Published on *The National Law Review* https://natlawreview.com

Generative AI Poses Unique Risks to Data Security, NIST Warns

Article By:

Generative artificial intelligence (AI) has opened a new front in the battle to keep confidential information secure. The National Institute of Standards and Technology (NIST) recently released a draft report highlighting the risk generative AI poses to data security. The report, entitled “Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile,” details generative AI’s potential data security pitfalls and suggests actions for generative AI management.

NIST identifies generative AI’s data security risk as “[l]eakage and unauthorized disclosure or de-anonymization of biometric, health, location, personally identifiable [information], or other sensitive data.” Training generative AI requires an enormous amount of data culled from the internet and other publicly available sources. For example, ChatGPT4 was trained with 570 gigabytes from books, web texts, articles, and other writing on the internet, which amounts to about 300 billion words residing in a generative AI database. Much of generative AI’s training data is personal, confidential, or sensitive information.

Generative AI systems have been known to disclose any information within its training data, including confidential information, upon request. During adversarial attacks, large language models have revealed private or sensitive information within their training data, including phone numbers, code, and conversations. The New York Times has sued ChatGPT’s creator, OpenAI, alleging in part that ChatGPT will furnish articles behind the Times paywall. This disclosure risk poses obvious data security issues.

Less obvious are the data security issues that generative AI’s capacity for predictive inference poses. With the vast quantity of data available to generative AI, it can correctly infer personal or sensitive information, including a person’s race, location, gender, or political leanings – even if that information is not within the AI’s training data. NIST warns that these AI models, or individuals using the models, might disclose this inferred information, use it to undermine privacy or apply it in a discriminatory manner. Already, we have seen a company settle an EEOC lawsuit alleging that it used AI to make discriminatory employment decisions. Generative AI threatens to increase this legal exposure.

From an AI governance perspective, NIST suggests several broad principles to mitigate the data privacy risk. Among other things, NIST recommends:

- Aligning generative AI use with applicable laws, including those related to data privacy and the use, publication, or distribution of intellectual property;
- Categorizing different types of generative AI content with associated data privacy risks;
- Develop an incident response plan specifically tailored to address breaches, and regularly test and update the incident response plan with feedback from external and third-party stakeholders;
- Establish incident response plans for third-party generative AI technologies deemed high-risk. As with all incident response plans, this incident response plan should include:
    - Communicating third-party generative AI incident response plans to all relevant AI actors;
    - Defining ownership of the incident response functions;
    - Rehearsing (or “table topping”) the incident response plans regularly;
    - Regular review of incident response plans for alignment with relevant breach reporting, data protection, data privacy, or other laws;
- Update and integrate due diligence processes for generative AI acquisition and procurement vendor assessments to include data privacy, security, and other risks; and
- Conduct periodic audits and monitor AI-generated content for privacy risks.

These actions will involve more than simply adding a reference to artificial intelligence to existing cybersecurity plans. They will involve carefully analyzing a company’s legal obligations, its contract obligations, and the company culture to design an AI governance plan that keeps confidential information out of the public domain and away from bad actors.

National Law Review, Volume XIV, Number 137