

Hybrid AI for Trustworthy AI

Fredrik Heintz

Dept. of Computer Science, Linköping University

fredrik.heintz@liu.se

@FredrikHeintz



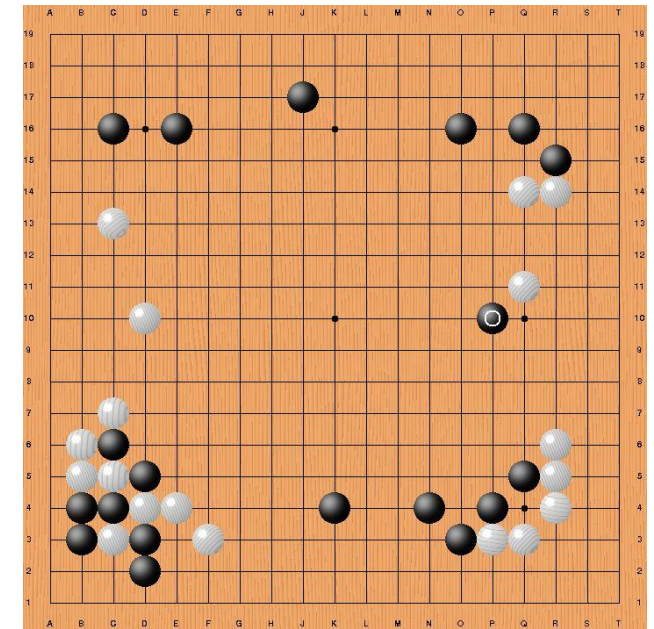
How to Evaluate AI Systems?



George Zarkadakis, Contributor
AI engineer and writer

Move 37, or how AI can change the world

11/26/2016 09:35 am ET



Ethics Guidelines for Trustworthy AI – Overview

Human-centric approach: AI as a means, not an end

Trustworthy AI as our foundational ambition, with three components

Lawful AI

Ethical AI

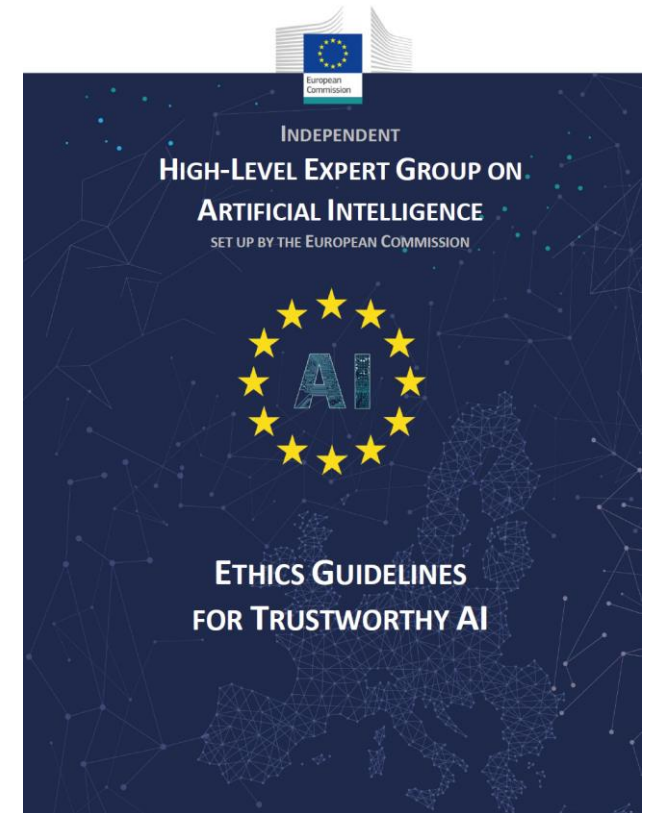
Robust AI

Three levels of abstraction

from principles
(Chapter I)

to requirements
(Chapter II)

to assessment
list (Chapter III)



Ethics Guidelines for Trustworthy AI – Principles

4 Ethical Principles based on fundamental rights



Respect for
human
autonomy

Augment, complement
and empower humans



Prevention of
harm

Safe and secure.
Protect physical and
mental integrity.



Fairness

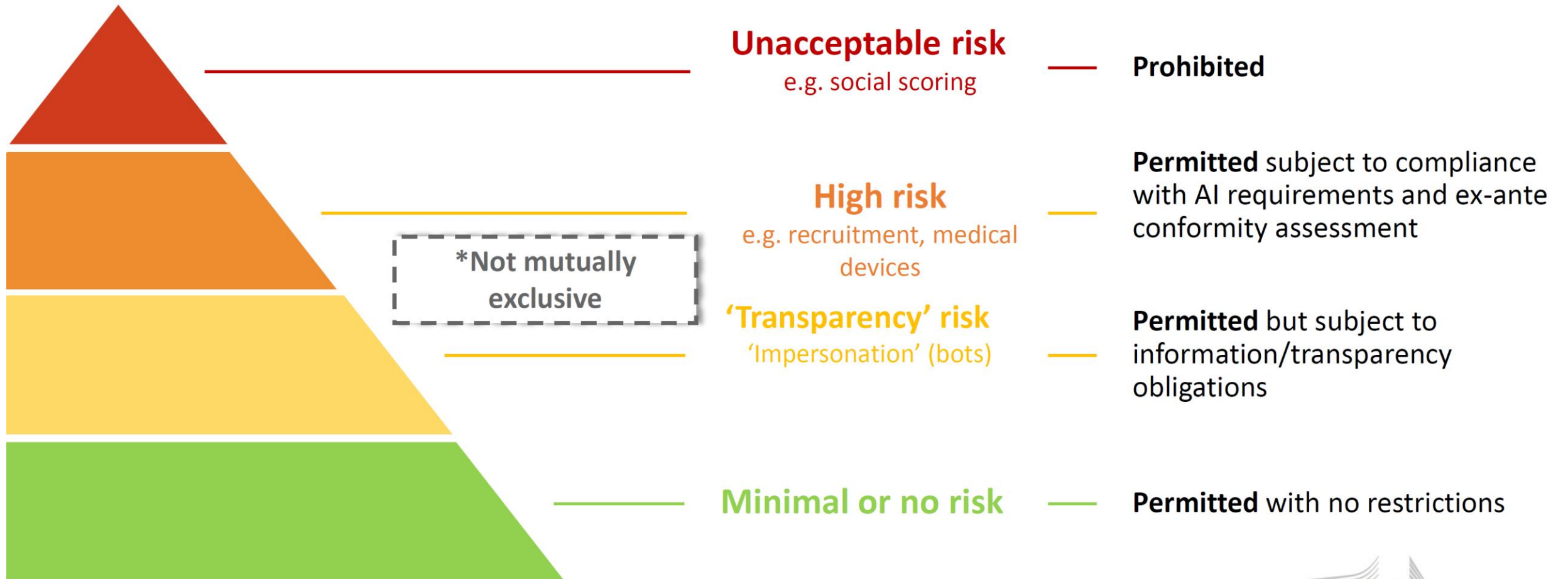
Equal and just
distribution of
benefits and costs.



Explicability

Transparent, open
with capabilities and
purposes, explanations

A risk-based approach



Requirements for high-risk AI systems (Title III, Chapter 2)



Establish and
implement **risk
management
system**
&
in light of the
**intended
purpose** of the
AI system

Use high-quality **training, validation and testing data** (relevant, representative etc.)

Draw up **technical documentation** & set up **logging capabilities** (traceability & auditability)

Ensure appropriate degree of **transparency** and provide users with **information** on capabilities and limitations of the system & how to use it

Ensure **human oversight** (measures built into the system and/or to be implemented by users)

Ensure **robustness, accuracy** and **cybersecurity**

TAILOR

Foundation of Trustworthy AI:
Integrating Learning, Optimisation and Reasoning



Fredrik Heintz

Dept. of Computer Science, Linköping University
fredrik.heintz@liu.se, @FredrikHeintz

TAILOR – Vision

Develop the scientific foundations for **Trustworthy AI** integrating learning, optimisation and reasoning realising the European Vision of human-centered trustworthy AI.

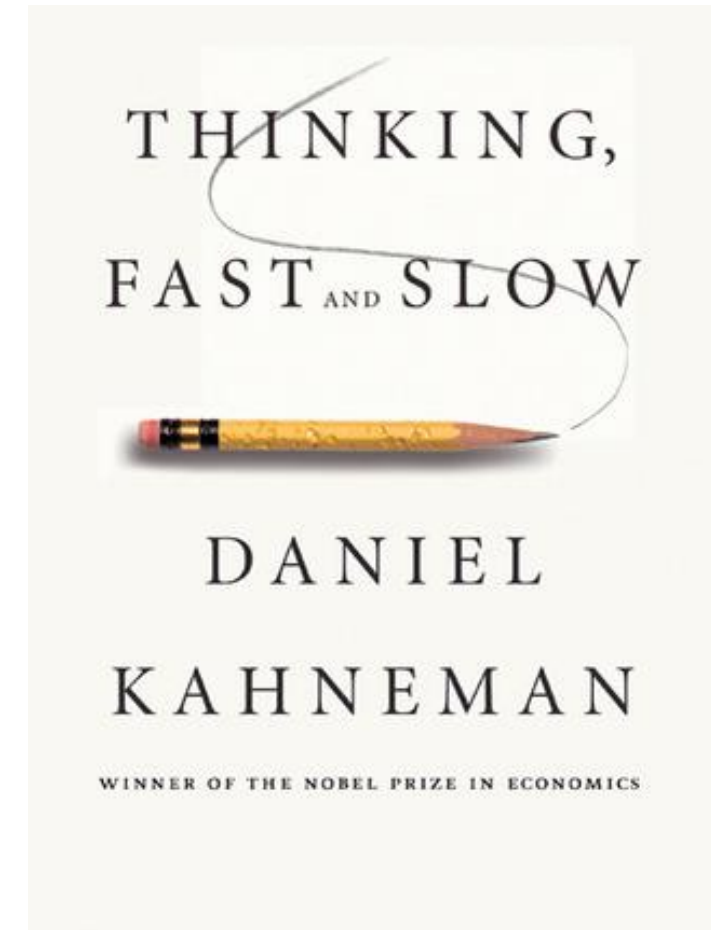
Boosting Capacity to Tackle Major Scientific Challenges

- A **core network** of outstanding AI research centres and major European companies (partners) plus **mechanisms for extending** the network (network members and connectivity fund) to be adaptive and inclusive.
- Five **virtual research environments** to address the **major scientific challenges** required to achieve Trustworthy AI supported by **AI-based network collaboration tools**.
- **Strategic** research and innovation **roadmap** to drive the long-term **scientific vision** combined with **bottom-up coordinated actions** collaboratively addressing specific research questions.

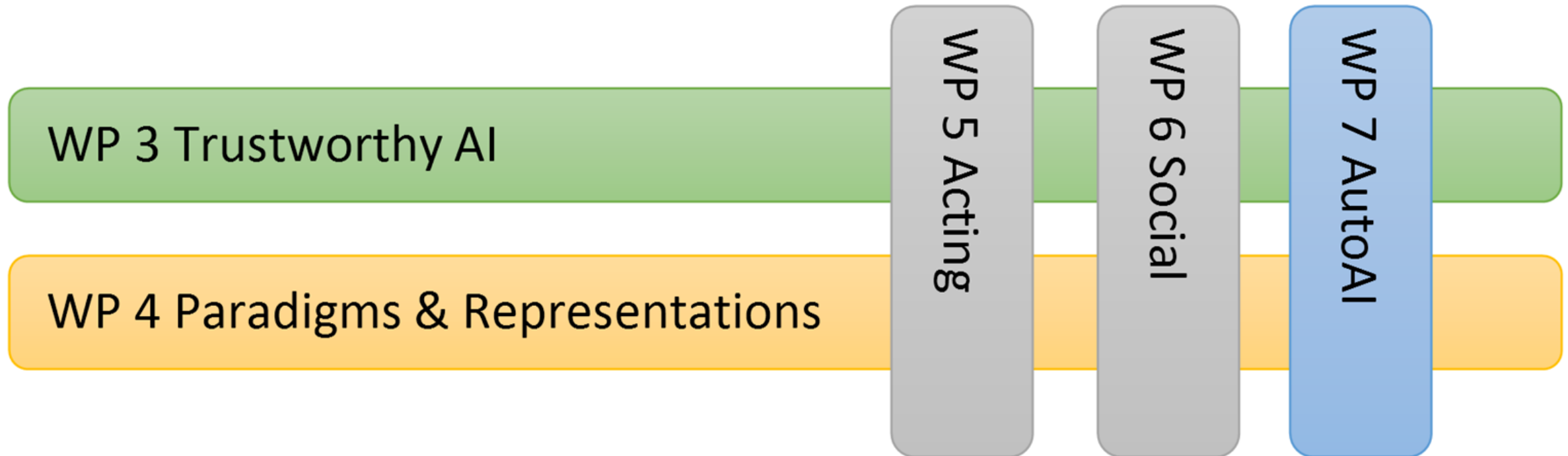
Human and Computational Thinking

Figure 1: A Comparison of System 1 and System 2 Thinking

System 1 "Fast"	System 2 "Slow"
DEFINING CHARACTERISTICS Unconscious Effortless Automatic	DEFINING CHARACTERISTICS Deliberate and conscious Effortful Controlled mental process
WITHOUT self-awareness or control "What you see is all there is."	WITH self-awareness or control Logical and skeptical
ROLE Assesses the situation Delivers updates	ROLE Seeks new/missing information Makes decisions

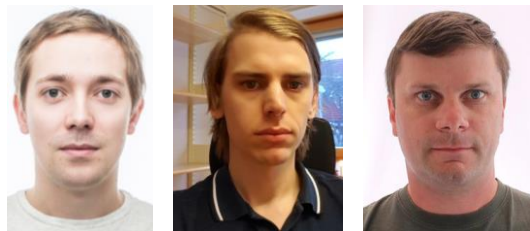


TAILOR – Basic Research Program



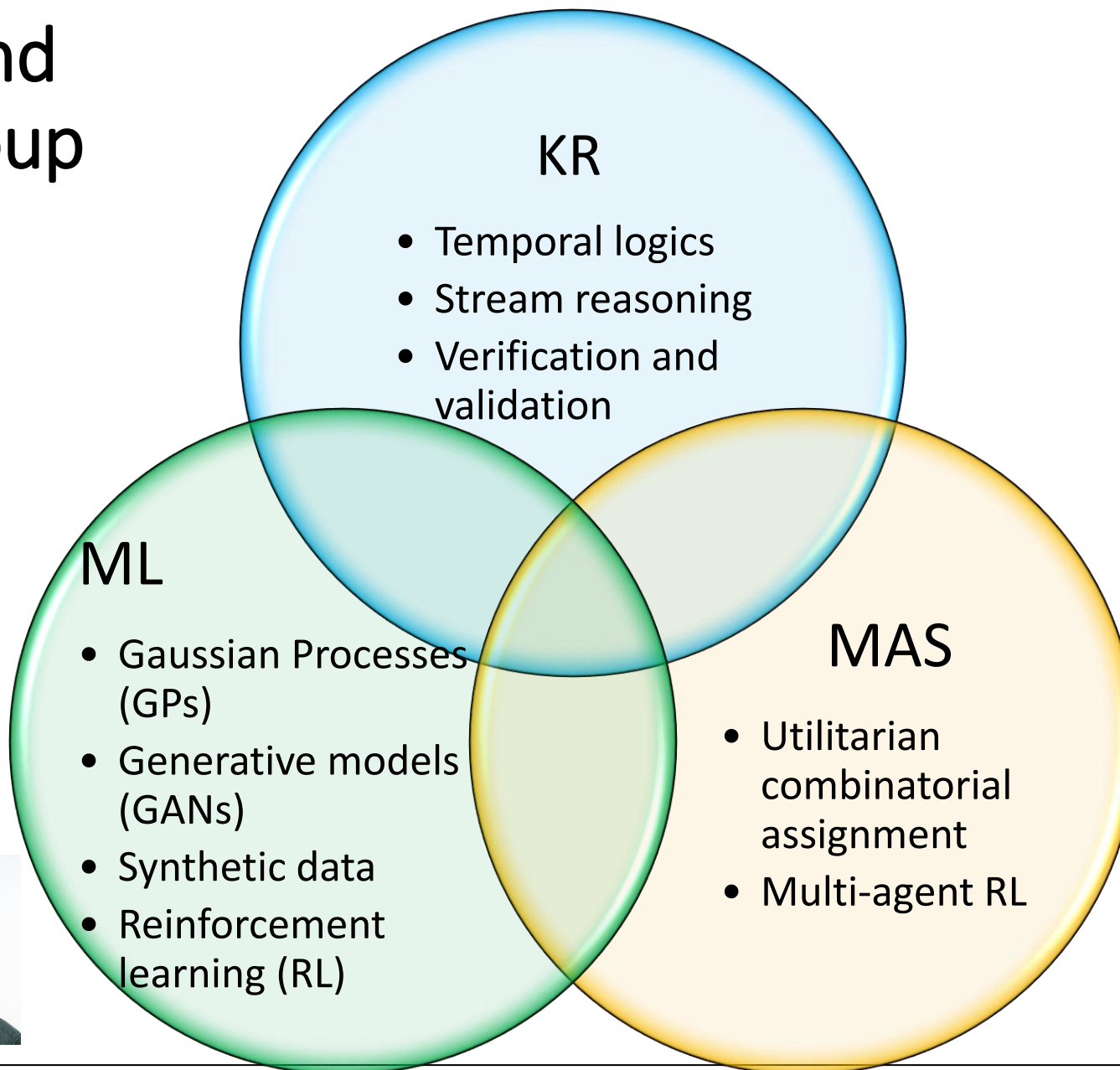
Reasoning and Learning Group

6 PhD students



2 Postdocs

1 Research engineer



*Knut and Alice
Wallenberg
Foundation*

WASP

STIFTELSEN
MARCUS OCH AMALIA
WALLENBERGS
MINNESFOND

VINNOVA
NFFP / UDI

 **ELLIIT**



TAILOR

WASP-HS

Research Overview



Learning generative models based on trajectory data



Probabilistic logical reasoning over observed and predicted trajectories



Utilitarian Combinatorial Assignment

Collaborative Unmanned Aircraft Systems

A principled approach to building collaborative intelligent autonomous systems for complex missions.



Motion Pattern Recognition

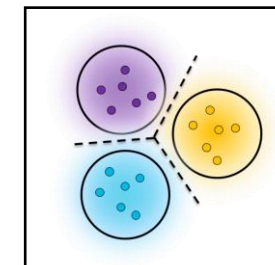
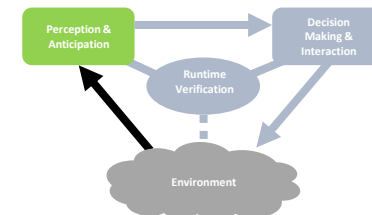
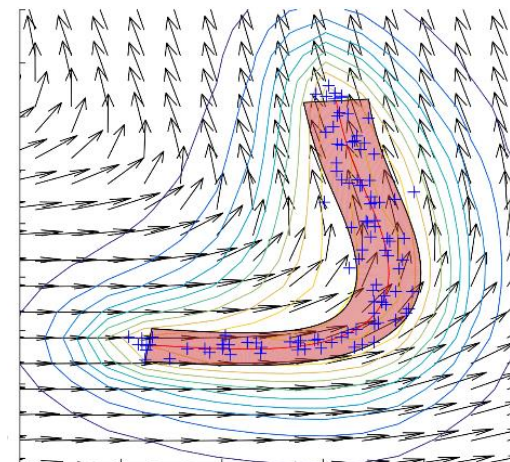
Generalized Motion Pattern Model

- Based on Gaussian processes
- Generative auto-encoder

Tracker

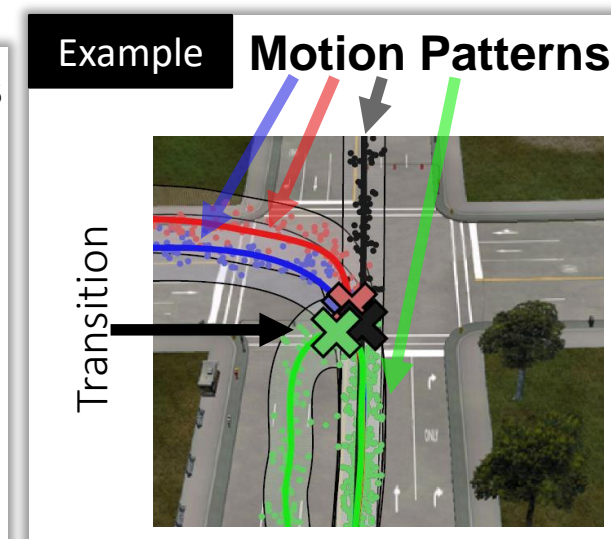
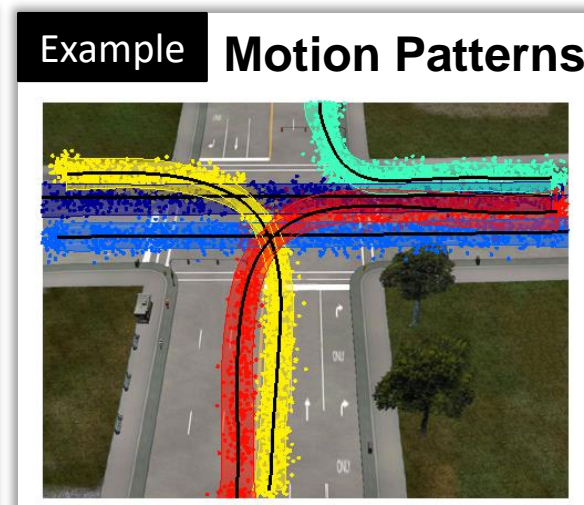
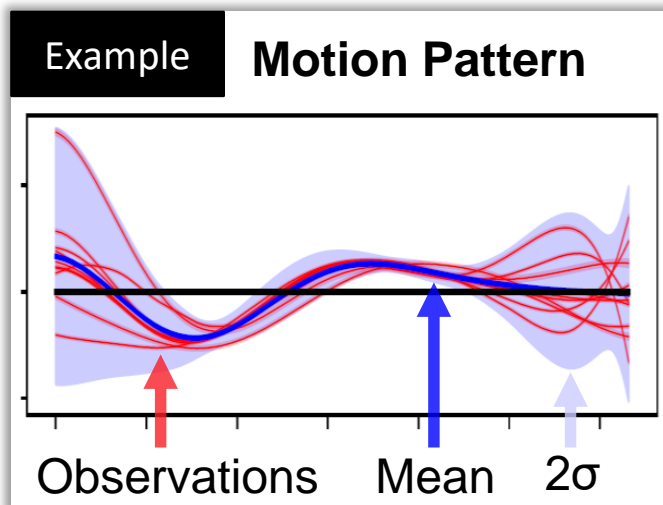
Trajectories

Motion Pattern



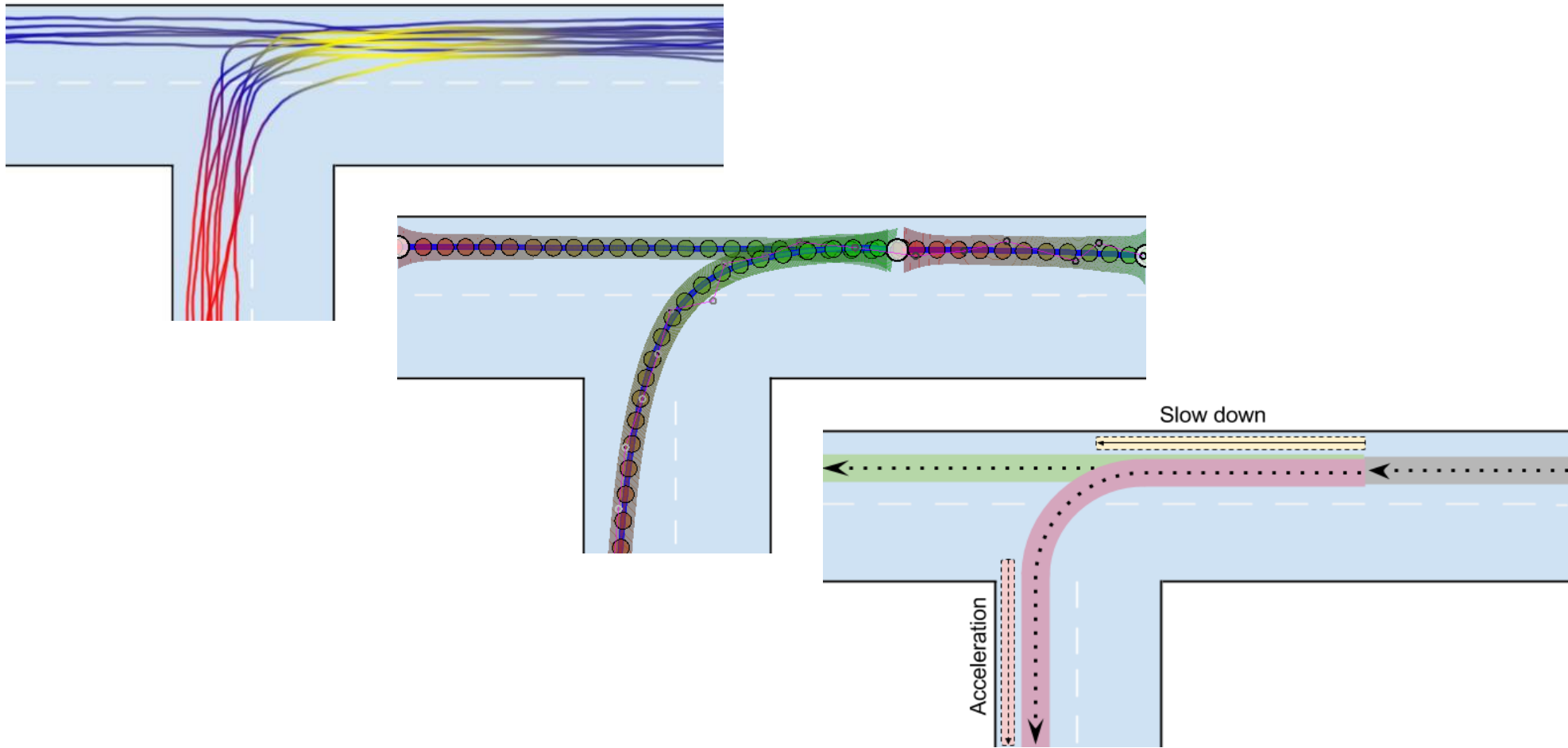
Multi-task

- One-class classification
(anomaly detection)
- Multi-class classification
- Predict continuation
- Predict sequence
- Temporally *align* trajectories



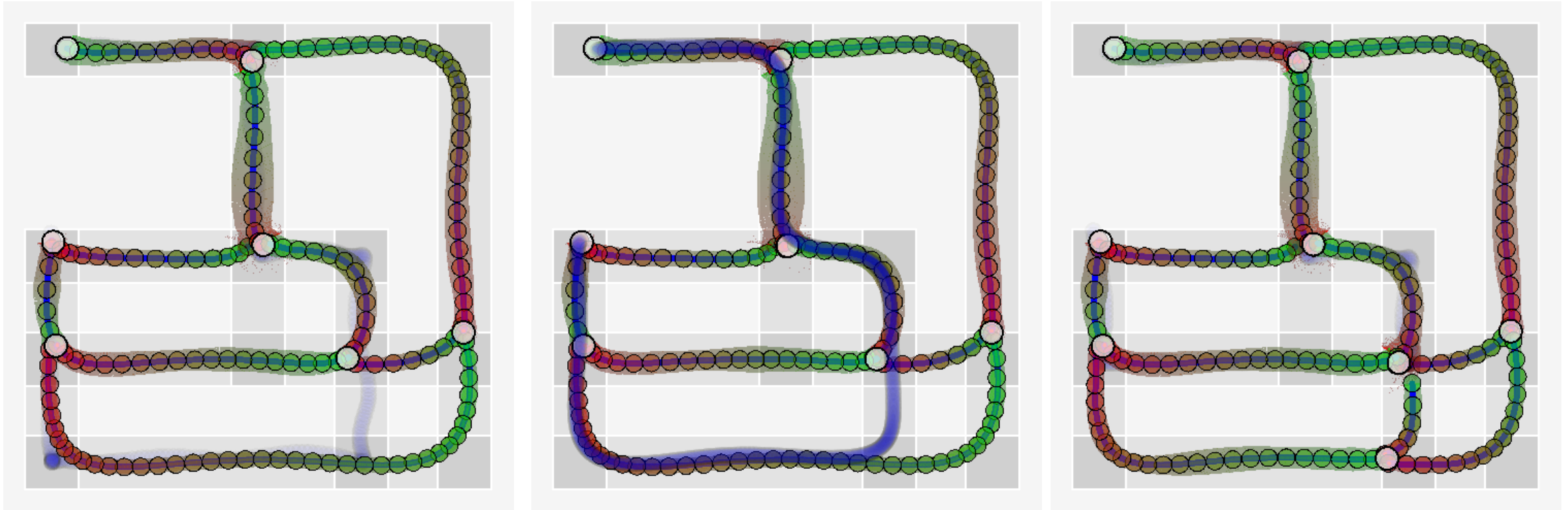
Learning generative models based on trajectory data

[Tiger and Heintz IV 2018, Tiger and Heintz FUSION 2015, Tiger and Heintz STAIRS 2014]



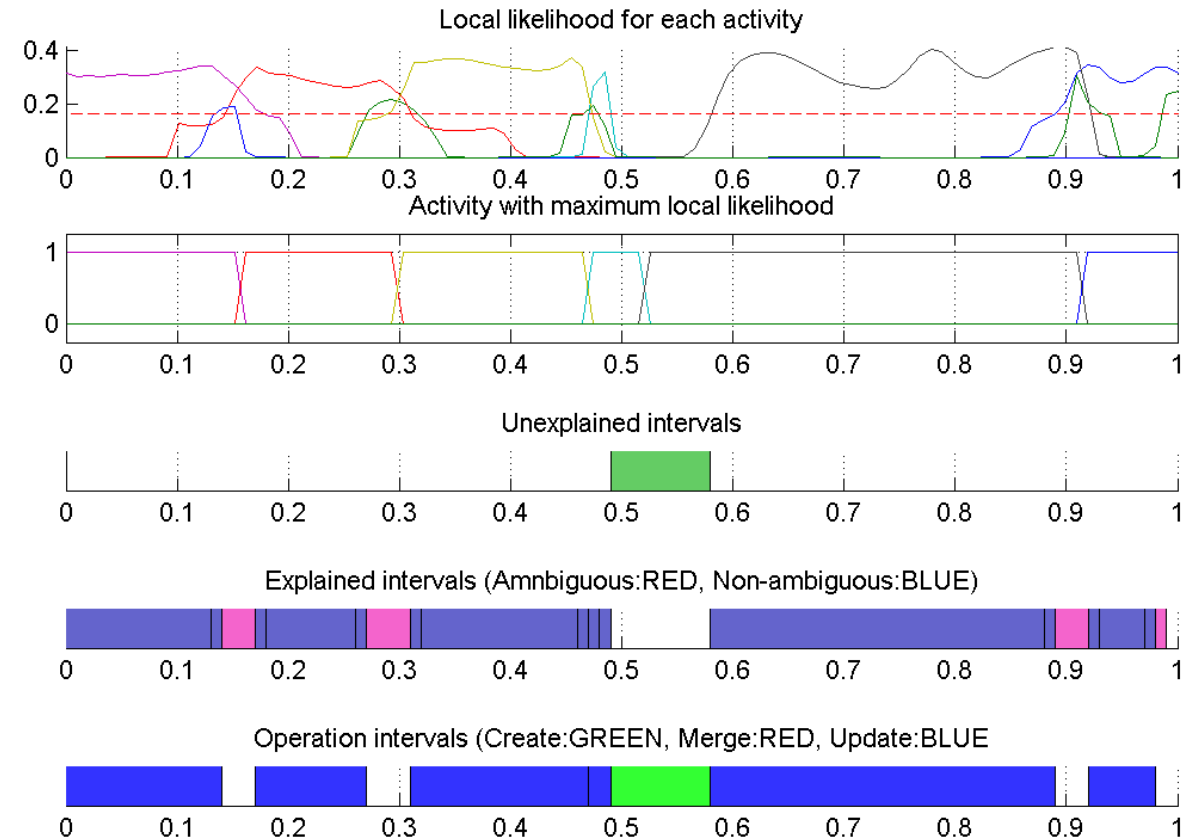
Learning generative models based on trajectory data

[Tiger and Heintz IV 2018, Tiger and Heintz FUSION 2015, Tiger and Heintz STAIRS 2014]



[Tiger and Heintz IV 2018, Tiger and Heintz FUSION 2015, Tiger and Heintz STAIRS 2014]

[Tiger and Heintz IV 2018, Tiger and Heintz FUSION 2015, Tiger and Heintz STAIRS 2014]



Learning generative models based on trajectory data

[Tiger and Heintz IV 2018, Tiger and Heintz FUSION 2015, Tiger and Heintz STAIRS 2014]

State-of-the-art **Flow Field** approach

$$(p_x, p_y) \rightarrow v_x, v_y$$

With two GP modelled latent functions:

$$\begin{bmatrix} v_x & v_y \end{bmatrix} = \begin{bmatrix} f_{v_x}(p_x, p_y) & f_{v_y}(p_x, p_y) \end{bmatrix}$$

Proposed **Inverse Mapping** approach

$$(p_x, p_y) \rightarrow \tau \rightarrow p_x, p_y, v_x, v_y$$

With five GP modelled latent functions:

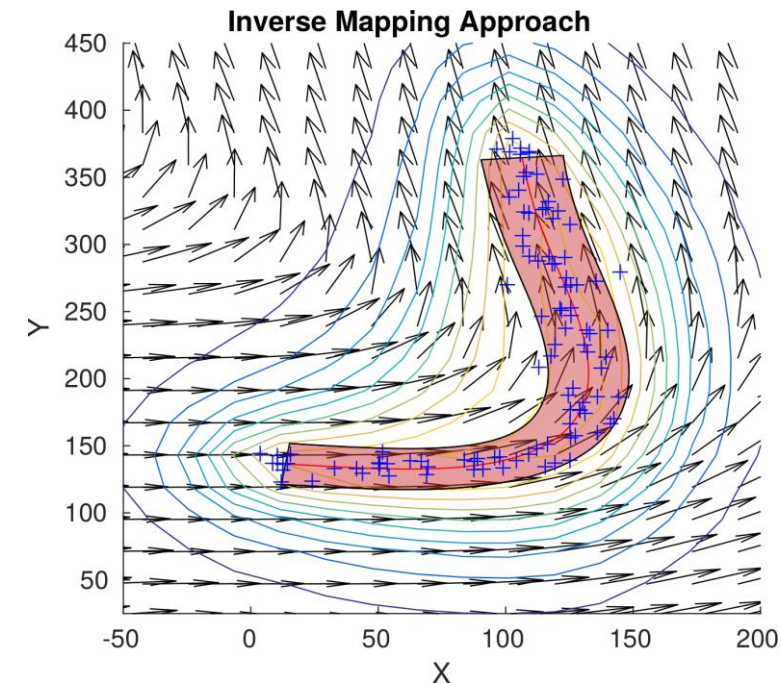
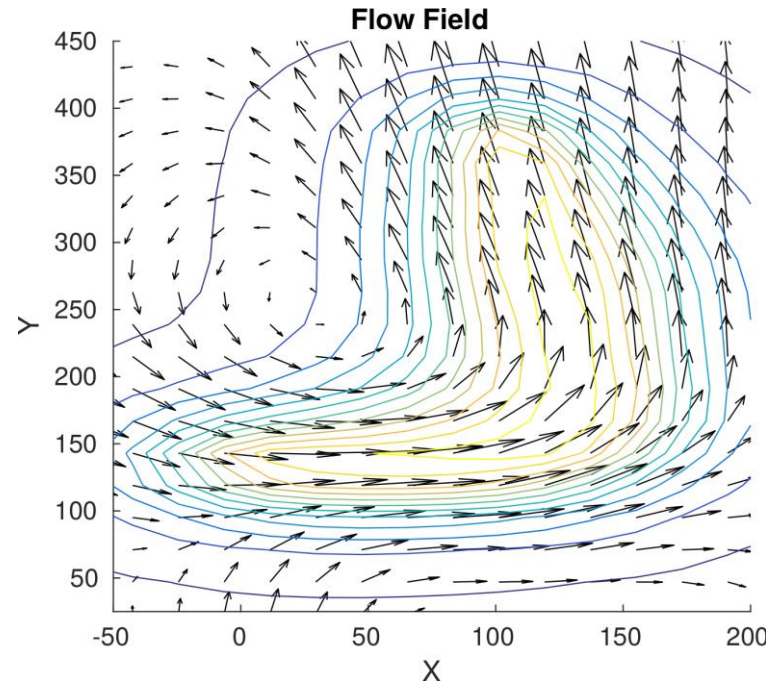
$$\begin{bmatrix} p_x & p_y \end{bmatrix} = \begin{bmatrix} f_{p_x}(\tau) & f_{p_y}(\tau) \end{bmatrix}$$

$$\begin{bmatrix} v_x & v_y \end{bmatrix} = \begin{bmatrix} f_{v_x}(\tau) & f_{v_y}(\tau) \end{bmatrix}$$

$$\tau = f_{\tau}(p_x, p_y)$$

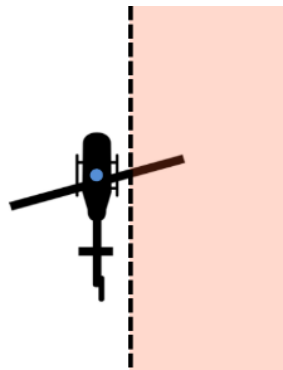
where $\tau \in [0, 1]$ is parametrized time
(motion pattern progression)

Models *flow*, *spatial extent*, *spatial locality* and *motion progression*.

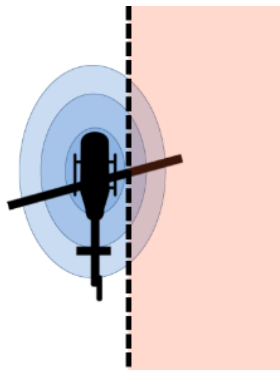


Probabilistic Predictive Stream Reasoning

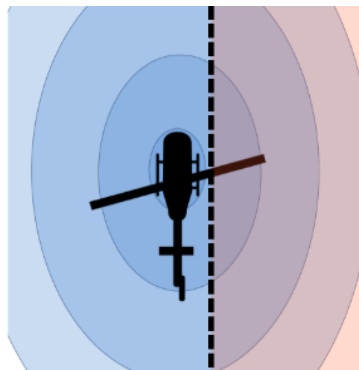
[Tiger and Heintz TIME 2016, IJAR 2020]



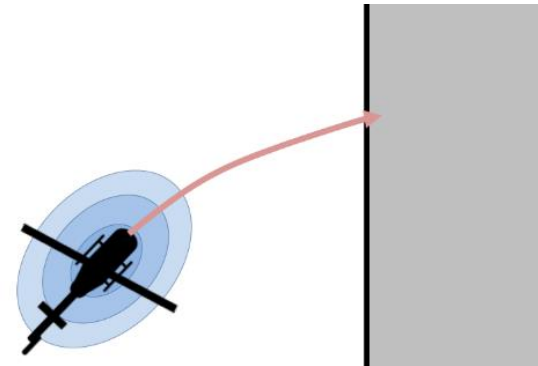
collision: false



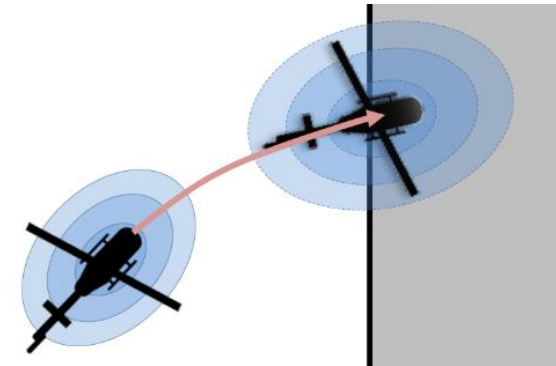
$\Pr(\text{collision}) = 0.1$



$\Pr(\text{collision}) = 0.4$



$\Pr(\text{collision now}) = 0.0\dots$



$\Pr(\text{collision soon}) = 0.5$

Reasoning over Uncertainty

Reasoning over Predictions

Mattias Tiger and Fredrik Heintz. 2020.

Incremental Reasoning in Probabilistic Signal Temporal Logic.

International Journal of Approximate Reasoning, **119**:325–352. Elsevier.

Probabilistic Predictive Stream Reasoning

[Tiger and Heintz TIME 2016, IJAR 2020]

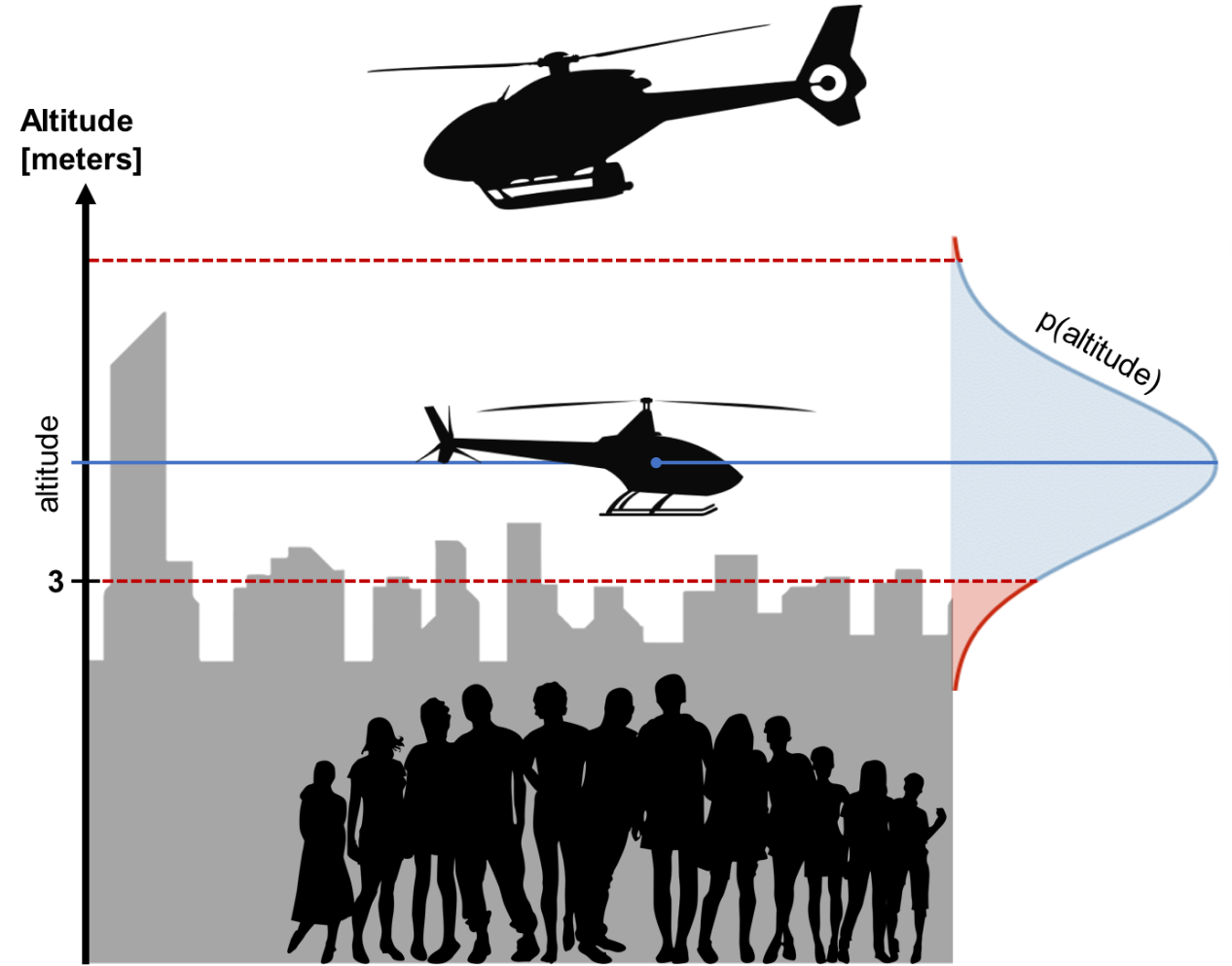
always ($\text{altitude}_0 > 3$)
true

always ($\Pr(\text{altitude}_{0|0} > 3) \geq 0.99$)
false

always ($\Pr(\text{altitude}_{2|0} > 3) \geq 0.99$)

Relative time to estimate

Relative time to estimate from

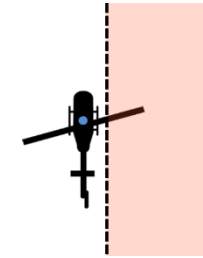


Probabilistic logical reasoning over observed and predicted trajectories

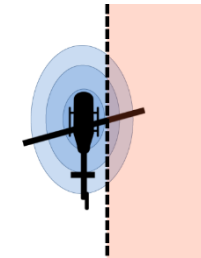
[Tiger and Heintz TIME 2016, IJAR 2020]

- Probabilistic
 - Is the UAV inside the no-fly-zone?
- Anticipatory
 - Will the UAV be colliding in the near future?
- Introspective
 - Is the prediction similar to the realization?

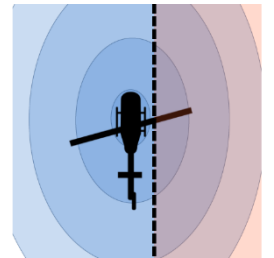
Reasoning over Uncertainty



collision: false

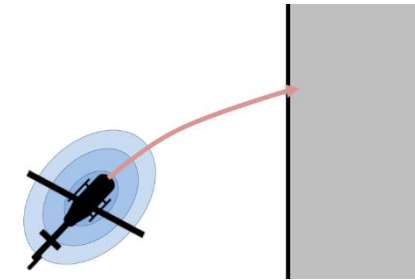


$\text{Pr}(\text{collision}) = 0.1$

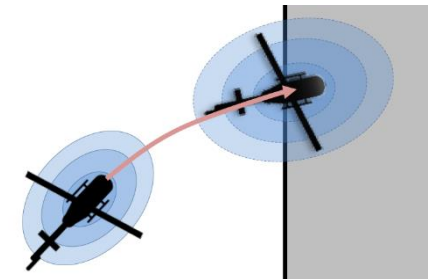


$\text{Pr}(\text{collision}) = 0.4$

Reasoning over Predictions

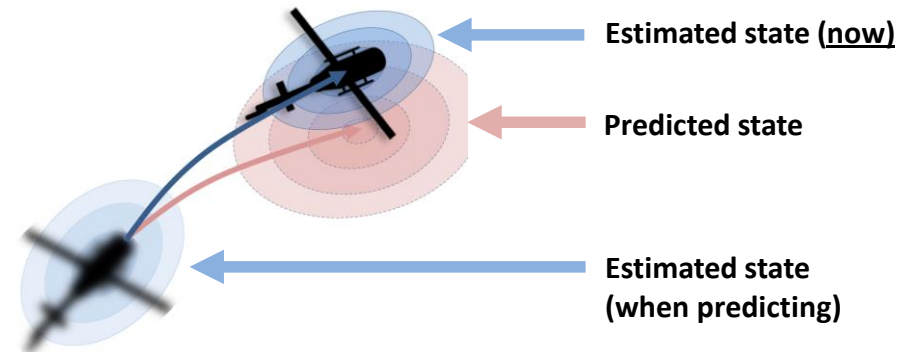


$\text{Pr}(\text{collision now}) = 0.0$



$\text{Pr}(\text{collision soon}) = 0.5$

Reasoning about Predictions

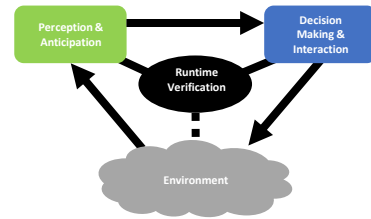
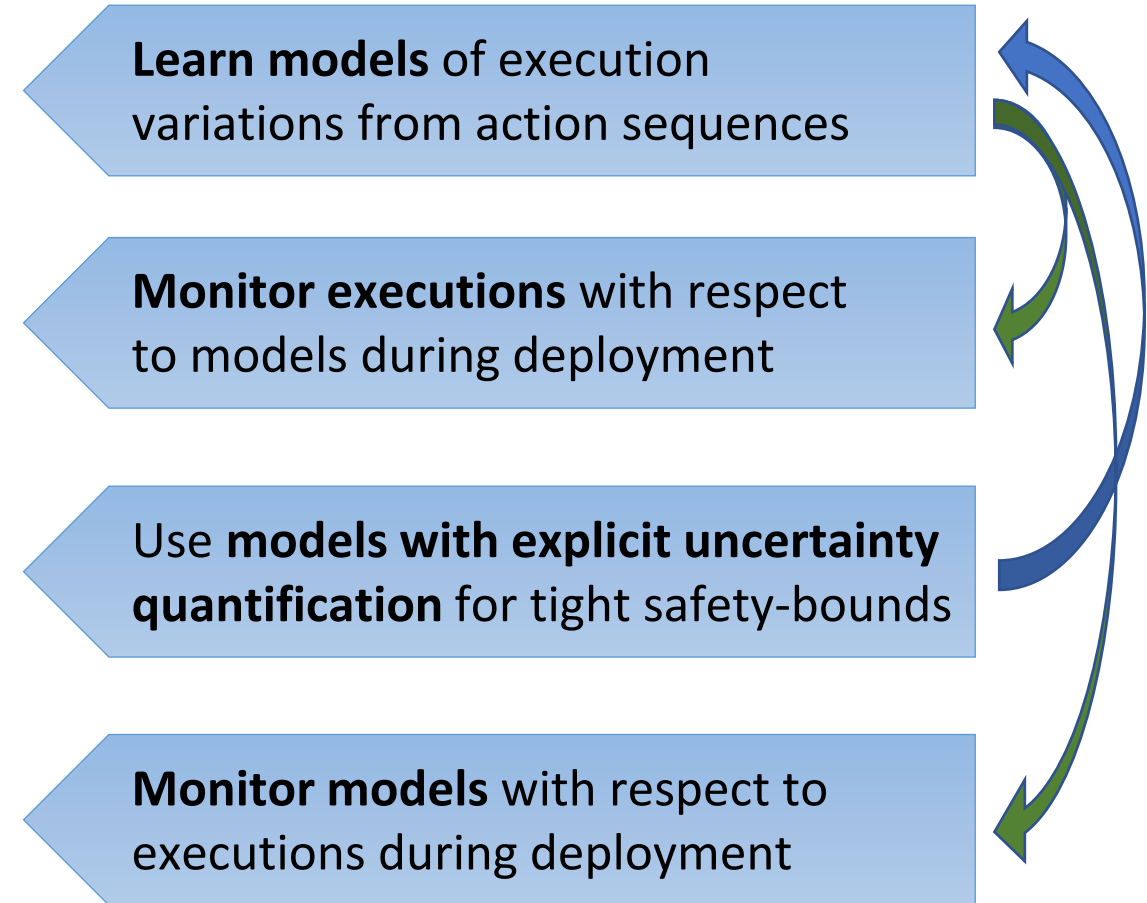


Introspective Motion Planning and Control

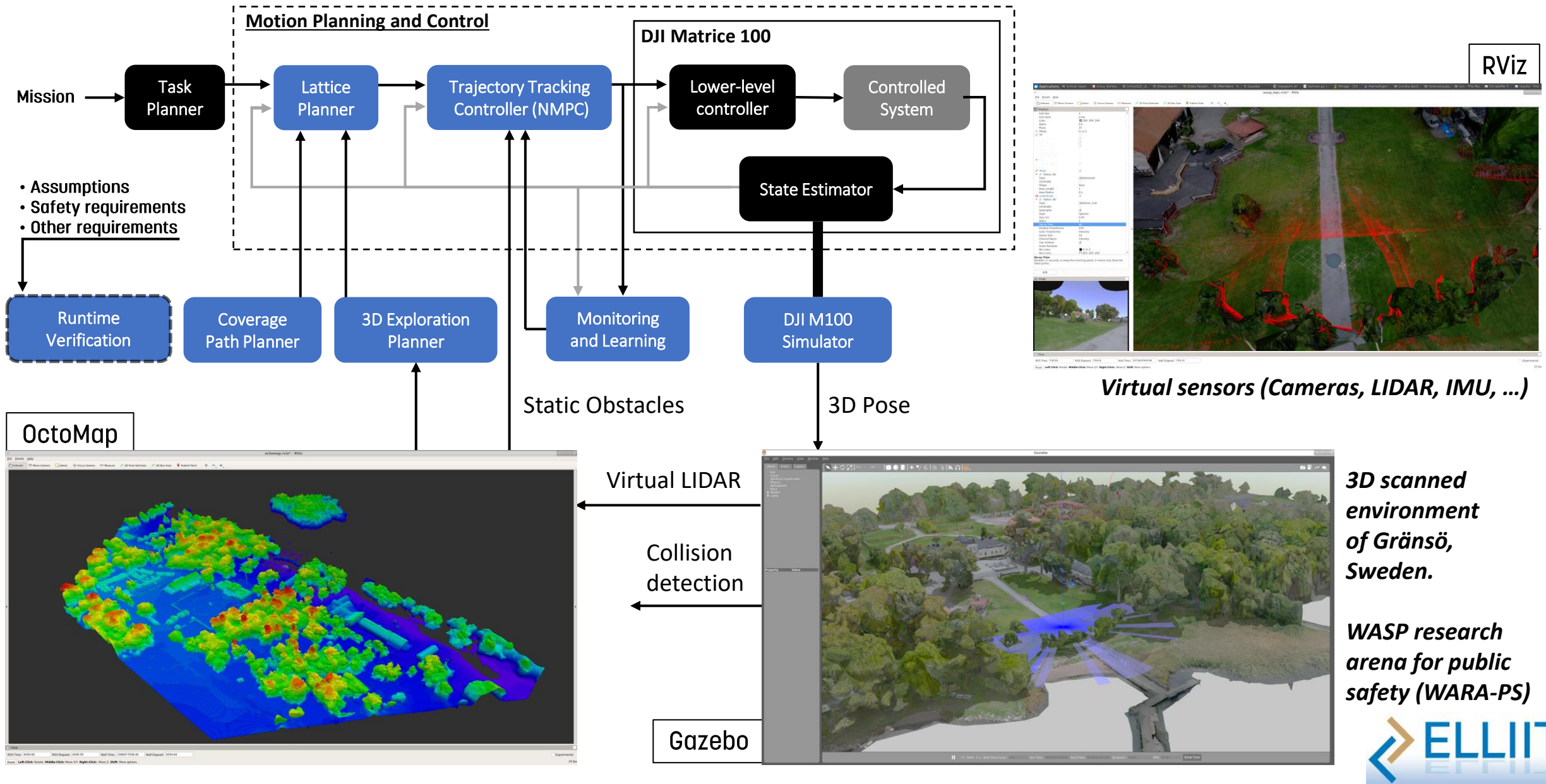
Challenges

- What is **normal** behavior?
- Is the robot **behaving normally**?
- **Safe**, but not task effective?
- Are **learned models** safe to use?

Our Approach



Example AI-Robotics Stack and Simulation Environment

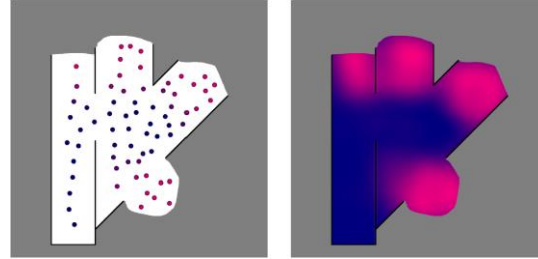


Motion Planning Applications

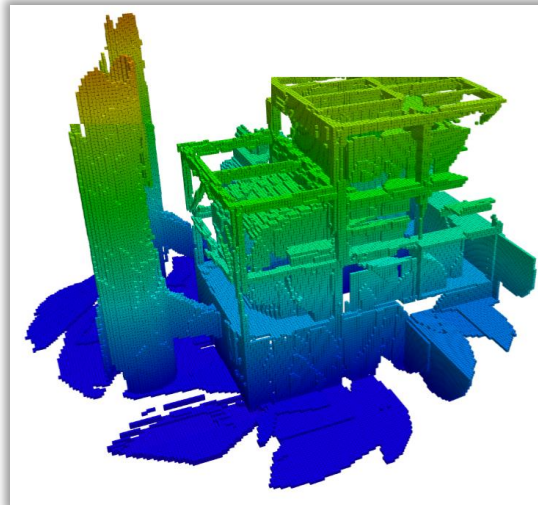
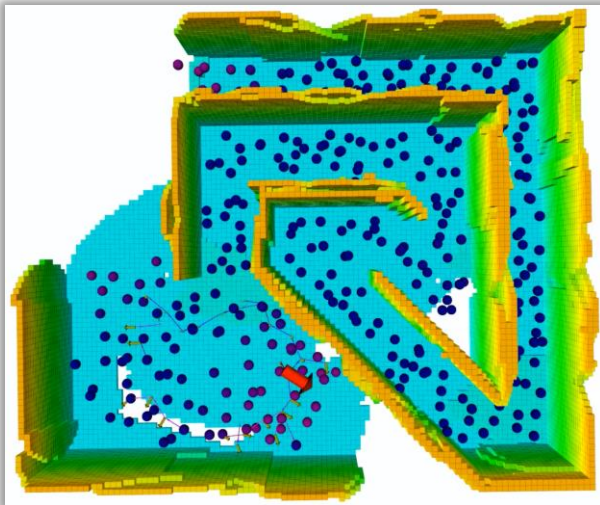
3D Exploration Planning | Coverage Path Planning

- Mapping
- Inspection
- Search for anomalies

Made efficient by Bayesian ML



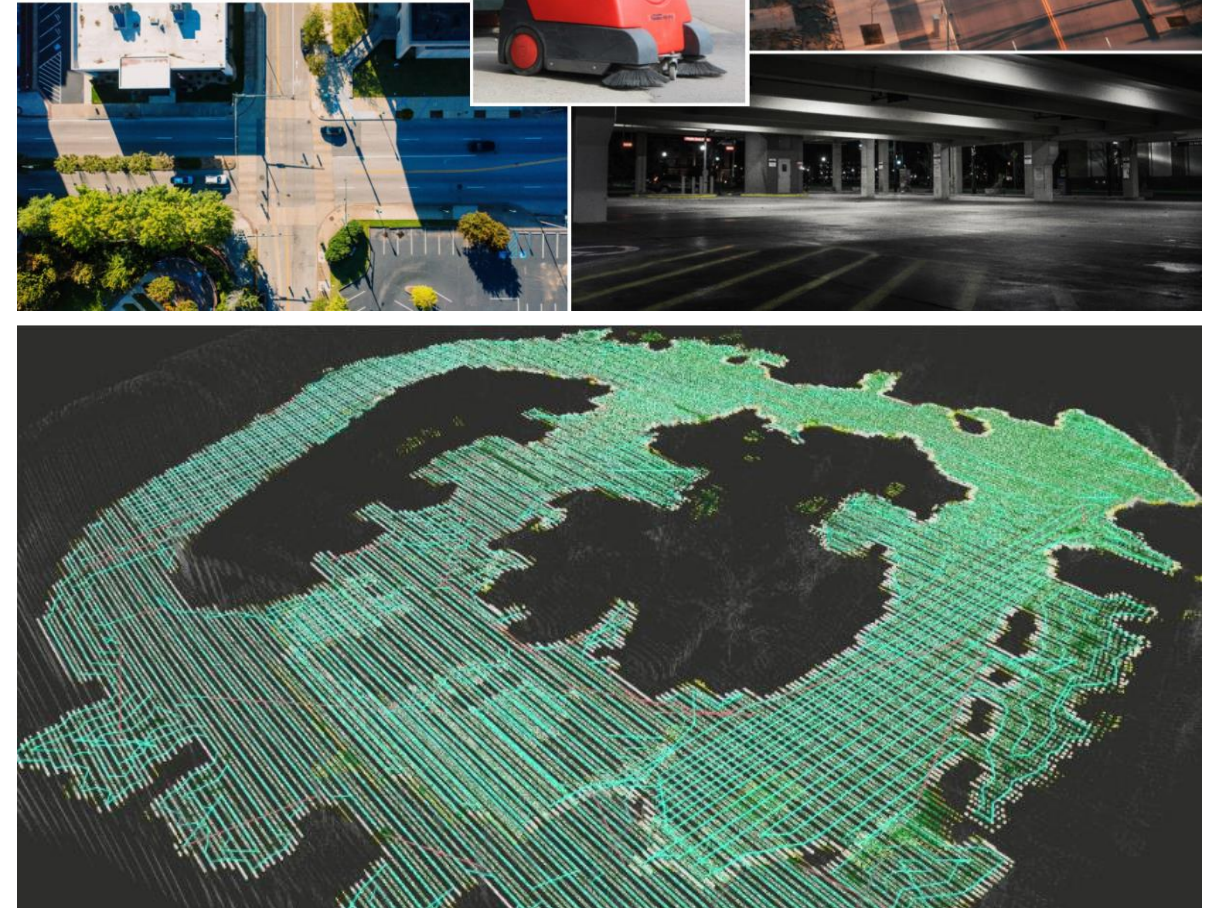
Large-scale, complex geometry, maze-like structures



Coverage Path Planning for Road Sweeping in Urban Environments.

Large-scale, complex geometry, multi-floor, slopes, uneven ground, non-trivial space division

Domain adaptation



[6] M. Selin, M. Tiger, D. Duberg, F. Heintz and P. Jensfelt. *Efficient autonomous exploration planning of large-scale 3D environments*. IEEE Robotics and Automation Letters 4.2 (2019): 1699-1706.

[7] D. Engelson, M. Tiger and F. Heintz. *Coverage Path Planning in Large-scale Multi-floor Urban Environments with Applications to Autonomous Road Sweeping*. IEEE ICRA (2022). (Submitted)

Many Publications Related to Different Components

Motion Planning

Receding-Horizon Lattice-based Motion Planning with Dynamic Obstacle Avoidance

Olov Andersson^{1*}, Oskar Ljungqvist², Mattias Tiger¹, Daniel Axelst¹, Fredrik Heintz¹

Abstract—A key requirement of autonomous vehicles is the capability to safely navigate in their environment. However, outside of controlled environments, such navigation is a very difficult problem. In particular, the real-world often contains both complex 3D structure, and dynamic obstacles such as people or other vehicles. Dynamic obstacles are particularly challenging, as a planned motion requires planning trajectories with respect to both vehicle dynamics, and the motion of the obstacles. Additionally, the requirements imposed by dynamic motion, coupled with real-world computational limitations, make standard optimization and complete search approaches difficult to satisfy. We present a unified optimization-based motion planning and control solution, that can navigate in the presence of both static and dynamic obstacles. By combining optimal and receding-horizon control, with longer receding-horizon planning, we can precompute optimal motion primitives, and allow real-time planning of physically-feasible trajectories in complex environments with dynamic obstacles. We demonstrate the framework by solving difficult indoor 3D quadcopter navigation scenarios, where it is necessary to plan in time, handling collisions on, and taking detours around, the motion of other people and quadcopters.

1. INTRODUCTION

Safe navigation for autonomous vehicles is an area under intense research. As autonomous companies are making vehicles towards full autonomy in structured street environments, unmanned aerial vehicles (UAVs) such as quadcopters are also increasingly being tasked for autonomous inspection, monitoring, search, and even delivery tasks. To efficiently solve such tasks in unstructured environments, often requires the capability to both safely navigate in cluttered environments, while at the same time taking into account other moving agents in the scene. Such dynamics in the environment may include, e.g., ground vehicles, UAVs and even people. This is a difficult motion planning problem, where a principled solution requires planning over time, with respect to both vehicle and obstacle dynamics. Additionally, moving obstacles will impose real-time constraints on planning. The whole real-world autonomous vehicles, have computational



Fig. 1: Example indoor 3D scenario with both static and dynamic obstacles, including humans (red) on the ground, and other quadcopters (orange) flying at varying altitudes. The failed spheres are predictions of future motion.

3D Exploration

Efficient Autonomous Exploration Planning of Large Scale 3D-Environments

Magnus Selin^{1,2}, Mattias Tiger¹, Daniel Odeh², Fredrik Heintz¹, Patrik Jensfelt²

Abstract—Exploration is an important aspect of robotics, whether it is for mapping, rescue missions or path planning in an unknown environment. Frontier Exploration planning (FEP) and Receding Horizon Next-Step View planning (RH-NVP) are two different approaches with different strengths and weaknesses. FEP explores a large environment consisting of separate regions with one, but it is slow at reaching full exploration due to moving back and forth between regions. RH-NVP shows great potential and efficiently explores individual regions, but has the disadvantage that it can get stuck in large environments not exploring all regions. In this work we present a method that combines both approaches, with FEP as a global exploration planner and RH-NVP for local exploration. We also present techniques to estimate potential information gain faster, to cache previously estimated values and to exploit them to efficiently estimate new queries.

Index Terms—Search and Rescue Robotics Motion and Path Planning Mapping

1. INTRODUCTION

IN this paper we study the problem of planning for exploring an unknown area. We propose a novel method, Autonomous Exploration Planner (AEP), which improves upon the state-of-the-art method Receding Horizon Next-Step View planning (RH-NVP) [1]. RH-NVP uses a sampling based approach to pick out the next best view point [2] in combination with Rapidly-exploring Random Trees (RRT) [3] to produce reasonable paths, weight the samples and evaluate the first edge before exploring again. The issue for each node in the RRT is the volume of unexplored space that would be covered by the sensor from the corresponding pose, weighted with the cost of going there.

In the work Receding Horizon Next-Step View planning is used as a local exploration strategy and is combined with Frontier exploration [4] for global exploration. When new information is available close to the agent, local exploration strategy is used.

This work was supported by the Swedish Research Council (VR) and the Swedish Space Agency (Svea). The authors would like to thank the anonymous reviewers for their comments. The authors would like to thank the anonymous reviewers for their comments. The authors would like to thank the anonymous reviewers for their comments.

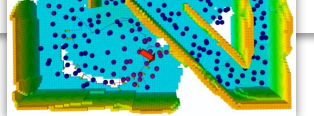


Figure 2: A 3D visualization of a large-scale environment being explored, showing a complex structure with many small cubes representing the explored volume.

Runtime Verification



Incremental Reasoning in Probabilistic Signal Temporal Logic

Mattias Tiger¹, Fredrik Heintz¹

Linköping University, Sweden

Email: mattias.tiger@liu.se

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Available online 24 January 2020

Motion Pattern Recognition

Online Sparse Gaussian Process Regression for Trajectory Modeling

Mattias Tiger¹, Department of Computer and Information Science, Linköping University, Sweden, Email: mattias.tiger@liu.se

Fredrik Heintz¹, Department of Computer and Information Science, Linköping University, Sweden, Email: fredrik.heintz@liu.se

Abstract—Trajectories are used in many target tracking and other fusion-related applications. In this paper we consider the problem of modeling trajectories as Gaussian processes and learning such models from sets of observed trajectories. We demonstrate that the traditional approach in Gaussian process regression is not suitable when modeling a set of trajectories. Instead, we introduce an approach to Gaussian process trajectory regression based on an alternative use of coupling two Gaussian process GP trajectory models and inverse GP regression. The benefit of our approach is that it works well online and efficiently supports online trajectory model manipulations such as merging and splitting of trajectory models. Splitting and merging are very useful in online trajectory modeling and learning where trajectory models are considered dynamic objects. The presented method and accompanying approximation algorithm have time and memory complexities comparable to the state-of-the-art of regular full and approximate GP regression, while having a more flexible model suitable for modeling trajectories. The novelty of our approach is in the very flexible and accurate model, especially for trajectories, and the proposed approximation method based on solving the inverse problem of Gaussian process regression.

1. INTRODUCTION

Gaussian processes (GP) is a flexible and powerful Bayesian non-parametric approach to modeling functions and performing inference on functions. They have been demonstrated to be practical and applicable to a wide variety of real-world statistical learning problems but also modeling, drawing and predicting sparse temporal trajectories such as

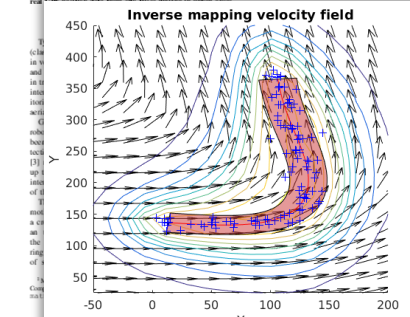
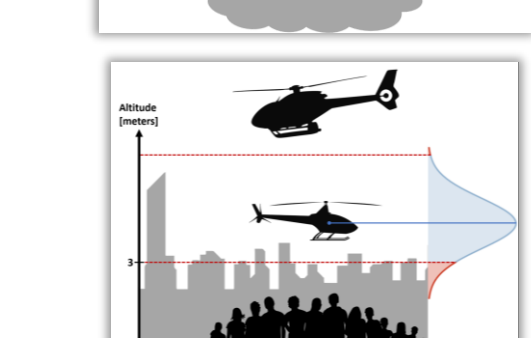
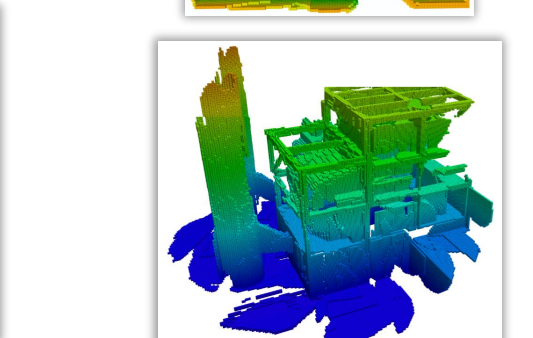
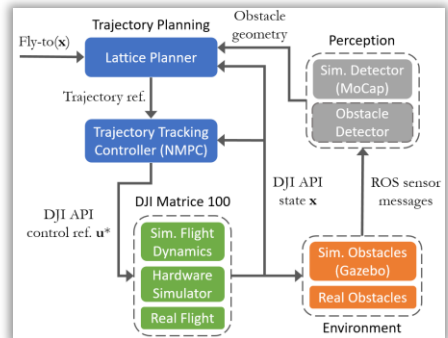
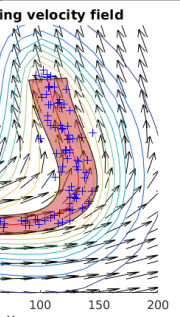
Gaussian Process Based Motion Pattern Recognition with Sequential Local Models

Mattias Tiger¹ and Fredrik Heintz¹

Abstract—Conventional trajectory-based vehicular traffic analysis approaches work well in simple environments such as a single crossing but they do not scale to more structurally complex environments such as networks of interconnected crossings (i.e. urban road networks). Local trajectory models are necessary to cope with the multi-modality of such structures, which in turn introduces new challenges. These larger and more complex environments increase the uncertainties of lack of motion and self-occlusions in observed trajectories which impose further challenges. In this paper we consider the problem of motion pattern recognition for the setting of sequential local motion pattern models. That is, classifying sub-trajectories from observed trajectories in accordance with which motion pattern that best explains it. We introduce a Gaussian process GP based modeling approach which outperforms the state-of-the-art GP based motion pattern approaches at this task. We investigate the impact of both a prior local model overlap and the length of the observed trajectory trace on the classification quality. We further show that introducing a pre-processing step filtering out steps from the training data significantly improves classification performance. The approach is evaluated using real-world data.

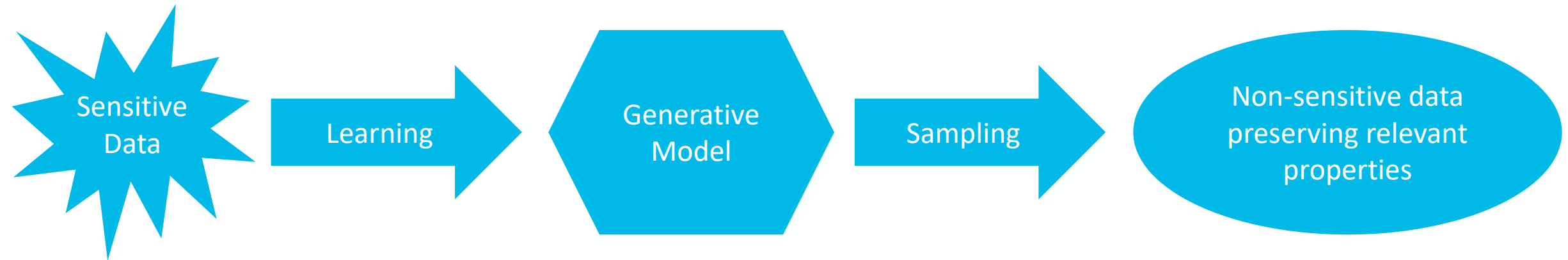
Fig. 1

Fig. 1: A long trajectory through a complex road structure. Problematic self-overlap is indicated in red. Self-overlap between lanes can occur when precision is low (dashed red).



Privacy-preserving synthetic data generation

[D. Bergström, Md F. Sikder, R. Ramachandranpillai]



1. Learn a generative model that captures the probability distribution of the sensitive data
2. Create a synthetic data set from the generative model that both captures the salient features of the original data set **and** is non-sensitive
3. Methods for verifying that the synthetic data set is accurate enough
4. Methods for verifying that the synthetic data set is non-sensitive

Synthetic Healthdata – Existing Models, Measures, and Problems

Models

- MedGAN[1]
- HealthGAN[2]
- Synthea[3]

Measures

- Privacy
- Utility
- Resemblance

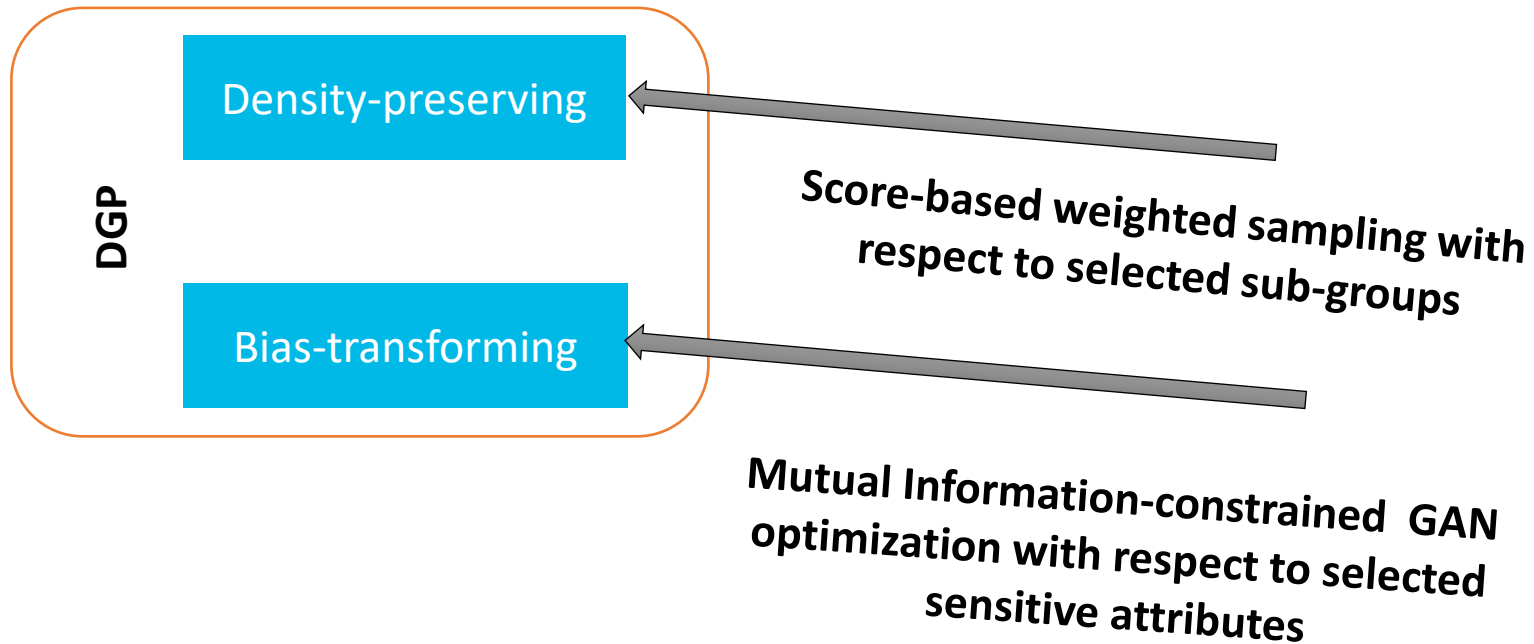
Research Gap

- Fairness? [4]

1. Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S. and Yang, B., 2020. MedGAN: Medical image translation using GANs. *Computerized medical imaging and graphics*, 79, p.101684.
2. Yale, A., Dash, S., Dutta, R., Guyon, I., Pavao, A. and Bennett, K.P., 2020. Generation and evaluation of privacy preserving synthetic health data. *Neurocomputing*, 416, pp.244-255.
3. Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., Duffett, C., Dube, K., Gallagher, T. and McLachlan, S., 2018. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association*, 25(3), pp.230-238.
4. Bhanot, K., Qi, M., Erickson, J.S., Guyon, I. and Bennett, K.P., 2021. The problem of fairness in synthetic healthcare data. *Entropy*, 23(9), p.1165.

Our Proposed Solution : Generate fair health data from biased data

- We define a **Fair Data Generation Process (FDGP)** in **Generative Adversarial Networks (GAN)** :

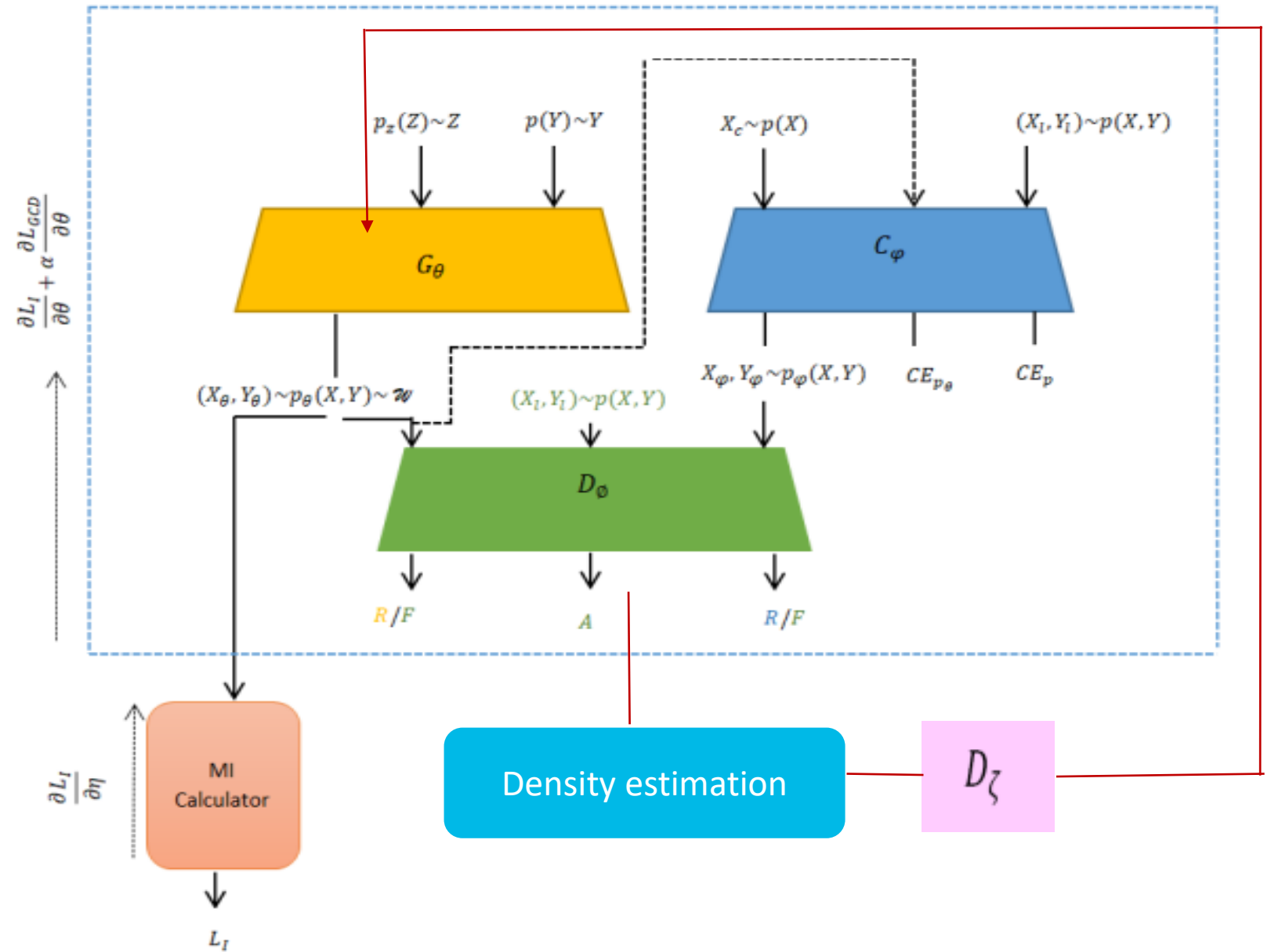


Goals:

- Formal equality with respect to some sub-groups
- Substantive equality with respect to some sensitive attributes
- Data utility

DeMISe

$$L_F = \underbrace{L_{GCD}}_{\text{Semi-supervised-generation}} + \underbrace{\alpha L_{MI}}_{\text{MIde-biasing}} \underbrace{\quad}_{\text{Fair-generation}}$$



Results and Discussions

Dataset	Medical Information Mart Intensive Care , Version 3 [4]
Benchmarks	HealthGAN and FairGAN [5]
Evaluation Metric	Data utility - Accuracy and F1 score Downstream fairness (substantive equality) - AUROC gap and Demographic Parity gap Fair resemblance - density score

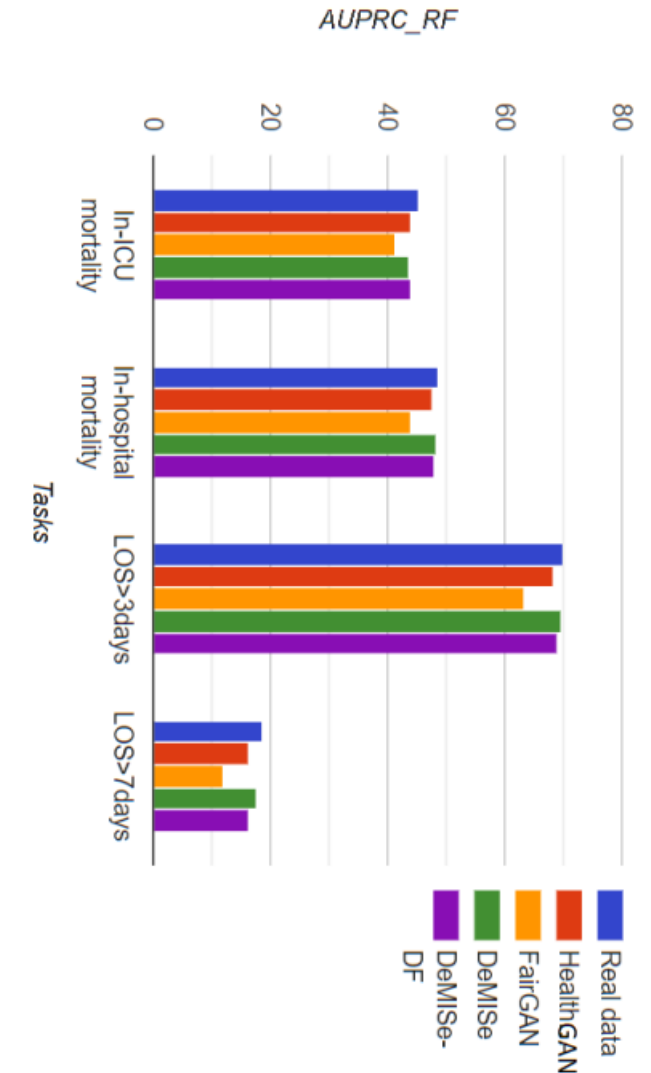
4. <https://physionet.org/content/mimiciii/1.4/>

5. Xu, D., Yuan, S., Zhang, L. and Wu, X., 2018, December. Fairgan: Fairness-aware generative adversarial networks. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 570-575). IEEE.

Data utility analysis

	Real data		Heal. GAN		Fair GAN		Ours	
	acc.	F1	acc.	F1	acc.	F1	acc.	F1
(a)	92.1	37.6	91.3	34.2	89.5	32.4	91.5	34.7
	91.8	12.1	91.1	12.0	88.3	11.6	90.9	11.7
(b)	71.2	59.9	69.4	58.3	67.1	56.3	68.1	57.3
	72.6	59	67.2	59.1	66.4	57.9	68.9	57.6
(c)	90.1	39.6	89.1	37	85.4	32.8	89.6	39.9
	89.3	17.9	88.3	15.8	86.3	14.3	90	18.1
(d)	89.9	7.0	87.9	8.5	86.1	4.3	88.4	6.8
	87.6	1.4	88.4	2.1	85.9	0.8	87.3	2.4

Accuracy and F1 on various prediction tasks with real data as reference point; (a) In-ICU mortality, (b) LOS > 3days,(c) In-hospital mortality, and (d)LOS > 7days. For each tasks, the first row denotes the predictions by LR and second row is the predictions by RF (higher is better for all the values).

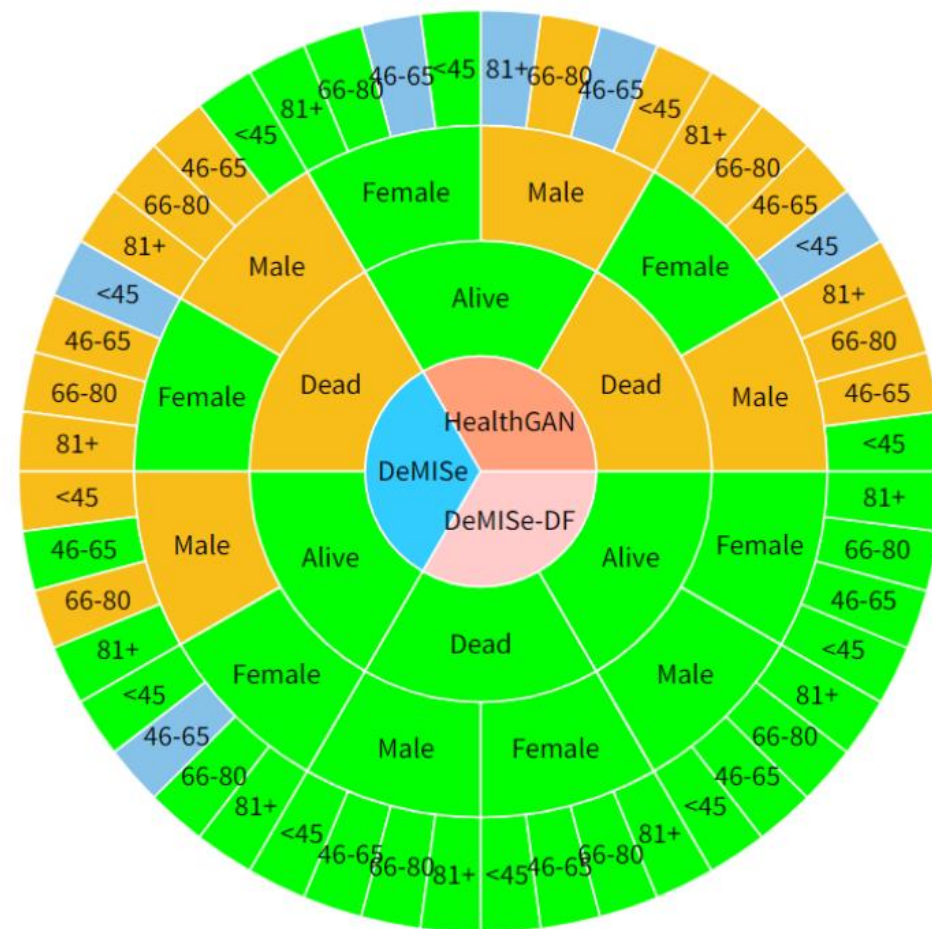
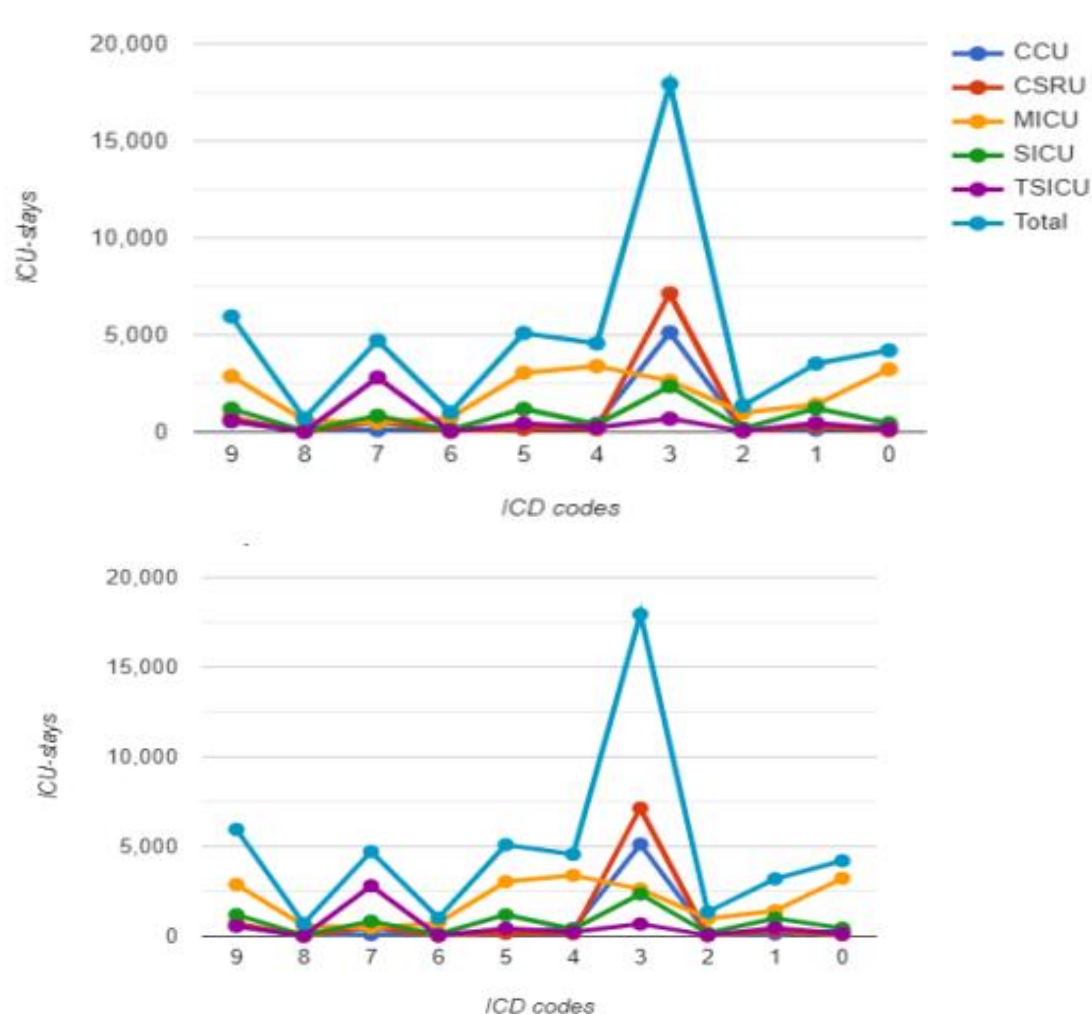


Downstream Fairness Analysis

Metrics	Prediction	Real Data	HealthGAN	FairGAN	DeMISe
AUROC gap	In-hospital mortality	0.043 ± 0.001 ,	0.082 ± 0.002	0.021 ± 0.001	0.001 ± 0.001
	In-ICU mortality	0.03 ± 0.007	0.15 ± 0.035	0.023 ± 0.064	0.012 ± 0.021
	LOS>3days	-0.003 ± 0.002	-0.104 ± 0.001	-0.003 ± 0.001	0.000 ± 0.001
	LOS>7days	-0.005 ± 0.002	-0.076 ± 0.002	$-0.061 \pm .001$	-0.013 ± 0.001
Parity gap	In-hospital mortality	-0.046 ± 0.018	-0.154 ± 0.010	-0.004 ± 0.014	0.000 ± 0.001
	In-ICU mortality	-0.031 ± 0.013	-0.331 ± 0.011	-0.005 ± 0.013	0.000 ± 0.000
	LOS>3days	0.022 ± 0.012	0.224 ± 0.012	0.022 ± 0.002	0.000 ± 0.001
	LOS>7days	-0.004 ± 0.002	-0.004 ± 0.002	-0.002 ± 0.001	-0.003 ± 0.001

The fairness gaps between White and Black patients across the different health care tasks, and models. Positive values represent a bias towards the white patients and negative values represent a bias towards the Black patients. The models are fair as the metric moves towards zero. The models are more unfair as the metric moves away from zero.

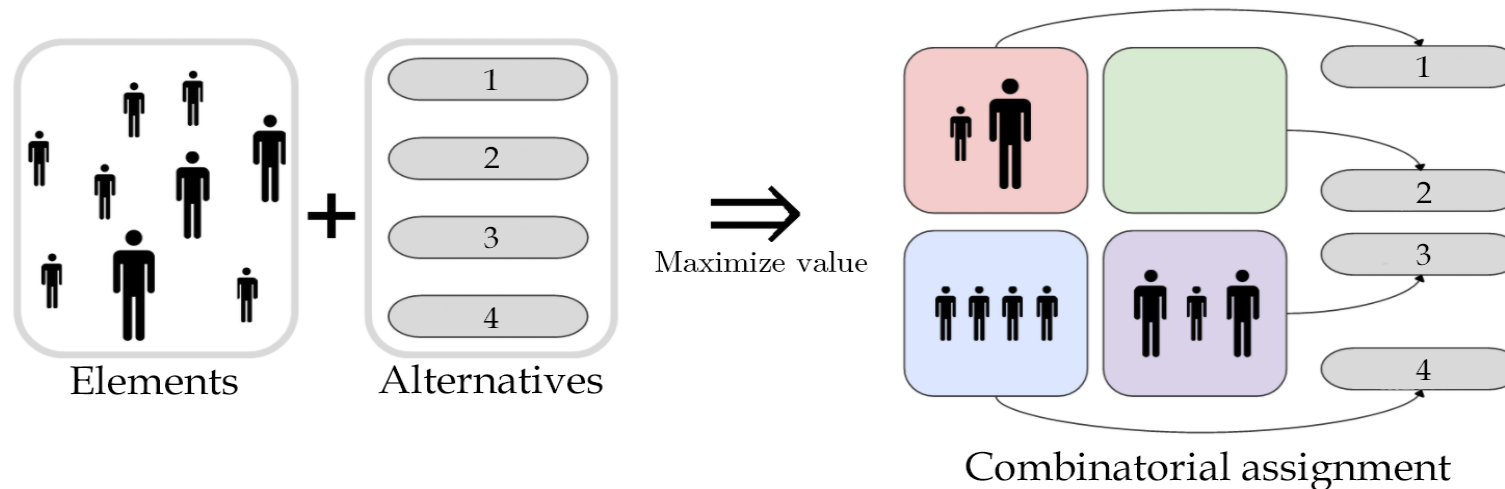
Fair Resemblance Analysis



The colors orange, blue and green respectively indicate under-represented, over-represented and adequately represented subgroups

Dividing the Indivisible to Maximize Value

We consider *combinatorial assignment*—the class of problems in which indivisible elements are partitioned into bundles among alternatives to maximize some notion of value (e.g., social welfare, expected utility).



The Combinatorial Assignment Problem

THE COMBINATORIAL ASSIGNMENT PROBLEM

Input: A set of n items N , a set of m alternatives M , and a function (called the *social welfare function*) $\Phi : \Pi_N^m \rightarrow \mathbb{R}$.

Output: A combinatorial assignment $C = \langle B_1, \dots, B_m \rangle$ over N that maximizes its *social welfare* $\Phi(C)$.

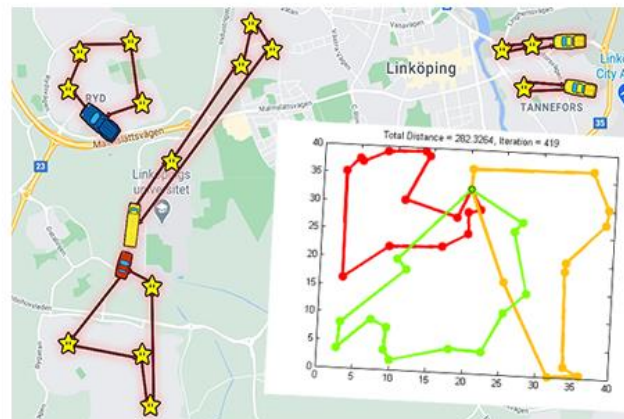
Important welfare function examples:

- *Utilitarian*: $\Phi(\langle B_1, \dots, B_m \rangle) = \sum_{i \in M} v(B_i, i)$.
Ex: Team formation; combinatorial auctions.
- *Egalitarian*: $\Phi(\langle B_1, \dots, B_m \rangle) = \min_{i \in M} v(B_i, i)$.
Ex: Minimize VRP makespan; “equitable” resource division.

Both are unfortunately APX-hard (hard to approximate within a constant factor) and the input size is exponential with respect to n .



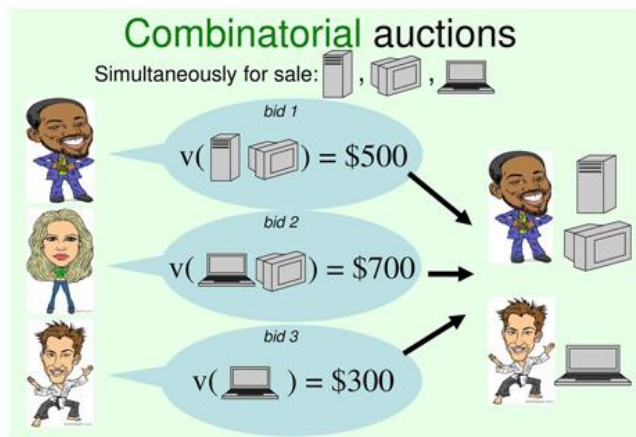
Assigning workers to jobs



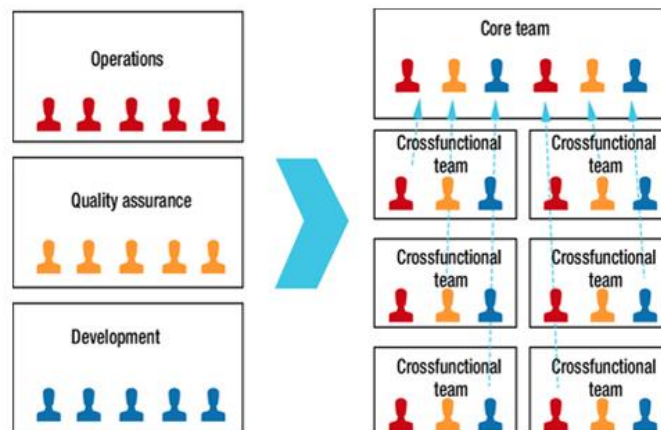
Multi-vehicle routing
(e.g., multiple TSP)



Multi-sensor
multi-target tracking



Combinatorial auctions



Team formation

Digit	Description
001-099	Service courses for nontechnical majors
100-199	Other service courses, basic undergraduate
200-299	Advanced undergraduate/beginning graduate
300-399	Advanced graduate
400-499	Experimental
500-599	Graduate seminars

Digit	Description
00-09	Introductory, miscellaneous
10-19	Hardware and Software Systems
20-39	Artificial Intelligence
40-49	Software Systems
50-59	Mathematical Foundations of Computing
60-69	Analysis of Algorithms

Course allocation

Utilitarian Combinatorial Assignment - Conclusions

- Important UCA problems include coalition structure generation, the winner determination for combinatorial auctions, and generalized assignment.
- Utilitarian combinatorial assignment is APX-hard—but under certain restrictions, the problem is solvable and/or approximable in polynomial time.
 - Synergy hypergraphs is an expressive concise representation that admits faster algorithms for many cases (AAMAS2022).
- In all of the experiments so far, our best optimal algorithm (a hybrid) finds optimum in (worst-case) $\approx 1\%$ of the time that the previous best method (IBM's CPLEX) needs.
 - It has been used successfully in a commercial setting (EU4).
- Machine learning can be used to generate heuristics that outperform conventional heuristics & Monte Carlo methods (AAMAS2022).

Select References

- [Fredrik Prántare and Fredrik Heintz \(2020\)](#). “An Anytime Algorithm for Optimal Simultaneous Coalition Structure Generation and Assignment”. In: [JAAMAS](#)
- [Fredrik Prántare and Fredrik Heintz \(2020\)](#). “Hybrid Dynamic Programming for Optimal Simultaneous Coalition Structure Generation and Assignment”. In: [PRIMA](#)
- [Fredrik Prántare, Herman Appelgren, and Fredrik Heintz \(2021\)](#). “Anytime Heuristic and Monte Carlo Methods for Large-Scale Simultaneous Coalition Structure Generation and Assignment”. In: [AAAI](#)
- [Fredrik Prántare, Mattias Tiger, David Bergström, Herman Appelgren, and Fredrik Heintz \(2022\)](#). “Learning Heuristics for Combinatorial Assignment by Optimally Solving Subproblems”. In: [AAMAS](#)
- [Fredrik Prántare, Leif Eriksson, and George Osipov \(2022\)](#). “Concise Representations and Complexity of Combinatorial Assignment Problems”. In: [AAMAS](#)



1. Analyze hardness



2. Optimal algorithms



3. Non-exact algorithms



4. Real-world applications

Research Overview



Learning generative models based on trajectory data

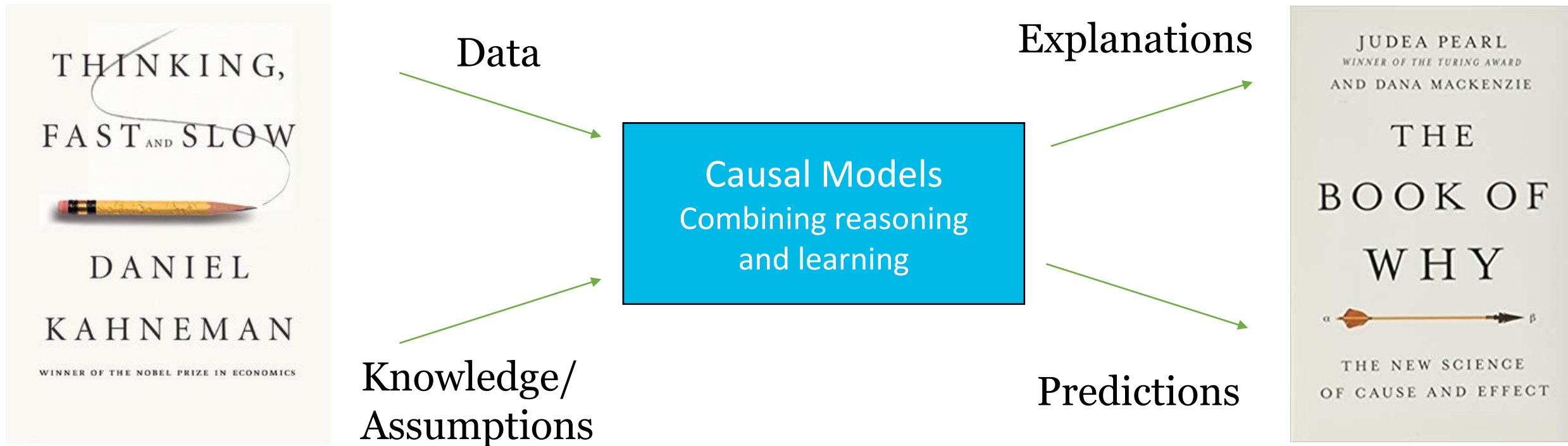


Probabilistic logical reasoning over observed and predicted trajectories



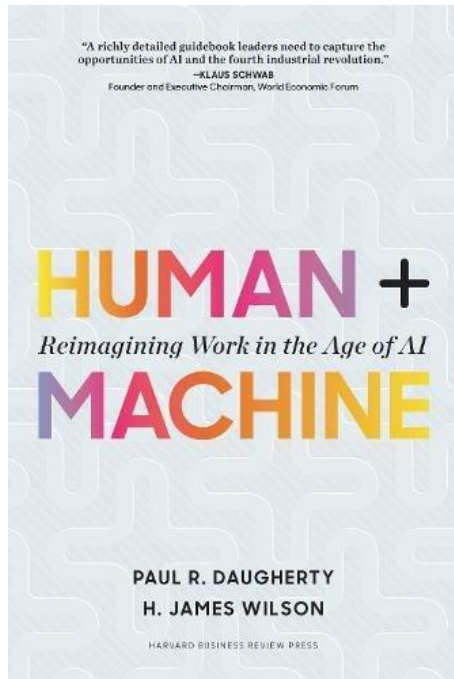
Utilitarian Combinatorial Assignment

The Way Forward



Other Components to Achieve Trustworthy AI

Humans + AI



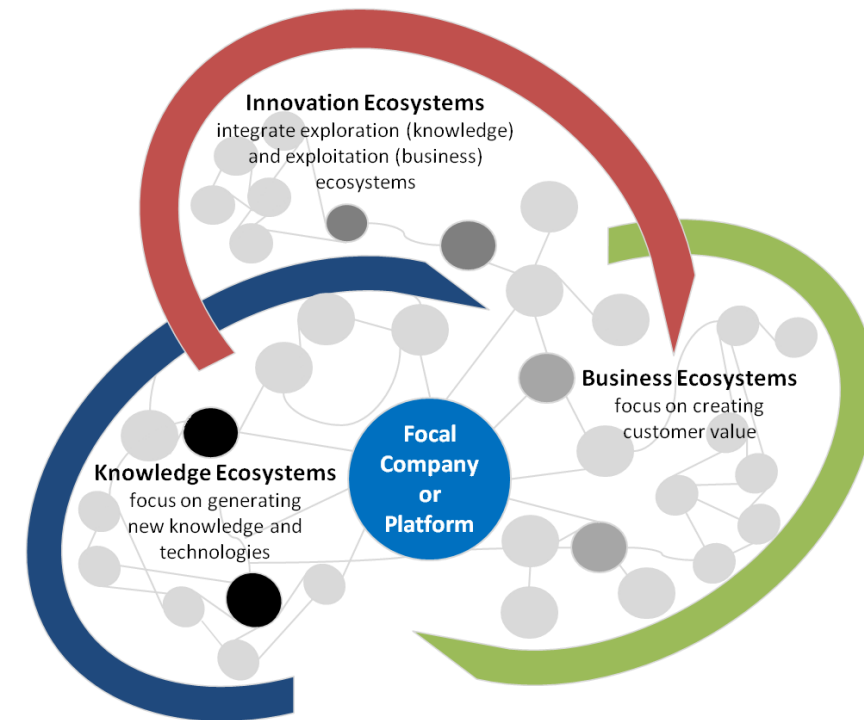
<https://knowledge.wharton.upenn.edu/article/reimagining-work-age-ai/>

Education



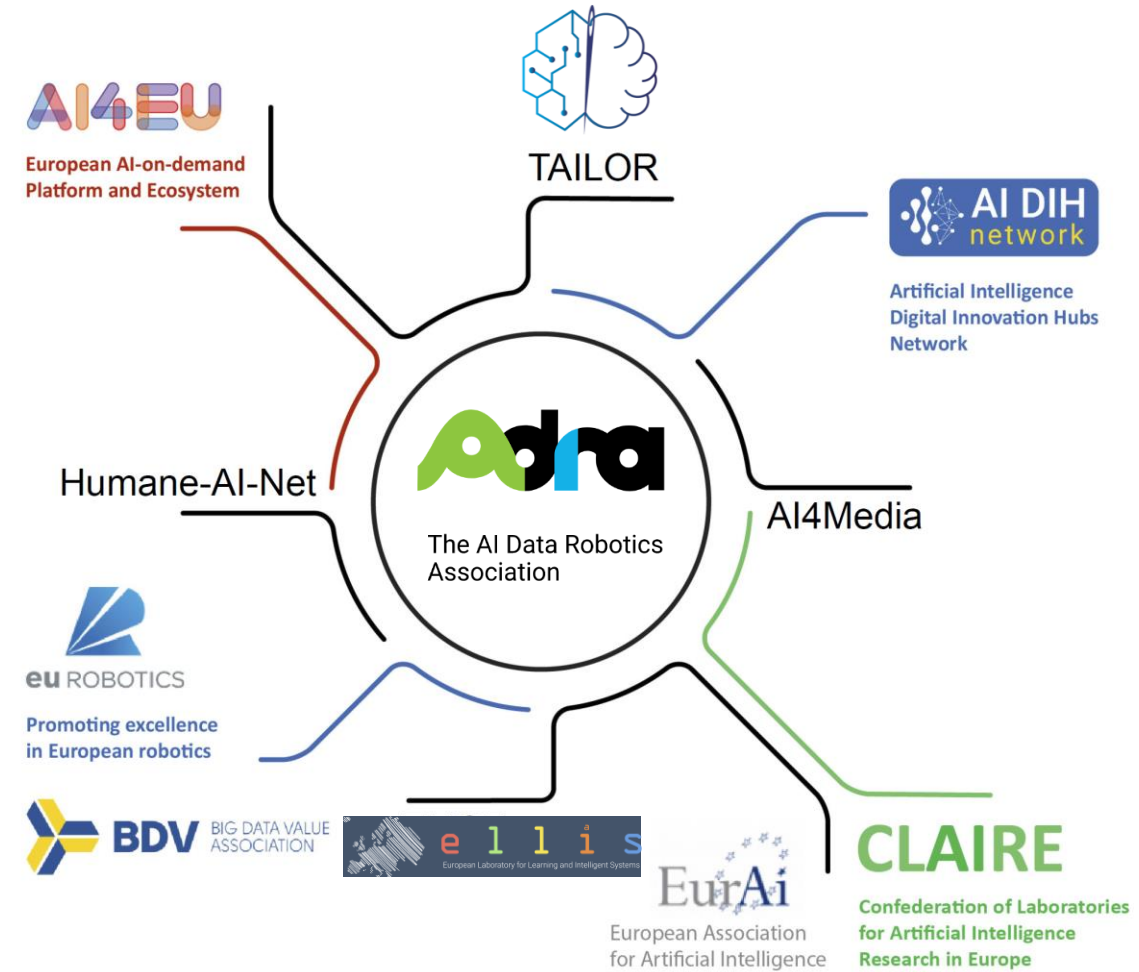
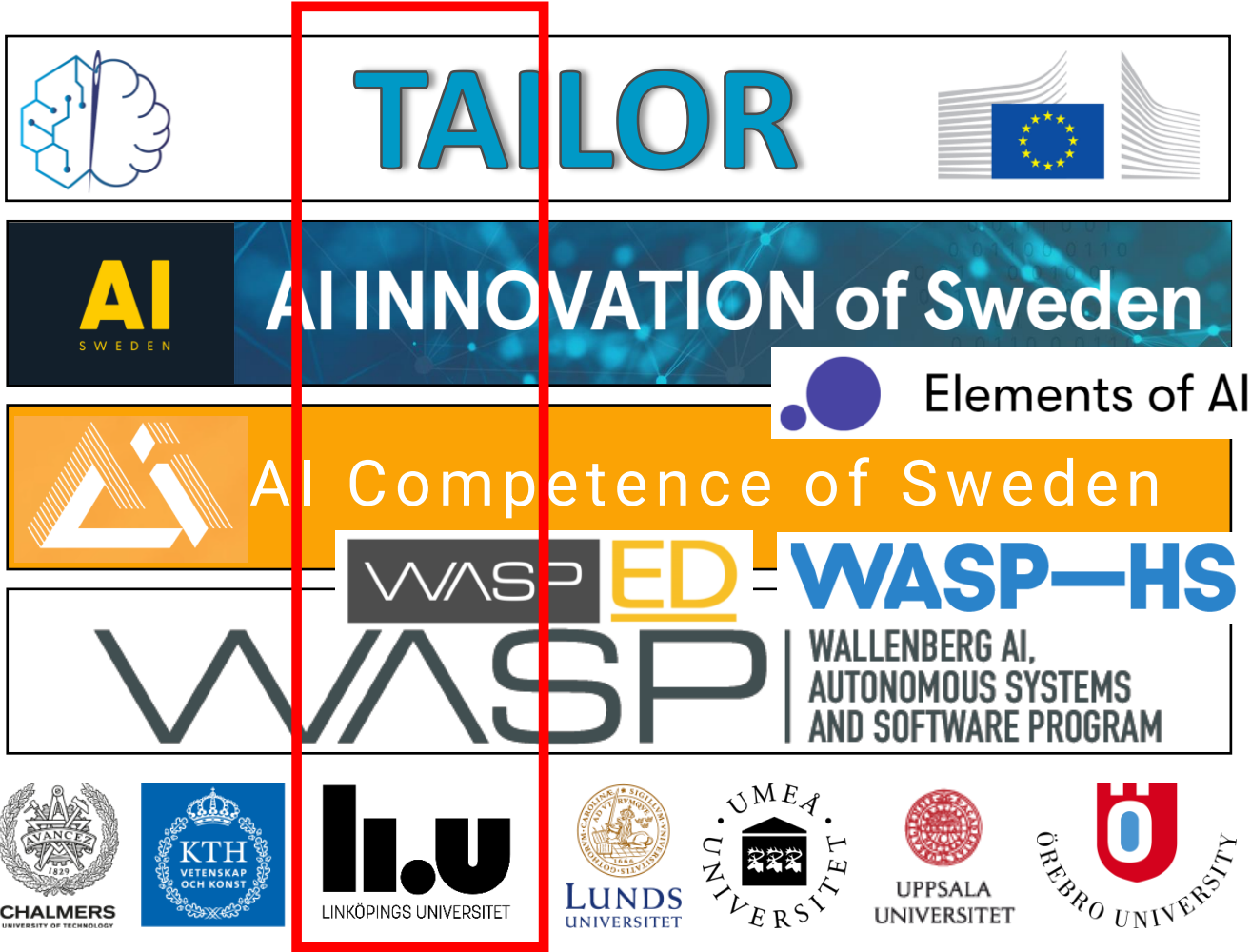
<https://elementsofai.se>

Ecosystems



<https://timreview.ca/article/919>

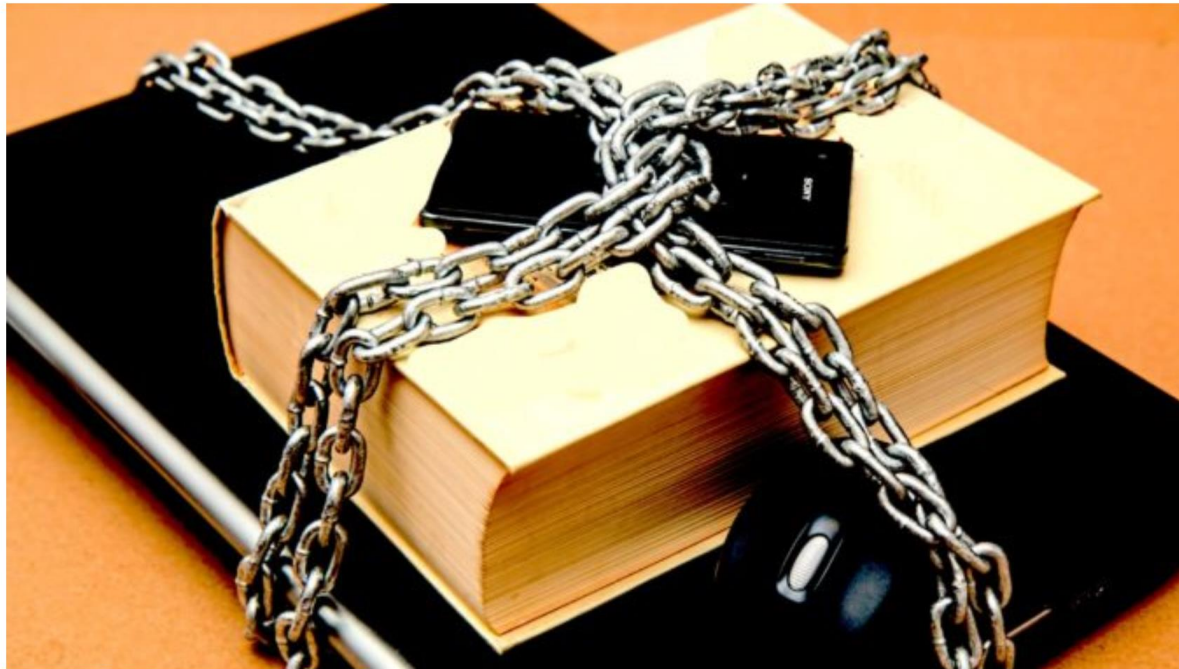
AI Innovation, Competence and Research Ecosystem



External Analysis of Human Decision Making

France Bans Judge Analytics, 5 Years In Prison For Rule Breakers

🕒 4th June 2019 👤 artificiallawyer ➡ Litigation Prediction 💬 52





“Weak human + machine + superior process was greater than a strong computer and, remarkably, greater than a strong human + machine with inferior process.”

Garry Kasparov

AI and Humans – Together

