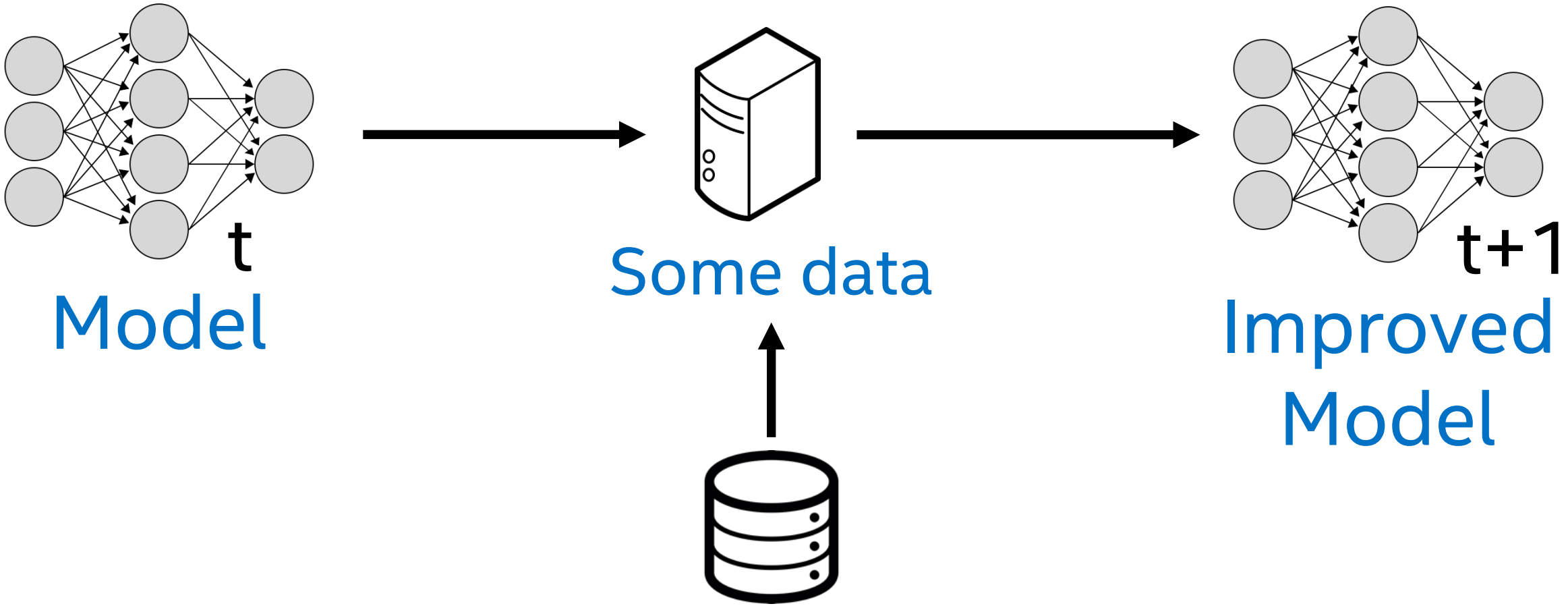
Federated Learning

Walter Riviera – AI TSS

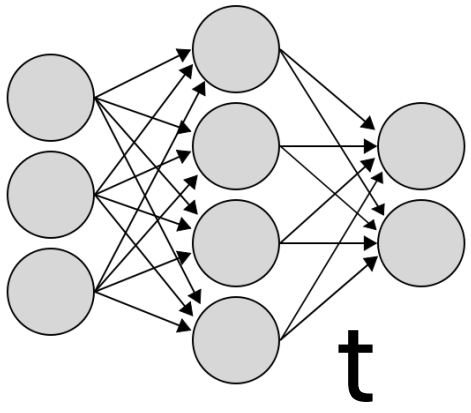
Datacenter group

Machine Learning/ Deep Learning



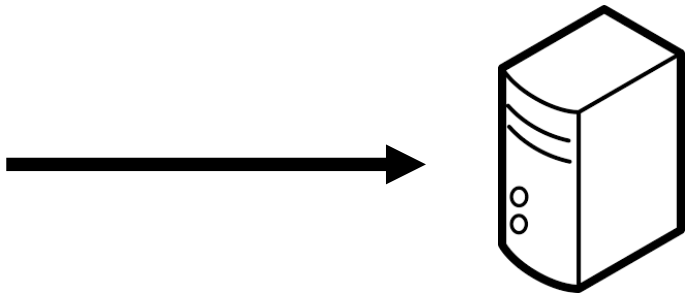
Eventually, we hit the limit of our dataset (information value)

Machine Learning/ Deep Learning

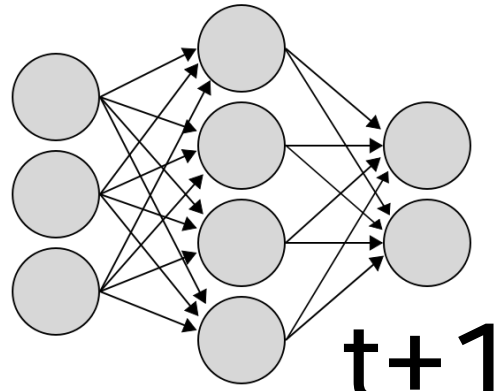


t

Model

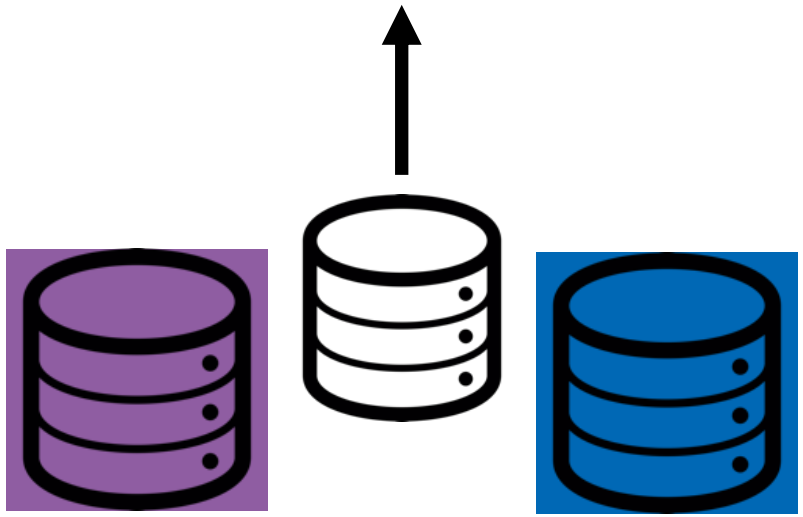


Some data



t+1

Improved Model

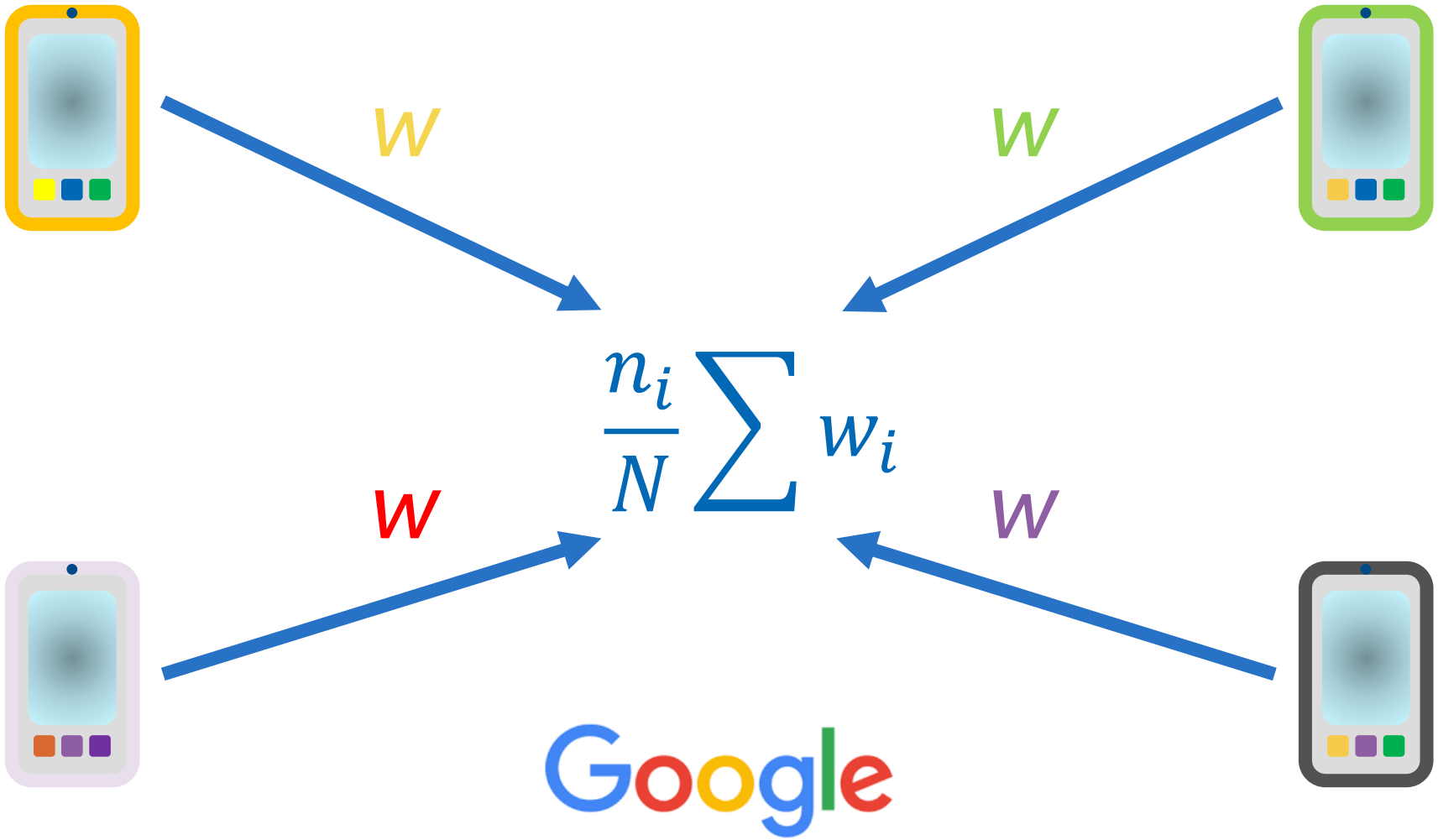


Can we add more data?

Challenges for Training AI Models?

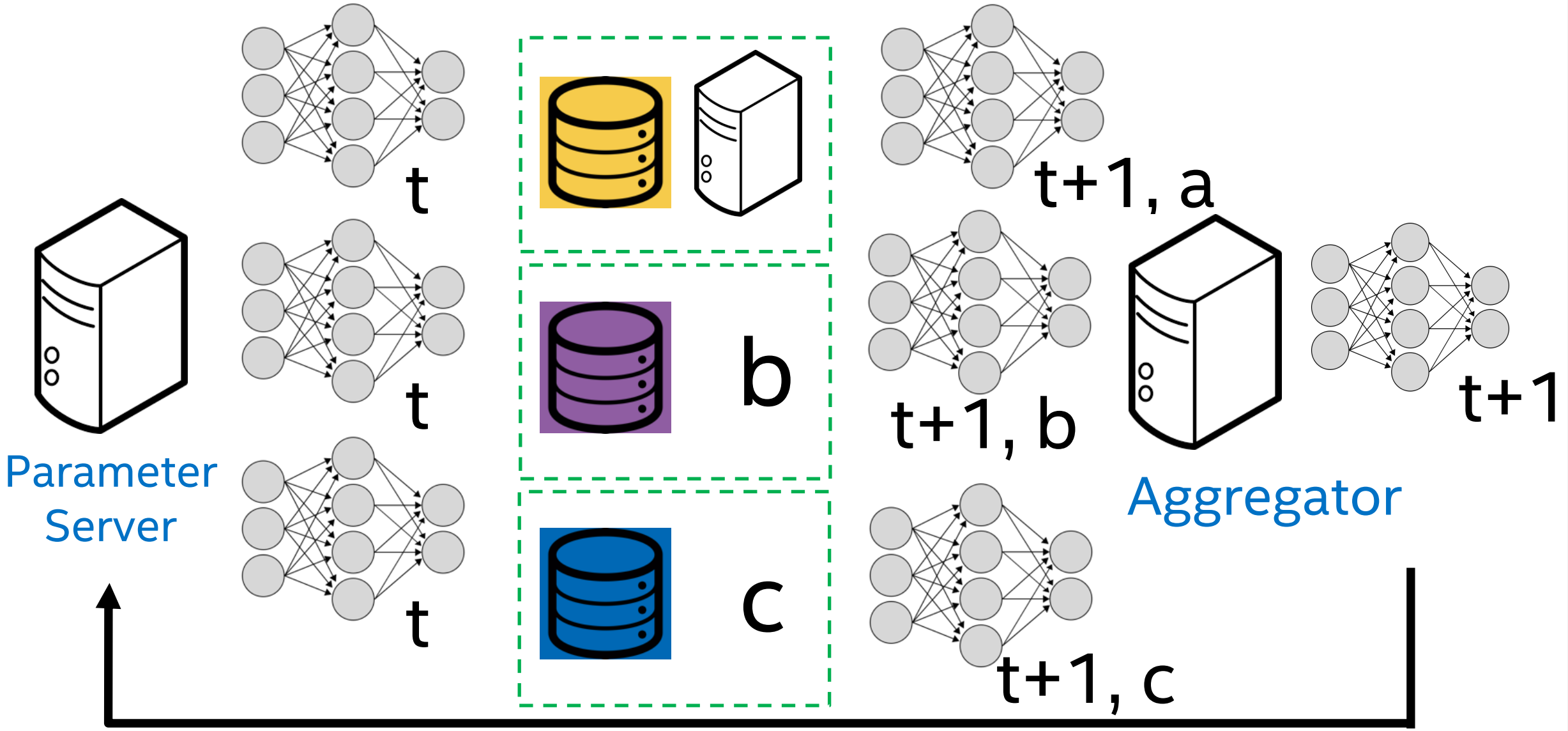
- Data is legally protected (HIPAA, GDPR, POPIA)
- Data is sensitive
- Data too valuable or value unknown
- Data too large to transmit

Federated Learning



https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Federated Learning



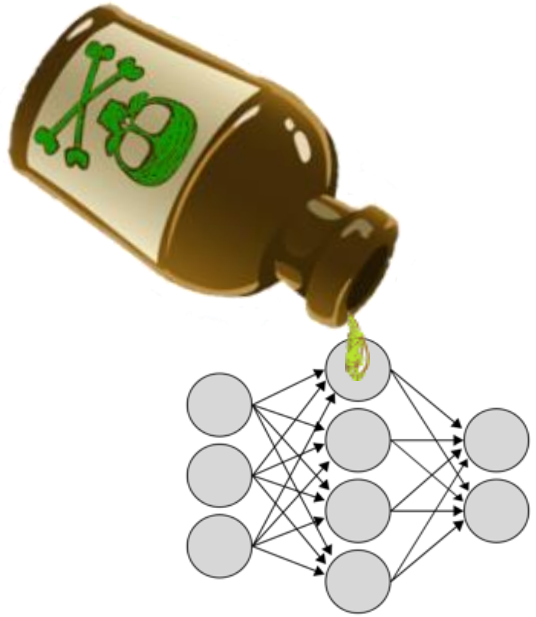
Key Takeaways about Federated Learning

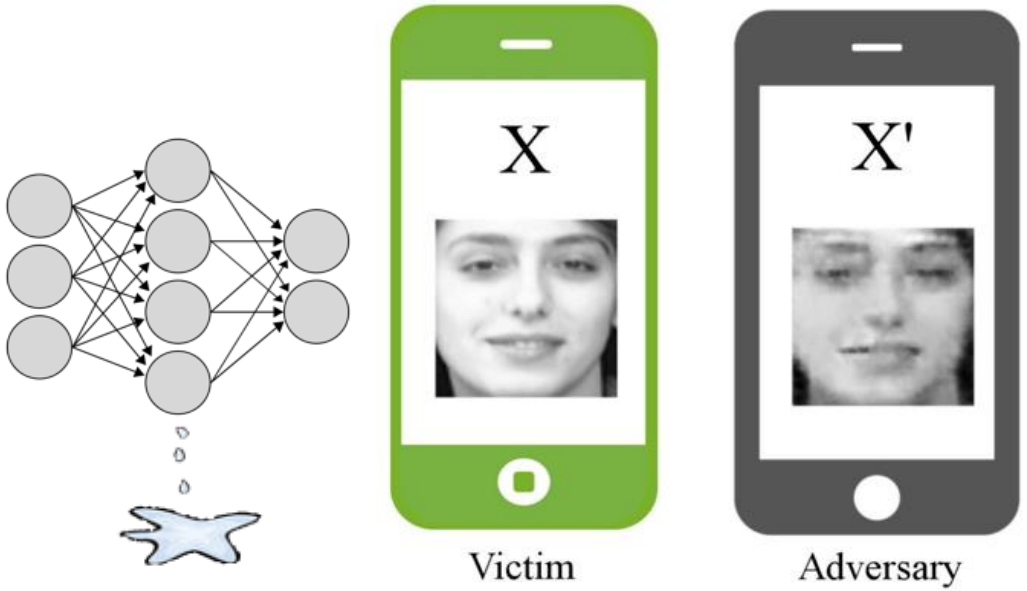
- FL solves a lot of data access problems.

But …

- FL doesn't protect the model.
- FL doesn't protect the training.
- FL doesn't completely protect the data.

FL Could Increase Security and Privacy Risks





Poisoning attacks may maliciously alter models.

Extraction attacks recover training data from models.

FL needs to have additional **security** to manage these risks

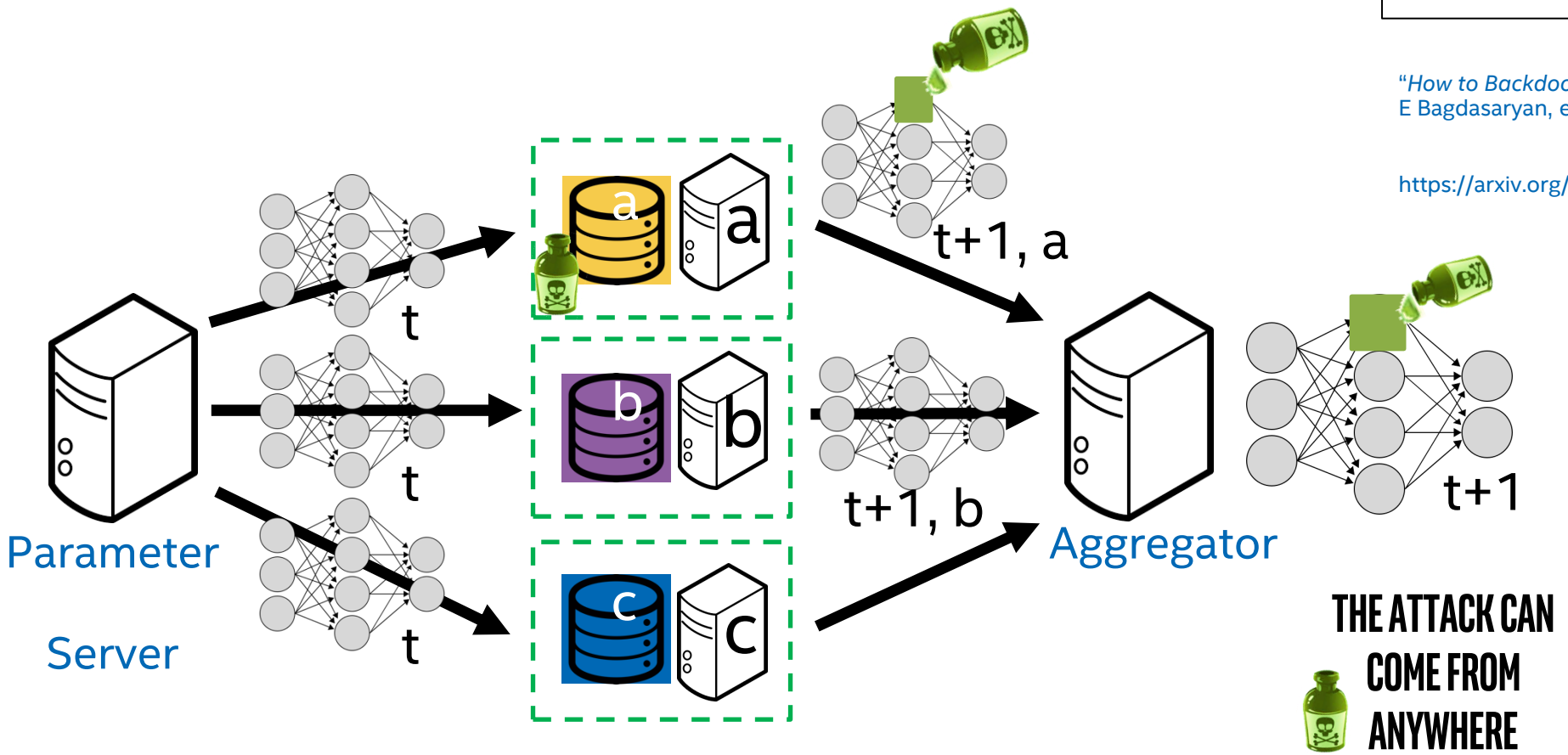
Models can be poisoned

"*Analyzing Federated Learning through an Adversarial Lens*", Bhahogi, et al.

https://arxiv.org/abs/1811.12470

"*How to Backdoor Federated Learning*", E Bagdasaryan, et al.

https://arxiv.org/pdf/1807.00459.pdf



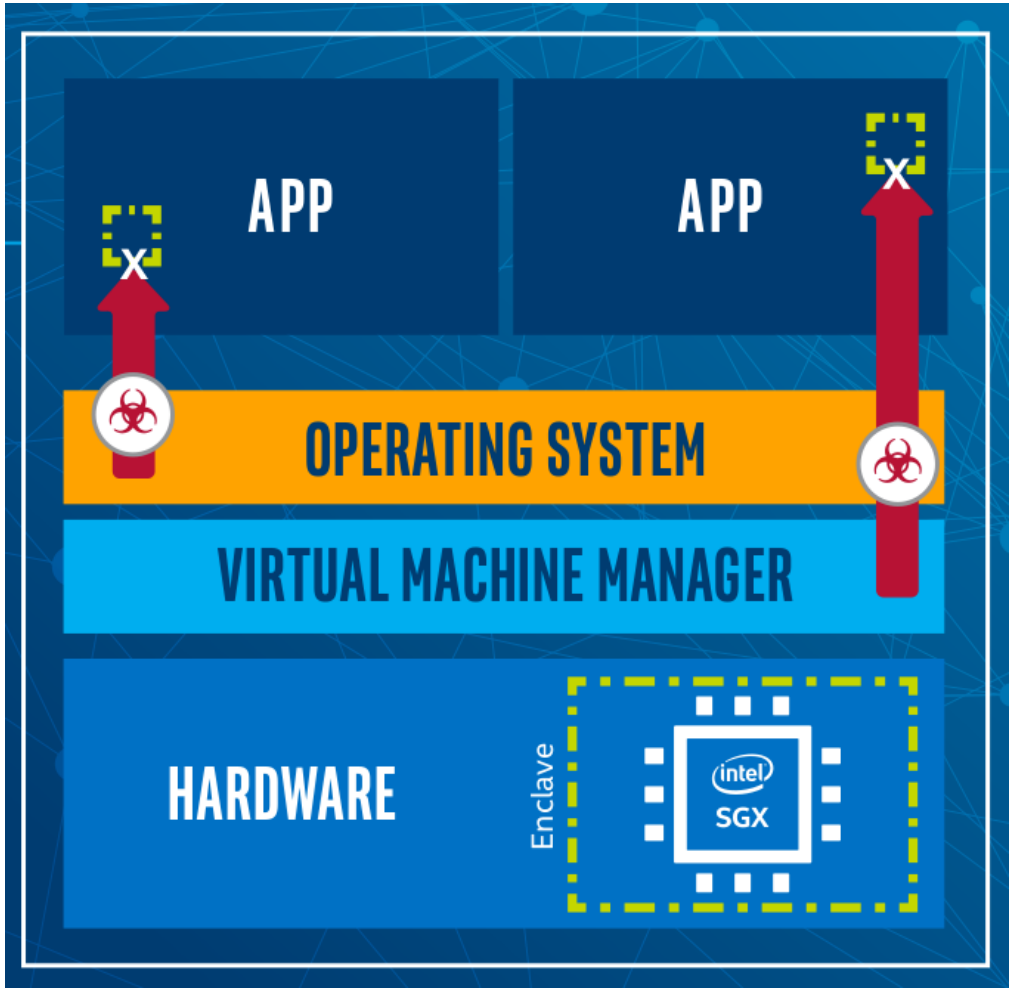
What Does it take to Secure FL?

- **Confidentiality** – Protecting data and models from the risk of exposure to untrusted parties during runtime
- **Execution Integrity** – Protecting the computation (i.e. training the model) from being changed at runtime
- **Attestation** – Validating that the software and hardware are genuine

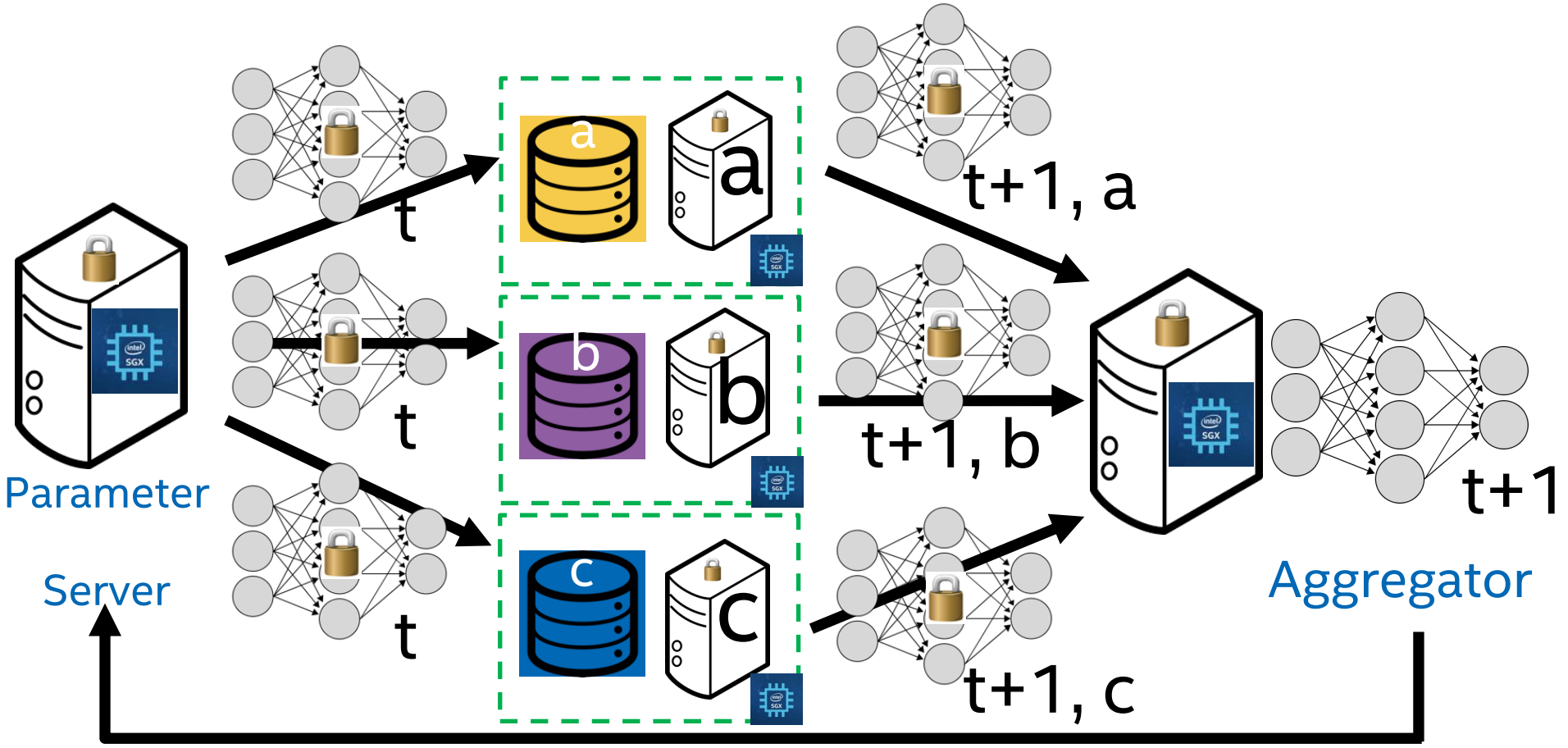
What is Intel® SGX?

Intel® Software Guard EXtensions is a set of CPU instructions that can be used by applications to set aside private regions of code and data.

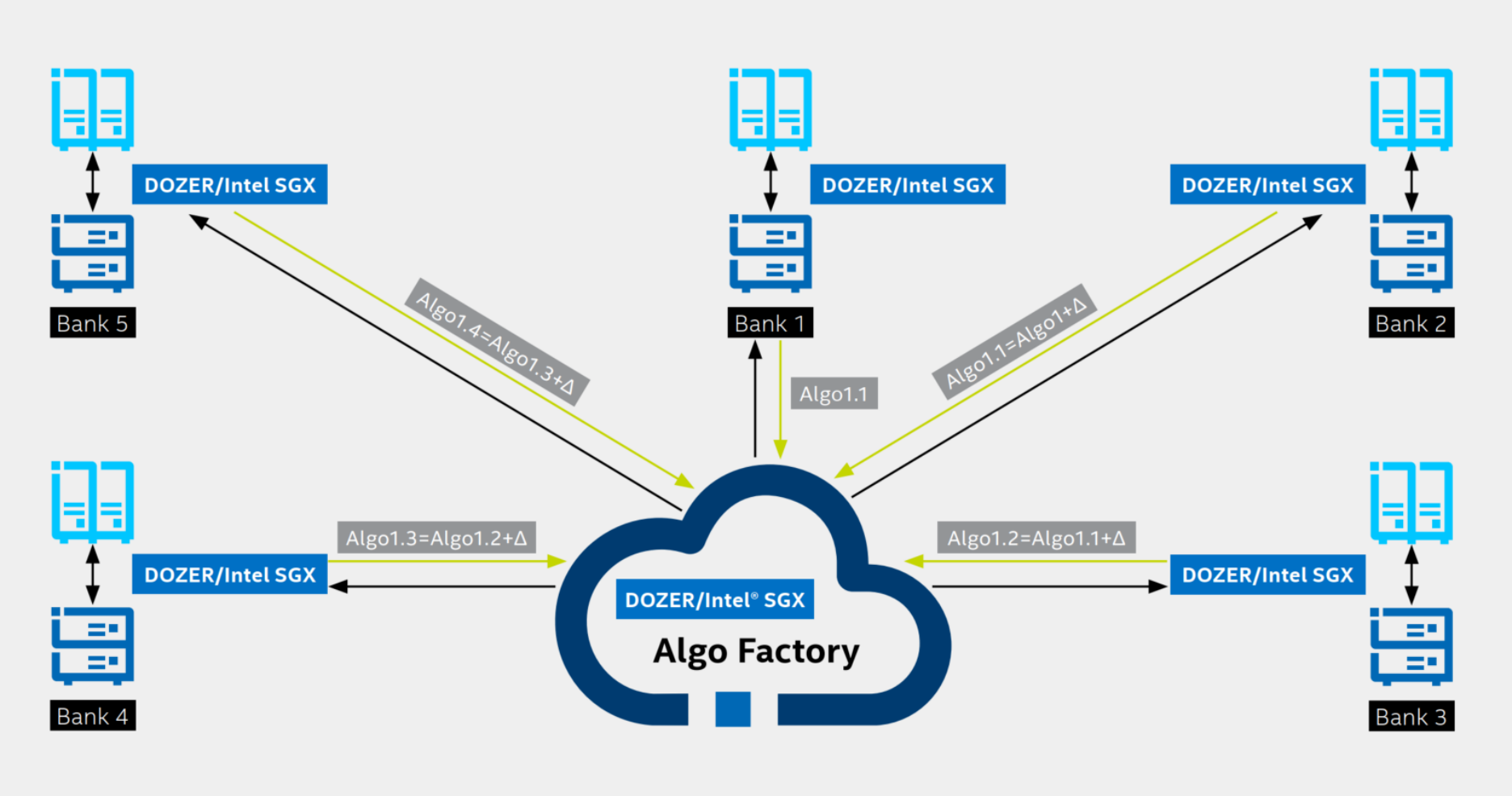
Security During Execution in Hardware



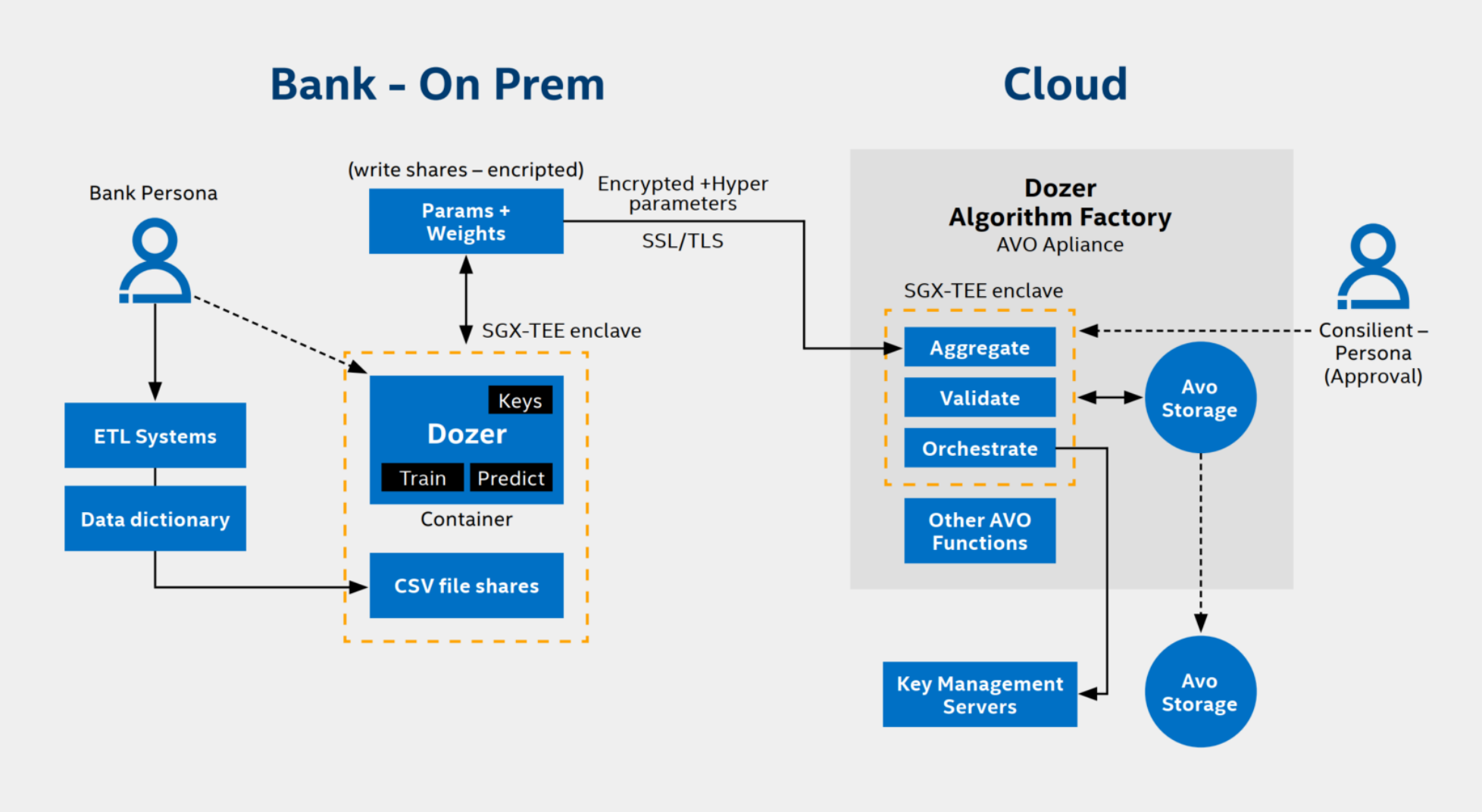
The security ally



Consilient use-case



Consilient use-case



What is the Opportunity?

- “AI healthcare algorithms take 16-30 months of development at $1.5M - $2.5M total cost (on validation)” [Bob Rogers, BeeKeeperAI]
- “The global federated learning market size is expected to grow from US$1.41 billion in 2017 to US$8.81 billion by 2022 at a CAGR of 44.1%” [industrywired.com Nov’20]
- By 2025, half of large organizations will implement privacy enhancing computation for processing data in untrusted environments and multiparty data analytics use cases." [Gartner Technology Trends for 2021]

Intel-UPenn Collaboration

Brain tumor segmentation finds tumors from MRIs

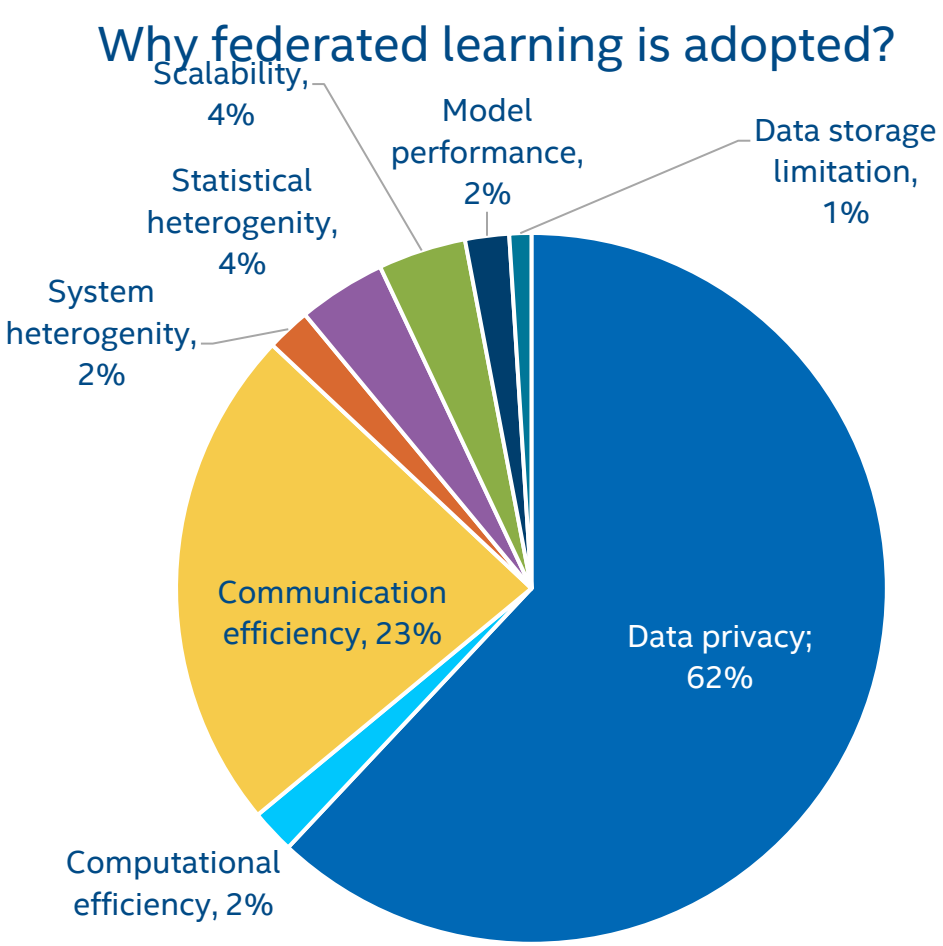
How much better does each institution do when training on the full data vs. just their own data?

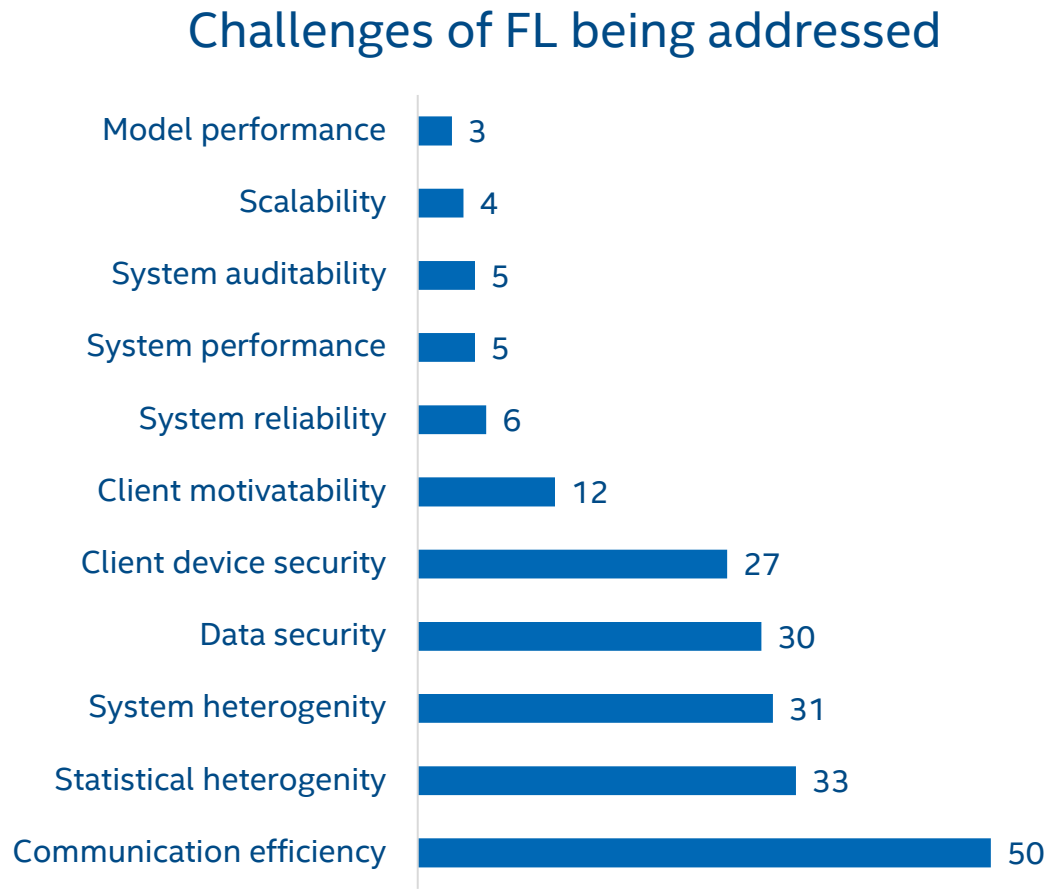
17% BETTER on the hold-out BraTS data

2.6% BETTER on their own validation data

Sheller, M.J., Edwards, B., Reina, G.A. *et al.* Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* **10,** 12598 (2020).

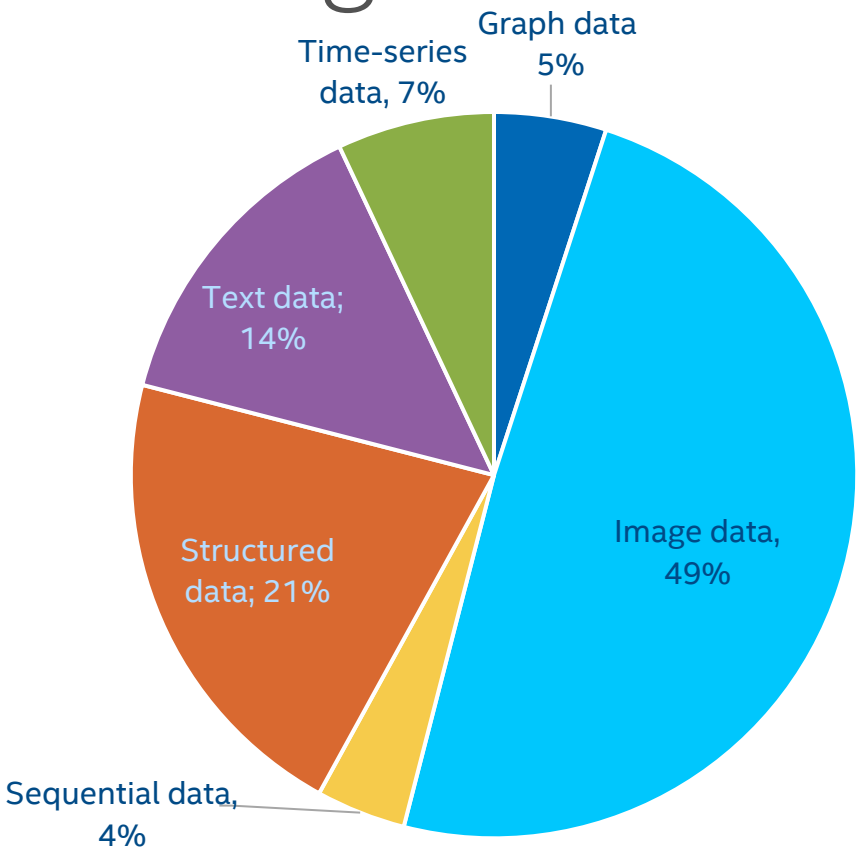
Why Federated Learning is adopted?





arxiv.org/abs/2007.11354

What are the applications of Federated Learning?

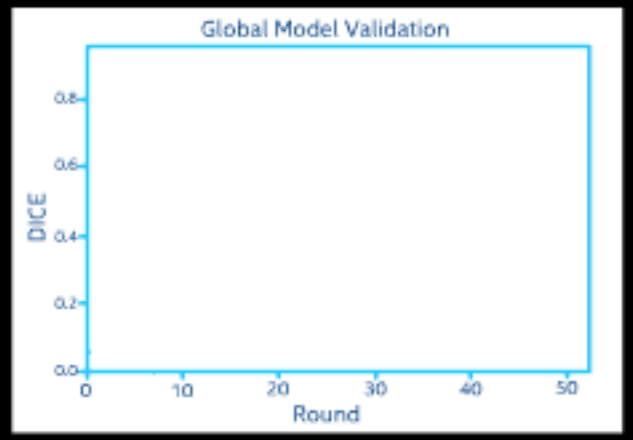


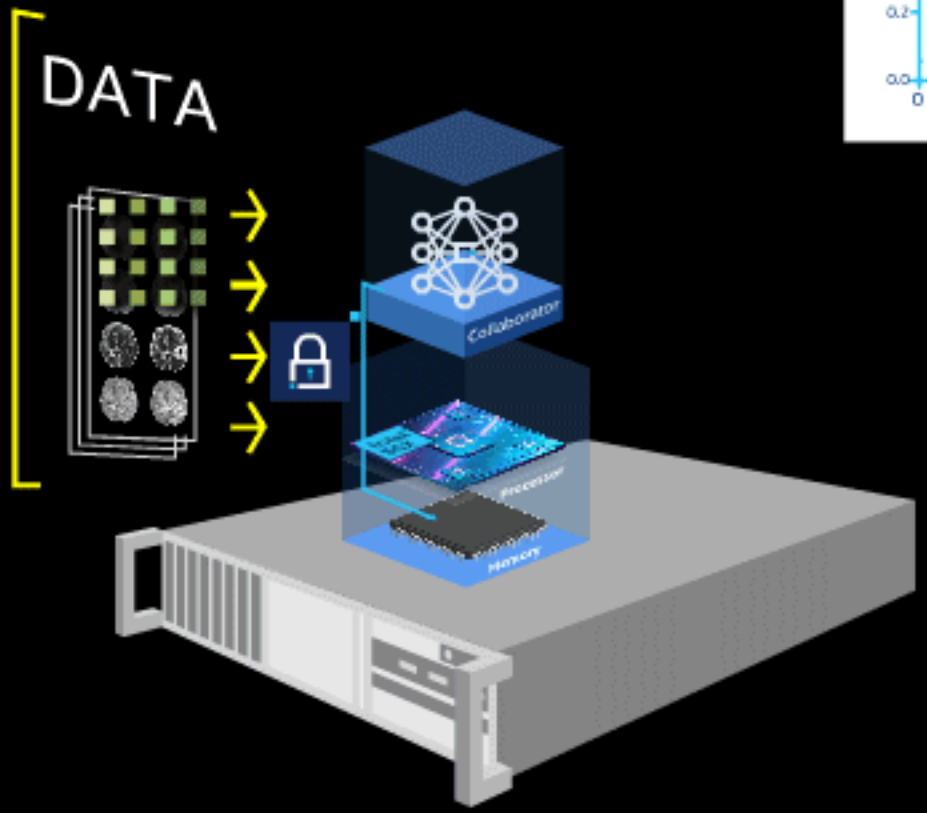
Data Types for Federated Learning

- **Image data (49%)**: Autonomous driving, Healthcare, Facial recognition, Handwritten character/digit recognition, Human action prediction
- **Structured data (21%)**: Healthcare, Credit card fraud detection, Bankruptcy prediction
- **Text data (14%)**: Keyboard suggestion, Sentiment analysis, Spam detection

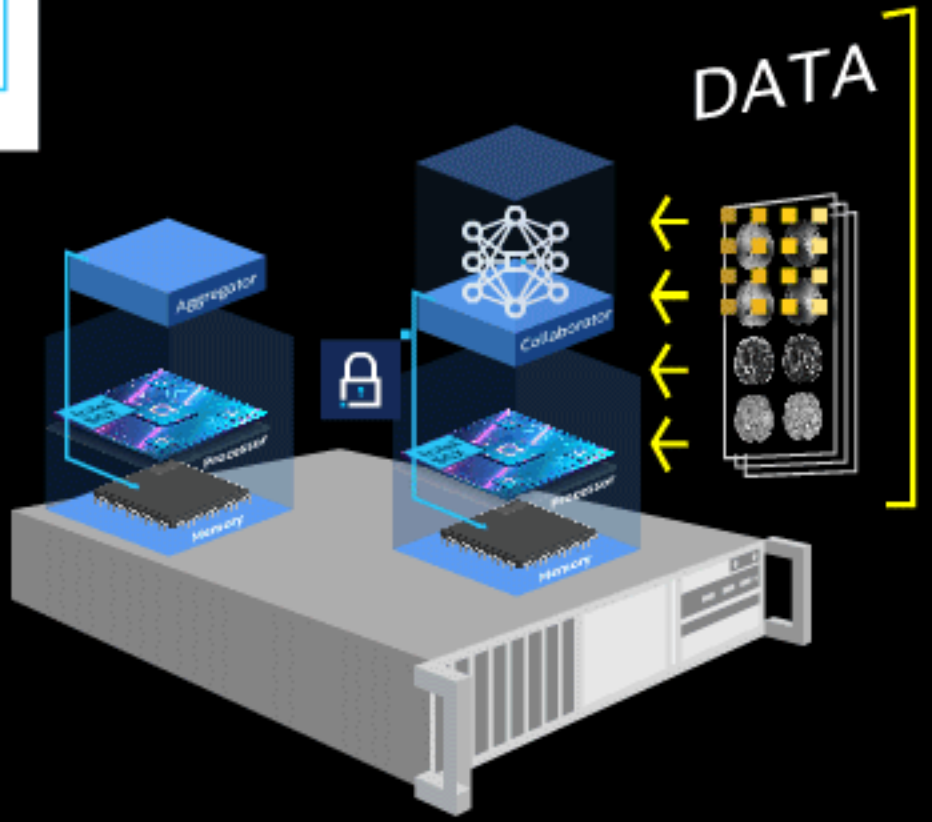
arxiv.org/abs/2007.11354

Federated Learning based on Intel® SGX



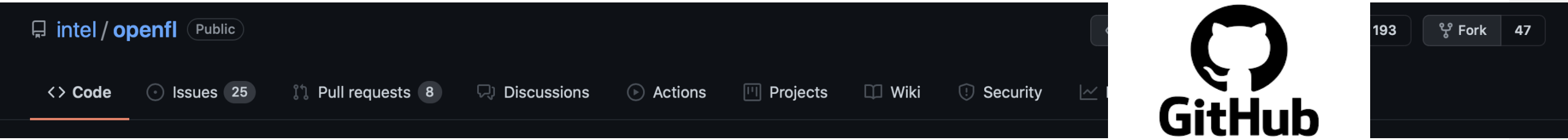


3rd Gen Intel® Xeon® Scalable Processors



3rd Gen Intel® Xeon® Scalable Processors

OpenFL: Open Federated Learning library

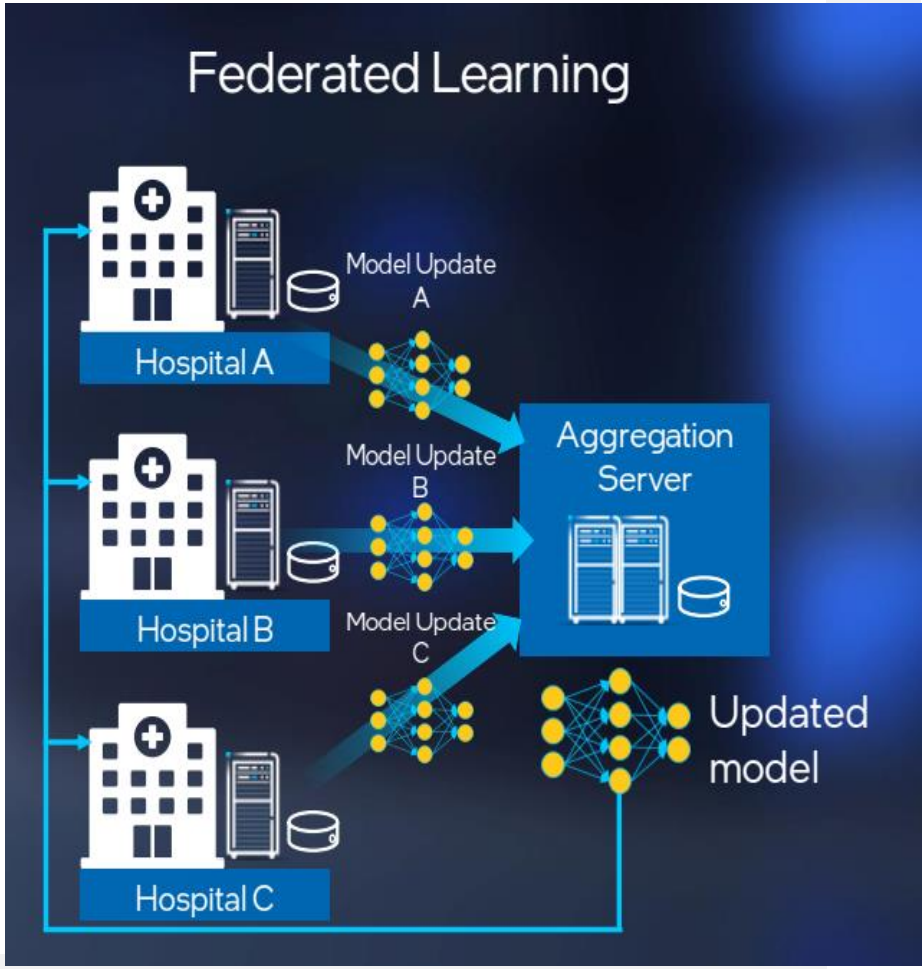


v1.3

- **Open Source Release (Apache v2, github.com/intel/openfl)**
- **For installation: PyPI is available!** >> pip install openfl
- PyTorch / TensorFlow 2 (Keras) / FastEstimator support
- Long-living services (Director/Envoy) – *set up a Federation once, and reuse it for multiple experiments*
- Interactive API (Jupyter) – *seamless play in Jupyter with remote data as it would be your local data*
- Handful examples: *UNet Polyp Segmentation, Brain Tumor Segmentation, Federated Fine-tuning for TinyImageNet, PyTorch Person Re-ID for Market, FedProx for MNIST, and even more…*
- Robust aggregation via FedProx & custom function aggregation interface

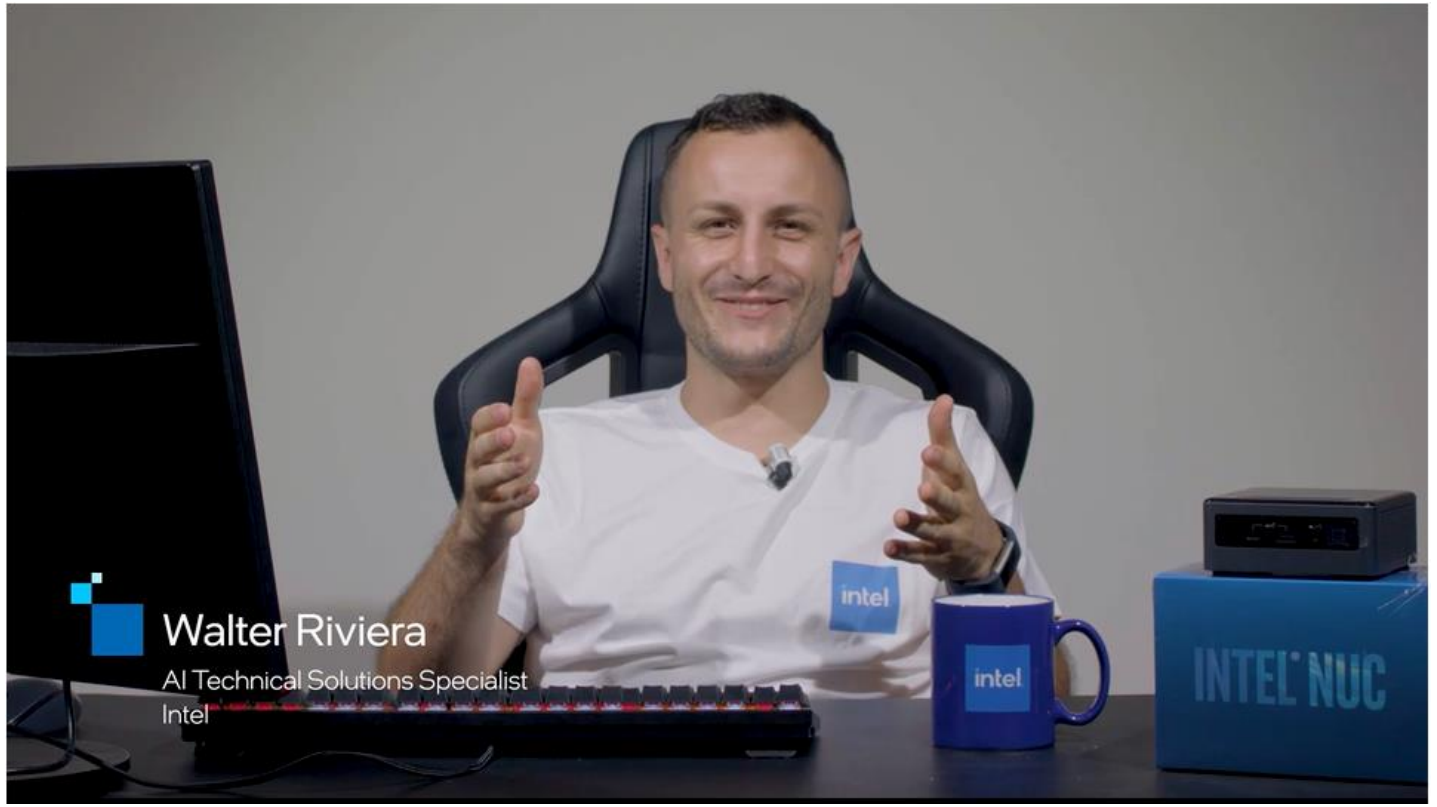
https://github.com/intel/openfl/tree/develop/openfl-tutorials/interactive_api

Docs: openfl.readthedocs.io/en/latest/index.html



Guided tutorial

FL video series with 4 episodes



https://www.intel.co.uk/content/www/uk/en/now/ai-video-series/building-the-federation.html


intel