Aalto University

The University of Manchester

Collaborative modelling, design and decision making with AI, Part I

Samuel Kaski

FCAI

Funded by:

ACADEMY OF FINLAND

UK Research and Innovation

Funded by the European Union

Nov 2022

FCAI

Finnish Center for Artificial Intelligence

Aalto University

UNIVERSITY OF HELSINKI

VTT

FLAGSHIP PROGRAMME

Finland for AI

Finland provides a favorable operating environment for the creation, development and utilization of AI technologies:

- Helsinki is among the top 50 global AI talent hubs, despite some heavy competition (*Harvard Business Review 2021*)
- Finland hosts LUMI, the fastest supercomputer in Europe and #3 in the world
- Finland #3 in Data Science (*Coursera's Global Skills Report 2022*)
- Finland is the 5th cited country in the EU (*OECD.AI 2022; Nordic Innovation 2022*)
- Finland #4 (*Government AI Readiness Index 2021*)
- AI adoption rate #3 in the EU (*European Commission 2022*)

FCAI: Finland's AI hub

What is FCAI?

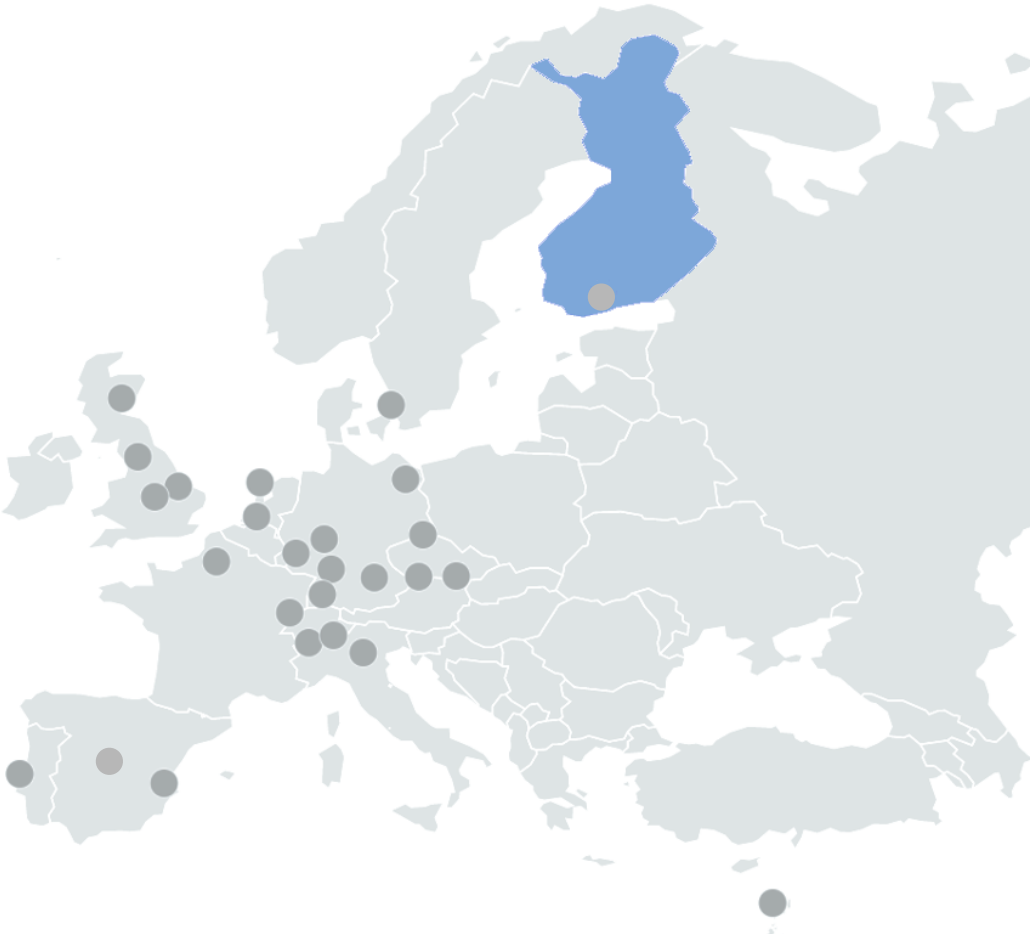
- Academy of Finland Flagship for 2019–26
- Built on the long track record of pioneering machine learning research and multidisciplinary collaboration
- Top-level research and wide impact together with industry and public sector
- 70 professors and their groups, volume 250 M€

In four years of operation

**) Combination of artificial intelligence, machine learning, human-computer-interaction, 2019–21*

Finland in the European ecosystem

- **Pioneer in AI and machine learning, world-class fundamental research since 1960’s**
- **Hosting one of the first ELLIS nodes**
- **Leading EU’s European Network of AI Excellence Centres — ELISE, which builds and executes a European Strategic Research Agenda in AI**



European Laboratory for Learning and Intelligent Systems

Contents

Virtual Laboratories: Transforming research with AI

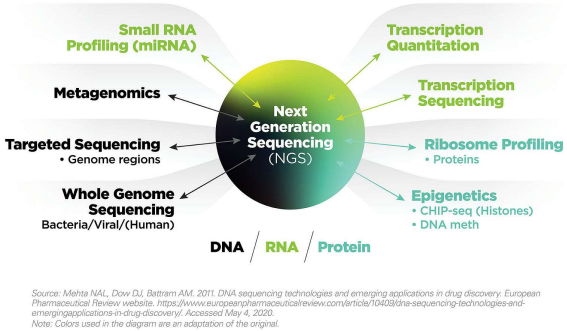
Arto Klami[1,3], **Theodoros Damoulas**[2,4], **Ola Engkvist**[5,6], **Patrick Rinke**[1,7], **Samuel Kaski**[1,2,8,9]

[1]*Finnish Center for Artificial Intelligence FCAI,* [2]*Alan Turing Institute,* [3]*Dept. of Computer Science, University of Helsinki,*
[4]*Depts. of Computer Science and Statistics, University of Warwick,* [5]*Molecular AI, Discovery Sciences, R&D, AstraZeneca,*
[6]*Dept. of Computer Science and Engineering, Chalmers University of Technology,* [7]*Dept. of Applied Physics, Aalto University,*
[8]*Dept. of Computer Science, Aalto University,* [9]*Dept. of Computer Science, University of Manchester*

DOI

10.36227/techrxiv.20412540.v1

How can we help with the global challenges?



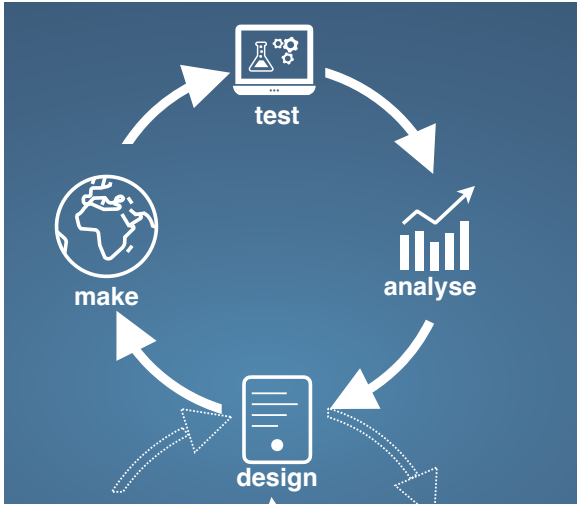
Source: Mehta NAL, Dow DJ, Battram AM. 2011. DNA sequencing technologies and emerging applications in drug discovery. European Pharmaceutical Review website. https://www.europeanpharmaceuticalreview.com/article/10409/dna-sequencing-technologies-and-emergingapplications-in-drug-discovery/. Accessed May 4, 2020.
Note: Colors used in the diagram are an adaptation of the original.

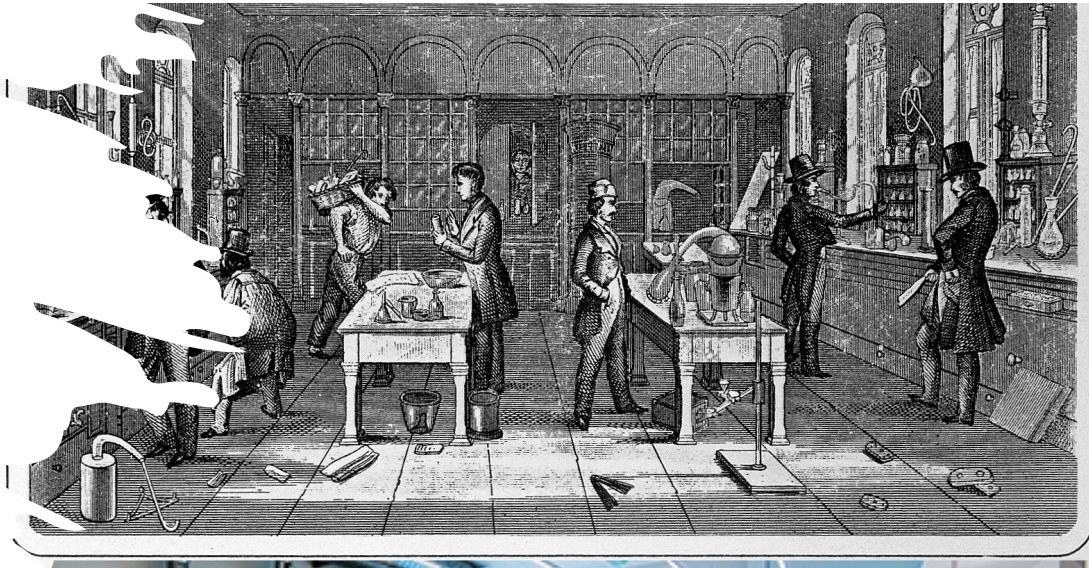
By becoming domain scientists? Good choice for some

Alternative: Improve the research process

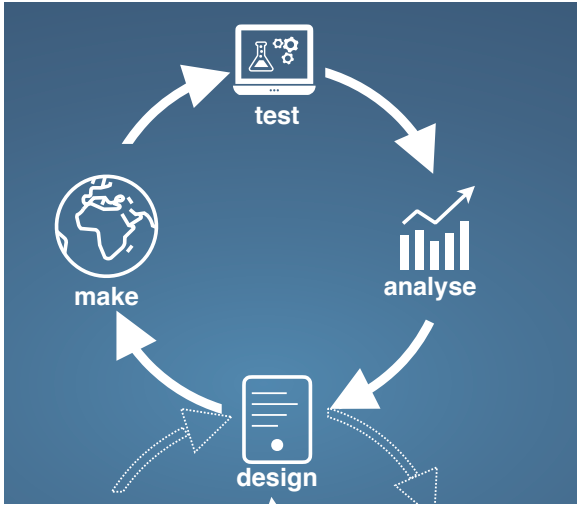
First part: Joint work with Arto Klami, Chris McGreavy, Carlos Sevilla Salcedo et al.

Research is still iterative refinement of hypotheses

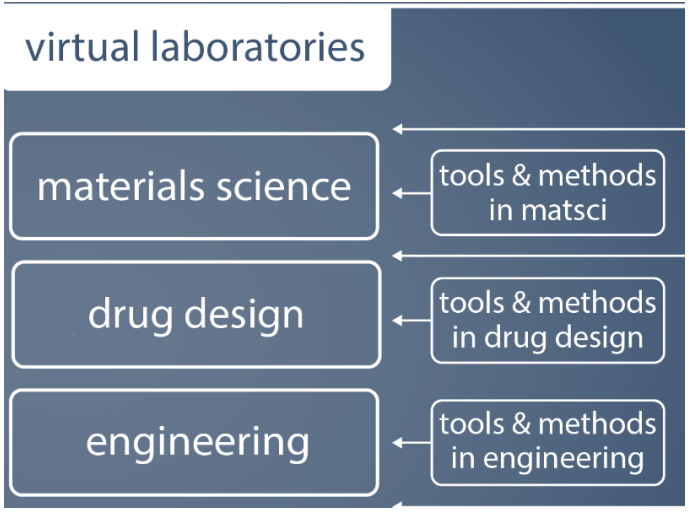




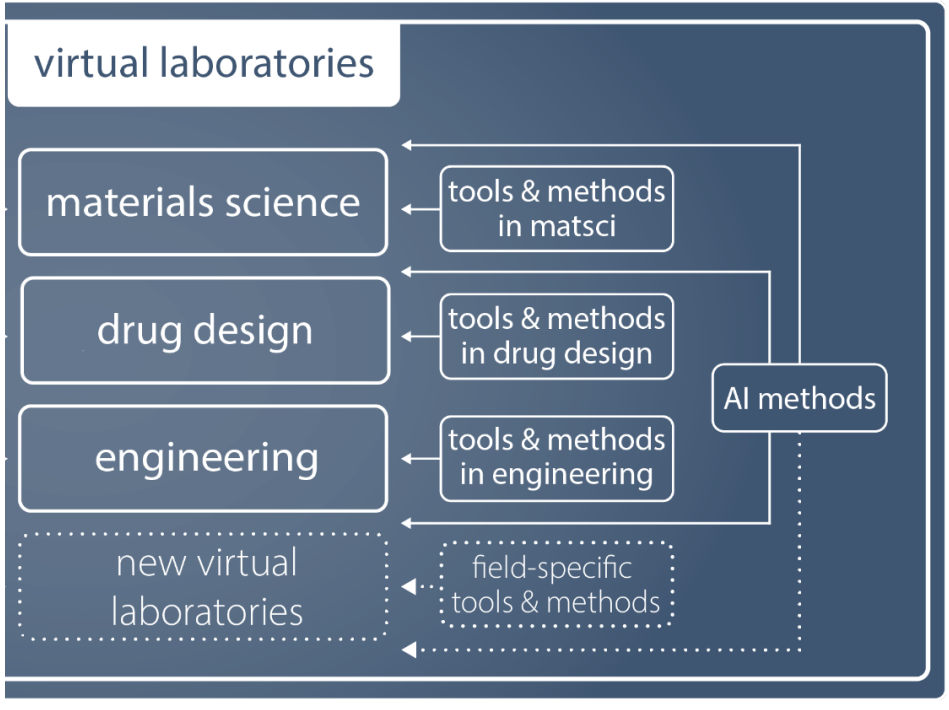
Research and design: Design-make-test-analyze loop



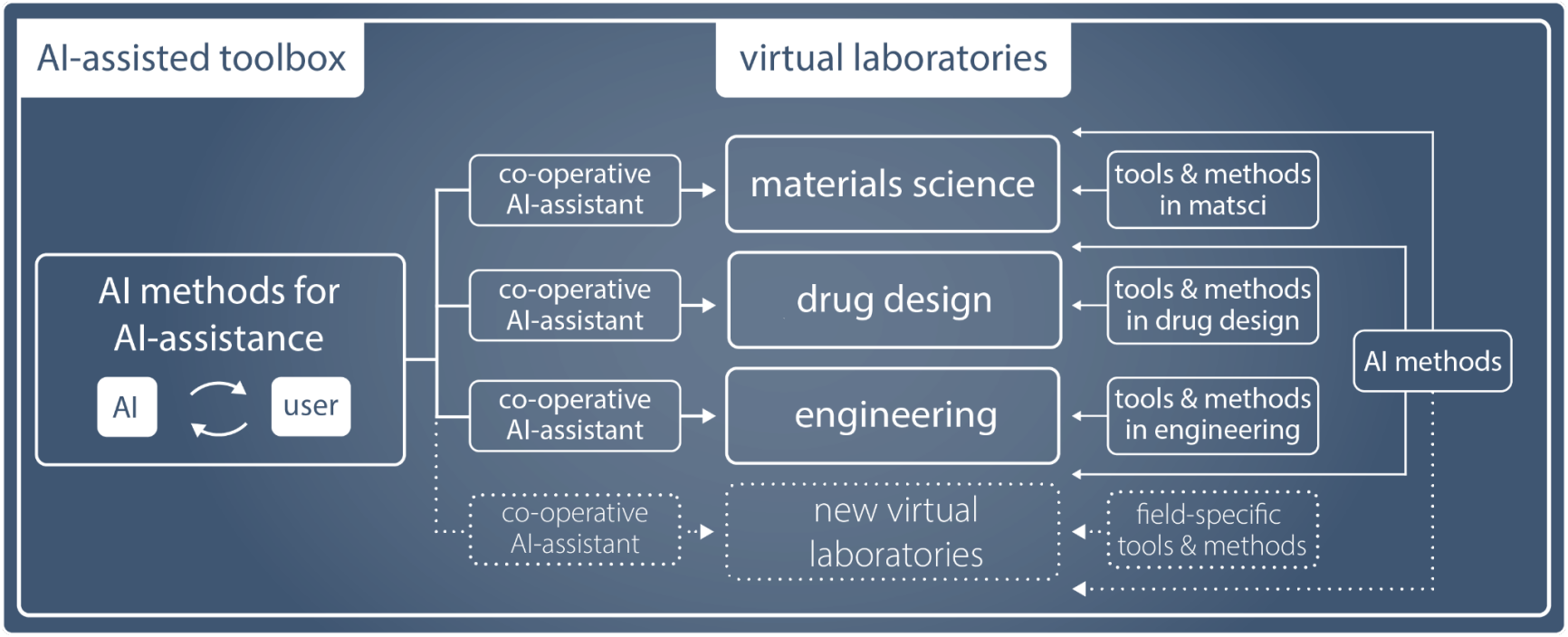
Virtual laboratories with field-specific tools



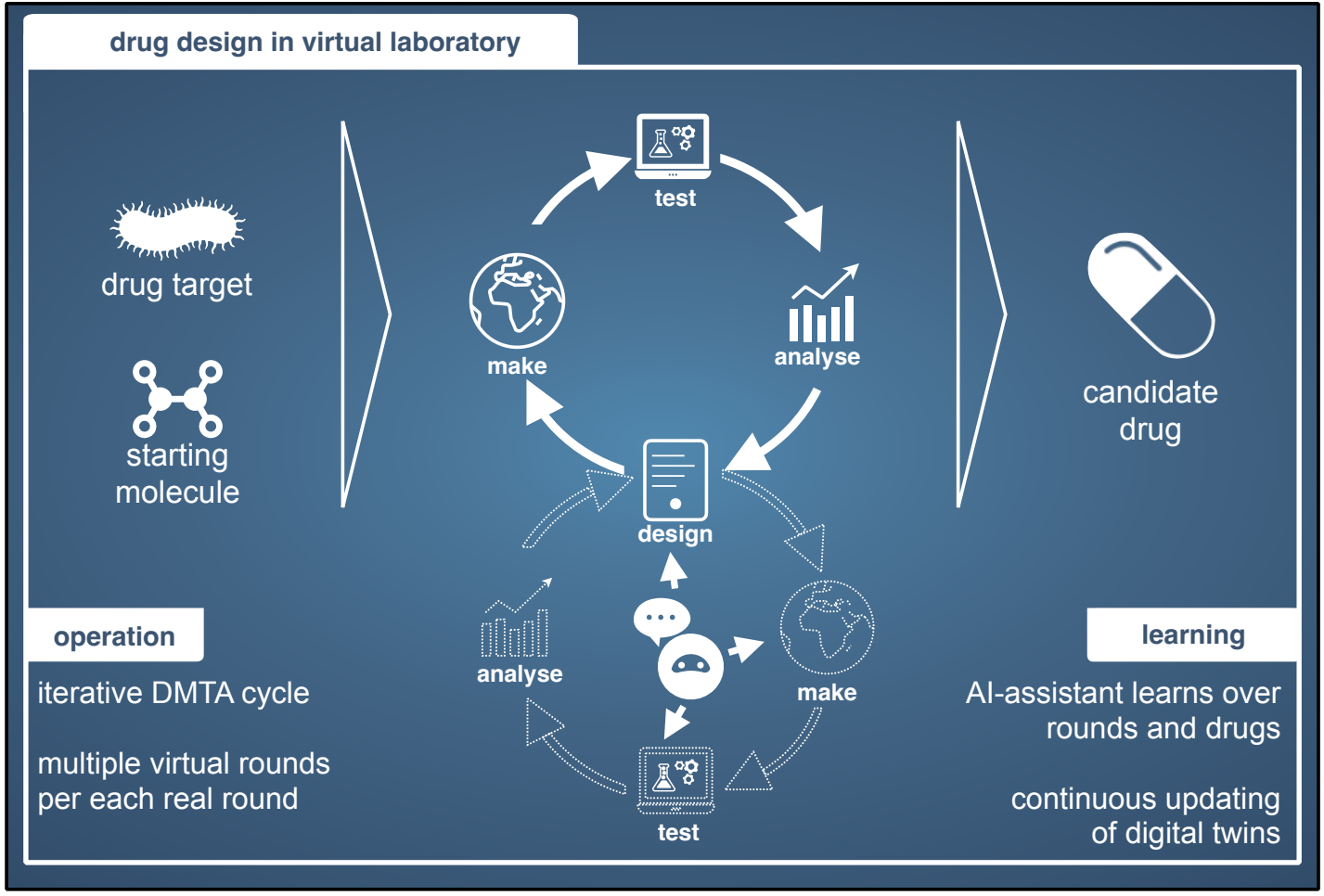
Scale advantage: AI tools across fields



Future: AI-assisted virtual laboratories



Klami et al, TechRxiv, 2022



Klami et al, TechRxiv, 2022

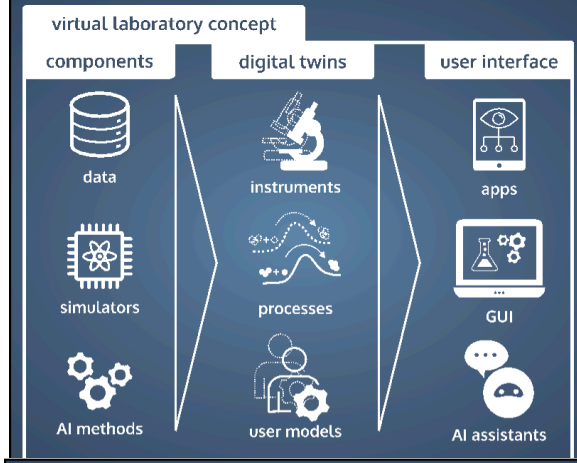
Virtual Laboratory

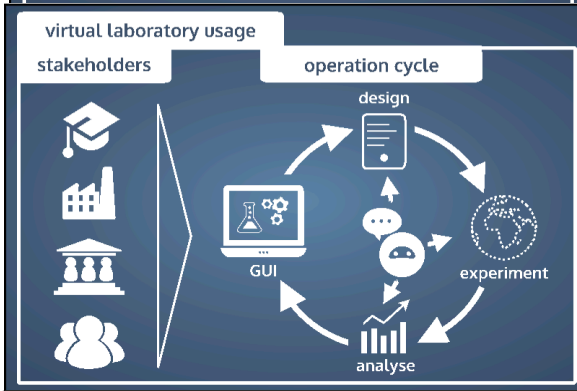
FCAI Finnish Center for Artificial Intelligence

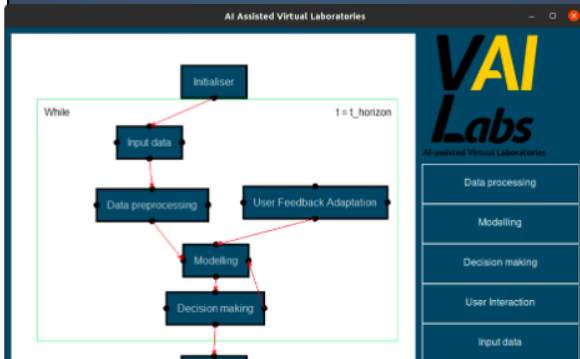
Rather than developing AI solutions for R&D of one field, we create them for the **R&D process itself**

A **Virtual Laboratory**

- Combines digital twins of instruments and processes with models of the researcher's intentions and capabilities
- Enables AI-assistance for supporting the researcher
- Klami et al. **Virtual Laboratories: Transforming research with AI**. https://doi.org/10.36227/techrxiv.20412540.v1
- First labs under way: material science, drug design, sustainble mobility
- Software: https://github.com/AaltoPML/VAI-Lab
- Open community - welcome to join







Why?

For VL hosts

- No need to create everything from scratch
- AI solutions for unimaginable problems, all advances easy to take in use
- Faster research cycle

For AI researchers

- Enables AI research with and without collaboration with VL hosts
- High-profile demonstrations
- Focus on your favorite part but benefit from what others have done

10% effort for ten times the impact

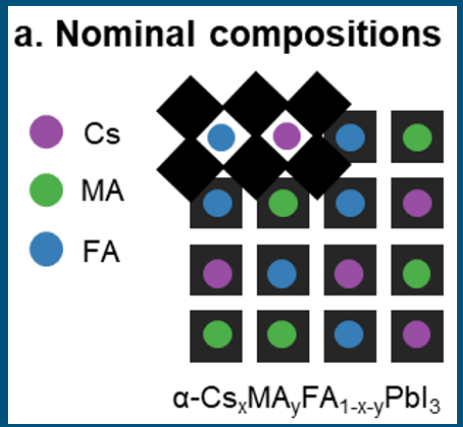
AI-assisted Virtual Laboratories (VAI Labs)

Chris McGreavy, Carlos Sevilla Salcedo, Samuel Kaski, Arto Klami

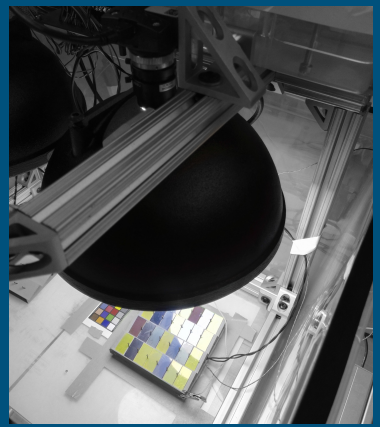
Material Stability using Bayesian Optimisation

Aim: Find crystal composition to maximise durability of solar panel material using rapid age-testing

Manufactured crystals



Composition Space



Age Testing w/image capture

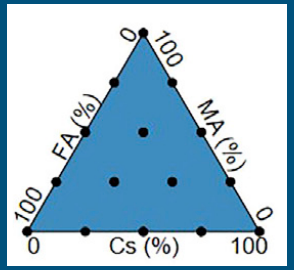
(5 days, 85°C, 85%, 0.15Sun)

Sun, S., Tiihonen, A., Oviedo, F., Liu, Z., Thapa, J., Zhao, Y., Hartono, N.T.P., Goyal, A., Heumueller, T., Batali, C. and Encinas, A., 2021. A data fusion approach to optimize compositional stability of halide perovskites. Matter, 4(4), pp.1305-1322.

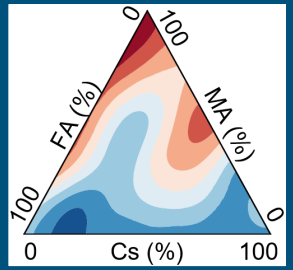
Material Stability using Bayesian Optimisation

Search space: 3 material components (as % of overall composition)

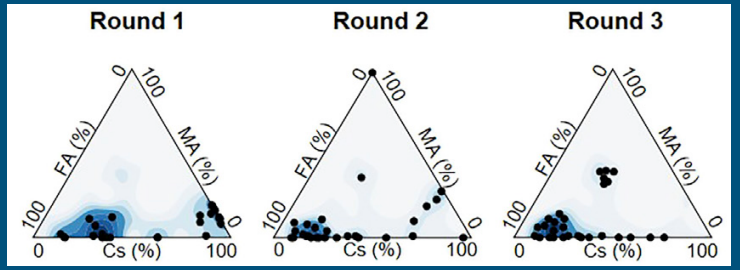
Method:



Uniformly sample search space



Generate ground truth model of stability (using GPR)

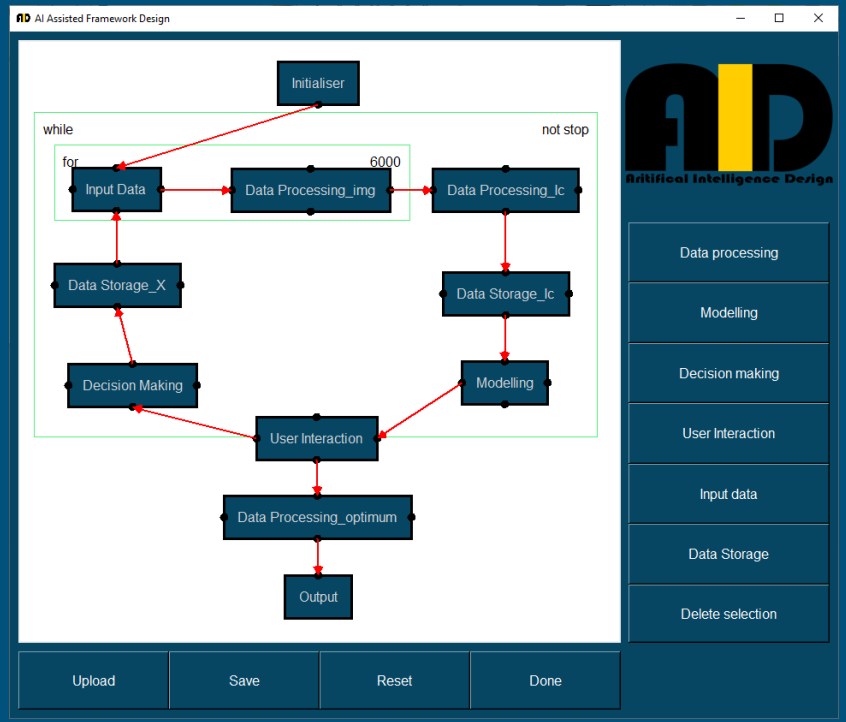


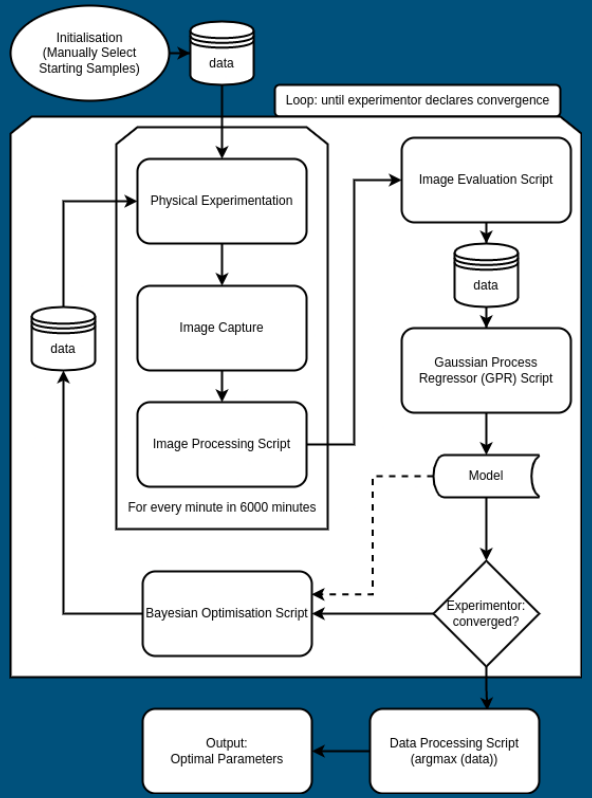
Bayesian Optimisation to search for optimally stable composition

(using ground truth model)

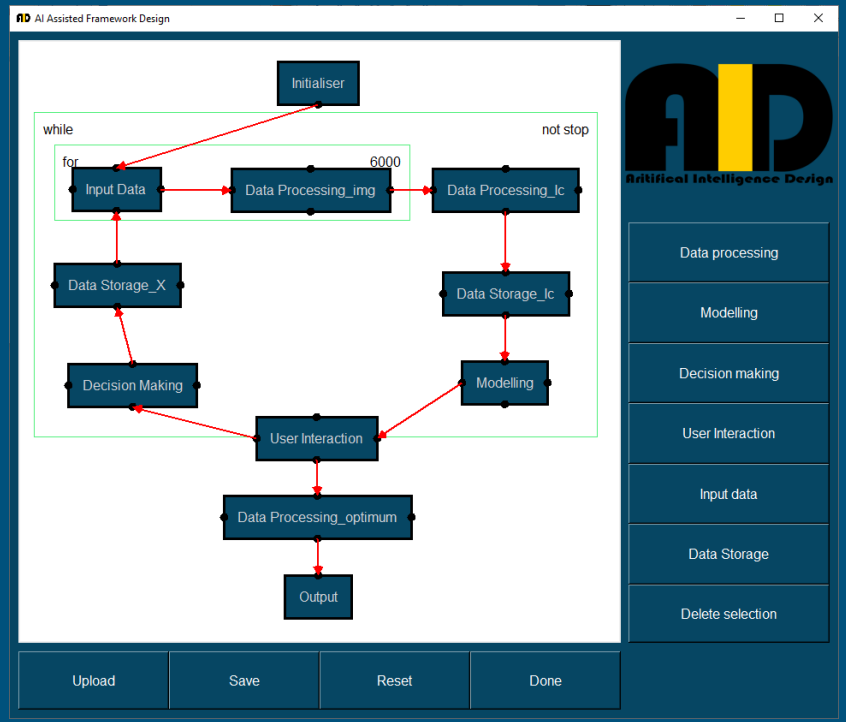
Sun, S., Tiihonen, A., Oviedo, F., Liu, Z., Thapa, J., Zhao, Y., Hartono, N.T.P., Goyal, A., Heumueller, T., Batali, C. and Encinas, A., 2021. A data fusion approach to optimize compositional stability of halide perovskites. Matter, 4(4), pp.1305-1322.

Material design - Use case





Material design - Use case




```
<pipeline>
    <Initialiser name="Initialiser">
        <inputdata>
            <X file="./X_tr.csv" />
            <Y file="./Y_tr.csv" />
            <X_test file="./X_tst.csv" />
            <Y_test file="./Y_tst.csv" />
        </inputdata>
    </Initialiser>
    <loop type="while" condition="not stop" name="loop0">
        <loop type="for" condition="6000" name="loop1">
            <relationships ... />
            <InputData name="Input Data">
                <relationships ... />
            </InputData>
            <DataProcessing name="Data Processing_img">
                <plugin type="getRGBval" />
                <relationships ... />
            </DataProcessing>
        </loop>
        <DataProcessing name="Data Processing_Ic">
            <plugin type="calcIc" />
            <relationships ... />
        </DataProcessing>
        <DataStorage name="Data Storage_Ic">
            <relationships ... />
        </DataStorage>
        <Modelling name="Modelling">
            <plugin type="GPR" />
            <relationships ... />
        </Modelling>
        <UserInteraction name="User Interaction">
            <plugin type="expertConvergenceDecision" />
            <relationships ... />
        </UserInteraction>
        <DecisionMaking name="Decision Making">
            <plugin type="BayesOpt" />
            <relationships ... />
        </DecisionMaking>
        <DataStorage name="Data Storage_X">
            <relationships ... />
        </DataStorage>
    </loop>
    <DataProcessing name="Data Processing_optimum">
        <plugin type="argmin" />
        <relationships ... />
    </DataProcessing>
    <Output name="Output" />
</pipeline>
```

Get involved!

Our aim is to build a community to adopt and develop this framework.

Find more information on the project at:

https://aaltopml.github.io/VAI-Lab/

Contact via email: carlos.sevillasalcedo@aalto.fi

Contents

1. Motivation: Virtual laboratories
2. How to engage with the domain expert
3. Simplified setting: knowledge elicitation
4. But: Scientist is not just a data source
5. General setting: AI-assisted decisions, design, modelling
6. What is still needed?

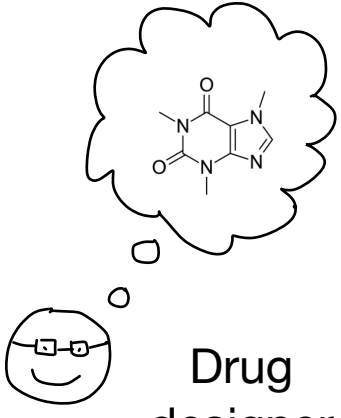
Approach - or dream

"Sidekick" intelligence: AI to help users achieve their goals
even when desired outcomes are tacit, uncertain or evolving

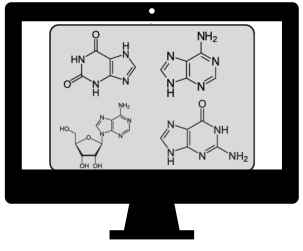
In short, needs to simultaneously model problem and expert

New types of AI assistants

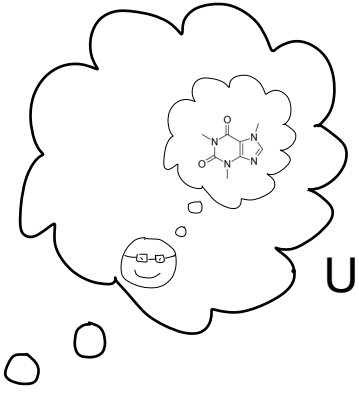
Drug design example



Drug designer



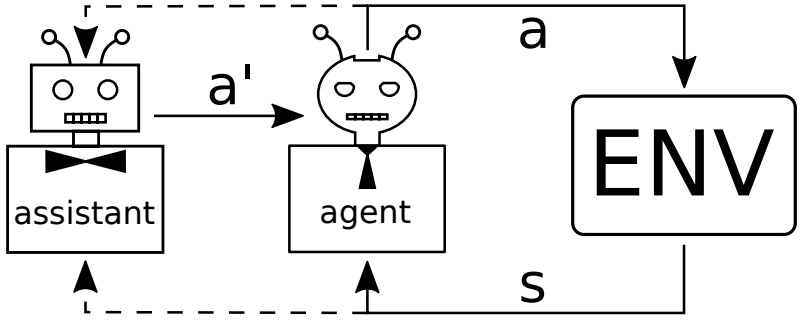
AI tool



User model

AI problem: Formulate an agent capable of learning to assist another agent

Setup:



For AI: essentially a delayed-reward decision problem;
for starters: in a simple game setting

Fig: Sebastiaan De Peuter

Contents

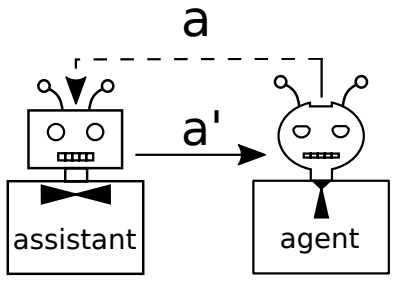
1. Motivation: Virtual laboratories
2. How to engage with the domain expert
3. Simplified setting: knowledge elicitation
4. But: Scientist is not just a data source
5. General setting: AI-assisted decisions, design, modelling
6. What is still needed?

Simplified setting: Knowledge elicitation

No external environment yet

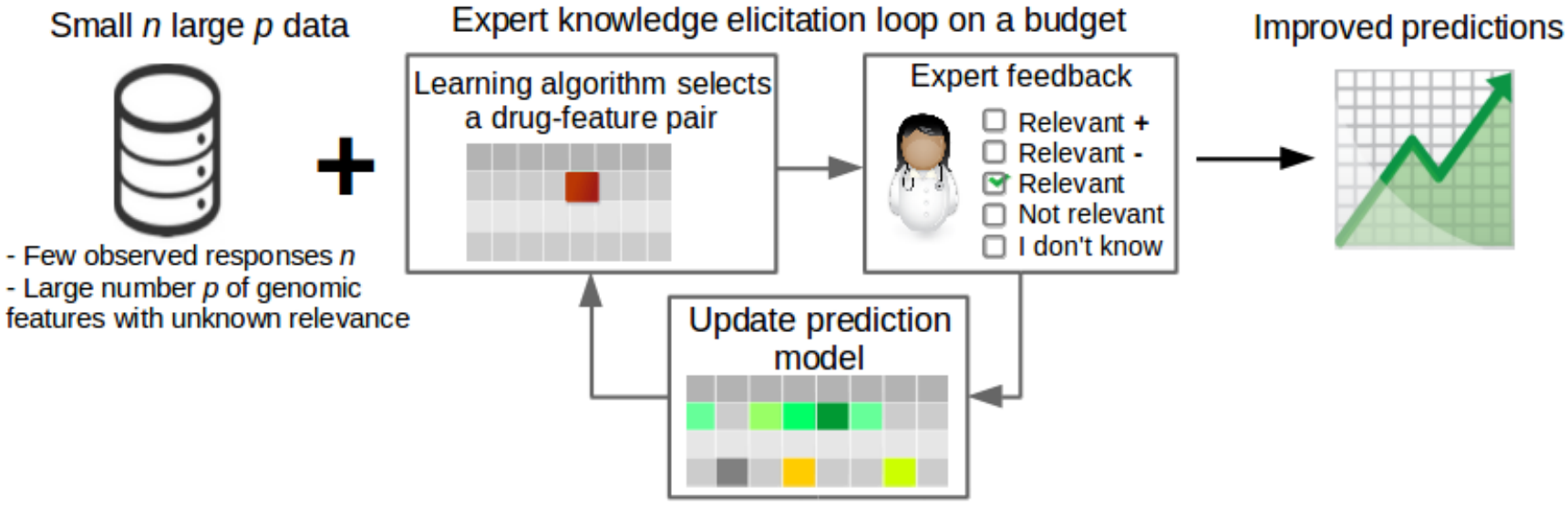
Assistant queries with a': suggestion, question, any other change such as displaying information

Agent responds with a: Answer to question, or other available action



For context: Prior knowledge elicitation: The past, present, and future. Mikkola et al., arXiv:2112.01380

Interactive relevance elicitation for prediction



Interactive system brings an expert to the loop

Sundin et al., Bioinformatics 2018

Task: prediction for high-dim data

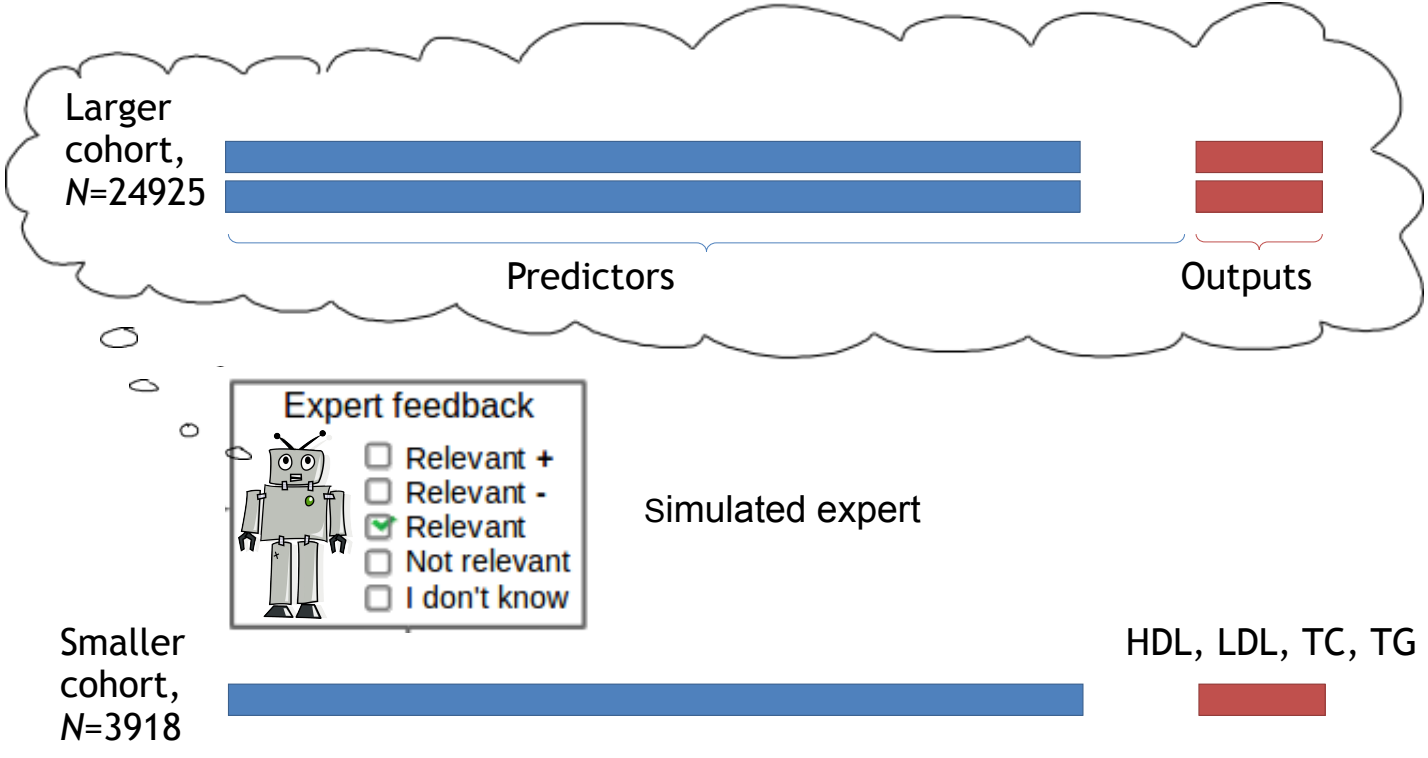
Predictors ***x*** Outputs ***y***

Precision medicine:	Molecular measurements and other data	Efficacy of treatments

Problem: too little data for estimating the predictor ("small *n*, large *p*")

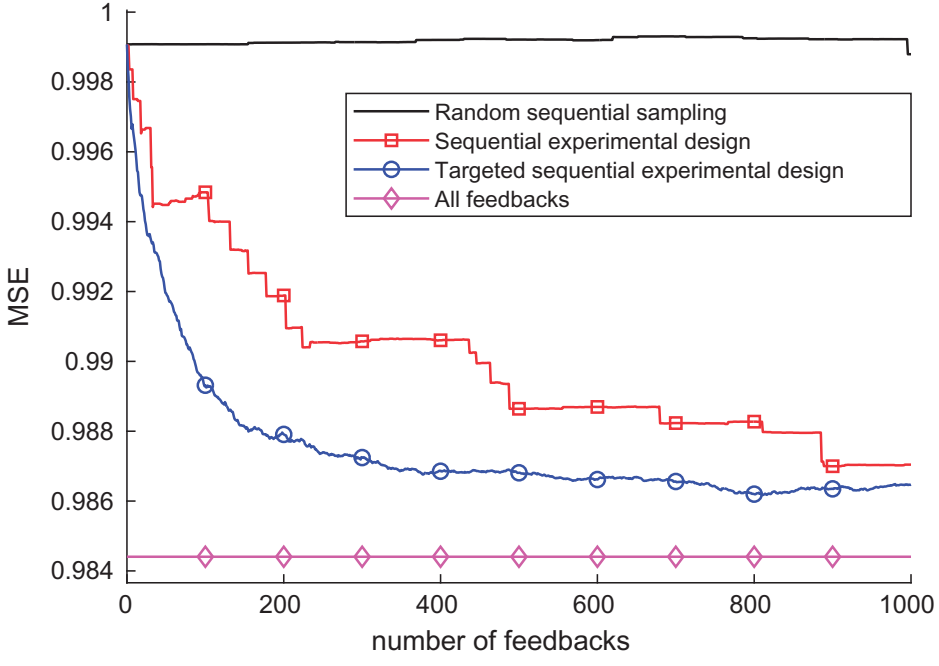
Sundin et al., Bioinformatics 2018

Case: Predict cholesterol levels



Sundin et al., Bioinformatics 2018

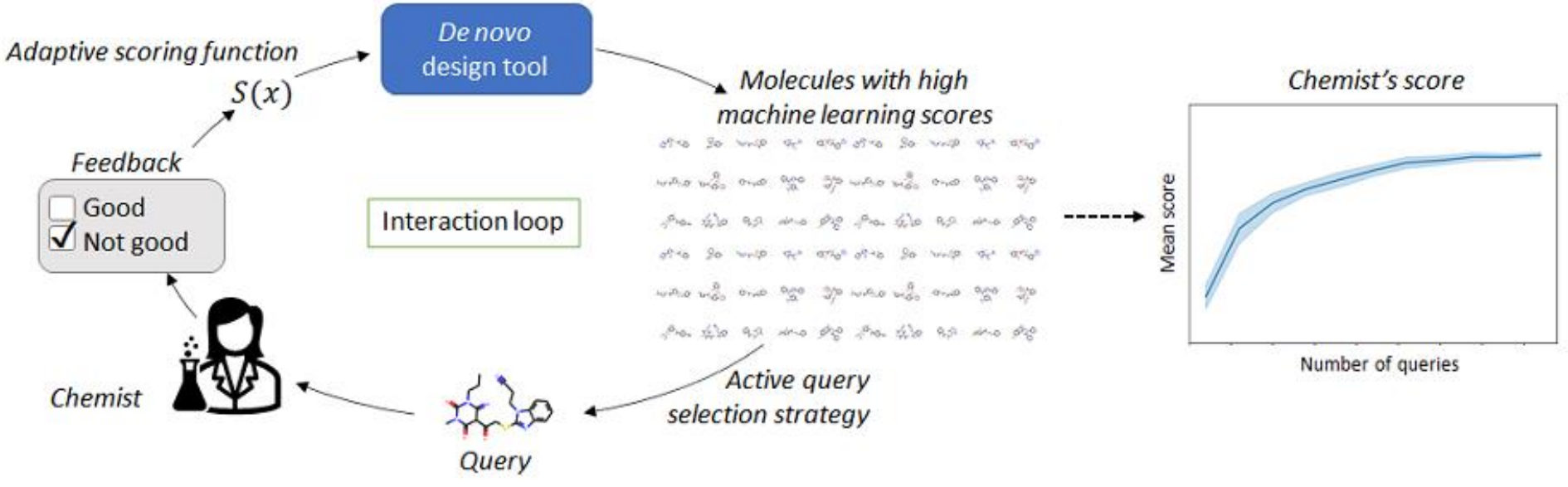
Case: Predict cholesterol levels



Sundin et al., Bioinformatics 2018

Drug Design

Active learning of feedback for a reinforcement learning engine



Sundin et al., Human-in-the-loop assisted *de novo* molecular design. *J. Cheminform.* 2022

What if the expert does not know?

Is human-in-the-loop ML a good idea when we do not know how much the particular human knows?

We can give guarantees that it is, for optimization, formulated as multi-fidelity Bayesian optimization robust towards adding unreliable information sources

---

Multi-Fidelity Bayesian Optimization with Unreliable Information Sources

---

Petrus Mikkola
Aalto University

Julien Martinelli
Aalto University

Louis Filstroff
ENSAI, CREST

Samuel Kaski
Aalto University
University of Manchester

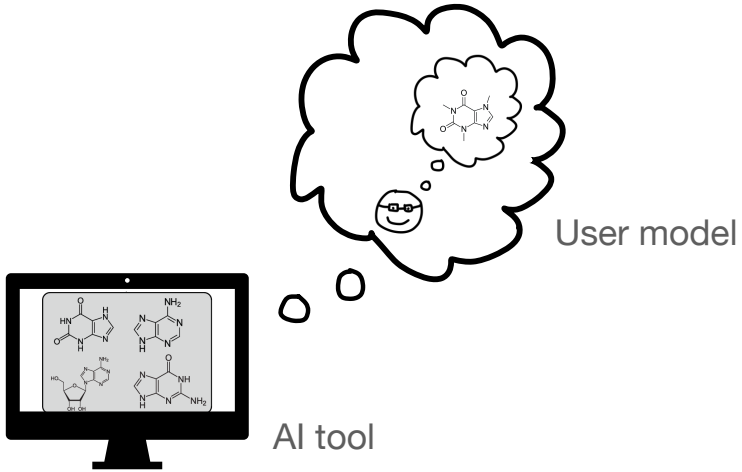
AISTATS 2023

Contents

1. Motivation: Virtual laboratories
2. How to engage with the domain expert
3. Simplified setting: knowledge elicitation
4. But: Scientist is not just a data source
5. General setting: AI-assisted decisions, design, modelling
6. What is still needed?

User Modelling

We need to combine prior knowledge from cognitive science with learning from data



COGNITIVE SCIENCE
A Multidisciplinary Journal

Parameter Inference for Computational Cognitive Models With Approximate Bayesian Computation

Antti Kangasrääsiö,[a] Jussi P. P. Jokinen,[b] Antti Oulasvirta,[b] Andrew Howes,[c] Samuel Kaski[a]

Useful formulation: Computational Rationality

Assumption: humans act rationally (+noise) *given their constraints and limitations*

Some of the constraints we get from cognitive science, some from the task definition. The rest needs to be learned from data.

Brute-force solution: Simulator-based inference with RL in the inner loop

Contents

1. Motivation: Virtual laboratories
2. How to engage with the domain expert
3. Simplified setting: knowledge elicitation
4. But: Scientist is not just a data source
5. General setting: AI-assisted decisions, design, modelling
6. What is still needed?

General setting: AI-assisted decision-making and design

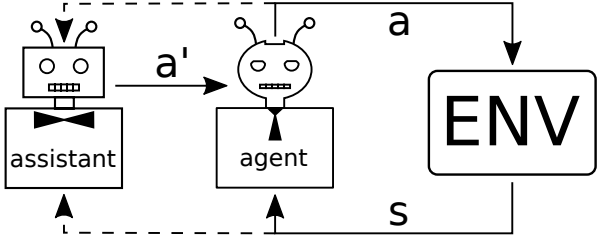
Decision-making

- a: which treatment
- a': suggestion + predicted effectiveness

Design

- a: design change
- a': suggestion + predicted properties of design

Utility/reward becomes only known further down the line

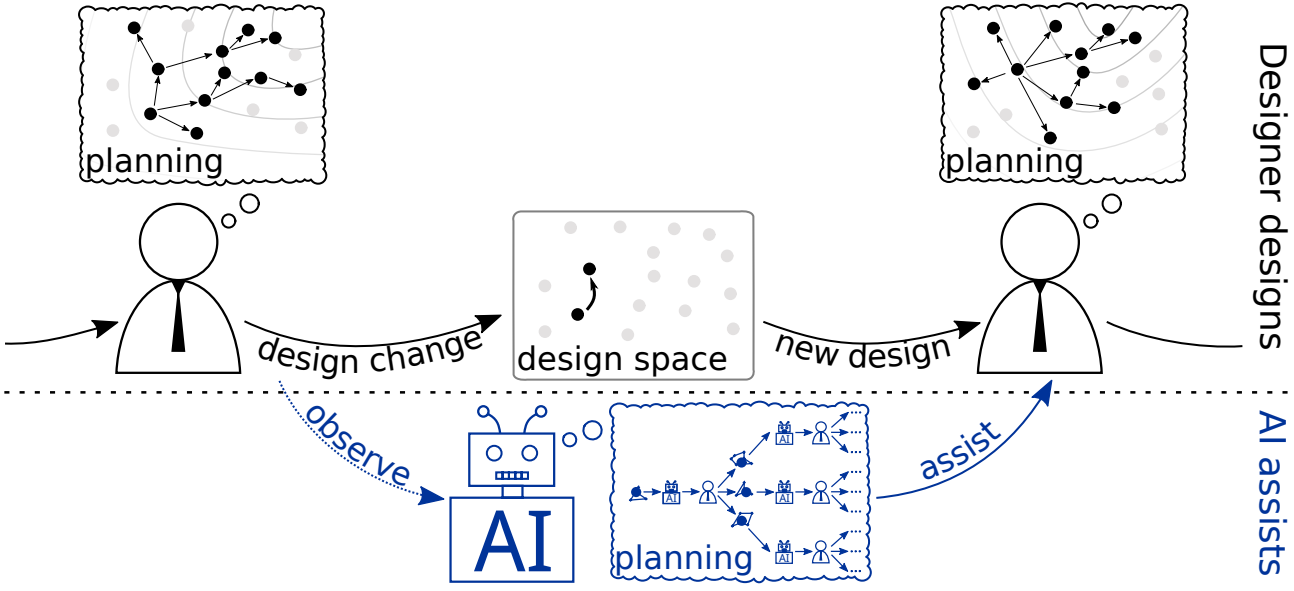


Zero-shot assistance

How to help an agent solve a sequential delayed-reward decision making task, when we know nothing about the reward function?

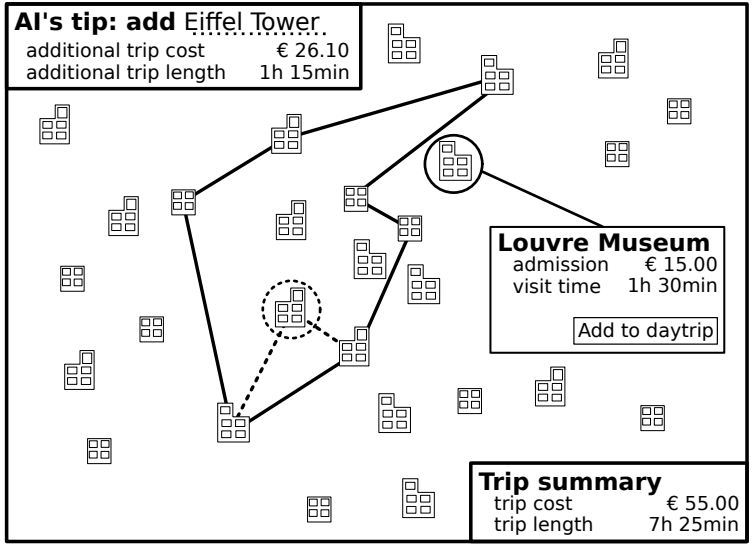
- agent cannot specify or communicate the reward
- agent has not solved it before, so no previous behavior to learn from

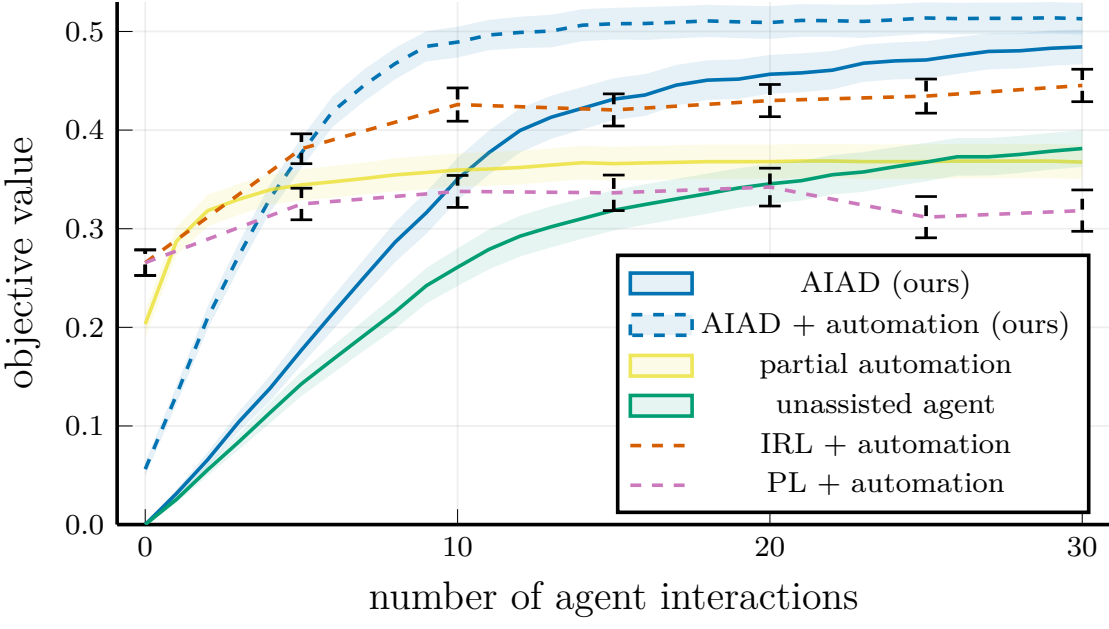
AI-assisted design



De Peuter, Oulasvirta, Kaski, AI Magazine 2023

AI-assisted design: Trip planning



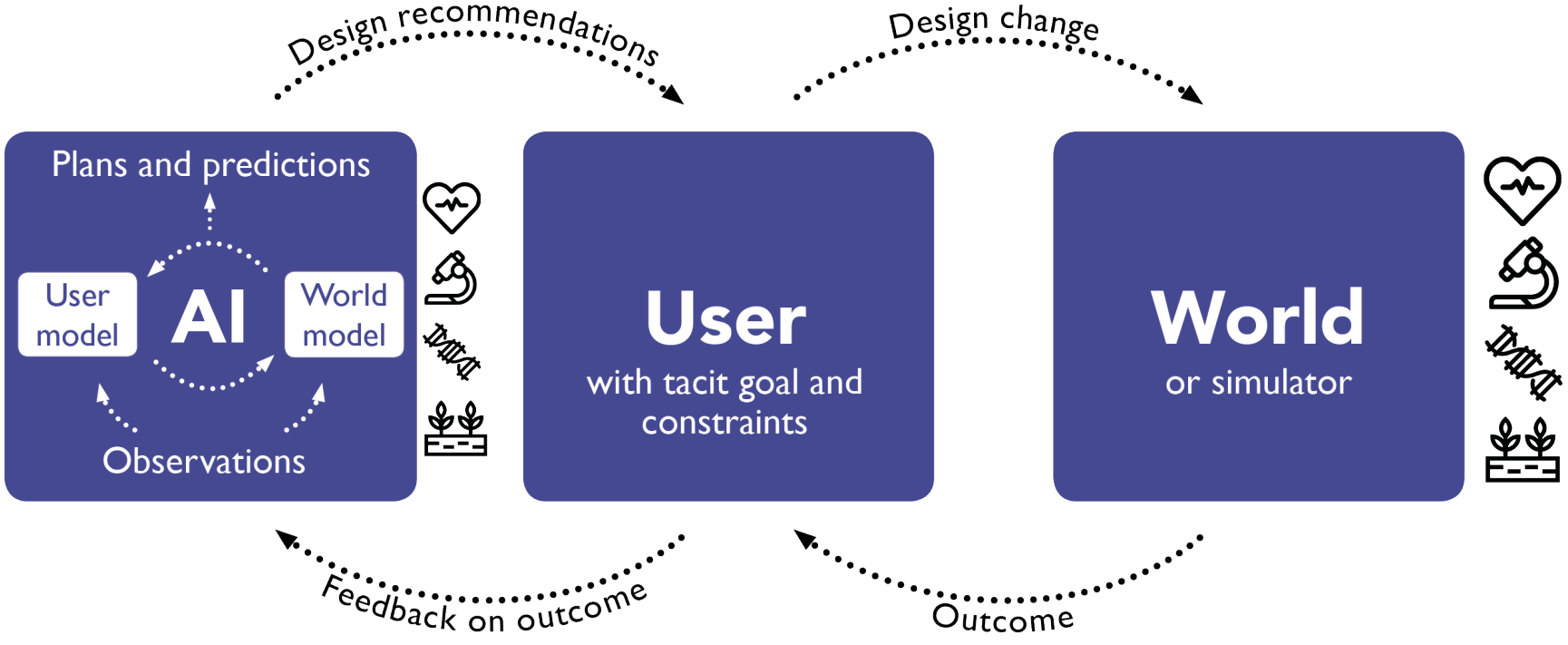


De Peuter, Kaski, AAAI 2023

Contents

1. Motivation: Virtual laboratories
2. How to engage with the domain expert
3. Simplified setting: knowledge elicitation
4. But: Scientist is not just a data source
5. General setting: AI-assisted decisions, design, modelling
6. What is still needed?

Collaborative decision making with AI



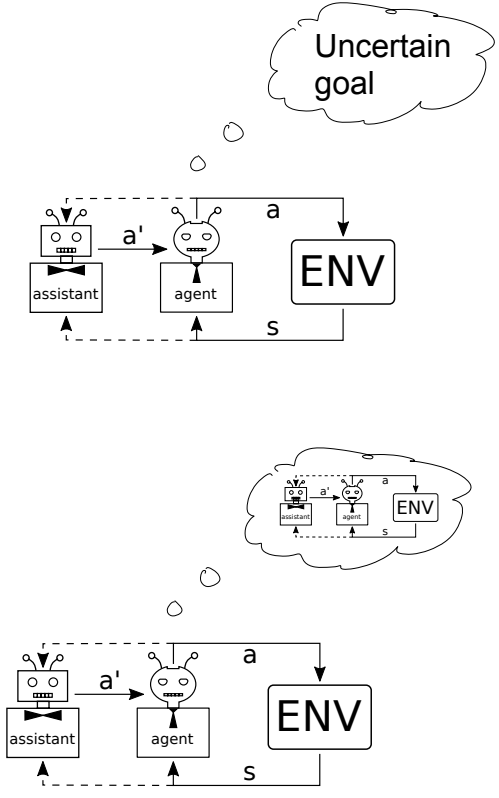
AI-assisted decision-making, design and modelling

What is needed of user modelling?

Generative model capable of producing:

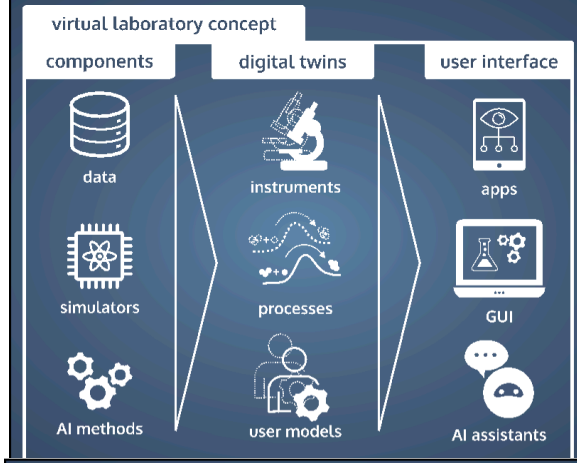
1. Goal-driven behaviour with constraints:
   Computational rationality

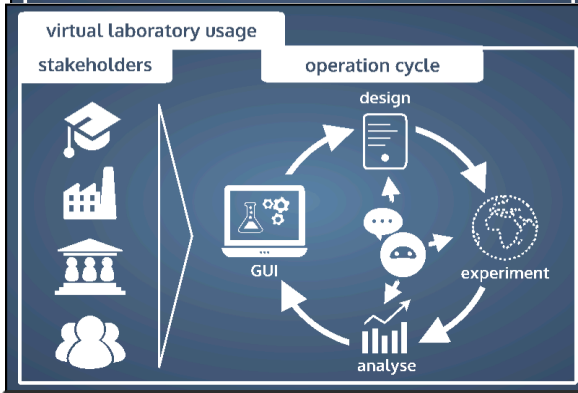
2. In an environment containing other goal-oriented
   agents, having representations of other agents:
   Theory of Mind

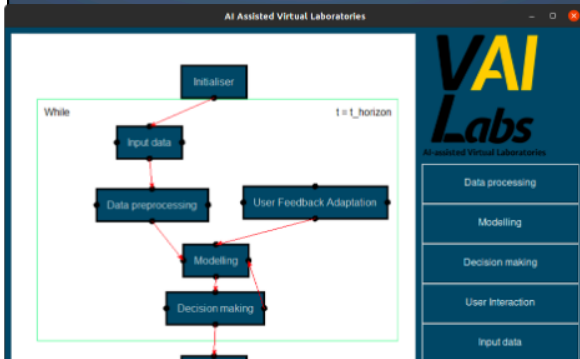


With these tools, close to AI-assisted Virtual Laboratories for Science and Engineering

- Concept: Klami et al. **Virtual Laboratories: Transforming research with AI**. https://doi.org/10.36227/techrxiv.20412540.v1
- Software: https://github.com/AaltoPML/VAI-Lab
- First labs under way, in collaboration of FCAI, Turing, a few companies: material science, drug design, sustainble mobility
- Open community - welcome to join!







Where?

ellis

FCAI

Thank you: To a number of group members and collaborators, who are authors of the mentioned papers