Evolution of Privacy-Preserving AI: From Protocols to Practical Implementations

Petr Emelianov

Year by year, artificial intelligence evolves and becomes more efficient for solving everyday human tasks. But at the same time, it increases the possibility of personal information misuse, reaching unprecedented levels of power and speed in analyzing and spreading individuals' data. In this article, I would like to take a closer look at the strong connection between artificial intelligence systems and machine learning and their use of increasingly private and sensitive data.

Together, we'll explore existing privacy risks, discuss traditional approaches to privacy in machine learning, and analyze ways to overcome security breaches.

Importance of Privacy in AI

It is no secret that today, AI is extensively used in many spheres, including marketing. NLP, or Natural Language Processing, interprets human language and is used in voice assistants and chatbots, understanding accents and emotions; it links social media content to engagement. Machine learning employs algorithms to analyze data, improve performance, and enable AI to make decisions without human intervention. Deep Learning relies on neural networks and uses extensive datasets for informed choices.

These AI types often collaborate, posing challenges to data privacy. AI collects data intentionally, where users provide information, or unintentionally, for instance, through facial recognition. The problem arises when unintentional data collection leads to unexpected uses, compromising privacy. For example, discussing pet food or more intimate purchases around a phone can lead to targeted ads, revealing unintentional data gathering. AI algorithms, while being intelligent, may inadvertently capture information and subject it to unauthorized use. Thus, video doorbells with facial identification intended for family recognition may unintentionally collect data about unrelated individuals, causing neighbors to worry about surveillance and data access.

Bearing in mind the above, it is crucially important to establish a framework for ethical decision-making regarding the use of new AI technologies. Addressing privacy challenges and contemplating the ethics of technology is imperative for the enduring success of AI. One of the main reasons for that is that finding a balance between technological innovation and privacy concerns will foster the development of socially responsible AI, contributing to the long-term creation of public value and private security.

Traditional Approach Risks

Before we proceed with efficient privacy-preserving techniques, let us take a look at traditional approaches and the problems they may face. Traditional approaches to privacy and machine learning are centered mainly around two concepts: user control and data protection. Users want to know who collects their data, for what purpose, and how long it will be stored. Data protection involves anonymized and encrypted data, but even here, the gaps are inevitable, especially in machine learning, where decryption is often necessary.

Another issue is that machine learning involves multiple stakeholders, creating a complex web of trust. Trust is crucial when sharing digital assets, such as training data, inference data, and machine learning models across different entities. Just imagine that there is an entity that owns the training data, while another set of entities may own the inference data. The third entity provides a machine learning server running on the inference, performed by a model owned by someone else. Additionally, it operates on infrastructure from an extensive supply chain involving many parties. Due to this, all the entities must demonstrate trust in each other within a complex chain. Managing this web of trust becomes increasingly difficult.

Examples of Security Breaches

As we rely more on communication technologies using machine learning, the chance of data breaches and unauthorized access goes up. Hackers might try to take advantage of vulnerabilities in these systems to get hold of personal data, such as name, address, and financial information, which can result in fund losses and identity theft.

A report on the malicious use of AI outlines three areas of security concern: expansion of existing threats, new attack methods, and changes in the typical character of threats. Examples of malicious AI use include BEC attacks using deepfake technology, contributing to social engineering tactics. AI-assisted cyber-attacks, demonstrated by IBM's DeepLocker, show how AI can enhance ransomware attacks by making decisions based on trends and patterns. Notably, TaskRabbit experienced an AI-assisted cyber-attack, where an AI-enabled botnet executed a DDoS attack, leading to a data breach which affected 3.75 million customers.

Moreover, increased online shopping is fueling card-not-present (CNP) fraud, combined with rising synthetic identity and identity theft issues. Predicted losses from it could reach $200 billion by 2024, with transaction volumes rising over 23%.

Privacy-Preserving Machine Learning

This is when privacy-preserving machine learning comes in with a solution. Among the most effective techniques are federated learning, homomorphic encryption, and differential privacy. Federated learning allows separate entities to collectively train a model without sharing explicit data. In turn, homomorphic encryption enables machine learning on encrypted data throughout the process and differential privacy ensures that calculation outputs cannot be tied to individual data presence. These techniques, combined with trusted execution environments, can effectively address the challenges at the intersection of privacy and machine learning.

Privacy Advantages of Federated Learning

As you can see, classical machine learning models lack the efficiency to implement AI systems and IoT practices securely when compared to privacy-preserving machine learning techniques, particularly federated learning. Being a decentralized version of machine learning, FL helps make AI security-preserving techniques more reliable. In traditional methods, sensitive user data is sent to centralized servers for training, posing numerous privacy concerns, and federated learning addresses this by allowing models to be trained locally on devices, ensuring user data security.

Enhanced Data Privacy and Security

> Federated learning, with its collaborative nature, treats each IoT device on the edge as a unique client, training models without transmitting raw data. This ensures that during the federated learning process, each IoT device only gathers the necessary information for its task. By keeping raw data on the device and sending only model updates to the central server, federated learning safeguards private information, minimizes the risk of personal data leakage, and ensures secure operations.

Improved Data Accuracy and Diversity

Another important issue is that **centralized data used to train a model may not accurately represent the full spectrum of data that the model will**

encounter. In contrast, training models on decentralized data from various sources and exposing them to a broader range of information enhances the model's ability to generalize new data, handle variations, and reduce bias.

Higher Adaptability

One more advantage federated learning models exhibit is a **notable capability to adapt to new situations without requiring retraining**, which provides extra security and reliability. Using insights from previous experiences, these models can make predictions and apply knowledge gained in one field to another. For instance, if the model becomes more proficient in predicting outcomes in a specific domain, it can seamlessly apply this knowledge to another field, enhancing efficiency, reducing costs, and expediting processes.

Encryption Techniques

To enhance privacy in FL, even more efficient encryption techniques are often used. Among them are homomorphic encryption and secure multi-party computation. These methods ensure that data stays encrypted and secure during communication and model aggregation.

The homomorphic encryption allows computations on encrypted data without decryption.

For example, if a user wants to upload data to a cloud-based server, they can encrypt it, turning it into ciphertext, and only after that upload it. The server would then process that data without decrypting it, and then the user would get it back. After that, the user would decrypt it with their secret key.

Multi-party computation, or MPC, enables multiple parties, each with their private data, to evaluate a computation without revealing any of the private data held by each party.

A multi-party computation protocol ensures both privacy and accuracy. The private information held by the parties cannot be inferred from the execution of the protocol.

If any party within the group decides to share information or deviates from the instructions during the protocol execution, the MPC will not allow it to force the other parties to output an incorrect result or leak any private information.

Final Considerations

Instead of the conclusion, I would like to stress the importance and urgency of embracing advanced security approaches in ML. For effective and long-term outcomes in AI safety and security, there should be coordinated efforts between the AI development community and legal and policy institutions. Building trust and establishing proactive channels for collaboration in developing norms, ethics, standards, and laws is crucial to avoid reactive and potentially ineffective responses from both the technical and policy sectors.

I would also like to quote the authors of the report mentioned above, who propose the following recommendations to face security challenges in AI:

1. Policymakers should collaborate closely with technical researchers to explore, prevent, and mitigate potential malicious applications of AI.
2. AI researchers and engineers should recognize the dual-use nature of their work, considering the potential for misuse and allowing such considerations to influence research priorities and norms. They should also proactively engage with relevant stakeholders when harmful applications are foreseeable.
3. Identify best practices from mature research areas, like computer security, and apply them to address dual-use concerns in AI.
4. Actively work towards broadening the involvement of stakeholders and domain experts in discussions addressing these challenges.

Hope this article encourages you to investigate the topic on your own, contributing to a more secure digital world.

AI Machine learning Data (computing) Protocol (object-oriented programming) security Data security

Opinions expressed by DZone contributors are their own.