What is federated learning?

Kim Martineau

Federated learning is a way to train AI models without anyone seeing or touching your data, offering a way to unlock information to feed new AI applications.

The spam filters, chatbots, and recommendation tools that have made artificial intelligence a fixture of modern life got there on data — mountains of training examples scraped from the web, or contributed by consumers in exchange for free email, music, and other perks.

Many of these AI applications were trained on data gathered and crunched in one place. But today's AI is shifting toward a decentralized approach. New AI models are being trained collaboratively on the edge, on data that never leave your mobile phone, laptop, or private server.

This new form of AI training is called federated learning, and it's becoming the standard for meeting a raft of new regulations for handling and storing private data. By processing data at their source, federated learning also offers a way to tap the raw data streaming from sensors on satellites, bridges, machines, and a growing number of smart devices at home and on our bodies.

To promote discussion and exchange ideas for advancing this nascent field, IBM is co-organizing a federated learning workshop at this year's NeurIPS, the world's top machine-learning conference.

Data and their discontents

Google introduced the term federated learning in 2016, at a time when the use and misuse of personal data was gaining global attention. The Cambridge Analytica scandal awakened users of Facebook and platforms like it to the dangers of sharing personal information online. It also sparked a wider debate on the pervasive tracking of people on the web, often without consent.

At the same time, a series of high-profile data breaches have further rattled public confidence in the ability of companies to safeguard personal information. In 2018, Europe passed its far-reaching data privacy law, GDPR, and California, the epicenter of digital platforms powered by advertising dollars, followed suit. Brazil, Argentina, and Canada have since proposed or passed their own digital privacy legislation.

Nathalie Baracaldo was finishing her PhD when Google coined the term federated learning in its landmark paper. It wasn't a new concept — people had been splitting data and computation loads across servers for years to accelerate AI training. But the term, with its comforting and possibly unintended associations with Star Trek's United Federation of Planets, stuck.

Baracaldo now heads IBM's AI privacy and security team, and recently co-edited a book on federated learning covering the latest techniques on a range of privacy and security topics.

How federated learning works

Under federated learning, multiple people remotely share their data to collaboratively train a single deep learning model, improving on it iteratively, like a team presentation or report. Each party downloads the model from a datacenter in the cloud, usually a pre-trained foundation model. They train it on their private data, then summarize and encrypt the model's new configuration. The model updates are sent back to the cloud, decrypted, averaged, and integrated into the centralized model. Iteration after iteration, the collaborative training continues until the model is fully trained.

This distributed, decentralized training process comes in three flavors. In horizontal federated learning, the central model is trained on similar datasets. In vertical federated learning, the data are complementary; movie and book reviews, for example, are combined to predict someone's music preferences. Finally, in federated transfer learning, a pre-trained foundation model designed to perform one task, like detecting cars, is trained on another dataset to do something else, like identify cats. Baracaldo and her colleagues are currently working to incorporate foundation models into federated learning. Under one potential application, banks could train an AI model to detect fraud, then repurpose itl for other use cases.

The benefits of breaking down data silos

To make useful predictions, deep learning models need tons of training data. But companies in heavily regulated industries are hesitant to take the risk of using or sharing sensitive data to build an AI model for the promise of uncertain rewards.

In health care, privacy laws and a fragmented market have kept the industry from reaping AI's full potential. Federated learning could allow companies to collaboratively train a decentralized model without sharing confidential medical records. From lung scans to brain MRIs, aggregating medical data and analyzing them at scale could lead to new ways of detecting and treating cancer, among other diseases.

Federated learning could also help in a range of other industries. Aggregating customer financial records could allow banks to generate more accurate customer credit scores or improve their ability to detect fraud. Pooling car-insurance claims could lead to new ideas for improving road and driver safety, and aggregate sound and image data from factory assembly lines could help with the detection of machine breakdowns or defective products.

As more computing shifts to mobile phones and other edge devices, federated learning also offers a way of harnessing the firehose of data streaming minute-by-minute from sensors on land, sea, and in space. Aggregating satellite images across countries could lead to better climate and sea-level rise predictions at regional scales. Local data from billions of internet-connected devices could tell us things we haven't yet thought to ask.

“Most of this data hasn't been used for any purpose,” said Shiqiang Wang, an IBM researcher focused on edge AI. “We can enable new applications while preserving privacy.”

Balancing the privacy-accuracy trade-off

Attackers will always look for ways to steal user data or hijack an AI model no matter what training method is used. In federated learning, the weakest link occurs when a data host trades their working model with the central server. Each exchange improves the model but leaves the data that helped train it open to inference attacks.

“When you're dealing with highly sensitive and regulated data, these risks can't be taken lightly,” said Baracaldo, whose book includes a chapter on strategies for preventing data leakage.

“The more rounds of information you exchange, the easier it is to infer information, particularly if the underlying information hasn’t changed much,” said Wang. “That’s especially true as you converge on a final model when the parameters don’t change much.”

“Legal and technology teams need to balance this trade-off between privacy and accuracy,” Wang added. “To train a distributed model you have to share something. But how do you make sure that what you’re sharing won’t violate privacy rules? It depends on the application.”

An AI tumor detector, for example, may need to be more accurate than a tool for predicting the next words you plan to type. But health care data also require stronger privacy and security guarantees. Much of the current research in federated learning, therefore, focuses on minimizing and neutralizing privacy threats.

Secure multi-party computation hides model updates through various encryption schemes to reduce the odds of a data leak or inference attack; differential privacy alters the precise values of some data points to generate noise designed to disorient the attacker.

Other challenges: efficiency, transparency, and incentives for good behavior

Training AI models collaboratively, in multiple places at once, is computationally intensive. It also requires high communication bandwidth. That’s especially true if data hosts are training their local models on-device.

To handle the bandwidth and computing constraints of federated learning, Wang and others at IBM are working to streamline communication and computation at the edge. Some of the proposed efficiency measures include pruning and compressing the locally trained model before it goes to the central server.

Transparency is another challenge for federated learning. Because training data are kept private, there needs to be a system for testing the accuracy, fairness, and potential biases in the model’s outputs, said Baracaldo. She and her colleagues at IBM have proposed an encryption framework called DeTrust that requires all parties to reach consensus on cryptographic keys before their model updates are aggregated.

“Adding a consensus algorithm ensures that important information is logged and can be reviewed by an auditor if needed,” Baracaldo said. “Documenting each stage in the pipeline provides transparency and accountability by allowing all parties to verify each other’s claims.”

Another challenge for federated learning is controlling what data go into the model, and how to delete them when a host leaves the federation. Because deep learning models are opaque, this problem has two parts: finding the host’s data, and then erasing their influence on the central model.

Currently, if data are deleted, the parties are obligated to retrain the model from scratch. To save computation, Baracaldo and her colleagues have proposed a method for unwinding the model only to the point at which the now-erased data were added.

A final challenge for federated learning is trust. Not everyone who contributes to the model may have good intentions. Researchers are looking at incentives to discourage parties from contributing phony data to sabotage the model, or dummy data to reap the model's benefits without putting their own data at risk.

"There needs to be an incentive to have everyone participate truthfully," Wang said.