorities, Congress may consider updating or providing additional specific authorities to federal agencies to address AI and related privacy issues. Additionally, Congress could consider what resources federal agencies may require to conduct additional oversight of AI and privacy issues.

Regulation of Data-Scraping

There are currently no federal laws that ban the scraping of publicly available data from the internet. The Computer Fraud and Abuse Act (CFAA, 18 U.S.C. §1030) imposes liability when a person “intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains ... information from any protected computer.”²⁵ Some court cases have held that this prohibition does not apply to public websites—meaning that scraping publicly accessible data from the internet does not violate the CFAA.²⁶

Scraping publicly available information from the internet has privacy implications beyond generative AI models. The facial recognition company Clearview AI has scraped over 20 billion images from the web, including social-media profile photos, which have been used for software and databases provided to law enforcement and other entities.²⁷ Some technology companies have also scraped publicly available data to amass large data repositories. Web-scraping may raise competition concerns since larger companies may block competitors from scraping data.

Many researchers, journalists, and civil society groups, among others, rely on scraping to conduct research that may be in the public interest. If Congress were to consider broad legislation to limit or provide guardrails for scraping information from the internet, it might consider implications for a range of activities that it may find beneficial.

Research and Development for Alternative Technical Approaches

Congress may wish to consider providing funds to federal agencies for intramural and extramural research to examine the development of alternative AI models or related technologies that may

AI,” press release, April 25, 2023, <https://www.ftc.gov/news-events/news/press-releases/2023/04/ftc-chair-khan-officials-doj-cfpb-eeoc-release-joint-statement-ai>.

²³ Adi Robertson, “The US Government Is Gearing up for an AI Antitrust Fight,” *The Verge*, March 28, 2023, <https://www.theverge.com/2023/3/28/23660101/ai-competition-ftc-doj-lina-khan-jonathan-kanter-antitrust-summit>.

²⁴ “ChatGPT and More: Large Scale AI Models Entrench Big Tech Power,” AI Now Institute, April 11, 2023, <https://ainowinstitute.org/publication/large-scale-ai-models>.

²⁵ The Computer Fraud and Abuse Act is codified at Title 18, United States Code, Section 1030.

²⁶ Zack Whittaker, “Web Scraping Is Legal, US Appeals Court Reaffirms,” *TechCrunch*, April 18, 2022, <https://techcrunch.com/2022/04/18/web-scraping-legal-court/>.

²⁷ Alex Hern, “TechScape: Clearview AI Was Fined £7.5m for Brazenly Harvesting Your Data—Does It Care?” *The Guardian*, May 25, 2022, <https://www.theguardian.com/technology/2022/may/25/techscape-clearview-ai-facial-recognition-fine>.

preserve individual privacy, such as privacy-enhancing technologies.²⁸ There are benefits and tradeoffs to some AI models under development that may have privacy implications. For example, smaller models that use less data or avoid transmitting and analyzing data in the cloud may minimize some privacy concerns but may amplify other issues, such as bias, by training on smaller datasets and potentially limiting the representativeness of data being used to train models.²⁹ Congress may consider directing agencies to conduct and fund research to support privacy-by-design³⁰ for AI and ML applications in order to both foster greater privacy for individuals and support the development of AI technologies and the global competitiveness of U.S. AI companies.

Author Information

Kristen E. Busch
Analyst in Science and Technology Policy

Disclaimer

This document was prepared by the Congressional Research Service (CRS). CRS serves as nonpartisan shared staff to congressional committees and Members of Congress. It operates solely at the behest of and under the direction of Congress. Information in a CRS Report should not be relied upon for purposes other than public understanding of information that has been provided by CRS to Members of Congress in connection with CRS's institutional role. CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.

²⁸ In a 2022 Request for Information, the White House Office of Science and Technology defined privacy-enhancing technologies as “a broad set of technologies that protect privacy.” Examples could include “privacy-preserving data sharing and analytics technologies, which describes the set of techniques and approaches that enable data sharing and analysis among participating parties while maintaining disassociability and confidentiality. Such technologies include, but are not limited to, secure multiparty computation, homomorphic encryption, zero-knowledge proofs, federated learning, secure enclaves, differential privacy, and synthetic data generation tools.” Office of Science and Technology Policy (OSTP), “Request for Information on Advancing Privacy-Enhancing Technologies,” 87 *Federal Register* 35250-35252, June 9, 2022.

²⁹ Kyle Wiggers, “The Emerging Types of Language Models and Why They Matter,” TechCrunch, April 28, 2022, <https://techcrunch.com/2022/04/28/the-emerging-types-of-language-models-and-why-they-matter/>.

³⁰ James Coker, “#DataPrivacyWeek Interview: Overcoming Privacy Challenges in AI,” *Infosecurity Magazine*, January 25, 2022, <https://www.infosecurity-magazine.com/interviews/data-privacy-week-privacy-ai/>.