Federated Learning in Bioinformatics
Welcome and Course Introduction

Laura Azzimonti

Dalle Molle Institute for Artificial Intelligence, SUPSI, Lugano, Switzerland
Swiss Institute of Bioinformatics, SIB, Switzerland

Machine Learning for Bioinformatics and Personalised Medicine group

Laura Azzimonti
Senior researcher and lecturer

Francesca Mangili
Senior researcher

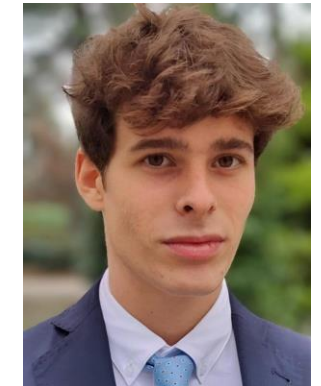**Daniele Malpetti**
Researcher and lecturer

Sandra Mitrović
Researcher

Lea Multerer
Researcher

Francesco Gualdi
Researcher

Tommaso Sommaruga
Bachelor student

Clara Galimberti
Researcher

Giovanni Angelotti
PhD student

Christian Berchtold
Master student

Manuel Acquistapace
Master student

Machine Learning for Bioinformatics and Personalised Medicine group

Laura Azzimonti
Senior researcher and lecturer

Francesca Mangili
Senior researcher

Daniele Malpetti
Researcher and lecturer

Sandra Mitrović
Researcher

Lea Multerer
Researcher

Francesco Gualdi
Researcher

Tommaso Sommaruga
Bachelor student

Clara Galimberti
Researcher

Giovanni Angelotti
PhD student

Christian Berchtold
Master student

Manuel Acquistapace
Master student

Dalle Molle Institute for Artificial Intelligence (IDSIA)

A research institute on AI founded in 1988 in Lugano providing a "bridge" between theory and applications

Thanks to the Italian philantropist Angelo Dalle Molle (1908-2002)

IDSIA is a common institute of USI & SUPSI since 2000

«The progress of science in general, and that of the emerging computer science in particular, should not subjugate people, but rather benefit them»

134 people
18 professors, 50+ researchers, 55+ PhD students

Six Research Areas

ALGORITHMS AND THEORY

2 professors, 4 researchers, 7PhDs

Approximation algorithms, combinatorial optimization, formal languages and automata theory, foundations of probability, foundational AI models

AUTONOMOUS ROBOTICS

2 professors, 7 researchers, 7 PhDs

Machine learning for robot perception, Interaction interfaces for humans and robots sharing spaces, Robotics for education

GEOMETRIC COMPUTING AND VISION

3 Professors, 3 researchers, 6 PhDs

Computational geometry and geometry processing, Computer graphics and computational fabrication, Machine vision

MACHINE LEARNING AND ARTIFICIAL NEURAL NETWORKS

3 professors, 19 researchers, 14 PhDs

Recurrent neural networks, Machine learning of temporal data, Security for ML, ML for security, causality in ML, reinforcement learning, Graph and geometric deep learning

INTELLIGENT CONTROL OF NETWORKED SYSTEMS

3 professors, 22 researchers, 9 PhDs

Graph-based reinforcement learning in structured environments, combinatorial optimization and AI for intelligent planning and control, Optimal control and self-tuning of industrial machines, relational spatio-temporal representations for prediction and control

INFORMATION RETRIEVAL AND NATURAL LANGUAGE PROCESSING

2 professors, 4 researchers, 5 Phds

Advanced text analysis and term weighting techniques; Mobile and conversational IR, Distributed word representations in NLP, Deep learning for NLP
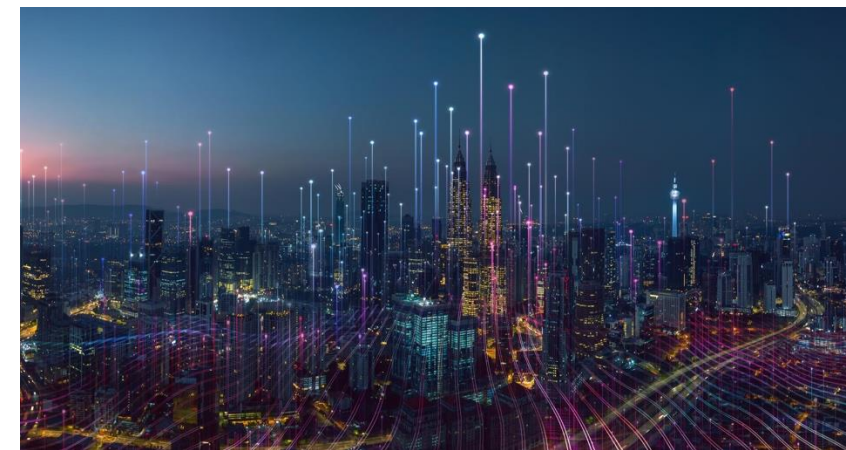
Some of our application domains

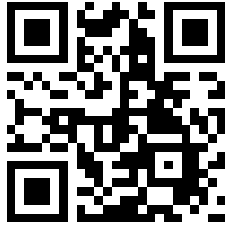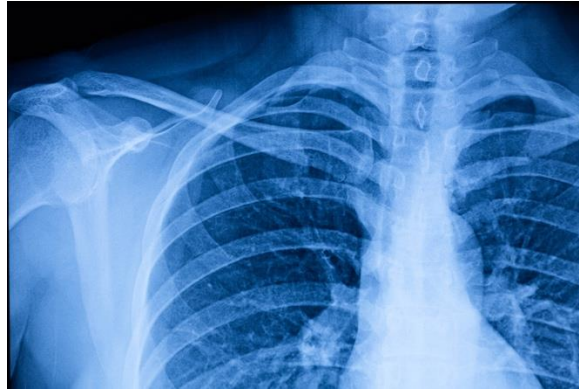AI for Industry 4.0   **AI for Health and Life Sciences**   AI for Energy and Sustainability
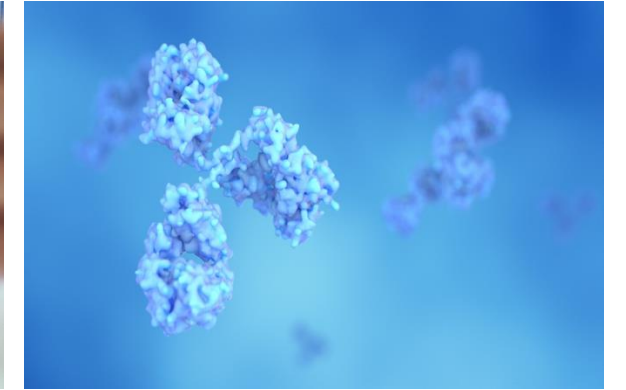
AI for Health and Life Sciences @IDSIA

ML for Personalised and Predictive Medicine

Computer Vision for Biomedical Imaging

NLP for Healthcare

Computational Biology

Genomics

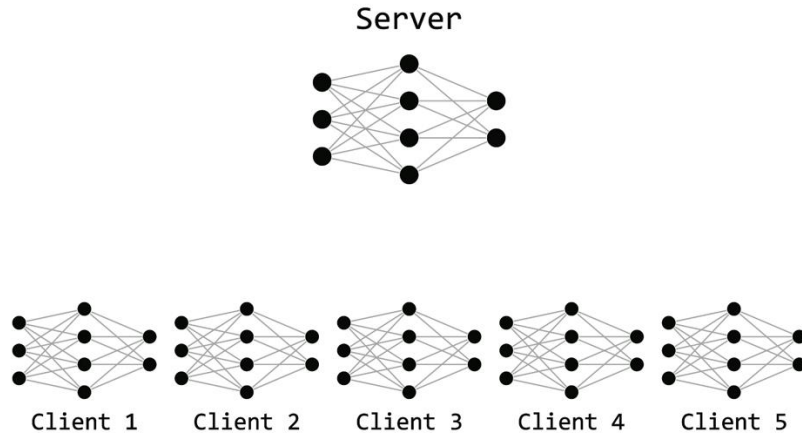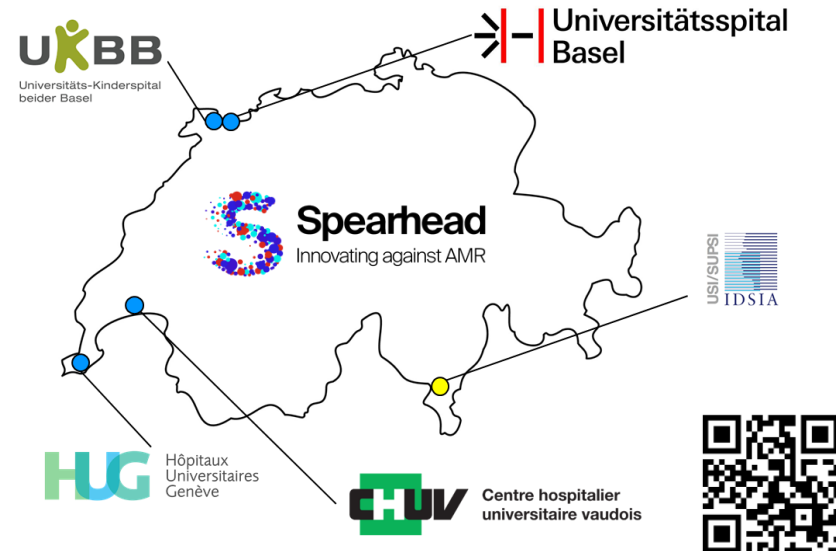Robotics and VR for rehabilitation

Biosignal processing

Ethics and AI

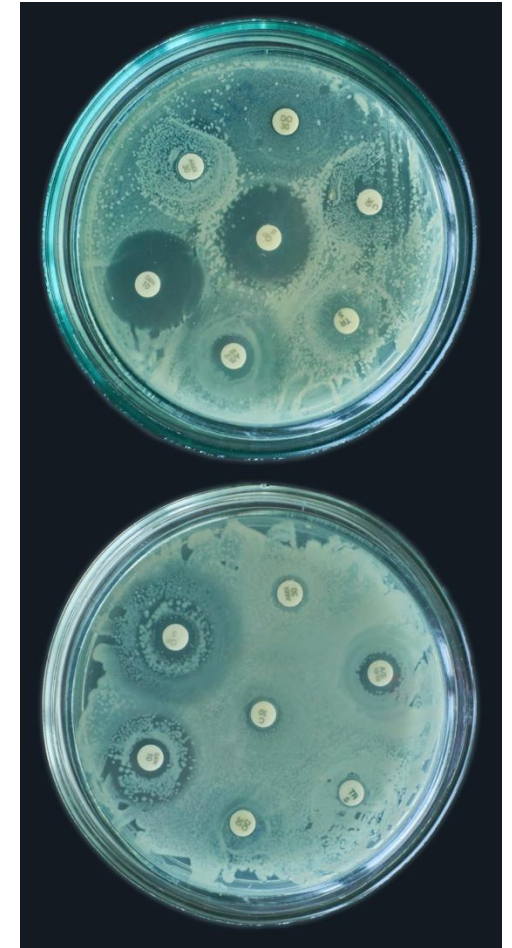# Federated Learning for antibiotic resistance prediction

Federated Learning permits collaborative model training **without** data exchange.

The consortium

Federated framework through **FLOTTA**. IDSIA's library for Federated Learning.

C. Bonesana et al., Flotta: a Secure and Flexible Spark-inspired Federated Learning Framework

Next Generation Tools for Genome Data Integration

Next generation tools for genome-centric multimodal data integration in personalized cardiovascular medicine.
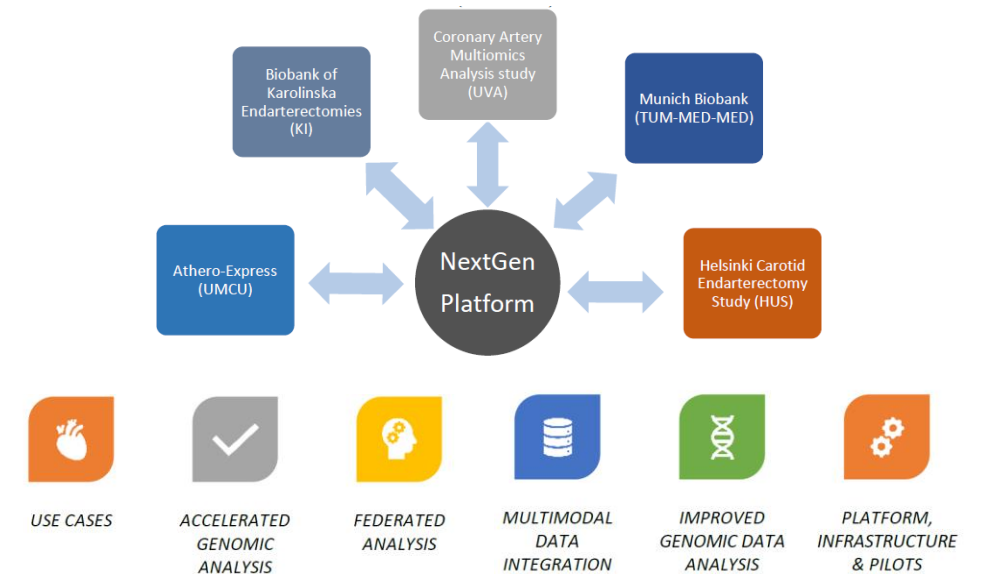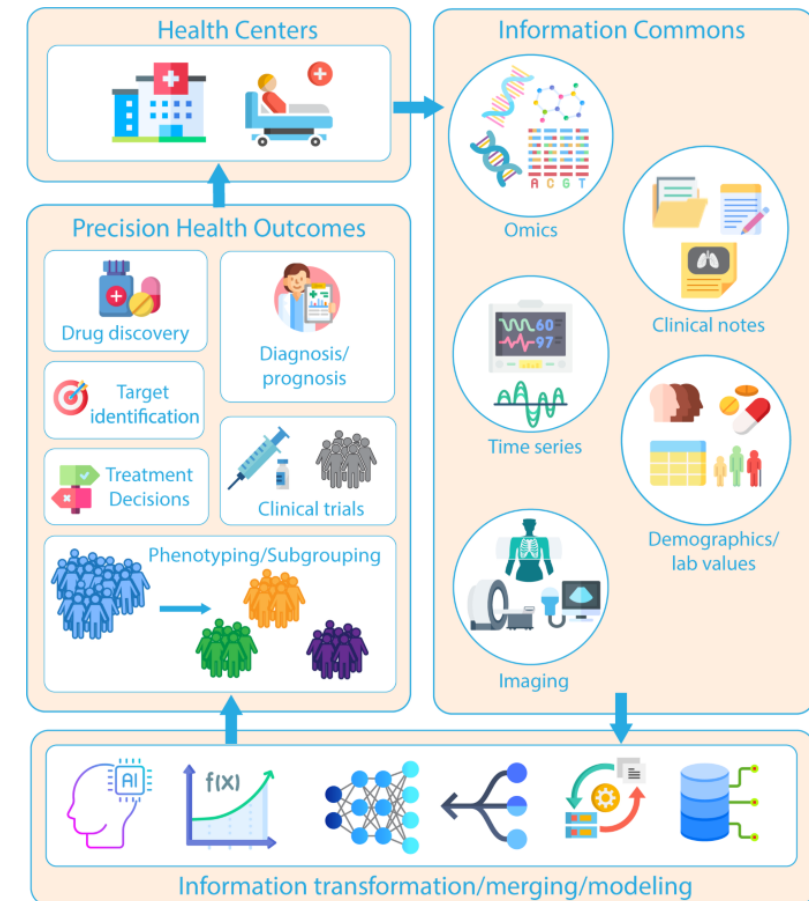
Next Generation Tools for Genome Data Integration

True portability of multimodal research

- Build tools to enable portable multimodal (multi-omics) research in cardiovascular medicine.

- Exploit distributed semantics for data discovery and harmonization

- Exploit AI for data integration by encoding

New federated genomic analyses

- Deploy a platform with federated data catalogues, machine learning and genomic analytics

- Build a pathfinder network with five demonstration site

Kline, Adrienne, et al. "Multimodal machine learning in precision health: A scoping review." *NPJ digital medicine* 5.1 (2022): 171.

Next Generation Tools for Genome Data Integration

Flower is the main FL framework adopted in NextGen

Federation in progress of multiple bioinformatics tools

- Genome Wide Association Studies

- Pathway-level information extractor (PLIER) for gene expression data

- Single-cell variational inference (SCVI)

- Machine learning for pathogenic variant prioritization models

Course Schedule

Start	End	Session
09:00	09:15	Welcome and course introduction
09:15	10:15	**Theory Block 1:** FL foundations
10:15	10:45	☕ Coffee break
10:45	11:30	**Guest Lecture:** Flower, a Friendly Federated AI Framework (Y. Gao & W. Lindskog, Flower Lab)
11:30	12:30	**Practical Block 1:** My first FL project
12:30	13:30	🍴 Lunch break
13:30	14:15	**Theory Block 2:** FL for Bioinformatics
14:15	14:30	☕ Coffee break
14:30	16:00	**Practical Block 2:** FL for single-cell data
16:00	16:30	Wrap-up & discussion

Practical sessions: Emulation of a realistic collaborative workflow.

Different institutions hold pancreas single-cell datasets, with each dataset generated using a different sequencing technology.

They collaborate to train an open-source model for technology-related batch-effect removal using a Federated Learning approach.

Then, a separate institution, which also has pancreas single-cell data, uses the released model to remove batch effects on their data and subsequently perform several downstream tasks.

Gene Expression UMAP

13