PakTech Today | Pakistan's Premier Technology Magazine
Vol. 01, Issue 17 • ISSN: 3136-2729

TECHNOLOGY FEATURE Industrial News / Technology Report / Research Communication

Federated Learning: Training AI Without Sharing Your Data

As data privacy concerns grow worldwide, a new approach to Artificial Intelligence is training models to learn from data without ever centralizing it. Federated Learning is changing the way we make smart systems in a world where privacy comes first.

By **Eman Shahid** | University of Agriculture, Faisalabad | Published: September, 2026
emanshahid278@email.com |Position: Computer Science Student

Introduction

Artificial Intelligence is now an essential component in modern technology. Recommendations, speech recognition, self-driving cars, and diagnostics in the medical industry are made possible by AI. However, all machine learning algorithms require massive amounts of data for training purposes that must be stored on large central servers. Although such practices have been fruitful in bringing about innovations, they have also created serious concerns regarding user privacy and freedom.

In recent years, there have been plenty of cases when data breaches affected user's lives negatively, making them careful about sharing their data. New global standards have been suggested through regulatory frameworks like the **European Union's General Data Protection Regulation (GDPR)**, demanding companies to reconsider their approach towards data processing. In Pakistan, where digital transformation is accelerating under initiatives such as Digital Pakistan, there are currently no laws that would encompass all types of personal data.

One potential solution that is worth exploring is **Federated Learning.** It takes a different approach by moving the model to the data rather than sending it to central server. It enables multiple devices to jointly train one model without disclosing any of their sensitive data. This may prove to be a disruptive technology for AI systems.

> ***" Federated Learning shifts the paradigm from bringing data to the model, to bringing the model to the data***

Technology Overview

Federated Learning is a machine learning approach that enables multiple client devices, such as smartphones, laptops, IoT sensors, or entire organizational servers, to jointly train a global model while keeping the data on their devices. A central coordination server begins the process by setting up a global model and sharing it with a small number of clients. Each device then trains the model using their own data; calculating model updates (gradients or weights) without revealing the data used to train the model.

The server only receives these encrypted model updates, not the raw data. The server then uses algorithms like Federated Averaging (FedAvg) to create an improved model. The process continues in several communication rounds until the model is good enough. This means that the original data remains on the device.

Three Different Types of Architecture

There exist three different categories for Federated Learning based on their distinct application domains. **Horizontal Federated Learning** occurs if each entity possesses information about various users but has the same feature space, as seen in typical consumer applications including mobile keyboard predictions. The **Vertical Federated Learning** occurs when all entities have data about the same users but have various features, which help different entities from various sectors engage in cooperation in developing, for instance, a credit scoring model that involves both a bank and an e-commerce website**.** Lastly, **Federated Transfer Learning** is performed with various datasets where the sample space and feature spaces do not intersect at all.

The transition from centralized to decentralized learning represents one of the most significant architectural changes in modern AI systems.

Real-World Deployments

One of the most prominent examples of Federated Learning is **Google's Gboard keyboard**, where typing suggestions and emojis are customized according to the user's actions without transmitting personal messages to external servers. Another popular application of FL has been seen in enhancing **Siri's voice recognition** ability by Apple, while NVIDIA uses federated learning in analyzing **medical imaging** at different hospitals.

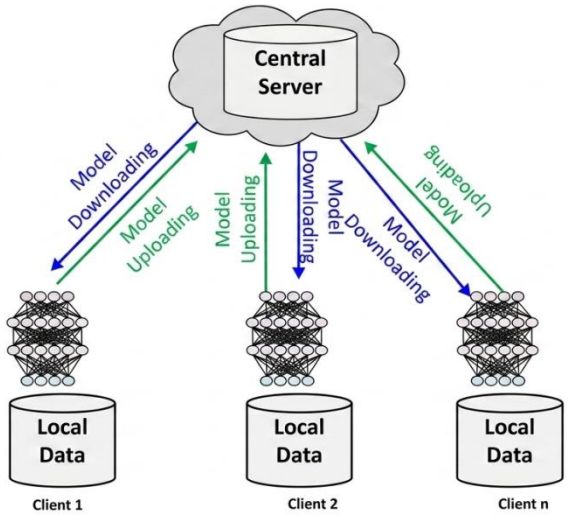

Figure 1: Federated Learning architecture showing decentralized model training, where client devices train locally on private data and only model updates are shared.

Relevance to Pakistan

In Pakistan, Federated Learning can become a groundbreaking technology for industries that face a significant challenge from data sensitivity. For example, in the healthcare sector, the use of federated learning would allow hospitals within Punjab and Sindh to train diagnostic models for diseases such as diabetes retinopathy or tuberculosis diagnosis through medical scans, without violating any ethical standards of medicine and new data protection laws through exchange of patient data. For the finance industry, banks and fintech firms could collaborate to enhance their fraud detection strategies without having to share customer transaction data. In agriculture, IoT sensors installed in various research stations could help create models that predict crop yields and identify pests and optimize irrigation use, but with decentralized farm-specific data.

Key Findings & Impact

The use of data in AI is being revolutionized by Federated Learning. It decentralizes data and brings the risk of any large-scale data breaches down significantly and enables practice of privacy-by-design. Data of the user will

be stored locally, reducing the risk of exposure to third party and remote servers. By only sharing model updates and not data, it also reduces network traffic. Most importantly, this approach is data protection by design and can help future data protection regulations in Pakistan if they are in place, which is a big plus for companies developing AI products that are compliant with regulations like GDPR.

Most importantly, Federated Learning enables new power to data that can't be shared or consolidated. Using medical data, financial transactions or even the behavior of personal devices for training without infringing on privacy or legal regulations. With potentially millions of edge devices available for training, the amount of data is vast.

Challenges and Open Problems

There are several challenges to Federated Learning. The most obvious one is that of statistical heterogeneity: various devices produce different types of quantities and qualities of data, which is referred to as non-IID (non-Independent and Identically Distributed). Such mismatches in distribution can lead to training instabilities, slow convergence, and lower final accuracy of the trained model when compared to idealized centralized training.

High communication costs are also an issue. Although only model updates are communicated, the rounds of communication with potentially millions of devices add delays and bandwidth overheads that researchers are actively working to reduce though gradient compression and quantization techniques. Finally, in terms of security, the aggregated model updates can also be revealed. For instance, model inversion attacks might try to recover training data from the gradients, and model poisoning attacks might involve bad actors injecting bad data into the global model. Defenses such as Differential Privacy, which adds appropriate noise to updates, and Secure Multi-Party Computation are being integrated to protect FL systems against these attacks.

The issue of systems heterogeneity adds another layer of complexity: computing capabilities, energy consumption, and connectivity differ greatly across devices in the network. The presence of stragglers can impact the speed of data aggregation. Creating a system of incentives that encourages stakeholders to share high-quality data and resources continues to be socio-technical challenge without a clear solution.

Impact on Pakistan's Tech Ecosystem

Federated Learning brings both opportunities and challenges for Pakistan's growing tech ecosystem, with expanding mobile adoption, thriving startups, and digital information initiatives by the government. There are both transformative opportunities and strategic challenges presented by federated learning. Startups based in Pakistan can leverage FL platforms such as TensorFlow Federated, PySyft, and FATE to develop private-sector applications in health-tech, agri-tech, and fintech domains.

Academic research teams are =well-equipped to experiment with FL's local applicability, for example, a collaboration between universities to build models for Urdu and other regional languages in NLP, trained on distributed text data collections without sharing sensitive linguistic data. This kind of research can simultaneously advance global AI science and address the unique needs of Pakistan.

What's Next?

Program Federated Learning to be more widely adopted across other domains and countries, and overcoming existing challenges, is the key to the future of Federated Learning. Researchers are quickly developing the ability to address non-IID data distributions, communication compression, and to secure training pipelines from adversarial attacks. Edge computing is already moving the field forward; bringing intelligence closer to the devices that create it and eliminating the need for central infrastructure.

For Pakistan, the time to act is now. As digitization becomes more rapid and more formalized are data protection regimes established by the relevant authorities, it is no longer an option, but a necessity to include private-centric approaches to AI technology in national strategies. Government, private sector, and the academic world must work together and play their part in developing expertise, research, and creating policy environments that are conducive to the development of expertise, research and policy.

Federated Learning helps keep data where it belongs; with the data owner, in a world of data, as the new oil. For a nation of 240 million people that is becoming more digital, the same goes for developing AI that Pakistanis can trust.

About the Author
Eman Shahid is a Computer Science student at the University of Agriculture, Faisalabad, with a strong interest in cyber security, systems design, and low-level computing. She focuses on bridging theoretical computer science with real-world applications in emerging digital ecosystems. She can be reached at emanshahid278@email.com.

Keywords: Artificial Intelligence; Cybersecurity; IoT; Federated Learning; Data Privacy; Machine Learning; Software Engineering; Pakistan Tech

References

[1] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS).* Google Research.

[2] Kairouz, P., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning,* 14(1–2), pp. 1–210.

[3] Bonawitz, K., et al. (2019). Towards federated learning at scale: A system design. *Proceedings of SysML 2019.* Google AI.

[4] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine,* 37(3), pp. 50–60.