Federated learning: Data for AI in the age of data regulations

Walter Riviera – EMEA AI Technical Lead

intel®

Challenges for Training AI Models

- Data is legally protected (HIPAA, GDPR)
- Data is sensitive
- Data is too valuable to share
- Data silo problem: data is too large to transmit

What is Federated Learning?

- Standard machine learning approaches require centralizing the training data on one machine or in a datacenter.
- Federated Learning enables training models on distributed and private datasets without the need to centralize them.
- Privacy and security are key considerations for data set owners participating in Federated Learning optimizations.

Financial Services and Insurance

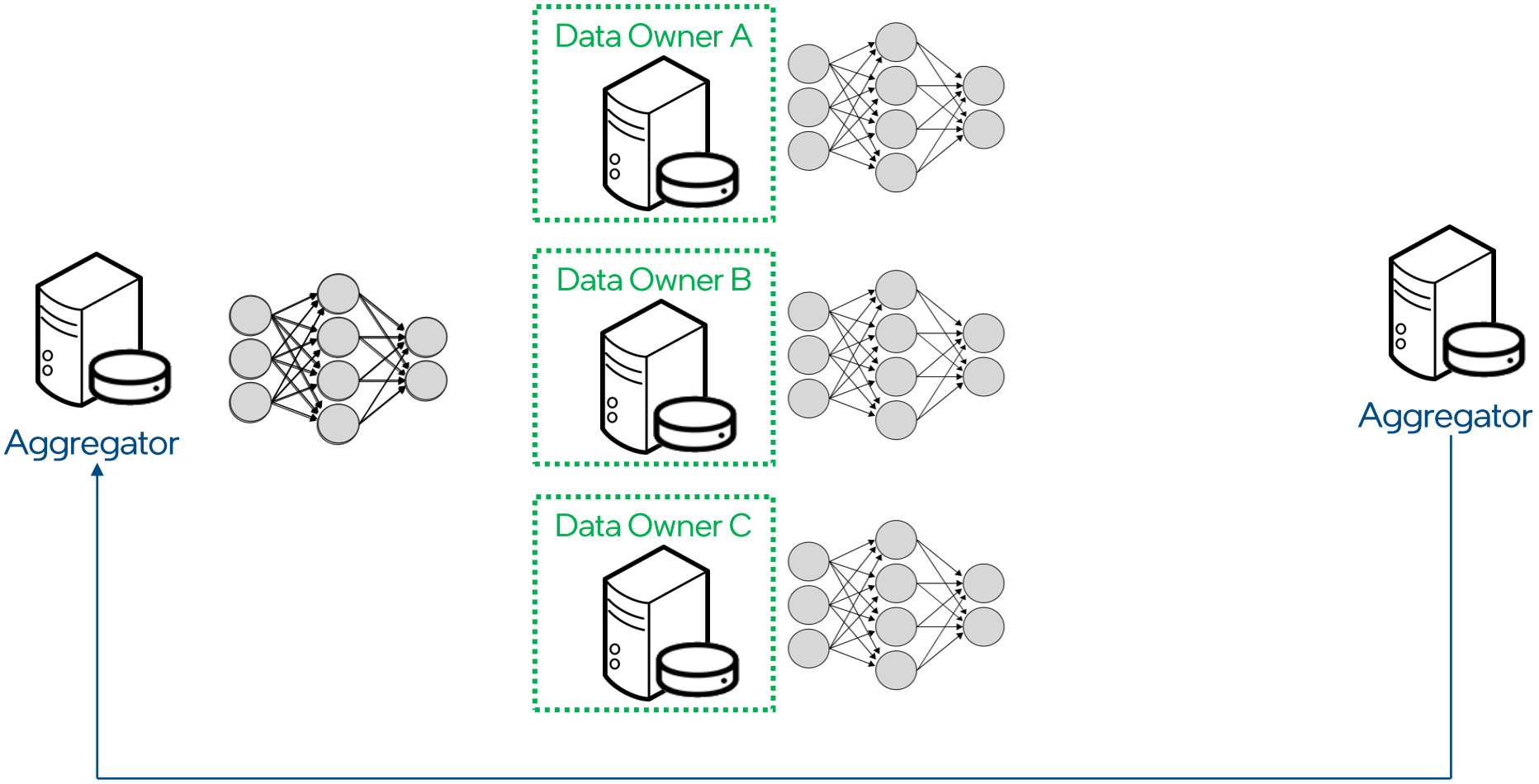
Healthcare

Manufacturing

Media

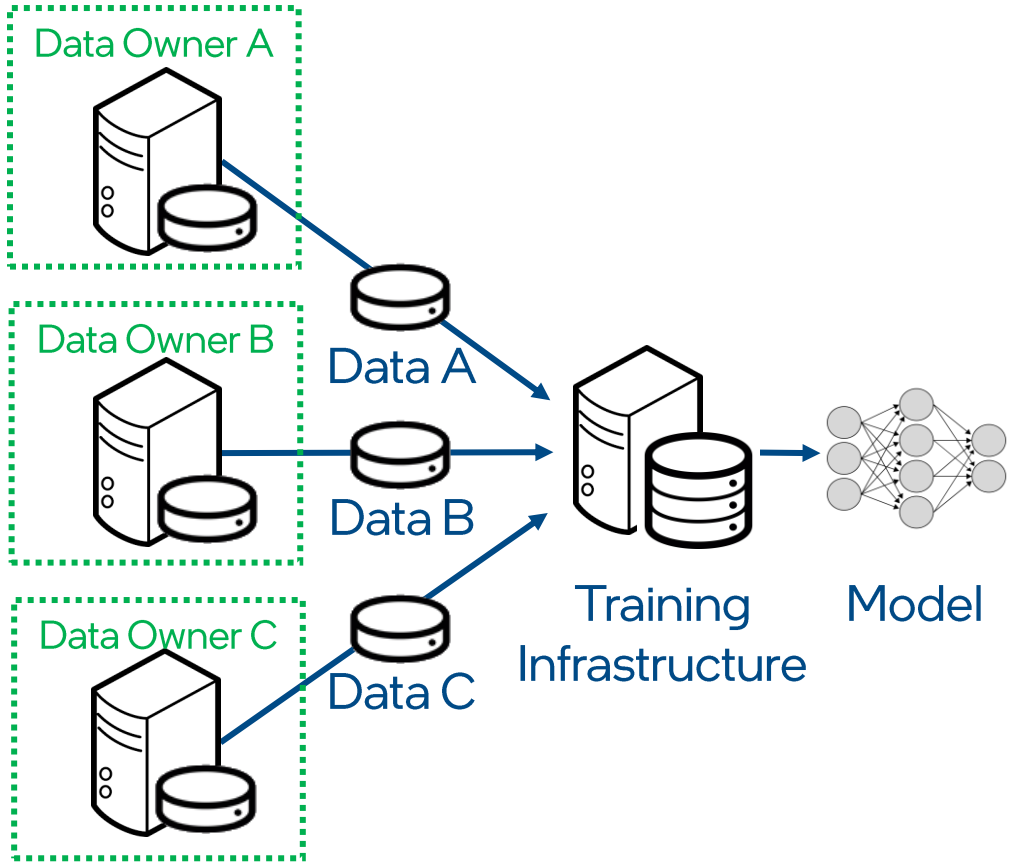
What is Federated Learning?

A dynamic example

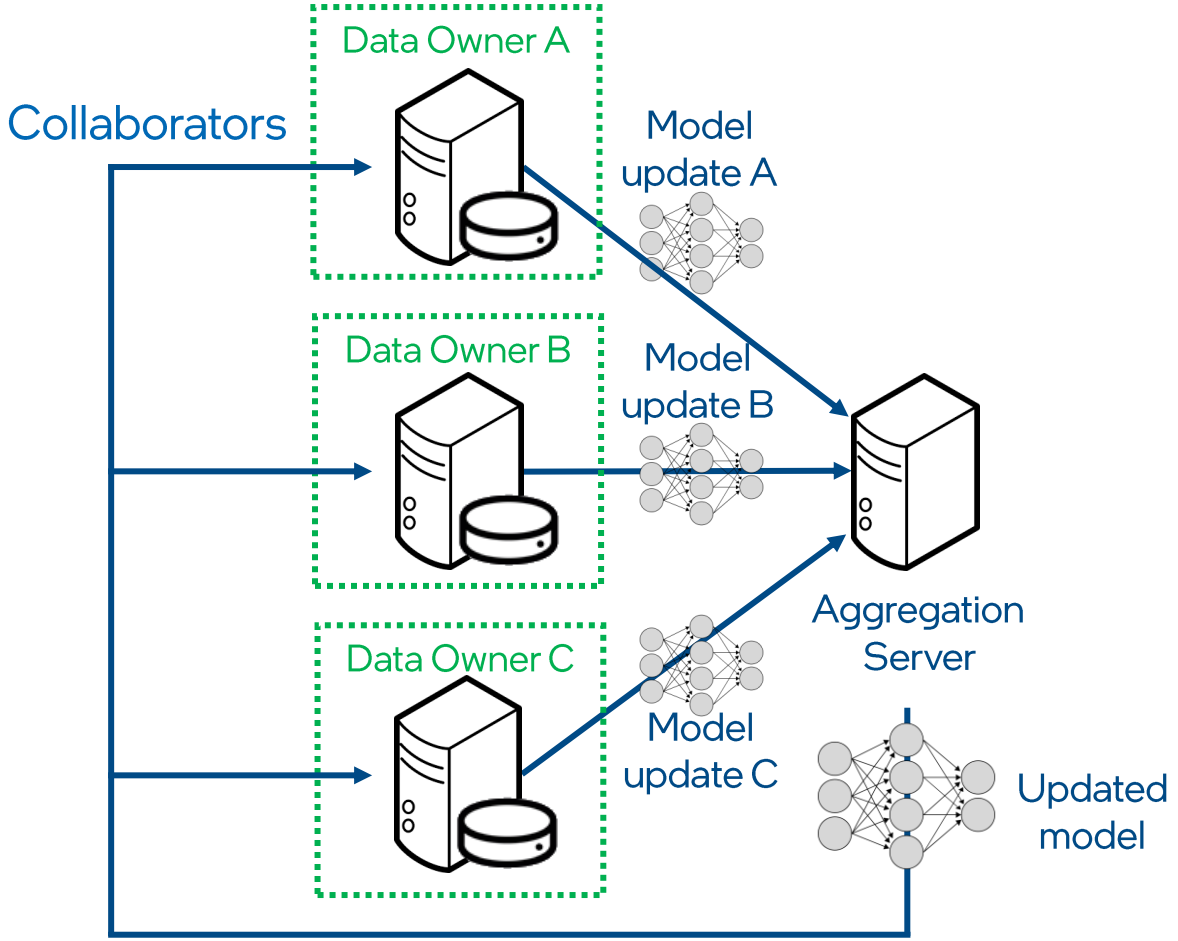


Centralized Learning versus Federated Learning

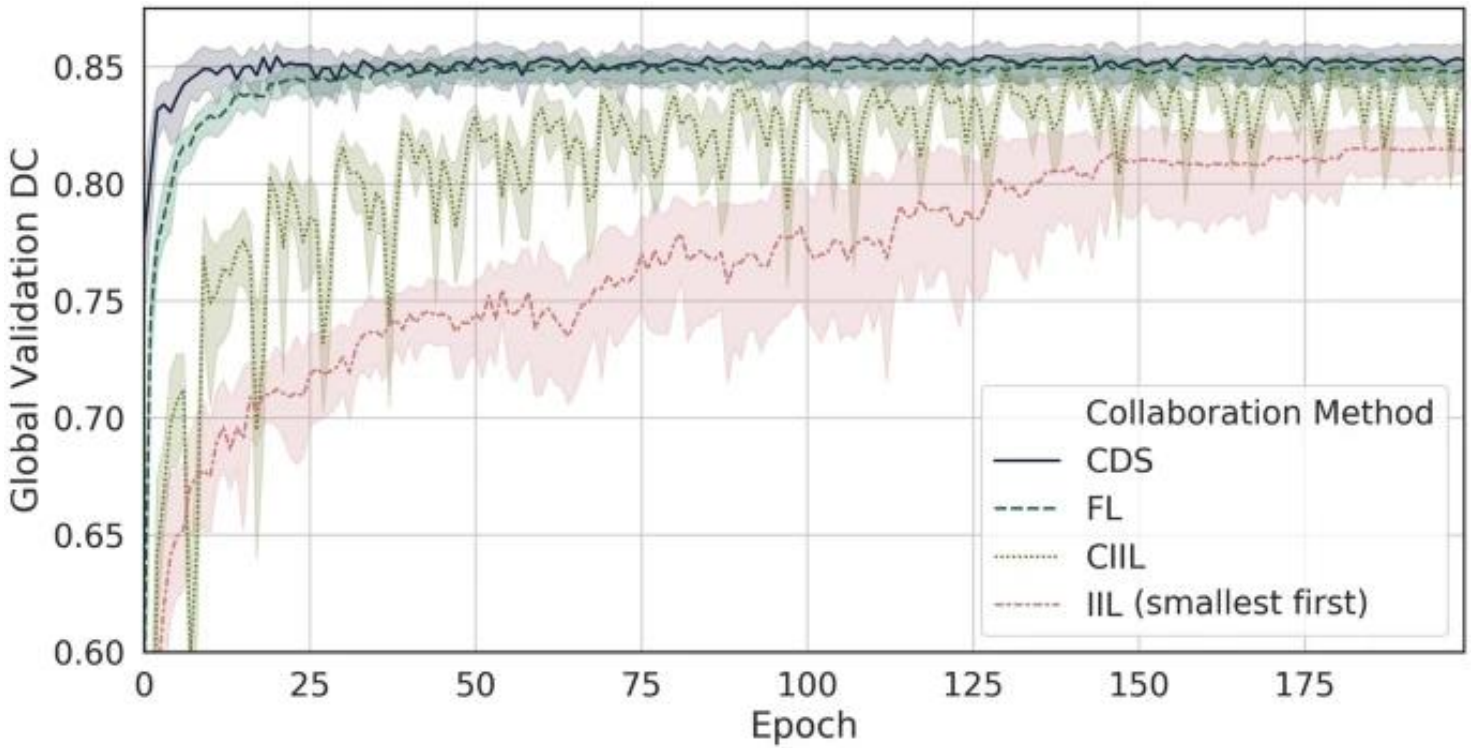
Centralized Learning



Federated Learning



Centralized Learning versus Federated Learning



scientific reports

Explore our content ⌄ Journal information ⌄

nature › scientific reports › articles › article

Article | Open Access | Published: 28 July 2020

Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data

Micah J. Sheller, Brandon Edwards, G. Anthony Reina, Jason Martin, Sarthak Pati, Aikaterini Kotrotsou, Mikhail Milchenko, Weilin Xu, Daniel Marcus, Rivka R. Colen & Spyridon Bakas ✉

Scientific Reports **10**, Article number: 12598 (2020) | Cite this article

3140 Accesses | **119** Altmetric | Metrics

Abstract

Several studies underscore the potential of deep learning in identifying complex patterns, leading to diagnostic and prognostic biomarkers. Identifying sufficiently large and diverse datasets, required for training, is a significant challenge in medicine and can rarely be found in

SCIENTIFIC REPORTS

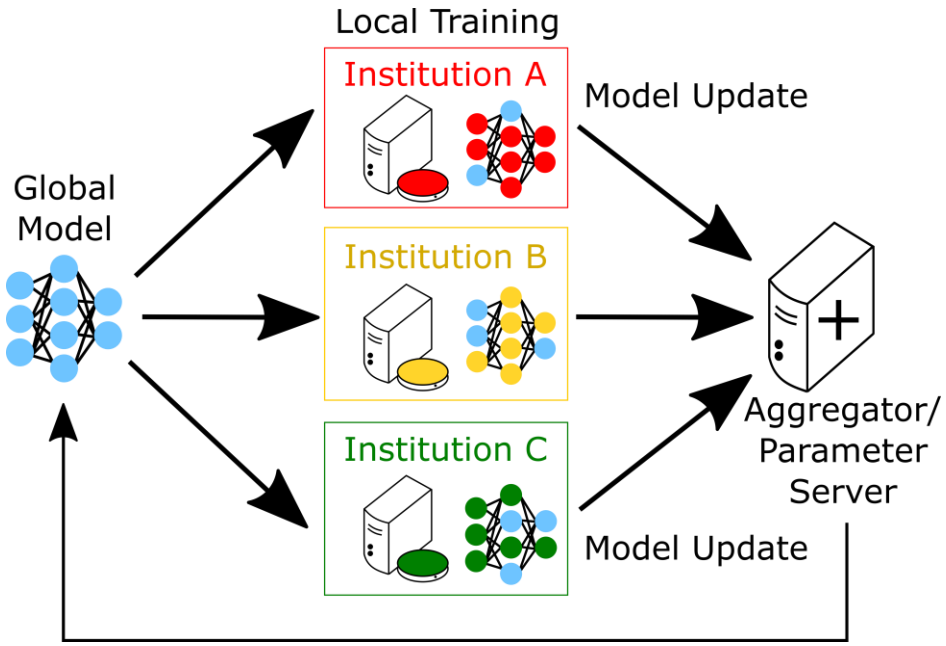
intel

Perelman School of Medicine University of Pennsylvania

nature.com/articles/s41598-020-69250-1

OpenFL: Federated Learning by Intel®

- OPEN-SOURCE SOFTWARE FOR A COMPLETE FEDERATED LEARNING SYSTEM ARCHITECTURE
- PRIVACY PRESERVED MACHINE LEARNING FOR DATA/MODEL IN TRANSIT, USE AND STORE
- OPEN FL IS EASY TO USE AND SCALABLE AND MANAGEABLE FOR LARGE FEDERATIONS



OpenFL Solves the data silo problem with software that accelerates time to market deployment of Federated Learning. It provides the greatest access to data through enabling secure, privacy preserved data.

OpenFL: how to get started

OpenFL is distributed through GitHub, PyPI and Docker Hub

github.com/intel/openfl

pypi.org/project/openfl

pip install openfl

hub.docker.com/r/intel/openfl

docker pull intel/openfl

OpenFL supports all the popular machine learning frameworks

OpenFL Keras Tutorial

OpenFL PyTorch Tutorial

OpenFL documentation

Our partners: Who is using OpenFL?

- University of Pennsylvania created the first real-life and largest federation of healthcare institutions.

- Federated Tumor Segmentation Challenge 2021 is the first federated learning competition. It focused on the task of brain tumor segmentation.

- Frontier Development Lab: NASA, Mayo Clinic and Intel used federated learning to understand the effect of cosmic radiation on humans

Federated Tumor Segmentation Initiative

FeTS is the largest international federation of healthcare institutions (~80 collaborating sites) from all over the world (~ 20 countries)

med.upenn.edu/cbica/fets/

Federated Tumor Segmentation Challenge 2021

- Brain tumor segmentation task fets-ai.github.io/Challenge/
- Information gathered from FeTS initiative (> 50 medical institutions), data used was representative for real-world use case and split into ~20 partitions
- The goal was to create effective weight aggregation methods for the creation of a consensus model

NASA & OpenFL use case

What's New: This summer, Frontier Development Lab (FDL) researchers conducted a landmark astronaut health study with Intel AI Mentors to better understand the physiological effects of radiation exposure on astronauts. Using Intel artificial intelligence (AI) technology, FDL created a first-of-its-kind algorithm to identify the biomarkers of cancer progression using a combination of mouse and human radiation exposure data.

This research leveraged Intel's Open Federated Learning (OpenFL) framework to make it possible to train and combine CRISP 2.0 models from institutions such as NASA and Mayo Clinic without moving the data to a central place.

This was crucial because, even though each organization had the necessary right to use the data, the data was private and the cost of transmitting data that could be generated aboard a spacecraft was high.

www.intel.com/content/www/us/en/newsroom/news/intel-ai-mentors-seek-improve-astronaut-health.html

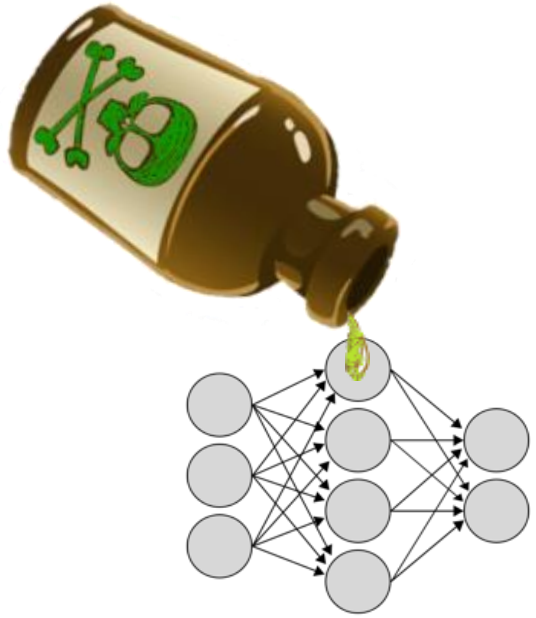
Key takeaways from Federated Learning

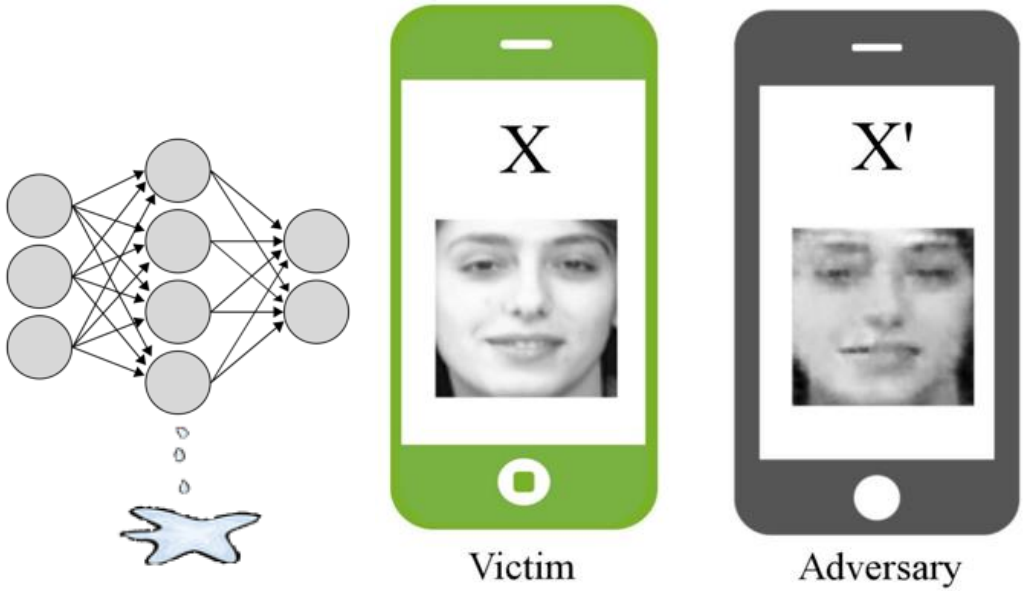
- FL solves a lot of data access problems.

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

What Does it take to Secure FL?

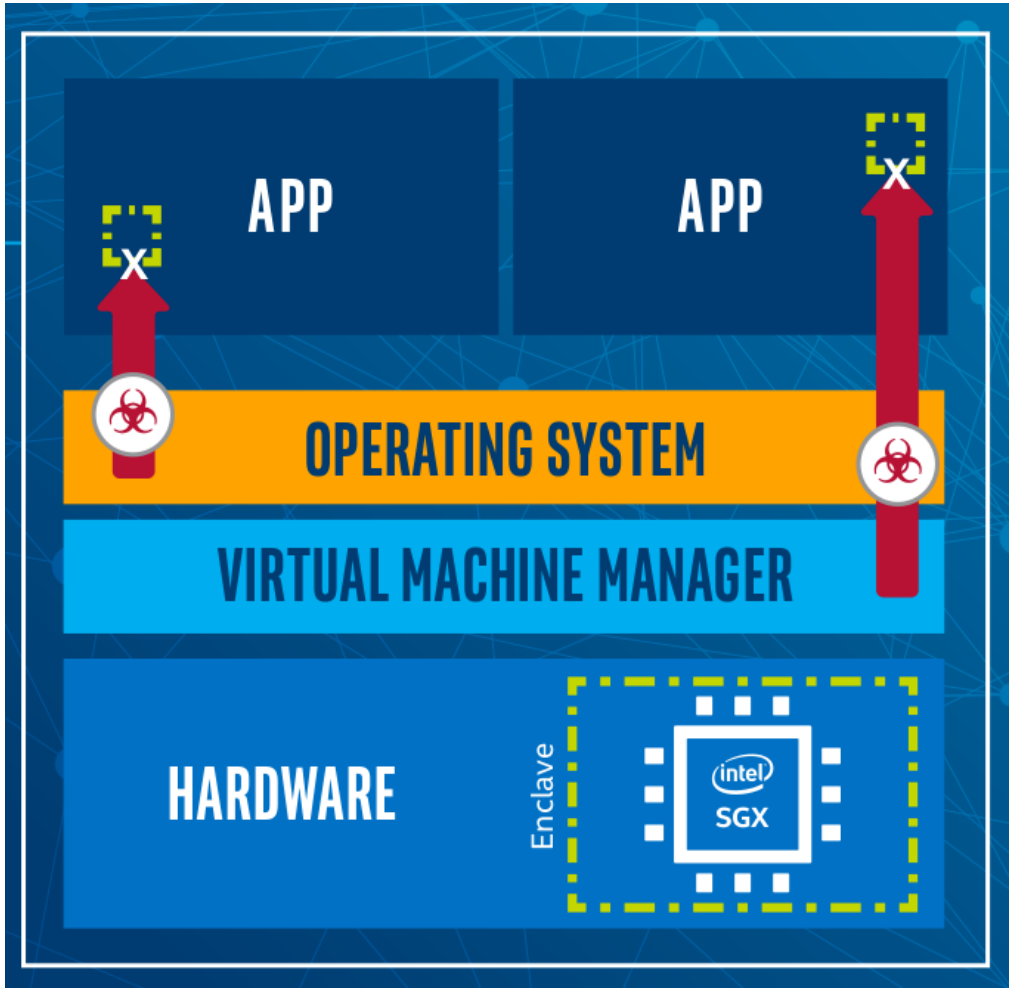
- **Confidentiality** – Protecting data and models from the risk of exposure to untrusted parties during runtime
- **Execution Integrity** – Protecting the computation (i.e. training the model) from being changed at runtime
- **Attestation** – Validating that the software and hardware are genuine

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.

https://www.intel.com/content/www/us/en/developer/tools/software-guard-extensions/overview.html

Security During Execution in Hardware



Why Federated Learning with Intel® SGX*

* SGX opensource integration with OpenFL will be added in next releases

CONFIDENTIALITY

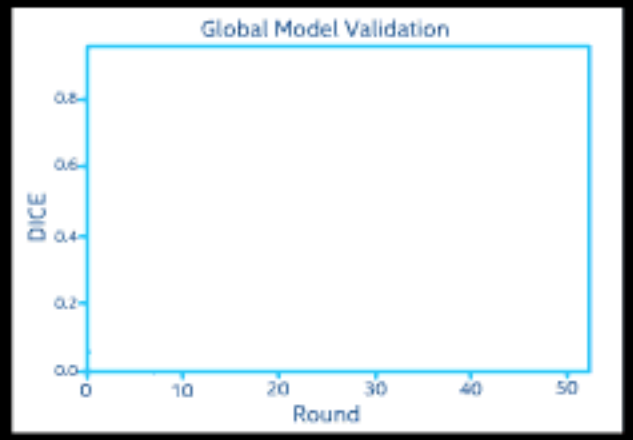
- Data never leave the premise of data owners.
- Model IP protected end-to-end in use and at rest.

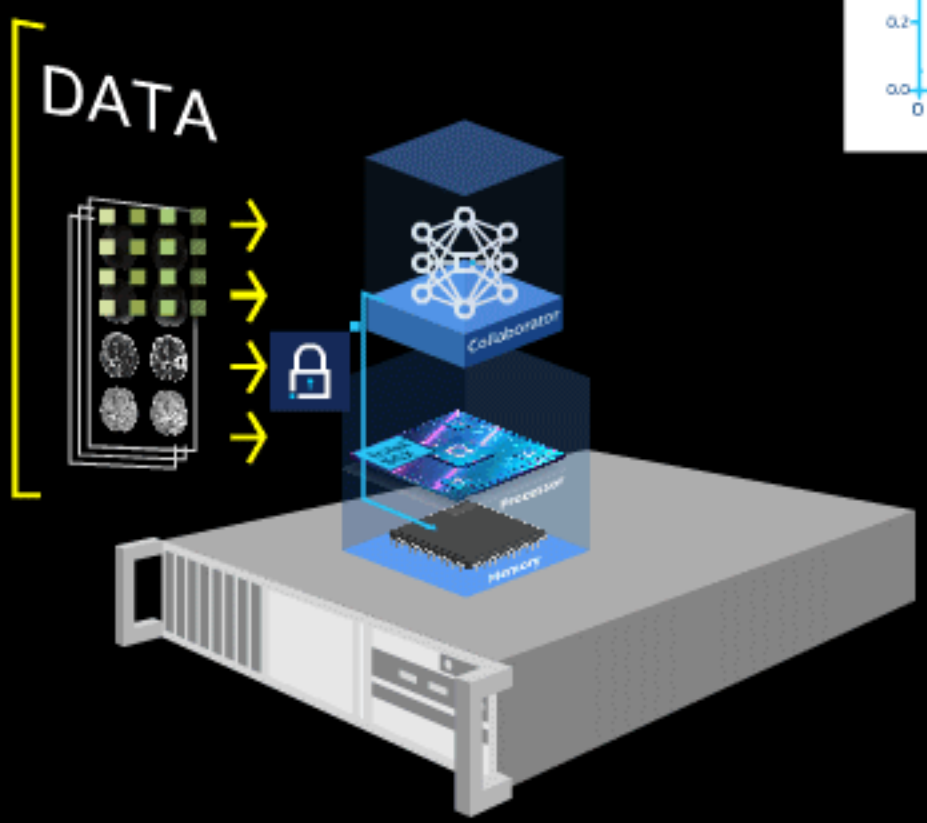
INTEGRITY & ATTESTATION

- Only verified and approved ML models.
- Participants can not insert unapproved code at any time.

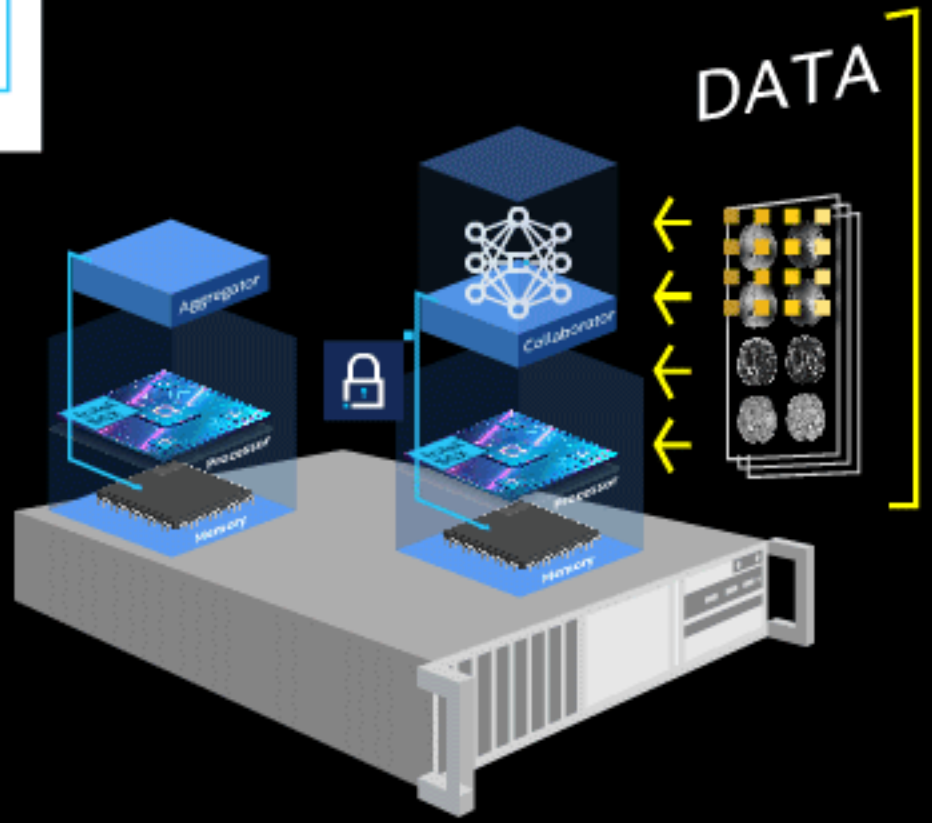
Provides a mechanism to prevent stealing the model or reverse-engineering data distribution.

Federated Learning based on Intel® SGX





3rd Gen Intel® Xeon® Scalable Processors



3rd Gen Intel® Xeon® Scalable Processors

Links & Materials

OpenFL opensource software: github.com/intel/openfl

OpenFL tutorials: github.com/intel/openfl/tree/master/openfl-tutorials

OpenFL PyPI: pypi.org/project/openfl/

OpenFL Docker Hub: hub.docker.com/r/intel/openfl

OpenFL white paper: arxiv.org/abs/2105.06413

Federated Learning on SGX – public demo 2020: player.vimeo.com/video/485722936#t=33m07s

virtual-demo.intel.com/virtual-demo/wp-content/themes/virtual-demo/demo/FL-SGX/

Federated Learning in Medicine – Nature 2020: www.nature.com/articles/s41598-020-69250-1

The Future with Digital Health – Nature, 2020: www.nature.com/articles/s41746-020-00323-1

The Federated Tumor Segmentation (FeTS) challenge 2021: www.med.upenn.edu/cbica/fets/miccai2021/

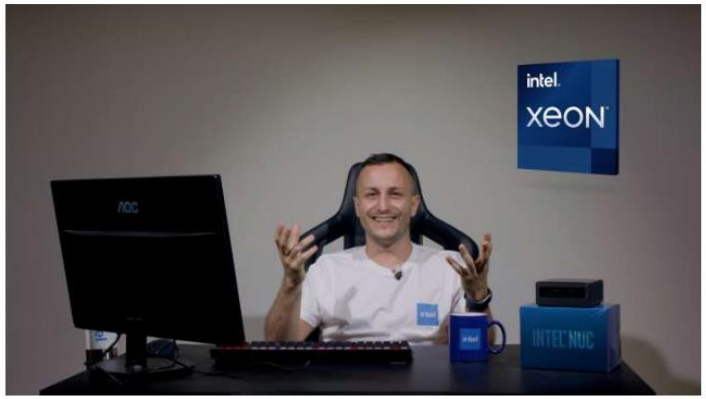
Federated Tumor Segmentation Initiative: www.med.upenn.edu/cbica/fets/

NASA use case with OpenFL 2021: www.intel.com/content/www/us/en/newsroom/news/intel-ai-mentors-seek-improve-astronaut-health.html

Links & Materials

Federated learning

Watch the Next Episode

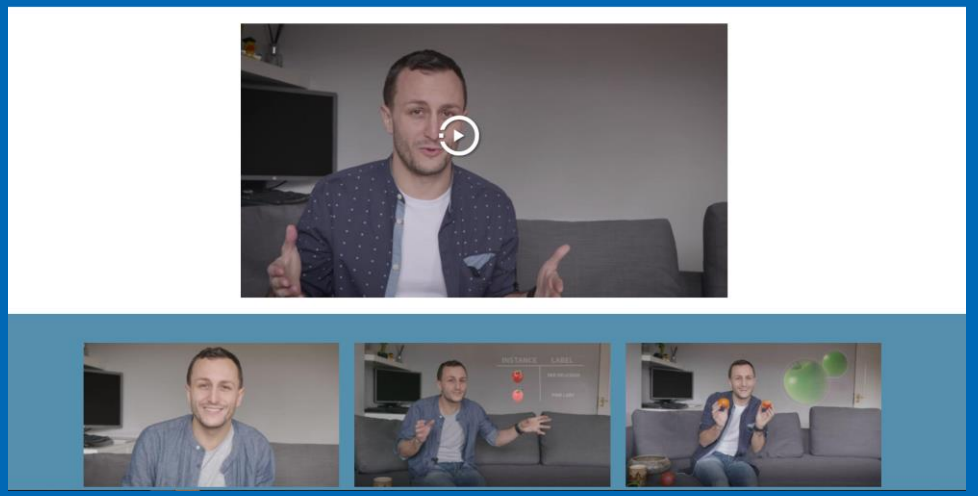


Building the Federation

Discover how to build the hardware and connect the infrastructure.

Watch Now →

Intro to AI



https://www.intel.co.uk/content/www/uk/en/now/ai-video-series/machine-learning-part-one.html

github.com/intel/openfl

bit.ly/2MKAyAv