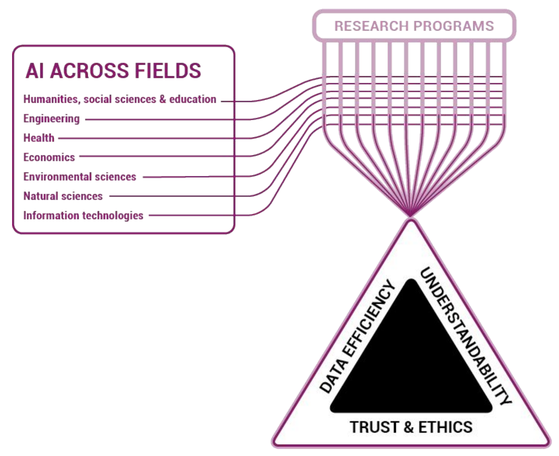
FCAI Finnish Center for Artificial Intelligence

fcai.fi

Privacy-preserving and secure AI (Research Program R4)

FCAI Research Program R4 studies trustworthy AI that can also operate in the presence of adversaries attempting to steal data or models, or evade models. For the time being, our focus is mainly on privacy, where we focus on differentially private machine learning.



Program objectives

Generally: developing rigorous and realistic models of adversaries and techniques for thwarting attacks.

Privacy:

- Differentially private ML algorithms and probabilistic modelling
- Differential privacy with distributed data and federated learning
- Differential privacy for data anonymisation
- Software tools for developed methodology

Methodologies

Differential privacy (DP) limits how much a single user's data can impact the outcome of a computation M: for all data sets *D* and *D'* differing by one sample, we have

$$\Pr(M(D) \in S) \leq e^{\epsilon} \Pr(M(D') \in S) + \delta$$

- Information-theoretic guarantee against privacy attacks
- Protects against adversaries with side information
- Degrades gracefully under repeated use

Example ("randomised response"): Answering a sensitive question by randomly flipping the answer with probability *p* is (ϵ,0)-DP with ϵ=log((1-*p*)/*p*)

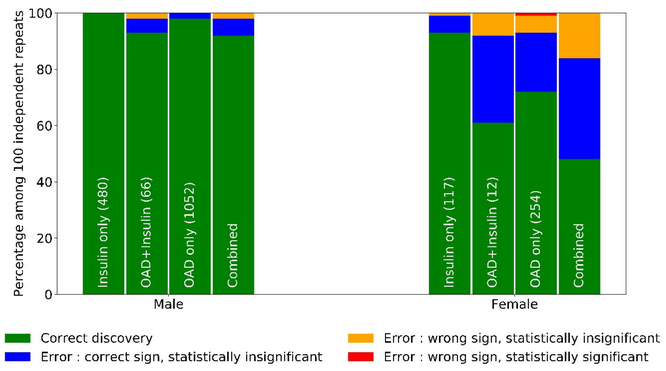
Research results

Differentially private data anonymisation

- Sharing strongly anonymised data under DP guarantees
- Idea: learn a generative model for the data under DP, generate synthetic data set from the model
- Useful for explorative analysis of data sets
- Can recreate models using the privatised data with very high accuracy, but smaller details may suffer
- Methods implemented in easy-to-use software package Twinify

(Jälkö et al., arXiv:1912.04439)

Findings of epidemiological research can be reproduced reliably on an anonymised data set when the number of cases is sufficient (left, 1598 cases) while small number of cases is difficult under strict privacy (right, 383 cases). (Both examples have 10 features, 226k total individuals for male, 208k total individuals for female, DP with ϵ=1.0.)

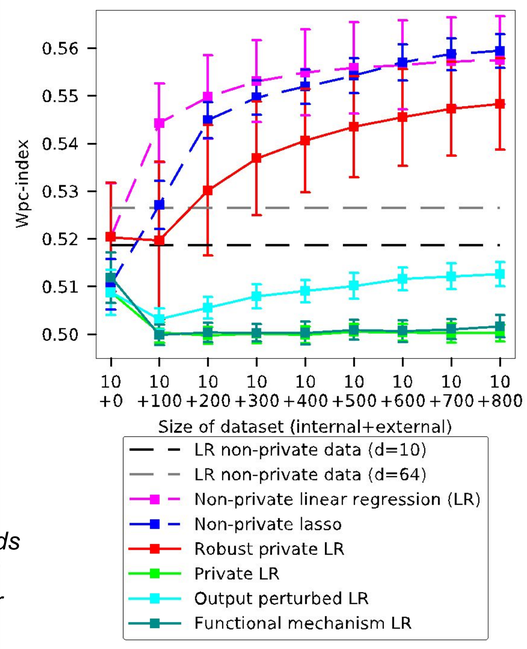


Differentially private drug sensitivity prediction

- First method to successfully predict the sensitivity of cancer cell lines to drugs using gene expression data under DP using limited training data
- Learning a low-dimensional representation for gene expression further improves prediction accuracy

(Honkela *et al.*, Biol Direct 2018; Niinimäki *et al.*, Bioinformatics 2019)

Prediction accuracy (higher is better) of different methods for drug sensitivity prediction. Our DP method (red solid line) yields similar results as non-private methods under strong privacy using only 4x more data.



Differentially private Bayesian inference

- Developing new algorithms enabling Bayesian inference with differential privacy guarantees:
- Differentially private variational inference (DPVI)
- Differentially private Markov chain Monte Carlo
- Noise-aware DP inference for generalised linear models

(Jälkö *et al.*, UAI 2017; Heikkilä *et al.*, NeurIPS 2019; Kulkarni *et al.*, under preparation)

Differentially private learning with distributed data

- Combining DP with secure multi-party computation allows efficient DP learning without relying on a single trusted party
- Learning by securely combining data of same individuals held by different parties (vertically partitioned data)

(Heikkilä *et al.*, NIPS 2017; Tajeddine *et al.*, under preparation)

Algorithms for differentially private machine learning

- Numerical methods for exact privacy accounting for DP stochastic gradient optimisation enable flexible and accurate evaluation of privacy loss in complex algorithms

(Koskela *et al.*, AISTATS 2020)

- Learning rate adaptation for DP stochastic gradient optimisation enables efficient use of computational resources and simplifies use of DP learning

(Koskela & Honkela, AISTATS 2020)

Coordinating professor

Antti Honkela

Associate professor of data science

University of Helsinki

antti.honkela@helsinki.fi

Finnish Center for Artificial Intelligence FCAI is a community of experts in artificial intelligence in Finland, initiated by Aalto University, University of Helsinki, and VTT Technical Research Centre of Finland. FCAI is one of the six Academy of Finland Finnish flagships.

Aalto University · UNIVERSITY OF HELSINKI · VTT · ACADEMY OF FINLAND