# Federated Learning

## Concepts, system design and responsible adoption

A complete guide to collaborative AI

# The learning journey

## 01 Foundations and applications

What FL does, how training works and where it is used.

## 02 Design and deployment

Data heterogeneity, readiness, configuration and evaluation.

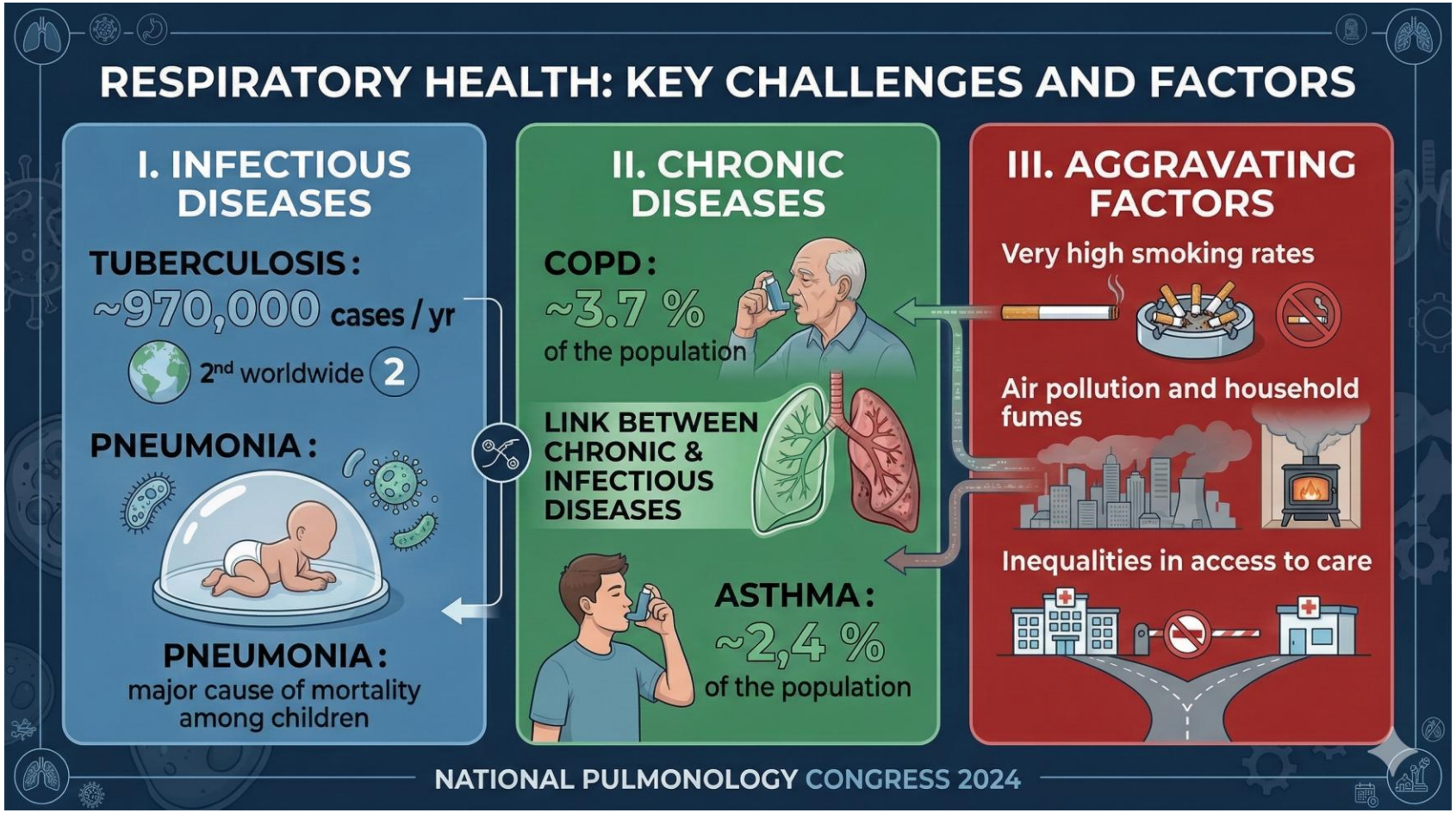
## 03 Protection and operation

Threats, safeguards, engineering costs and governance.

## 04 Federated analytics and resources

Distributed statistics, practical tools and next steps.

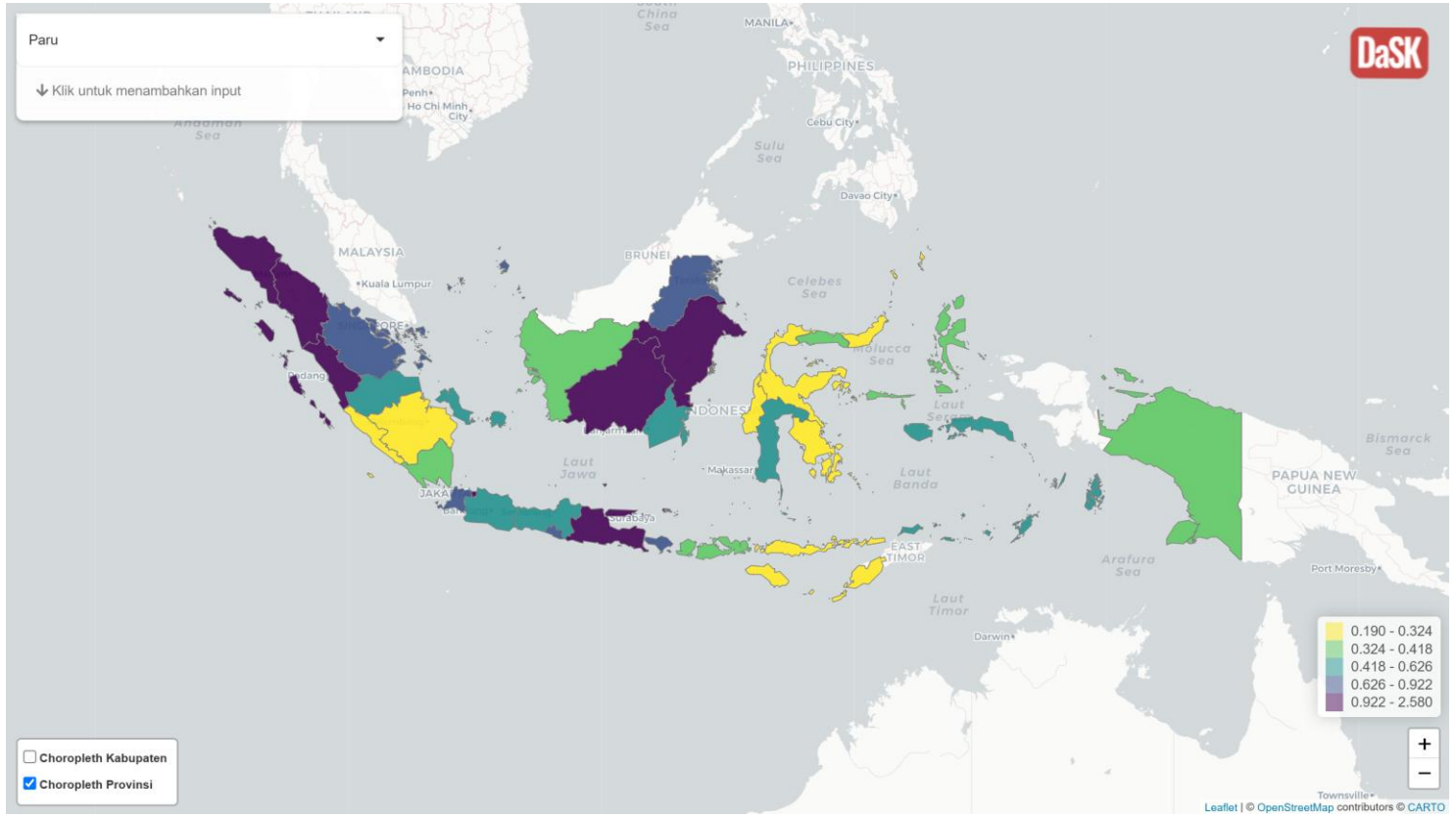
# Lung Diseases



# Healthcare in Indonesia

Indonesia’s healthcare system faces several structural challenges:

1. a population spread across many islands
2. an uneven distribution of medical specialists
3. hospital infrastructure that varies across regions
4. Gouvernement Regulation




In some remote areas, access to specialists may be limited.
AI systems can help **reduce these inequalities in access to diagnosis**.

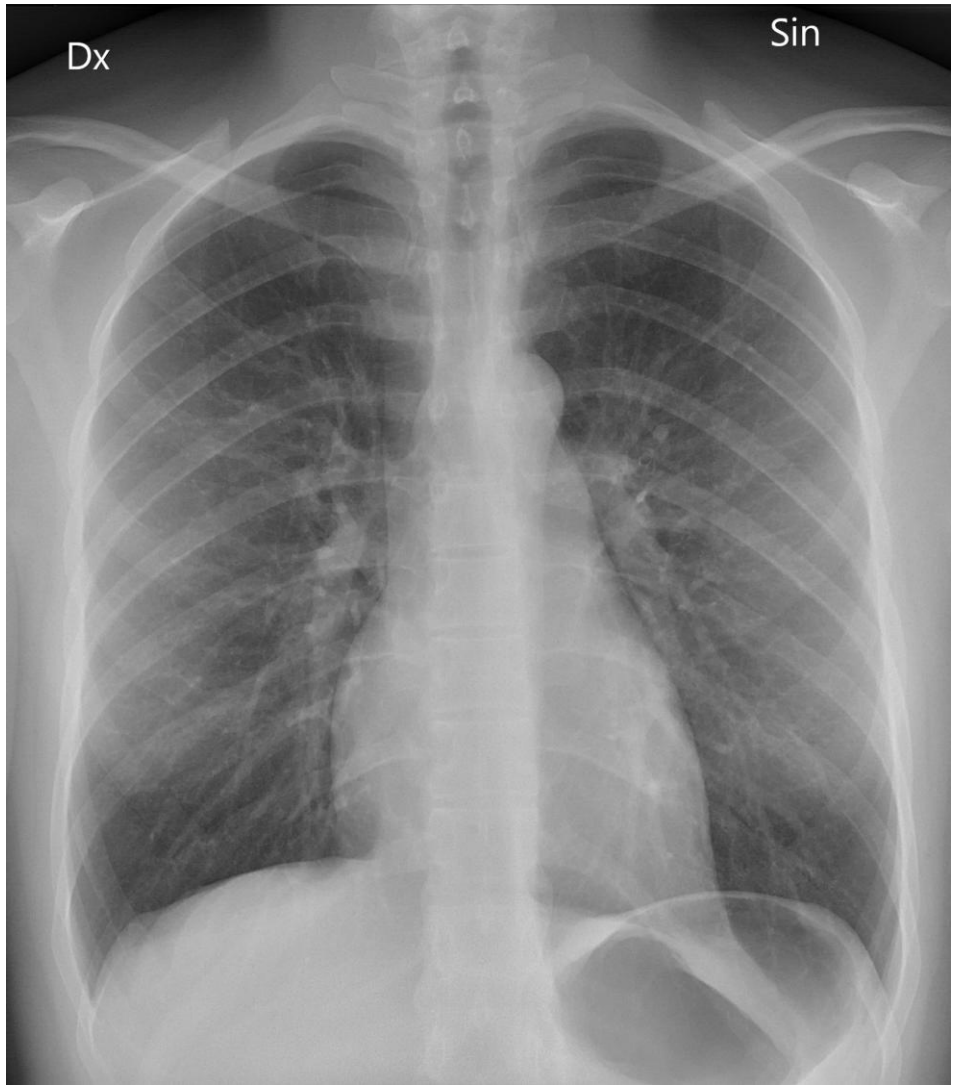
# AI in Medical Imaging

Artificial intelligence plays an increasingly important role in medical image analysis. Deep learning methods can help clinicians:

- detect abnormalities in medical images
- improve diagnostic consistency
- reduce analysis time

Among the different imaging modalities, **chest radiography** is one of the most widely used worldwide.

This makes it particularly well suited to developing AI tools for diagnostic support.

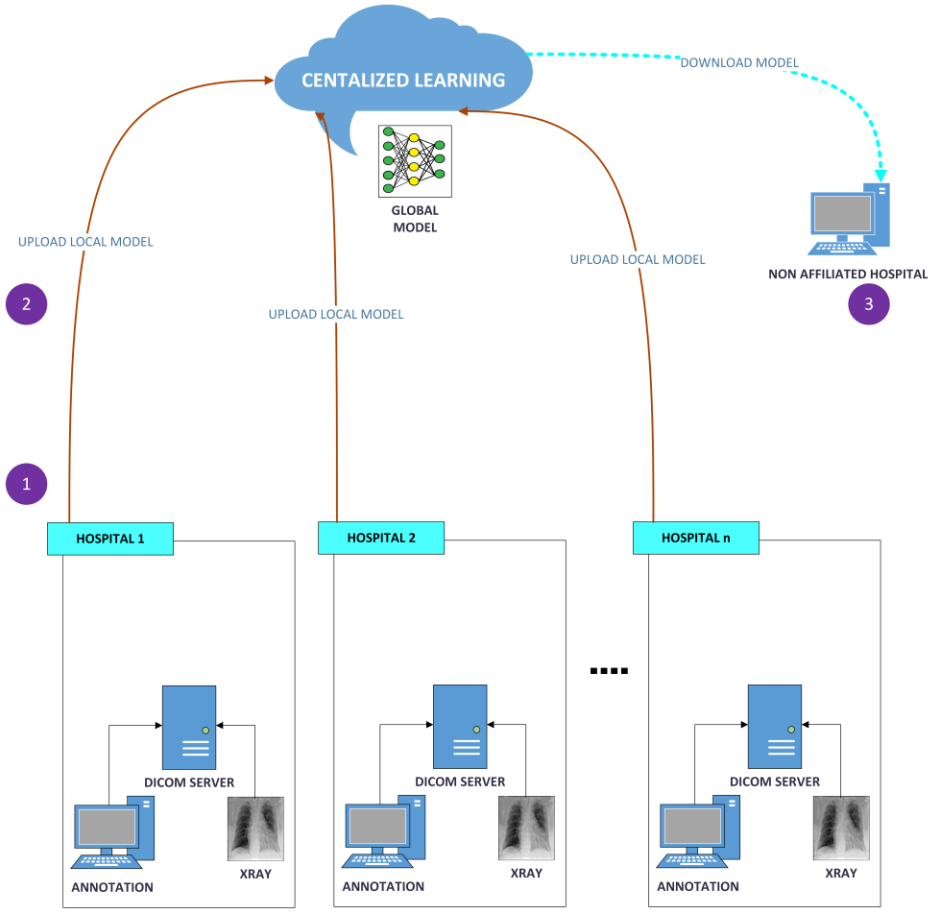


# Centralized Learning

Traditionally, AI models are trained using a **centralized approach**.
The process is straightforward:

- Data from multiple hospitals are collected and transferred to a **central repository**
- The AI model is trained using the **combined dataset**

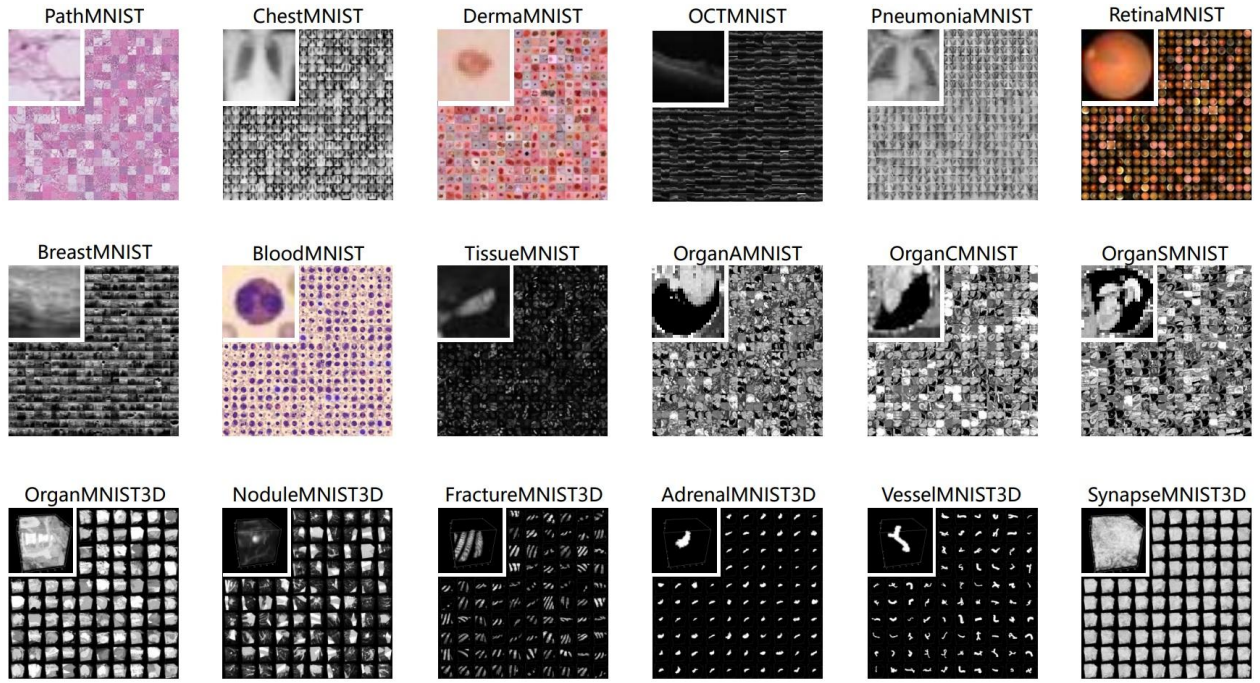
**Advantages:**
Different organisations hold **large amount and** complementary data. Collaboration may improve coverage and generalization to obtain **better model performance**



# Medical Data Challenge

AI models need **large amounts of data** to perform effectively.
Ideally, data should come from **multiple hospitals** to increase data diversity and improve model robustness.
However, sharing medical data raises several concerns:

- patient privacy
- legal and regulatory requirements
- cybersecurity risks



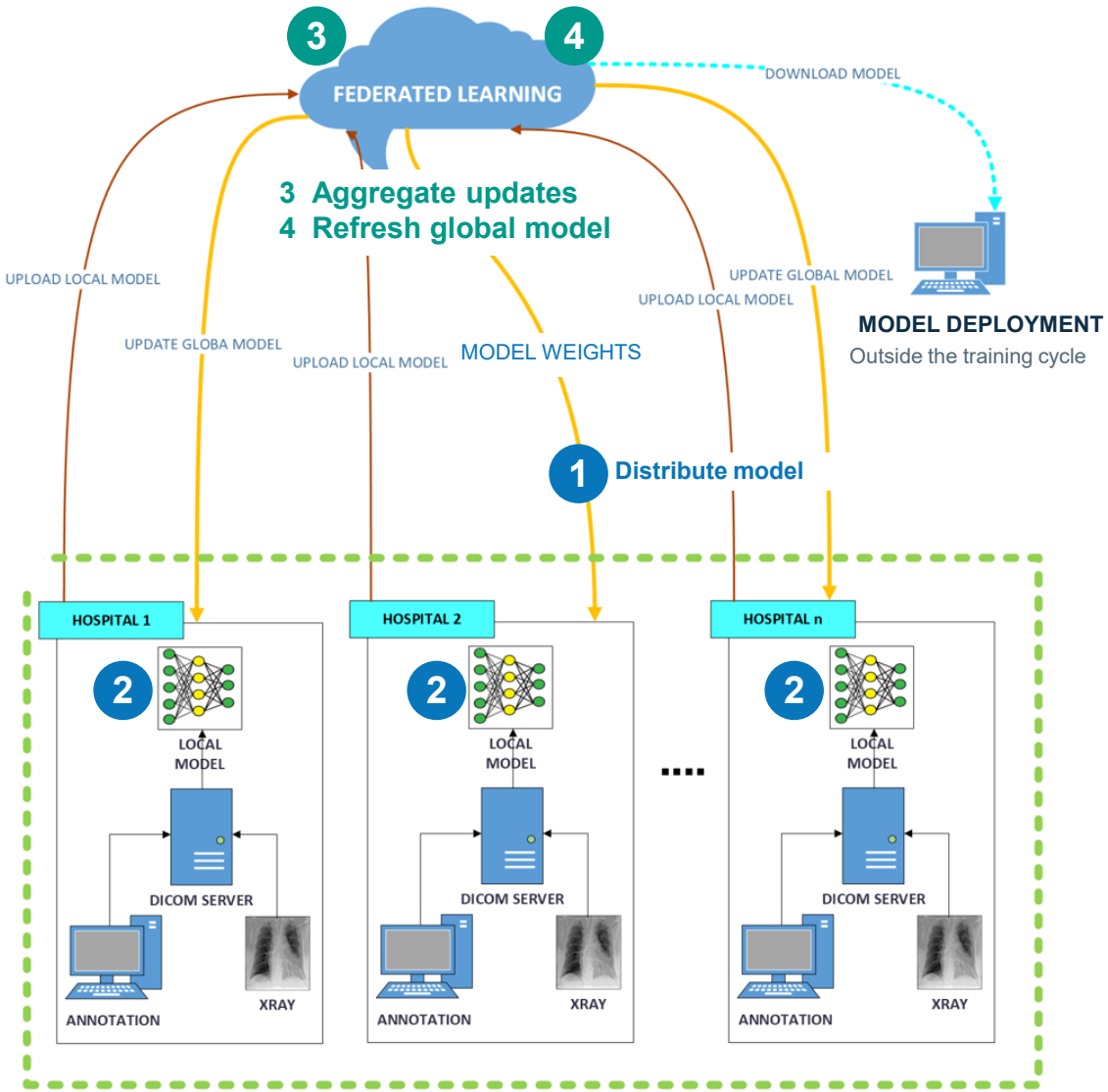
# Federated Learning

Federated learning is a machine learning setting where **multiple entities (clients) collaborate** in solving a machine learning problem, under the coordination of a central server or service provider. Each **client's raw data is stored locally** and not exchanged or transferred; instead, **focused updates intended for immediate aggregation** are used to achieve the learning objective.

Advances and Open Problems in Federated Learning (arxiv/1912.04977)

# The federated learning training cycle

Model parameters move between sites. Patient data remain local.




### 1 Select and distribute

Select participating hospitals. Send model weights and training instructions.

### 2 Train locally

Each selected participant learns from its own data.

### 3 Combine and aggregate

Collect and combine model updates at the server. Apply timeout policies.

### 4 Synchronise

Apply the aggregate and refresh the shared global model.

**Repeat steps 1–4 until the stopping criterion is met.**

**Only model parameters or updates leave each hospital; patient data stay local.**

# Why organisations consider federated learning

1) Sending the data may be too costly :
   - Self-driving cars are expected to generate several TBs of data a day,
   - Some wireless devices (IOT) have limited bandwidth/power
2) Data may be considered too sensitive to be shared
   - Growing public awareness and regulations on data privacy
   - Keeping control of data can give a competitive advantage in business.

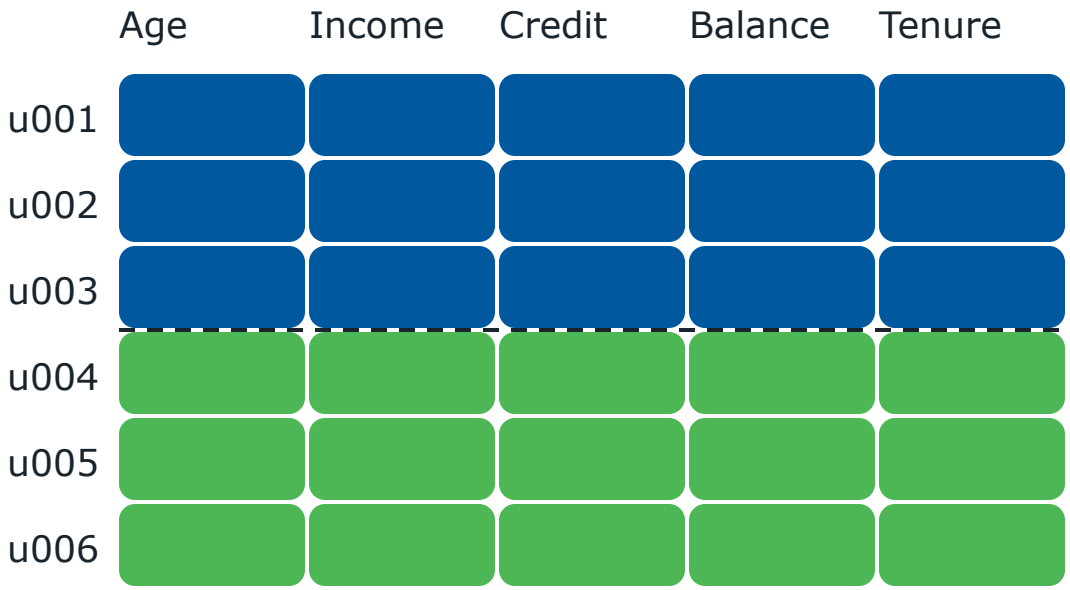
Importantly, **data always remain within the originating servers** and are never directly shared with the central server or other institutions

**FL brings training to the data and exchanges model updates.**

# Horizontal Federated Learning

## HFL

Horizontal Federated Learning




**Same features**
**Different samples**

**Real-world example**

Two regional banks train on different customers using an identical feature schema.

**Bank A**
Rows 1–3

**Bank B**
Rows 4–6

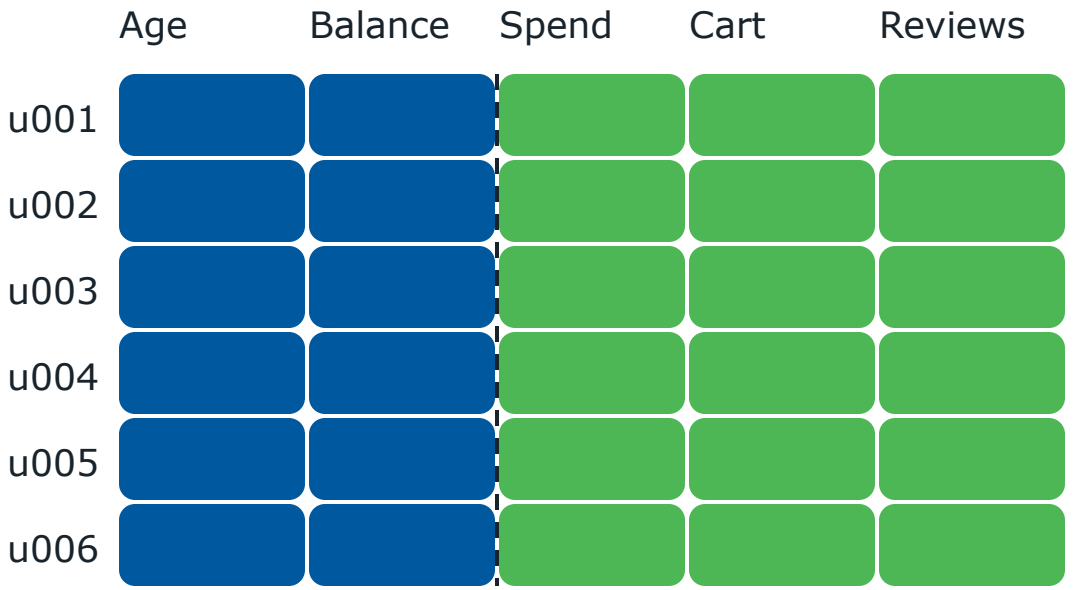
Split row-wise (by sample)

**Participants share a feature schema, but hold different customers.**

# Vertical Federated Learning

## VFL

Vertical Federated Learning




**Same samples**
**Different features**

**Real-world example**

A bank and an e-commerce company collaborate on shared customers, with different features at each party.

**Bank**
Columns 1–2

**E-commerce**
Columns 3–5

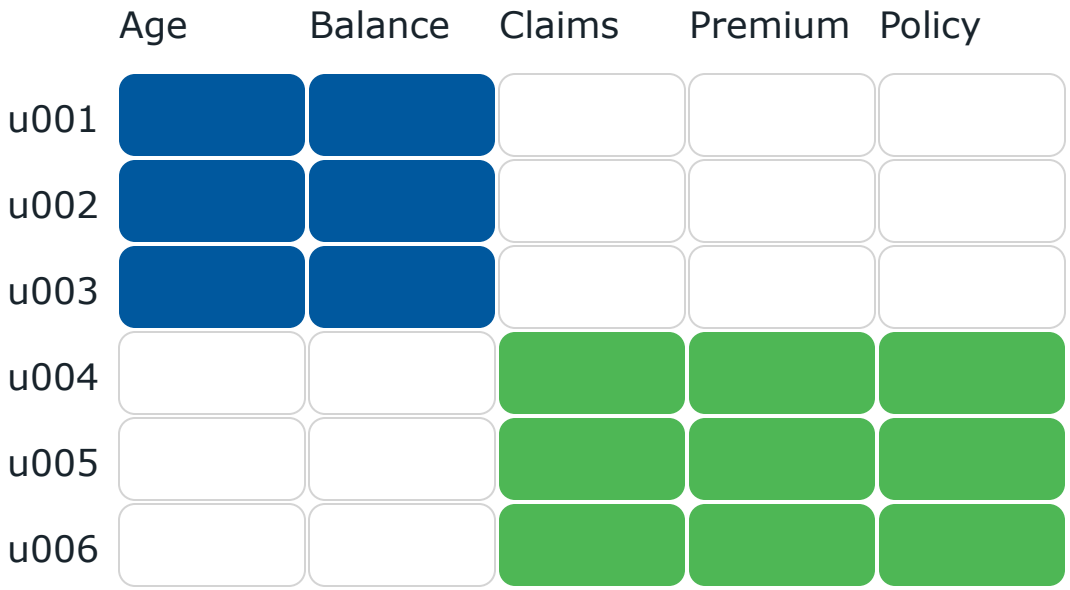
Split column-wise (by feature)

**Participants align shared customers and contribute different attributes.**

# Federated Transfer Learning

## FTL

Federated Transfer Learning




**Different samples**
**Different features**

**Real-world example**

A bank in country A transfers knowledge to an insurer in country B across different domains.

**Bank**
Rows 1–3

**Insurer**
Rows 4–6

Disjoint on rows AND columns

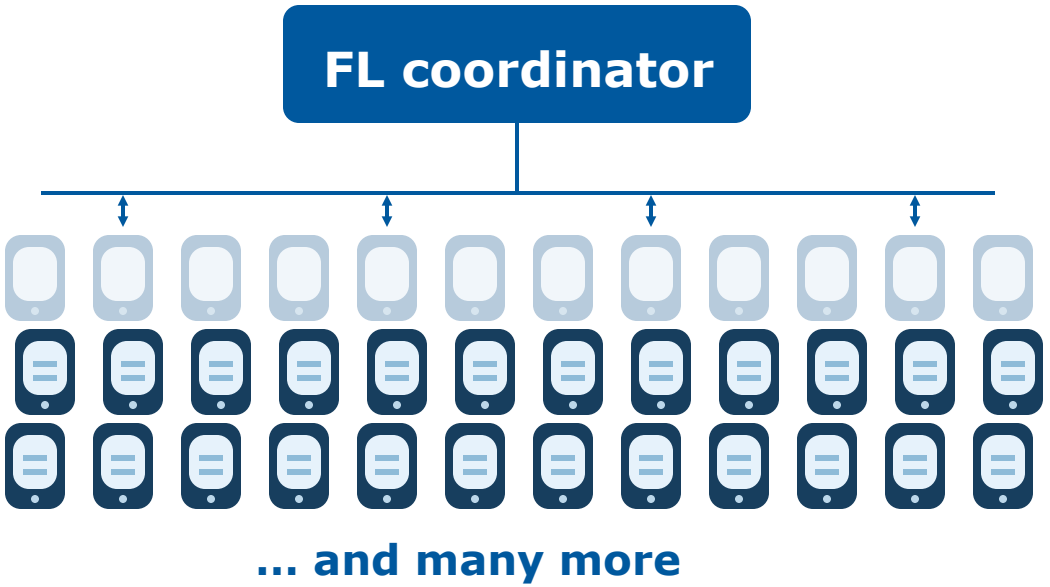
**Transfer knowledge across domains with different samples and features.**

# Architectural Approach

Same principle: local training, shared model updates

## CROSS-DEVICE

Model + instructions / local updates

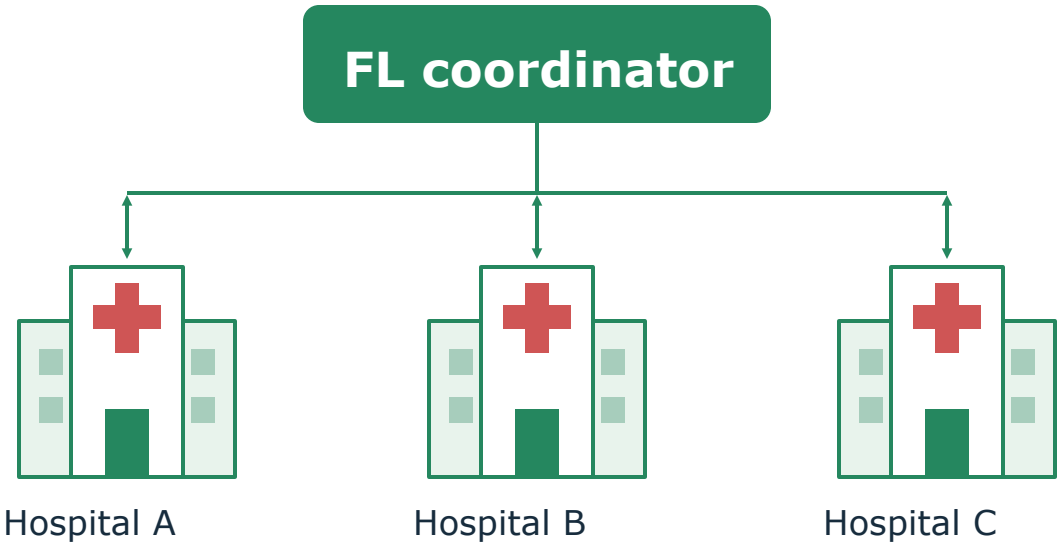



### Potentially millions of devices

A selected subset joins each round. Devices may be offline or unavailable.

## CROSS-SILO

Model + instructions / local updates

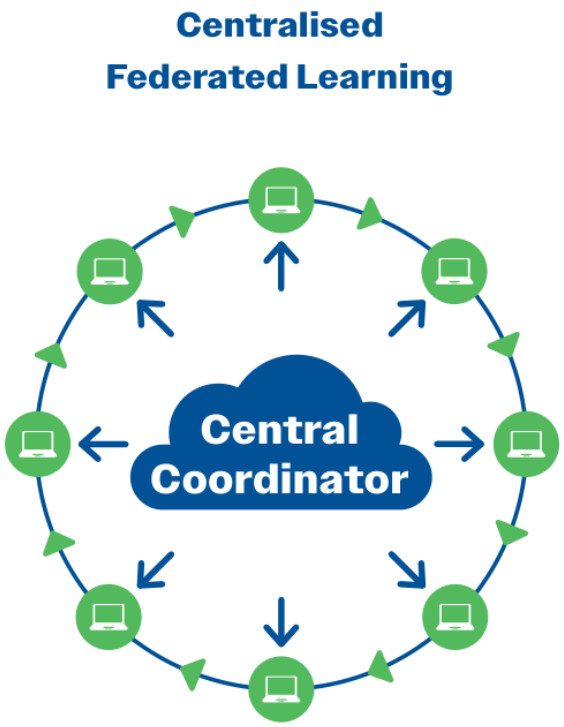



### Fewer participating organisations

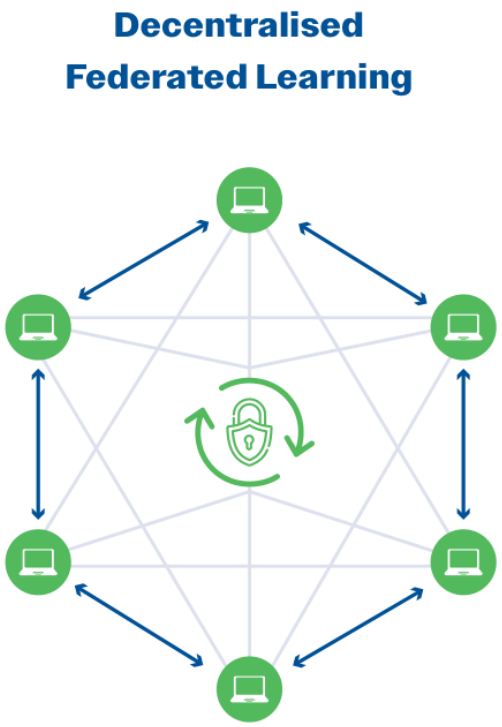
Institution-managed compute; often more stable availability.

**Raw data stays with each participant. Privacy still requires appropriate safeguards.**

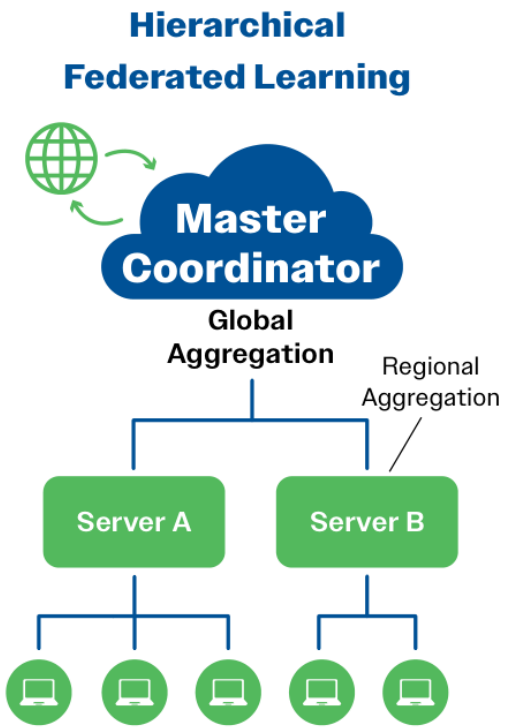
# Coordination topologies




Central control




Peer coordination




Regional aggregation

# FedAvg: a weighted model average

## Global model = weighted average of participating local models

$$w^{t+1} = \frac{1}{n}\sum_{k=1}^{K} w_k^{t+1}$$

where:

$w^{t+1}$: Updated global model parameters after round

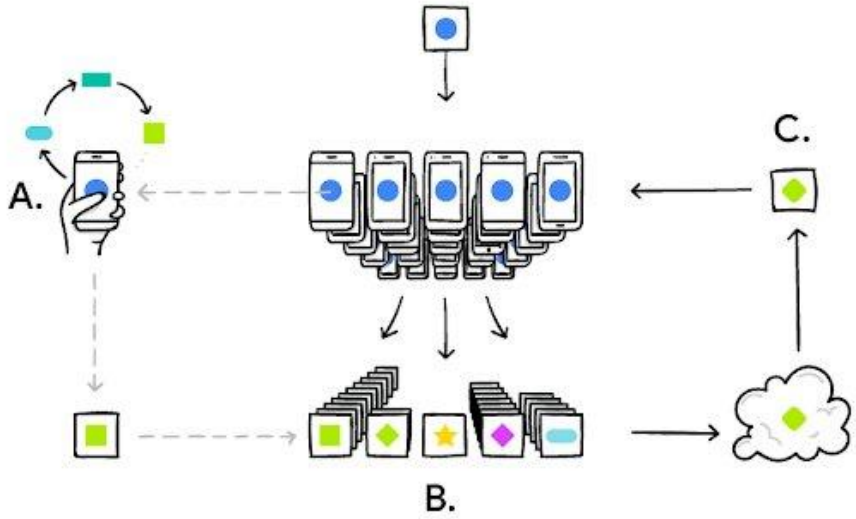
$n$: Total number of clients.

$w_k^{t+1}$: Updated model parameters from client $k$ after local training

# GBoard: learning from private text

## Problem

Keyboard prediction benefits from diverse language patterns, but keystrokes are sensitive.




## Approach

Devices train locally and secure aggregation combines their updates.

## Related capability

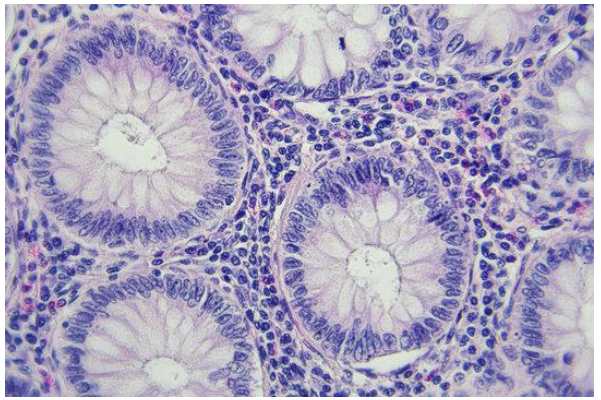
The system learns aggregated language patterns while keeping individual user text private, reducing the risk of centralised data breaches

**Learning and analytics can use different privacy-preserving workflows.**

# Dana-Farber: Oncology research across institutions

## Problem

Robust cancer classification models requires diverse datasets from multiple institutions, but privacy, security, and IP restrictions on pathology data hinder sharing and collaborative research.



## Approach

FL enables healthcare institutions to collaboratively train a global AI model by securely aggregating encrypted local model updates in Google Cloud while keeping raw patient data on-site.

## Key Benefit

FL enables collaboration across clinical centres, improving AI accuracy and generalisability for cancer diagnosis and personalised treatment while keeping patient data local.

**Confidential execution can complement data locality.**

https://www.dana-farber.org/newsroom/news-releases/2025/dana-farber-joins-leading-cancers-centers-in-launching-new-ai-learning-platform-to-accelerate-cancer-research

# Customer engagement prediction with VFL

## Fragmented features

Ant International and partner wallets hinder accurate voucher redemption predictions, as privacy and business constraints prevent combining complementary customer data.

## Privacy-preserving alignment

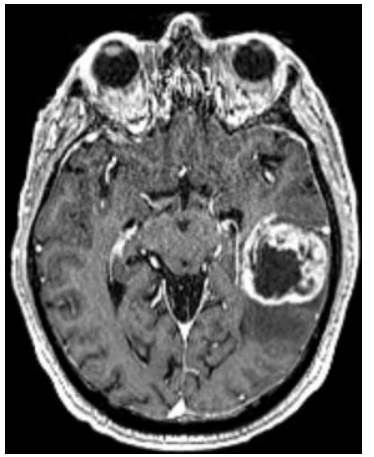
uses Private Set Intersection to match shared customers, enabling VFL with encryption to predict voucher redemptions without exposing either party's raw data.

## Collaborative prediction

The proof of concept improved customer targeting by enabling collaborative learning from fragmented datasets while preserving data privacy.

## Complementary information can support a shared prediction task.

# Global FL for Rare Brain Tumor Segmentation



## Challenge

Individual hospitals have limited glioblastoma data, while privacy, ownership and regulatory constraints prevent pooling MRI scans—limiting model generalizability.

## Solution

Train a shared model across institutions while keeping MRI scans local. Standardised preprocessing addresses scanner differences, and Intel SGX trusted execution environments protect model updates during aggregation.

## Impact

**33% improvement** in tumour core delineation, **23% improvement** in complete tumour extent delineation and better generalisation to patients from unseen institutions

**Complementary information can support a shared prediction task.**

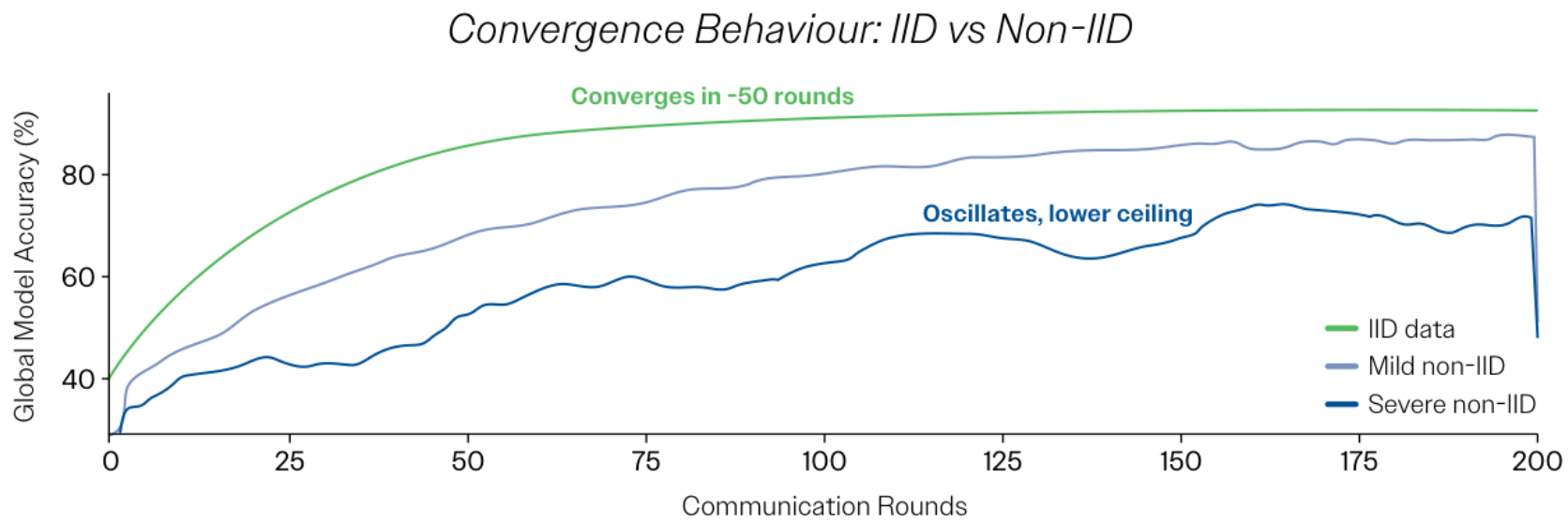
Pati, S., Baid, U., et al (2022). Federated learning enables big data for rare cancer boundary detection. Nature Communications, 13(1), 7346. https://doi.org/10.1038/s41467-022-33407-5

# What non-IID data looks like

| Difference | Meaning | Illustrative example |
|---|---|---|
| Label skew | Different class proportions | Specialist versus general hospital |
| Feature skew | Different input distributions | Different scanners or populations |
| Quantity skew | Unequal dataset sizes | Large and small institutions |
| Temporal skew / gaps | Different periods or missing classes | Recent versus older observations |

*) Independent and Identically Distributed

# Heterogeneity can destabilise convergence




**Slower convergence**
**Client drift**
**Uneven site performance**

Illustrative curves from the guide, not a measured benchmark

**Global averages can hide weak performance at individual sites.**

# Responses to heterogeneous data

## Optimisation

Use drift-aware local training or adaptive server optimization (FedProx as alternative).

## Personalisation and clustering

Adapt models locally or group participants with similar distributions.

## Data preparation

Harmonise features and consider augmentation or suitable shared reference data.

## Architecture-specific work

Improve entity alignment for VFL and assess domain similarity for transfer FL.

# When FL is a suitable choice ?

## Reasons to consider FL

Raw data must stay local. Partners have useful complementary data. Collaboration can address a real modelling gap.

## Reasons to reconsider CL

Centralisation is feasible, analytics alone is sufficient, or a simpler privacy approach meets the objective.

**A measurable collaboration benefit should justify the added complexity.**

# Configuration before training

## Architecture and trust

Choose the federation type, topology and coordinator trust assumptions.

## Model and platform

Set model scale, initial weights and framework requirements.

## Privacy mechanisms

Select safeguards that address the threat model and workload.

## Access policies

Define who can use or inspect intermediate and final models.

# Four risk outcomes to manage

## Private information leakage

Updates or released models reveal sensitive information.

## Model integrity compromise

Faulty or malicious contributions change model behaviour.

## Training disruption

Participants or coordinating services cannot complete training.

## Supply-chain compromise

Software, configurations, artefacts or logs bypass intended protections.

# Confidentiality attacks

## Membership inference

Was a particular record used in training?

## Property inference

What characteristics does a participant's dataset contain?

## Reconstruction

Can an attacker recover inputs from gradients or updates?

## Model extraction

Can repeated predictions reveal the model's functionality?

# Integrity attacks

## Data poisoning

Corrupted examples or labels influence local training.

## Model poisoning

A participant directly manipulates submitted gradients or parameters.

## Contaminated shared sources

Benign participants can unknowingly train on poisoned public data.

# What to remember

## Collaboration

FL can train across data that remains under local control.

## Evidence

Data heterogeneity, systems and privacy controls affect the result.

## Responsibility

Deployment requires shared governance and layered safeguards.

**The right starting point is a clear use case and a measurable pilot.**