FCAI
Finnish Center for Artificial Intelligence
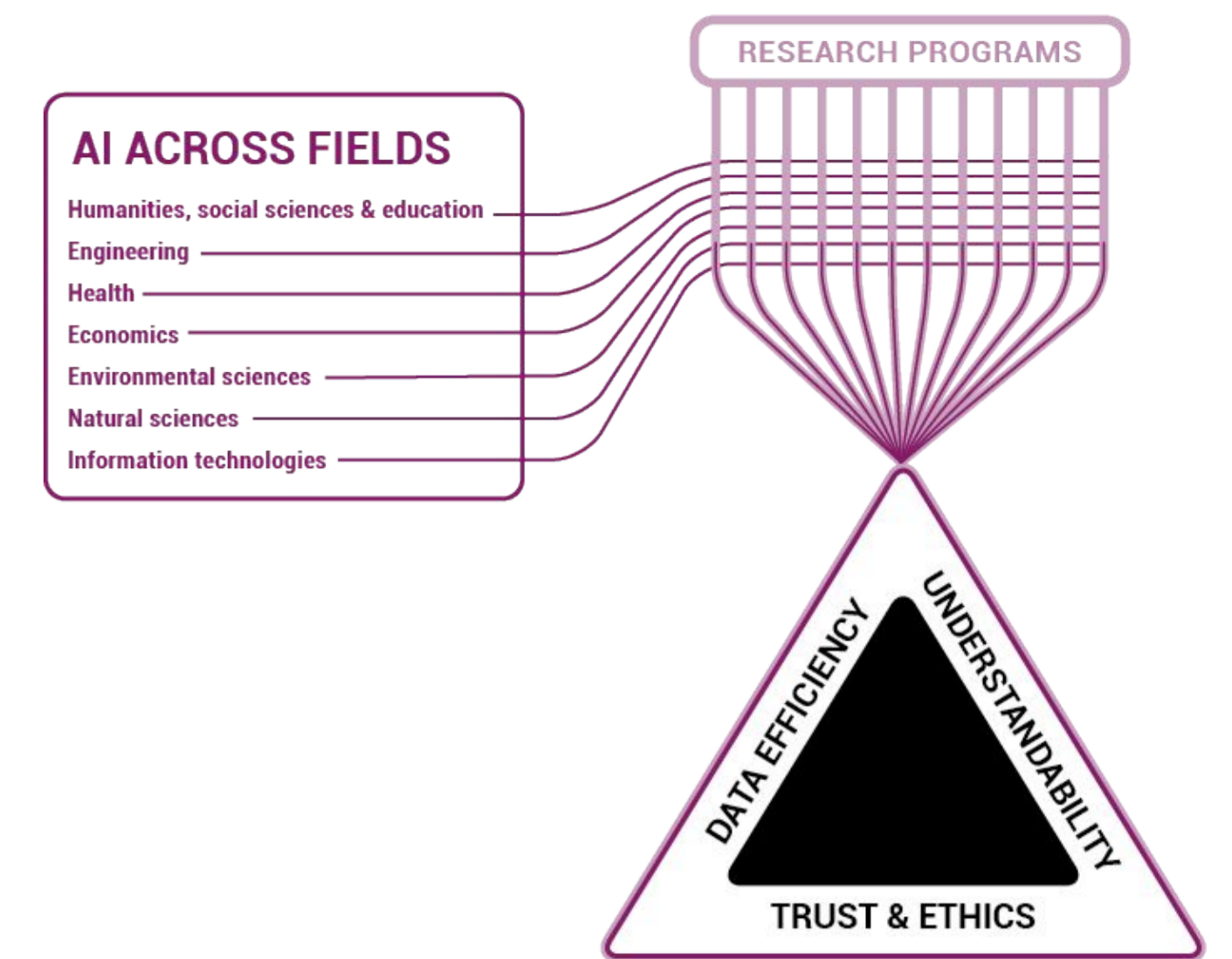
fcai.fi

Easy and privacy-preserving modeling tools (Highlight A)

Developing AI is still difficult and time-consuming, especially for probabilistic approaches. Many of FCAI's research contributions are about making it easier. FCAI Highlight Program A exists to make sure this research has impact for people building AI in industry and academia.

Highlight objectives

AI has uniquely narrow gap between fundamental and applied research: the advances in AI methods can rapidly be taken into use in industry and other research fields. This requires AI that is

1. **Easy enough** to be used by **wide audience**
2. Easily **accessible** as (open-source) software
3. Computationally **reliable** and addressing the **real needs**

FCAI takes this seriously and supports all stages of AI-assisted design, modeling and decision-making in form of fundamental research and practical tools. For more details, see the related research programs:

R1: Reliable, replicable and efficient Bayesian workflow to support model specification

R2: Efficient and accurate algorithms for simulator-based inference

R4: Differential privacy guarantees for probabilistic inference

R5: Human-AI collaboration in model design and human aspects in modeling workflows
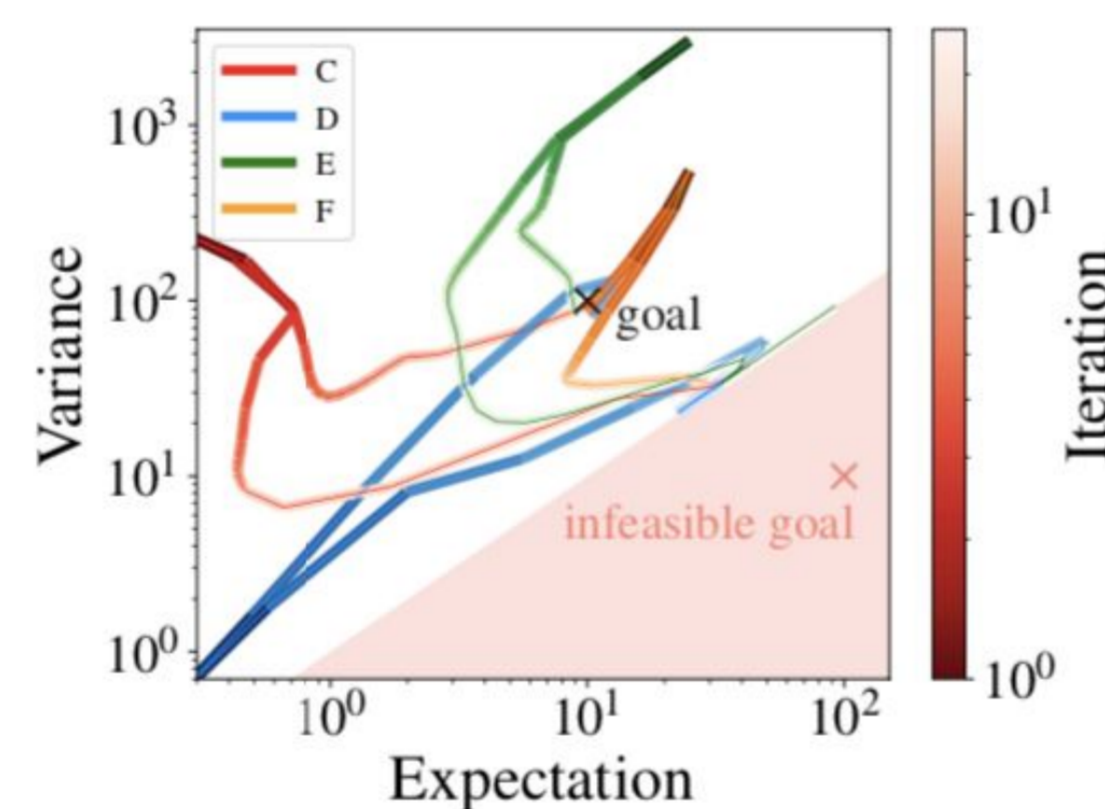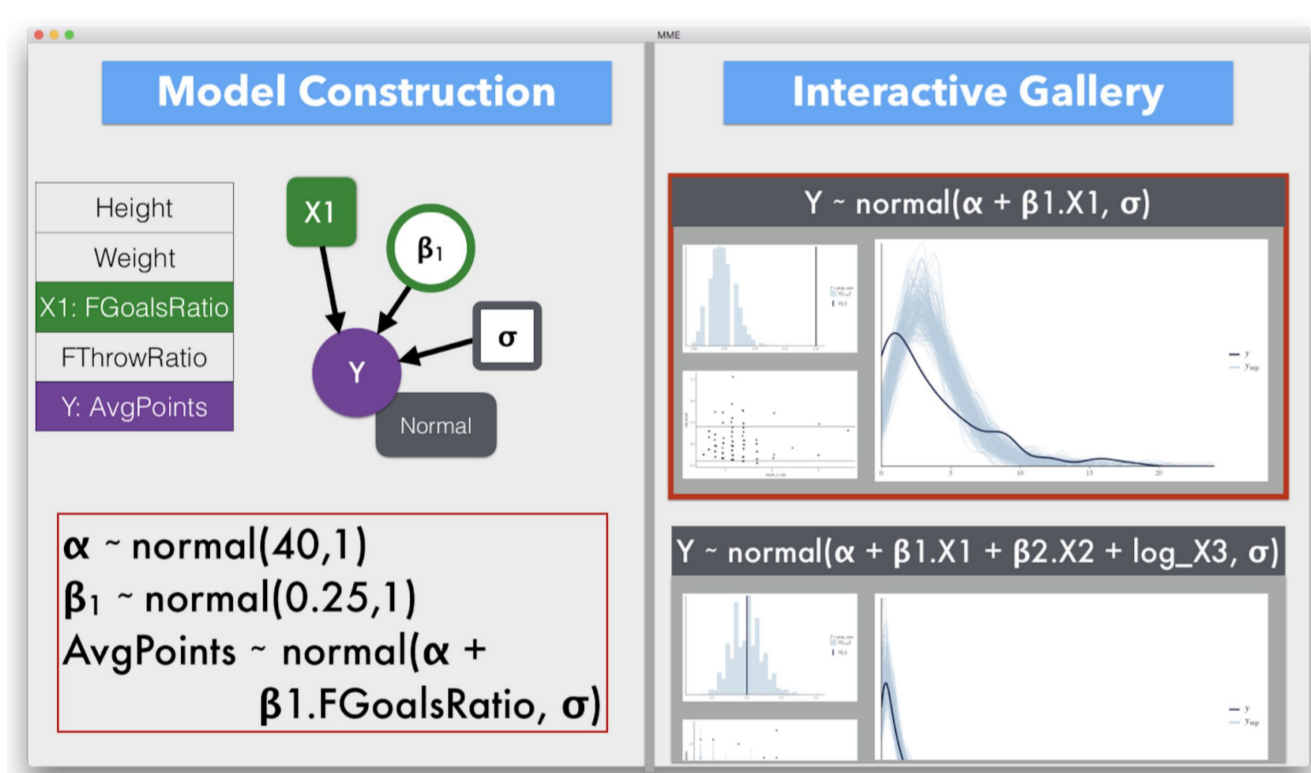
Modeling made easy (R5)

Interactive tool for domain experts without modelling expertise. AI assists by intelligent querying, feedback, and suggestions, to extract knowledge, facilitate model exploration and construction, and support learning and understandability.

Prior elicitation (R1)

Automatic selection of prior distributions for predictive Bayesian models and interactive elicitation of expert's knowledge for interpretable models, based on prior predictive distribution.

Open-source software

FCAI devotes significant effort to make tools for the community by

1. Contributing to established software libraries with large user bases
2. Leading the development of dedicated tools building on core FCAI research activities

Besides supporting the development, we also measure and maximize the **impact** of these tools:

1. Are the tools being used?
2. Do they help solving important problems? (in REAL WORLD)
3. How strong is the community involvement?
4. Are they actually easy to use? (for REAL PEOPLE)
5. Are they ethically sustainable?

Key software activities

Stan (https://mc-stan.org/)

FCAI contributes directly to Stan, the leading probabilistic programming platform with large user-base. Stan is frequently used for solving fundamental science problems in top forums (e.g. *Nature* and *Science*).

In probabilistic programming we lead or contribute significantly to several software components, highlighted by

1. **brms** (https://cran.r-project.org/package=brms) Generalized multi-level models. 1500+ citations for the main paper (Bürkner, 2017), received the *2020 SIPS Mission Award* for improving methods and practices in psychological science
2. **loo** (https://cloud.r-project.org/package=loo) Efficient evaluation of probabilistic models. 1300+ citations for the main paper (Vehtari et al. 2017)
3. **posterior** (https://github.com/stan-dev/posterior/) Tools for analysis of posterior distributions.
4. **ArviZ** (https://github.com/arviz-devs/arviz) Exploratory analysis for Bayesian models, with support for multiple inference engines.

ELFI (https://elfi.readthedocs.io/)

Engine for Likelihood Free Inference (ELFI) is FCAI-lead project providing flexible support for posterior inference for simulator-based models.

- Supports multiple likelihood-free inference algorithms (see R2)
- ~200 stars, ~50 forks in github despite early stage (v0.8)
- Used for top-level scientific research e.g. in *Nature Microbiology* and *Cognitive Science* in 2020.

Differential privacy

We provide the tools for adding differential privacy in probabilistic programming. While there are several ongoing software activities for securing data and user privacy, tools that ingerate into part of a regular probabilistic modeling workflow have been missing. We lead development of

1. d3p (https://github.com/DPBayes/dppp) Differential privacy support for NumPyro, supporting DP stochastic variational inference and privacy accounting
2. Twinify (https://github.com/DPBayes/twinify/) Easy-to-user tool for creating synthetic data that retains privacy by generative modeling

Coordinating professor

Arto Klami

Assistant professor in computer science

University of Helsinki

arto.klami@helsinki.fi

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