Generative AI Poses Security Risks

Generative AI can be a great productivity booster, but concerns over protecting the data used to train LLMs -- even the public-facing data -- are driving enterprises to carefully consider how the technology is used. Opaque Systems' VP of product, Jay Harel, explains.

Upside: Generative AI seems to have taken the world by storm, but many issues remain, among them security and privacy.

Jay Harel: There are three main issues when it comes to generative AI and privacy, especially in terms of large language models (LLMs). The first issue surrounds queries. Any LLM provider can have visibility into the queries entered by their users, which may include sensitive information such as proprietary code or personal identifiable information (PII). This may lead to information loss or privacy-regulation infringements.

The second issue comes with learning and training AI models. To train AI models, the provider must use fresh training data. This data is retained by the AI model, increasing the chances of sensitive information leaking or landing in the wrong hands.

Finally, there are IP issues for organizations with proprietary models. Organizations want to fine-tune their models on company data, but this re-

quires either giving their proprietary LLM provider access to their data or allowing the provider to deploy the proprietary model within the customer organization. This becomes a privacy issue because private and sensitive data will be accessible to people outside of the internal organization, leaving room for data breaches.

Is this a problem that involves all company data -- for example, should a company be worried about ChatGPT using the results of that company's publicly distributed survey -- or is the concern only about confidential data (such as PII or internal data such as sales figures)?

The main concern is around confidential data (such as PII) or internal data (such as sales figures). However, in the example of a company's publicly distributed survey, despite that survey being public, the data being input into the survey might be private or sensitive -- for example, email addresses. Although respondents are typically required to agree to a privacy statement before filling out a survey, this likely does not give the company permission to use PII for model training and inference.

Are companies aware of the problem or the security measures that are needed to protect data in this emerging age of generative AI?

Companies have a general awareness of possible privacy and security concerns, but because generative AI is still in the emerging stages and people are learning as they go, many have gaps in their understanding of the risks and how to mitigate them. That's why we're releasing our Confidential AI solution while it's still in the testing stages so that people can start understanding the possible solutions for remedying these concerns.

What, to your knowledge, is happening at generative AI companies, such as OpenAI, to address security and privacy issues?

Generally, companies seem to be more concerned about bringing their own models to the market to remain competitive. However, those who are trying to address these issues, such as OpenAI, are leaning heavily on lobbying for regulation and guidelines to help people use this technology safely and ethically.

Until these issues are solved, what best practices do you recommend for enhancing security and privacy?

Lacking a proven privacy and security solution, LLM users must exercise caution. Organizations should set and enforce usage guidelines and educate employees about the potential pitfalls. Case in point: Alphabet, Google's parent company, recently warned employees about using confidential information on AI chatbots. A full ban is not practical given the proliferation of tools, with more AI models and applications becoming available every day. Education remains the most effective way to reduce the risk until solutions such as Opaque's become widely available.