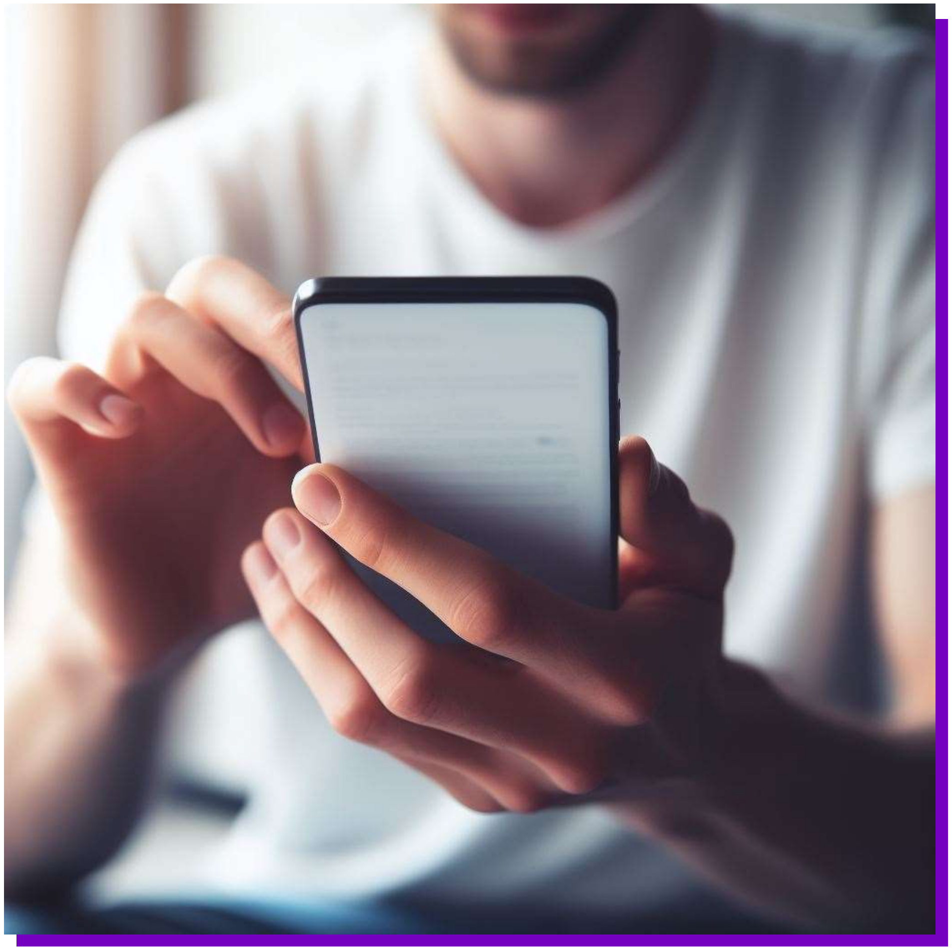
Federated Learning – its potential & paradigm shift in clinical research and development

Alexandros Giannakis

November 2023

New Science. Transformative patient outcomes.

Federated Learning has been in our daily lives since 2018



Federated Learning has been in our daily lives since 2019

Federated Learning has been in our daily lives since 2020

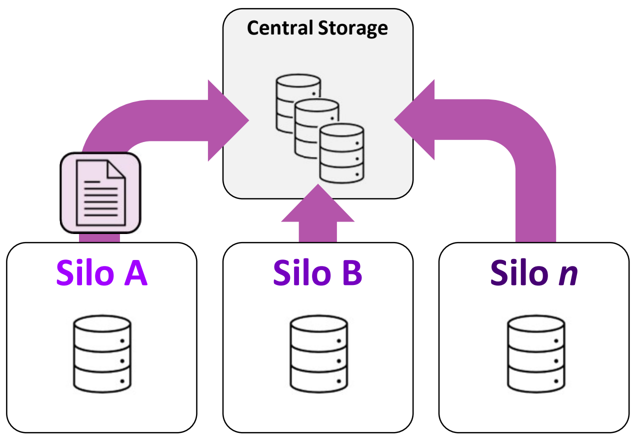
Federated Learning enables similar benefits of centralized analytics, but without sharing or transferring data

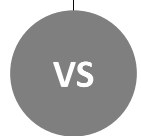
Centralized analytics

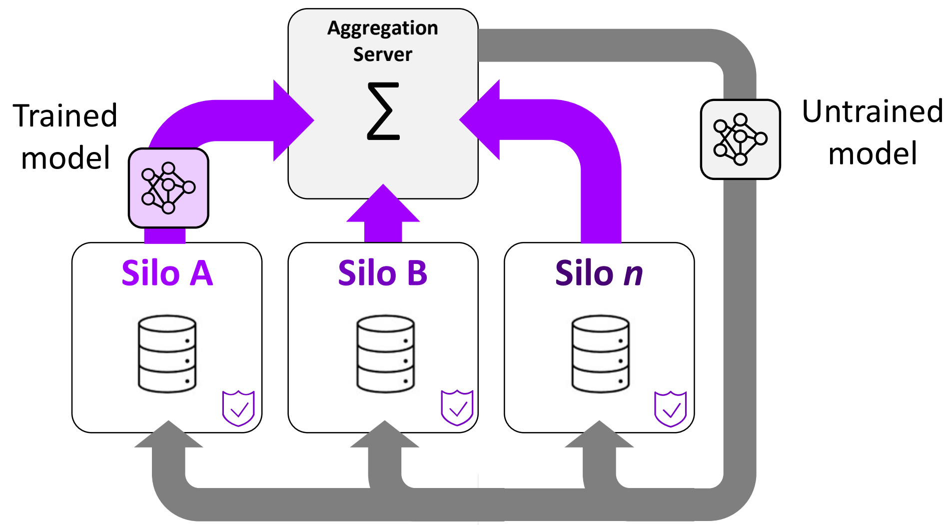
An approach where AI model is trained on a central server using data collected and stored centrally from various sources, resulting in direct access and visibility of the data.

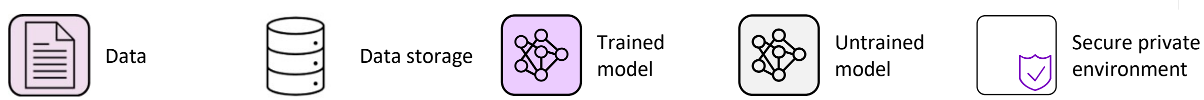
Federated Learning (FL)

A machine learning approach where a model is trained across multiple decentralized devices or servers holding local data samples, without exchanging them.

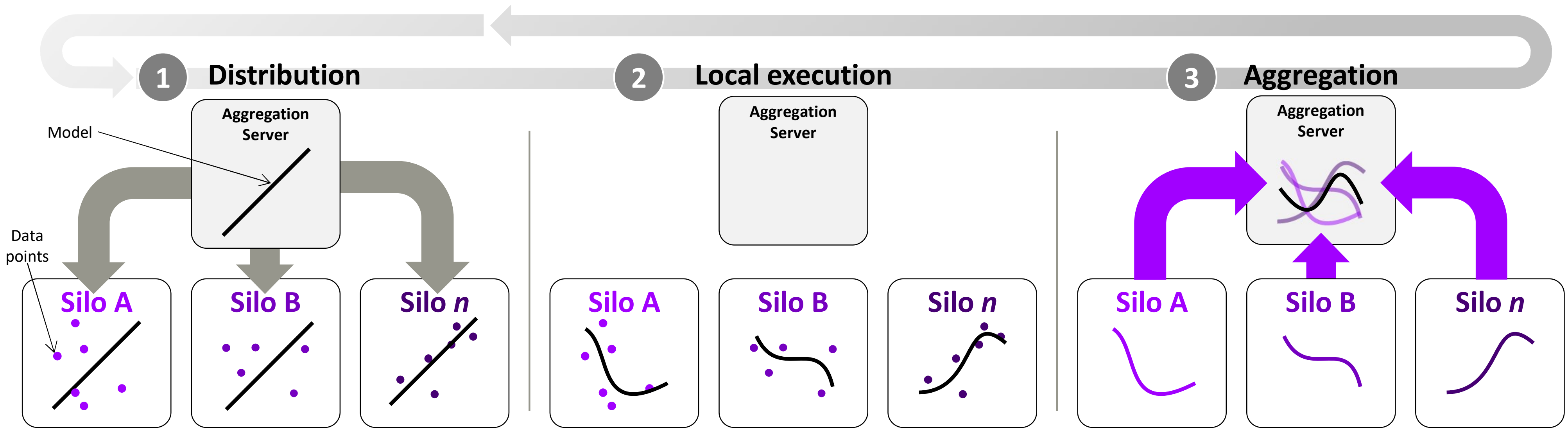








Underlying concepts of federated approach drive the adoption



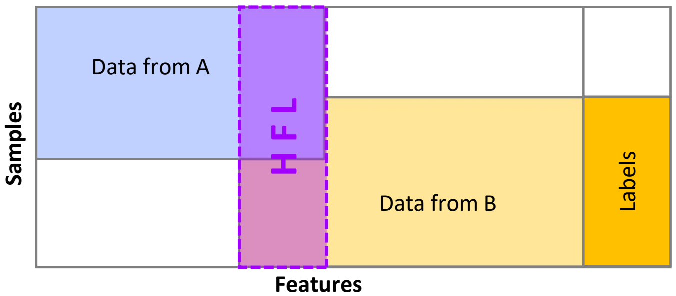
Motivations

Enabling **data utilization** while **preserving privacy**

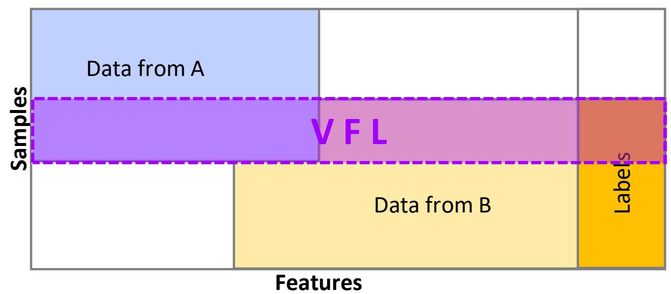
Increasing **cost efficiency** for **continuous** algorithm **improvements** and/or new insight generation

Improving **collaboration** between stakeholders **in diverse regulatory environments**

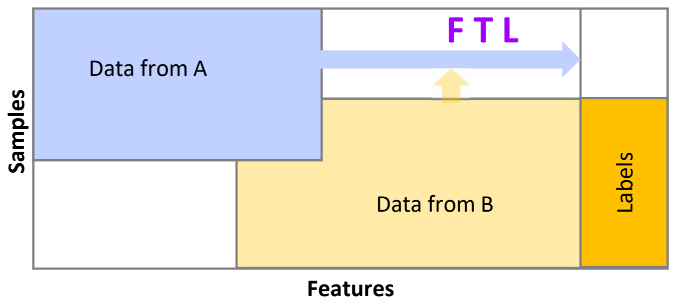
Different types of Federated Learning focuses on diverse set of applications – data strategies



Horizontal Federated Learning is utilized when different entities possess data **samples of the same features but from different individuals**. It enables model training across these datasets without centralizing the data, thus preserving privacy.



Vertical Federated Learning involves training models on datasets that share the **same data samples but have different features**. It's suitable when different entities possess different sets of features for the same individuals without directly sharing the sensitive data.

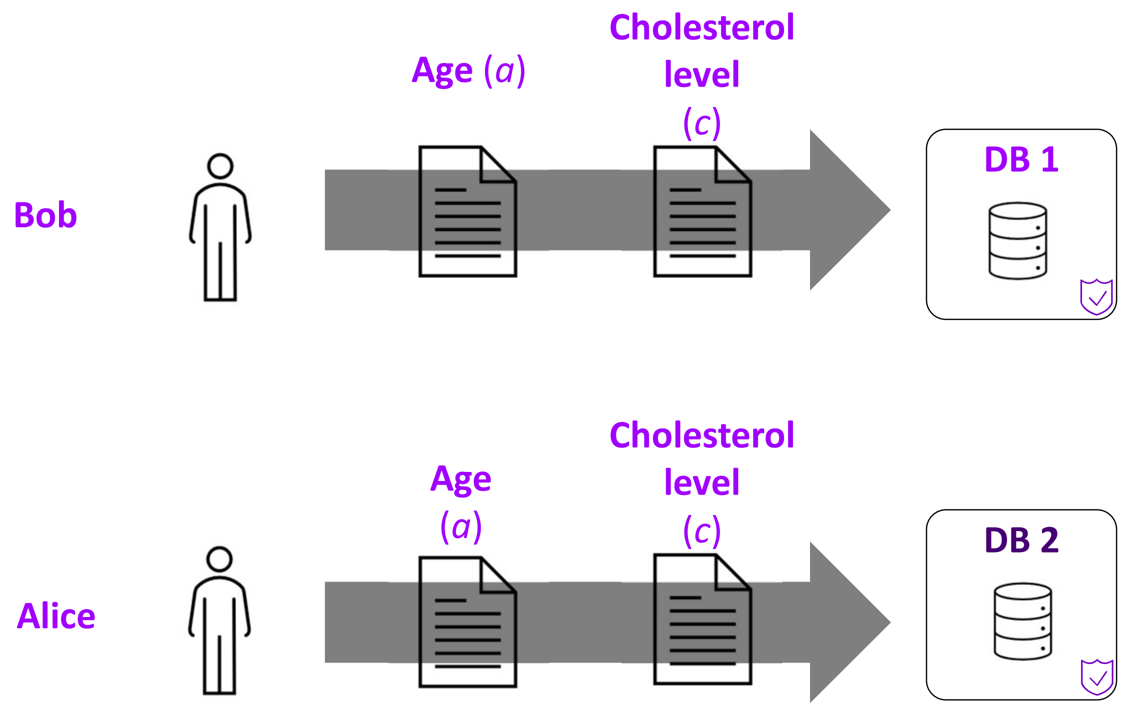


Federated Transfer Learning is the practice of **applying knowledge gained while solving one problem to a different but related problem**. It involves training a model on one task and utilizing the learned features or fine-tuning the model for a different task, thereby saving computational resources and time.

Simple example of horizontal Federated Learning usage in SaMD development

Creating an ML tool which is assessing the cardiovascular disease (CVD) risk for two patients, Bob and Alice, using their age and cholesterol level respectively, without sharing this sensitive data directly.

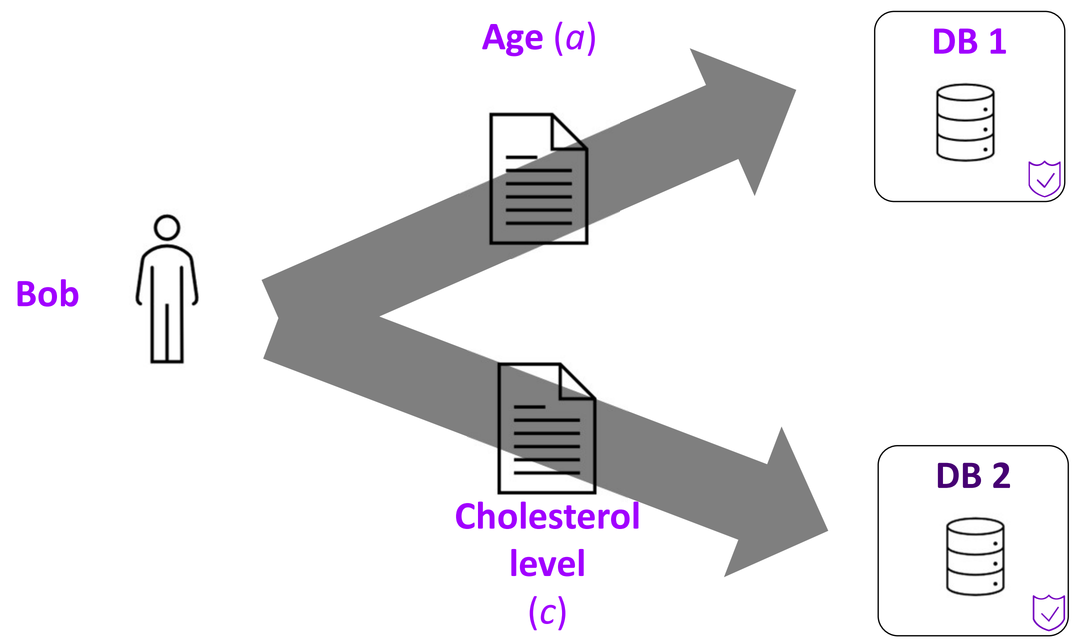
- Bob's age and cholesterol level data resides at DB 1.
- Alice’s age and cholesterol level data resides at DB 2.



Simple example of vertical Federated Learning usage in SaMD development

Creating an ML tool which is assessing the cardiovascular disease (CVD) risk for a patient, Bob, using his age and cholesterol level respectively, without sharing this sensitive data directly.

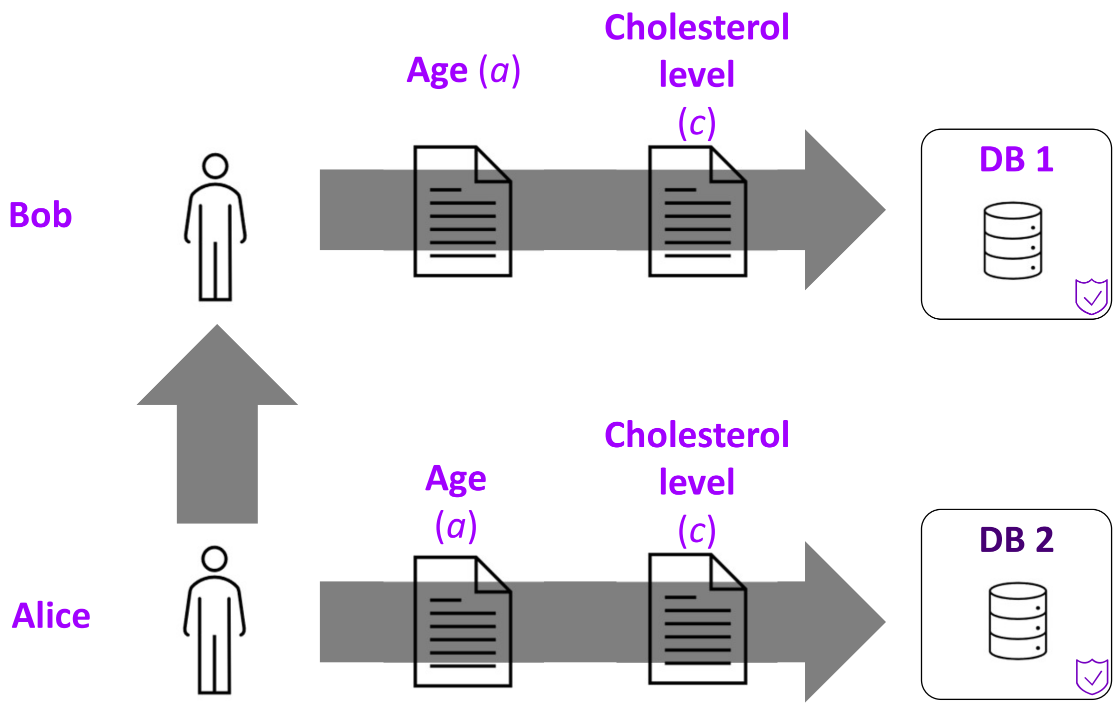
- Bob's age data resides at DB 1.
- Bob's cholesterol level data resides at DB 2.



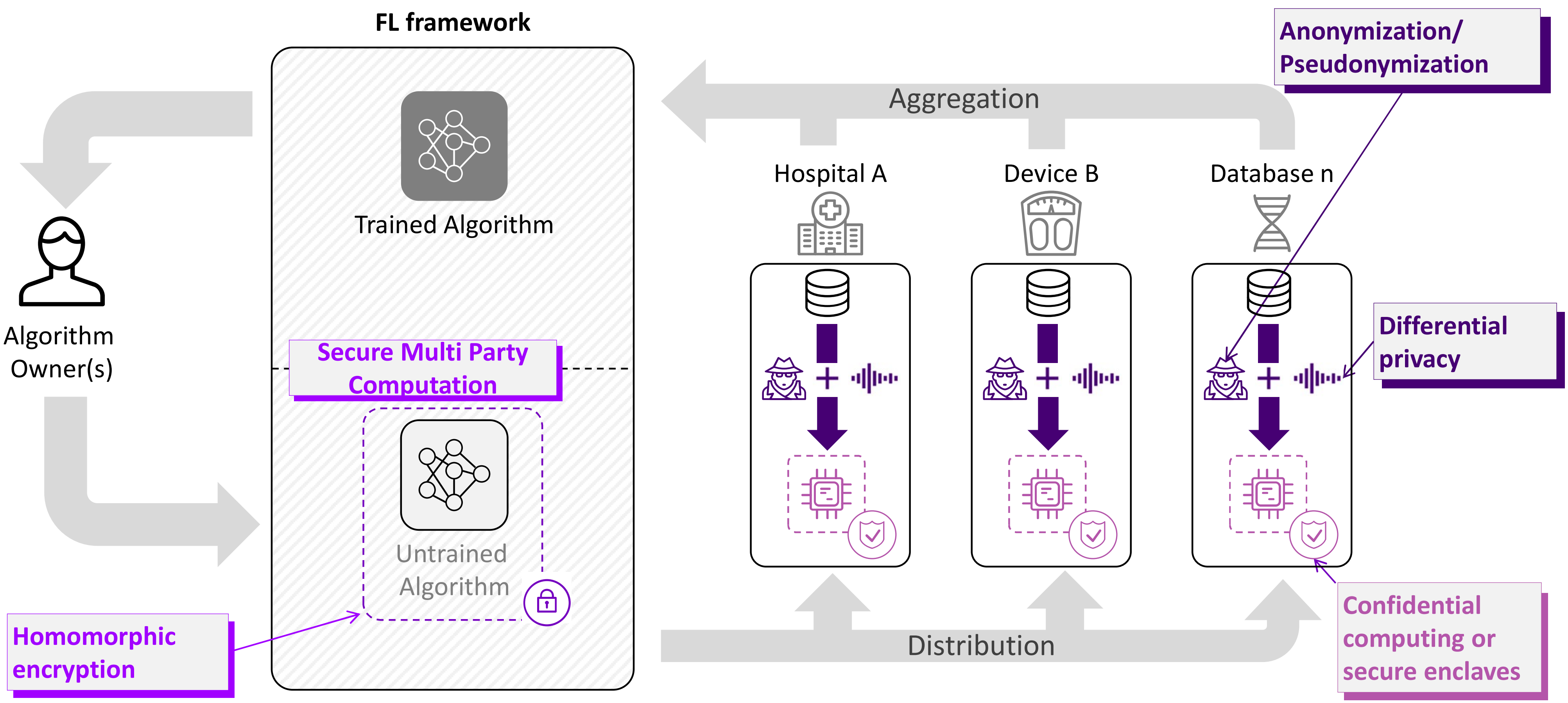
Simple example of Federated transfer Learning usage in SaMD development

Creating an ML tool which is assessing the diabetes melitus disease risk for patients, e.g. Bob, which has been developed based on an ML tool which is assessing the CVD risk of other patients, e.g. Alice, using their age and cholesterol level in both cases.

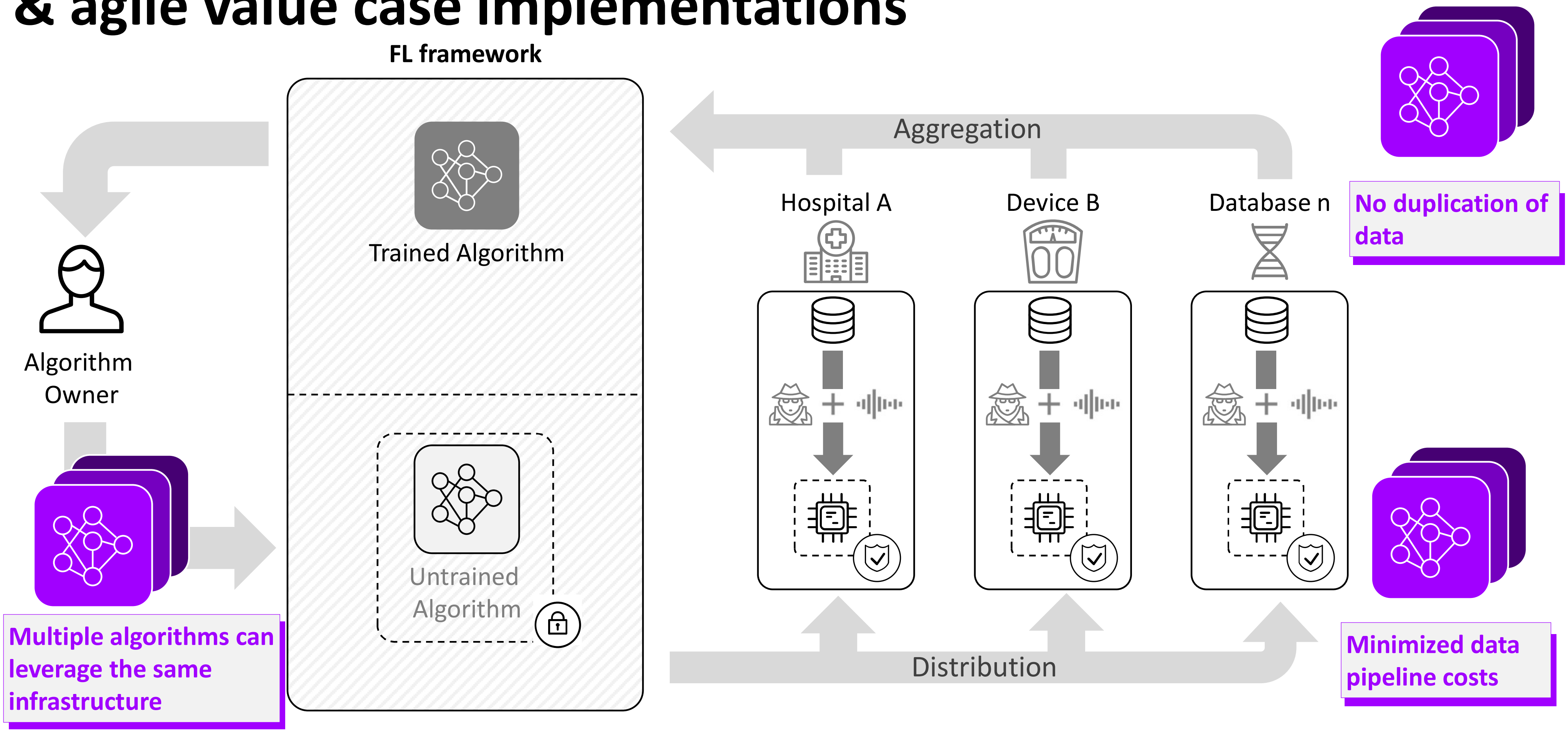
- Bob's age and cholesterol level data resides at DB 1.
- Alice’s age and cholesterol level data resides at DB 2.



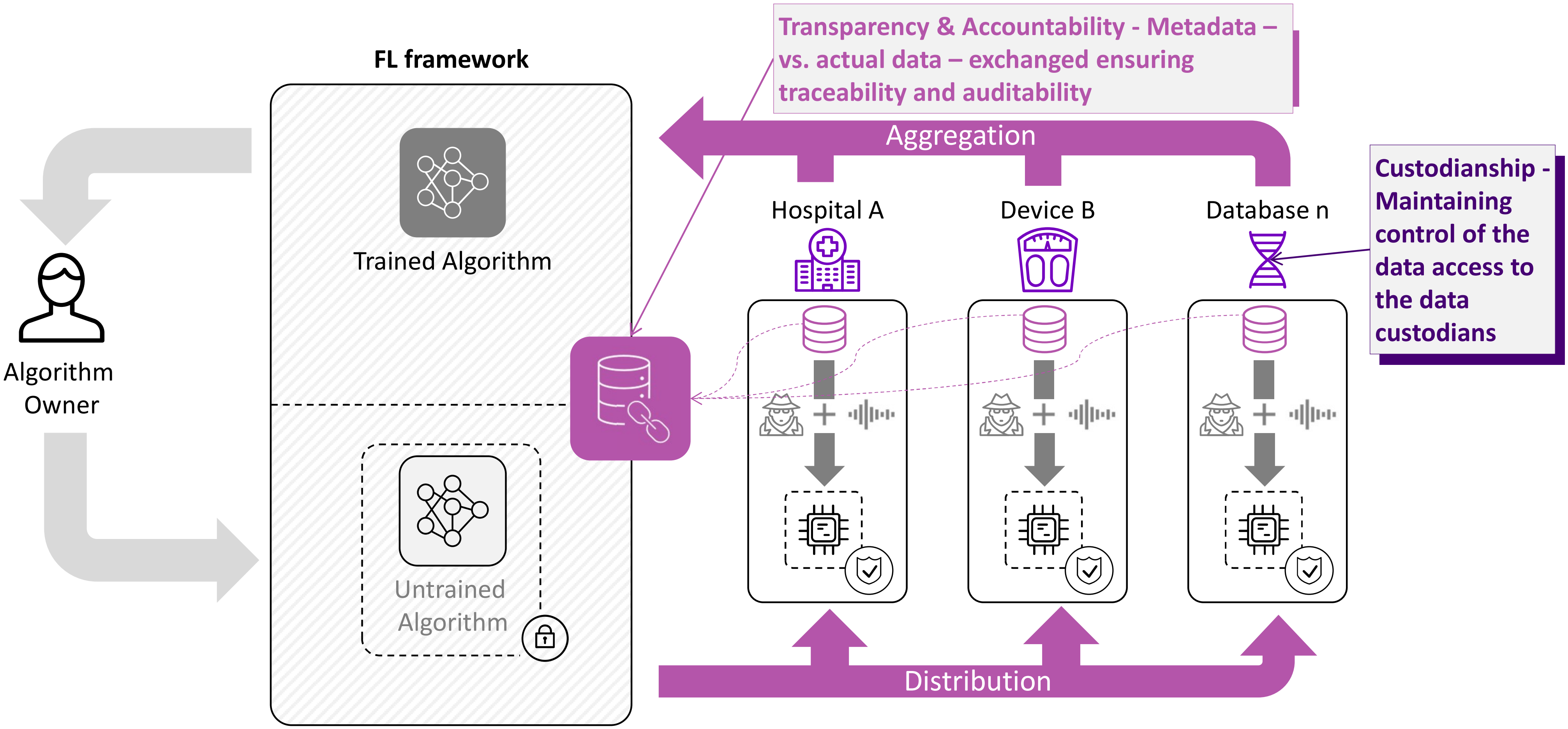
The art of possible with Federated Learning – Privacy



The art of possible with Federated Learning – Cost efficiency & agile value case implementations



The art of possible with Federated Learning – Responsible AI



Opportunities empowered by Federated Learning

Non-Exhaustive

Use cases that can benefit from FL	Imp. status
Clinical trial patient selection and recruitment	★
Adverse event prediction and surveillance	★★
Healthcare outcome research	★★★
Clinical outcome analysis & healthcare resource allocation	★★
Drug interaction analysis	★★
Genomic data analysis	★★
Personalized treatment plans	★★
Drug discovery	★★★
Medical image analysis	★★★

★ Conceptual ★★ Academic Research ★★★ Industrial Deployment

Many of these use cases can also be pursued using traditional centralized machine learning (AI) methods. However, FL offers unique advantages that can address specific challenges and unlock opportunities that traditional AI may struggle with.

Opportunities enabled by Federated Learning

Non-Exhaustive

Collaborative Research without Centralization: Life Science companies can engage in collaborative research on complex questions without the need to centralize or share sensitive data.

Competitive Advantage Preservation: Life Sciences companies can form partnerships and consortia while safeguarding their competitive edge.

Bridging the Interdisciplinary Gap: Federated Learning facilitates the exchange of knowledge between medical researchers and data scientists, bridging the gap between AI and clinical practice.

Diverse Research Exploration: Thanks to the adaptable nature of federated learning frameworks, collaborations can span various research domains, fostering innovative solutions more rapidly than conventional machine learning methods.

Compliance and Infrastructure Adaptation for Global Collaboration: FL can conform to region-specific privacy regulations, provide essential IT infrastructure, standardize code-sharing practices, and establish equitable compensation models for participating partners.

Scientific and technical challenges in Federated Learning

Algorithmic & model challenges

- **Optimization challenges**: developing algorithms to tackle communication constraints and non-convex loss landscapes.
- **Non-IID data robustness**: designing algorithms that remain robust amidst non-IID data distributions.
- **Fairness and bias**: mitigating bias arising from uneven data distribution.

Technical & infrastructural challenges

- **Infrastructure variability**: adapting to diverse device setups and node configurations.
- **Data standardization**: harmonizing data formats to enable cohesive training.
- **Scalability**: managing a large number of nodes and large models efficiently.

Privacy, security & evaluation challenges

- **Privacy-Preserving technologies**: balancing privacy preservation, training efficiency, and model accuracy.
- **Evaluation metrics**: establishing metrics to comprehensively assess performance, privacy, and other crucial aspects.

Depending on the challenge at hand, the various prominent solutions can be an ideal or less optimal fit

Challenges / Frameworks	MSR FLUTE FL	NVIDIA FLAIR	PySyft	DataSHIELD	intel/openfl
Algorithmic & Model Challenges					
Optimization Challenges	★★	★★★	★★	★	★★
Non-IID Data Robustness	★★★	★★★	★★	★★	★
Fairness and Bias	★	★★	★★	★★	★
Technical & Infrastructural Challenges					
Infrastructure Variability	★★	★★	★★	★	★★
Data Standardization	★★★	★	★★	★★	★★
Scalability	★★	★★★	★★	★	★★★
Privacy, Security & Evaluation Challenges					
Privacy-Preserving Technologies	★	★★	★★★	★	★★
Evaluation Metrics	★★	★★	★★	★★	★★

★ Fair ★★ Good ★★★ Excellent

Case Studies of Federated Learning

Over the last five years, we have witnessed an acceleration in the pace of creation of new public and private initiatives, showing the general need and trend for collaboration with Federated Learning

FeTS: A project that employs federated learning to enable different institutions to collaboratively develop brain tumor segmentation models from MRI scans without centralizing patient data.

AI4BH: A federated learning initiative aimed at improving cardiovascular health by enabling multiple healthcare providers to collaboratively enhance AI-driven diagnostic and prognostic models for heart disease.

MELLODDY: A collaborative project that applies federated learning to allow multiple pharmaceutical companies to work together on drug discovery without sharing sensitive chemical data.

SAIL DATABANK 2006
23andMe 2006
biobank uk 2006
the YODA PROJECT Forging a unified scientific community 2011
ANSWER ALS 2013
FDA NIH 2014
healthverity 2014
HEALTH DATA COLLABORATIVE 2015
health bank 2015
MIDATA 2015
LONGENESIS 2017
luna 2017
SALUS COOP COOPERATIVA CIUDADANA DE DATOS PARA LA INVESTIGACIÓN EN SALUD 2017
PHT-meDIC 2018
MAYO CLINIC 2019
KAISER PERMANENTE. RESEARCH BANK 2019
MELLODDY 2019
TRUVETA 2020
2020
bigpicture 2021
AI4BetterHearts Cardiovascular Data Collaborative 2022
NHS 2022
Labelia IA responsable et de confiance
SAFE
NFLC

2006 2018 2023

q&a

thank you