Proskauer

Privacy and Data Security

June 2023

The very definition of generative AI suggests the creation of new content based on a program training on existing data, a recipe that necessarily raises potential U.S. and EU data privacy issues, not to mention related ethical and societal concerns over how to protect against unintended consequences and responsibly regulate the technology. The third Webinar in our series explored various privacy and ethical issues related to AI. The presenters first highlighted some of the relevant GDPR issues both developers and users of GAI products must consider (and briefly touched on the upcoming EU AI Regulation), and then discussed the patchwork of state data privacy laws that may regulate automated decision making and how a number of federal agencies (including the FTC) have made AI an enforcement target going forward. The U.S. legal discussion also highlighted the emerging issues related to AI and healthcare and how the testing of new AI technologies to create healthcare breakthroughs can be a case study as to what privacy issues may arise in the development of AI products going forward. The Webinar continued with a section on cybersecurity concerns and remarks about how certain privacy-related tenets and principles will likely be employed when thinking about or legislating the responsible and ethical use of AI. The session concluded with privacy law takeaways to consider when developing or using an AI product.

EU and GDPR Concerns

The main item grabbing attention in the last few months related to GAI and data protection in the EU was the Italian data protection authority's temporary ban of ChatGPT in Italy and OpenAI's subsequent measures that sought to allay the Italian DPA's concerns and prompted OpenAI to eventually restore service to Italian users on April 28, 2023. Specifically, the Italian DPA issued a temporary limitation on the processing of Italian users data by OpenAI, citing a March 20th breach involving a misconfiguration in its cache system that allowed some users to see information from others' chat histories, as well as what the Italian DPA considered shortcomings in ChatGPT's lack of age verification protections, among other things. The geoblock on Italian users was lifted on April 28th, with OpenAI making several changes for EU users, including rolling out an updated privacy policy and privacy notices and allowing users to opt-out of the reuse of inputs for AI training purposes, among other things.

The first step that entities developing or using GAI can take to gauge GDPR compliance is to determine their role, as defined under the GDPR, whether a data controller or processor. An entity's role will determine its obligations and responsibilities surrounding data protection compliance. Under the GDPR, a controller, for example, is an organization that determines the purposes and means of processing; a processor, on the other hand, is an entity that acts on the instructions of the controller. Organizations might ask several questions to pinpoint their role and figure out how the GDPR applies to it, including: What are you inputting into the AI system? What is the output? What is your level of control (e.g., are you using an API or plug-in, with a service provider in between you and the users, or are you directing engaging the GAI application, or are you the GAI provider itself?). As the presenters stated, broadly speaking, the GAI customer would likely be the data controller of the input and output data because it is likely controlling what data is going in and is using the output for its own purposes and the GAI developer is likely to be the data processor of the input data and the output (but as to the training dataset it previously used or is collecting concurrent with the use of the AI system, the GAI developer would likely be deemed the controller of that data).

There is a myriad of potential compliance issues under the GDPR for both GAI developers and customers. For example, to process personal data, there must be a lawful basis under the GDPR, and in the presenters' views, the obvious candidates that a GAI platform could rely on would be consent or legitimate interests, which, of course, depend on how the GAI system is operated and what the parties are doing with the data. Transparency as to privacy practices and data usage is also a key requirement. Users and developers also need to be thinking about the collection of personal data related to the datasets used to train the AI – Was personal data obtained? How was it obtained? Was it obtained lawfully? – questions that relate to the method of collection and the content of privacy notices and related contracts. Another GDPR issue is whether parties must produce a data protection impact assessment (DPIA), which is required where there is a high risk to the rights and freedoms of individuals; one could envision how a DPIA could be required in certain cases due to the usage of an AI platform and whether there are any risks to individuals (and if such risks can be mitigated), all subject to change as usage and new risks emerge as the technology becomes more powerful or is used in a different manner. Lastly, the issue of automated decision-making may arise; if there is no human involvement and there is an automated decision that has a legal or significant effect, individuals may have further rights under the GDPR (e.g., further information or right to human intervention in the decision-making process).

Beyond the GDPR, the presenters outlined some open questions that may complicate GDPR (and other legal) compliance. These include:

- **The "black box" issue**: Developers may not know the precise extent of what's in a dataset, and because machine learning involving multiple data points to create an output, it's likely not possible to know if a piece of personal data was collected and used to process a particular output.
- **Consents**: Can an AI developer, like OpenAI, prove that its prior collection of certain website data for AI training was done with proper user consents or else prove it had a "legitimate interest"? Similar concerns with using user prompts to train AI (though that issue was perhaps resolved by OpenAI's incognito mode offering).
- **Data rights**: How does a data subject employ their rights under the GDPR with respect to data that may be in a large language model dataset (even if such data could be identified, could one person's "data" even be removed from an AI training dataset? Is "machine unlearning" possible?

Beyond GDPR compliance, the presenters reminded viewers of the forthcoming EU AI Regulation, which is expected to be enacted at some time this year. The draft Regulation suggests that the EU will take a risk-based approach to regulating AI applications, prompting many to ask whether GAI applications be classified as high-risk applications and garner more scrutiny.

U.S. Developments

There is no overarching federal AI or data privacy law, leaving a patchwork of laws that might implicate AI, including various state data privacy and AI-related privacy laws and the enforcement reach of the FTC and other consumer agencies. There has been some push for AI-related legislation in Congress following the White House's release of guiding principles (e.g., the "Blueprint for an AI Bill of Rights"), but nothing concrete has yet emerged. Thus, the legal landscape currently consists of voluntary frameworks, executive orders against algorithmic discrimination, unfair business and anti-discrimination laws as enforced by the FTC (and other agencies), and a patchwork of state laws that may have some specific AI-related components. For example, beyond state laws that regulate the use of AI in the employment context, it should be noted that, originally, California's CCPA was silent on automated decision-making issue, but since the CPRA amendments, the CCPA now contains a provision that addresses automated decision-making technology. At this point, the laws or subsequent regulations are not necessarily clearly written in a way that might aid compliance, but the main point is that we are starting to see state data privacy laws and regulations being drafted with a focus or acknowledgement of AI. This follows a trend where privacy has become a natural place for addressing a lot of concerns around AI, given that data privacy considerations share some of the same principles and components that might address the regulation of AI more broadly, such as the privacy-by-design concept that suggests developers should think of privacy issues and ethical approaches to data from the very beginning of the development process.

The presenters noted that the FTC and other regulators are starting to focus on AI. For example, the FTC, along with other agencies, recently issued a joint statement that reiterated their stance that existing legal authorities apply to the use of automated systems and innovative new technologies just as they apply to other practices. It should be noted that in the past, the FTC has brought several past enforcement actions related to what the agency considered unfair or deceptive use of algorithms that processed consumer data. One common thread in these enforcements was the FTC's remedy, namely, imposing an algorithmic destruction remedy requiring the settling defendants to delete models and algorithms the companies developed by allegedly unlawfully using the consumer data at issue. Thus, the presenters stressed that when a company develops or uses AI that processes data obtained from the past, one important issue that regulators may focus on is examining whether you had consent for the use of such data and for the secondary use of that data at the time the data was collected.

The presenters brought up some questions that organizations should consider at the senior management and board level, as these questions have real impact to ongoing operations:

- How transparent does a company have to be regarding its use of generative AI applications when it comes to processing consumer data?
- Are there more privacy compliance concerns re: AI automated decision-making in products or services?
- Should every company that intends to use generative AI and related technologies update an existing privacy program to meet the privacy questions of AI?
- Should a company monitor internal use of generative AI (mapping its uses and cataloguing them) to prepare for privacy issues or potential compliance issues in the future (e.g., knowing the extent and types of business uses)?

They also suggested that implementing AI tools might bring efficiencies and benefits, but with come more privacy compliance concerns that might require updating existing privacy programs and statements and establishing a procedure to monitor the internal usage of GAI tools.

Ethical Considerations

AI and GAI offer a multitude of benefits – write faster, code faster, solve intractable societal problems like climate change and improve the nature of work – but it also has the potential power and access and ability that raised ethical concerns. On a basic level, ethical questions concern not what AI products are permitted to do from a legal standpoint, but what they should not be doing and where guardrails and boundaries should be drawn. The presenters noted that in recent years, international organizations and even the White House and federal agencies have released voluntary principles on how ethical considerations can, will and should be part of AI product development and usage. Thus, the presenters noted that, generally speaking, AI tools should engender trust and transparency to the extent possible, minimize algorithmic bias and discriminatory outcomes, and consider public safety concerns.

Cybersecurity

OpenAI recently acknowledged that due to a bug in an open source library there was a data leak on March 20th that allowed some users to see titles from another active user's chat history, and, in some cases, payment-related information of some ChatGPT Plus subscribers. There have also been news reports of at least one company banning employees from using GAI program after it was discovered that an employee inputted sensitive code into the GAI program. Thus, when it comes to cybersecurity and GAI, the issue is no longer merely a theoretical concern.

Companies and governments are also expressing concern over how GAI can be used to create deepfake videos, voice clones, fake websites and social media profile content, and code for malware attacks, all in furtherance of financial or cyber crimes. The use of voice clones is particularly concerning, given the use of voice as an authentication tool in finance. These new realities have prompted organizations to reexamine cybersecurity response readiness to prepare for these types of attacks.

Final Takeaways

- The AI procurement process should entail some due diligence and close scrutiny of the terms of enterprise licenses. At minimum, organizations will want to ensure that customers can opt-out of the reuse of data for training purposes.
- When the use of AI involving consumer-facing platforms or the collection of consumer data, considerations should be made for such measures as notices, terms of use and disclaimers.

- Data privacy and cybersecurity should not be overlooked. As with many organizations' current compliance measures, this may be a global concern as AI-related regulation is developing in the U.S, EU, and elsewhere.
- AI is moving at a fast pace and governments and regulators are now trying to catch up, so it's important to stay informed of legal developments.

Related Professionals

- **Kelly M. McMullon**
  Special International Labor, Employment & Data Protection Counsel