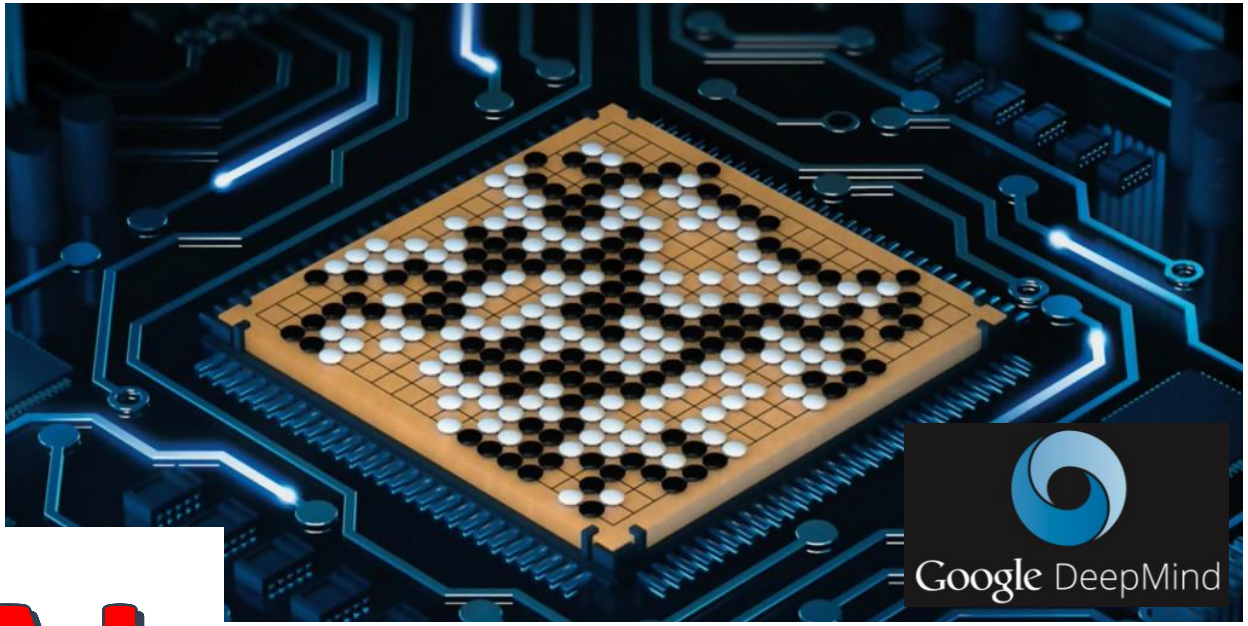
Towards Trustworthy AI
- Integrating Reasoning and Learning

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

LINKÖPINGS UNIVERSITET

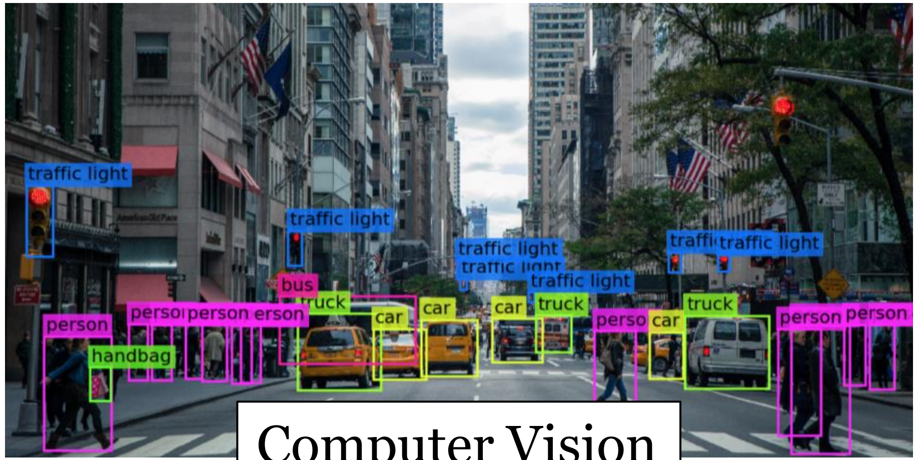

Google DeepMind

AI


THINK
$300,000
$1,000,000
$200,000
KEN
WATSON
BRAD

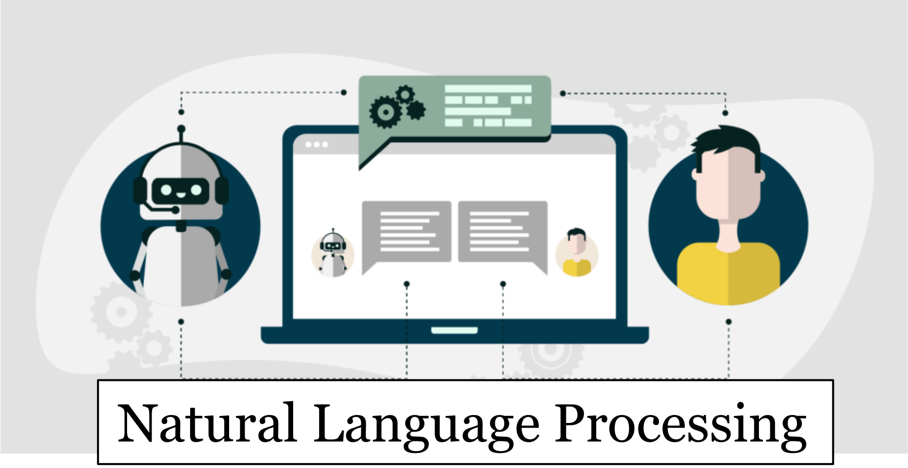

BostonDynamics
Boston Dynamics

Applications of AI

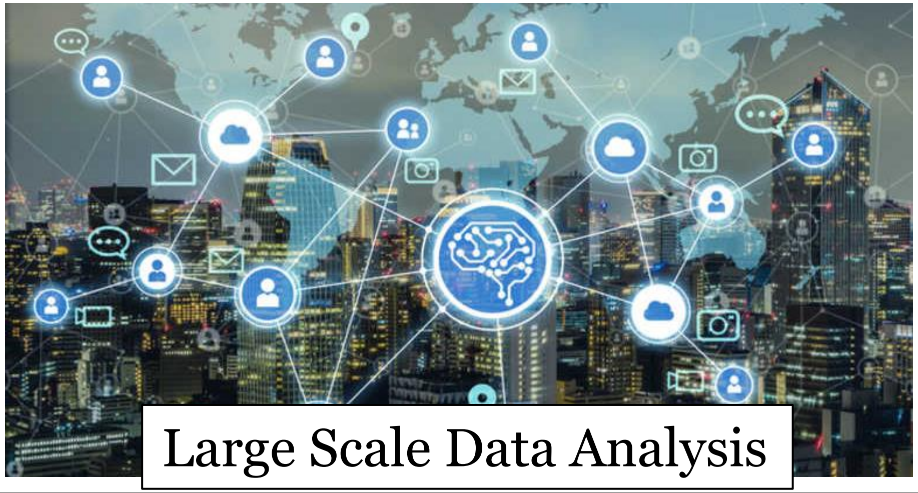


Computer Vision

Voice Recognition



Natural Language Processing



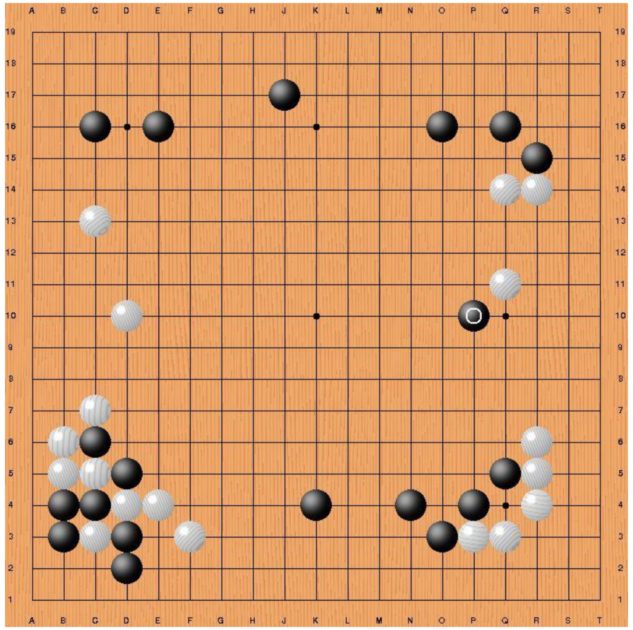
Large Scale Data Analysis

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

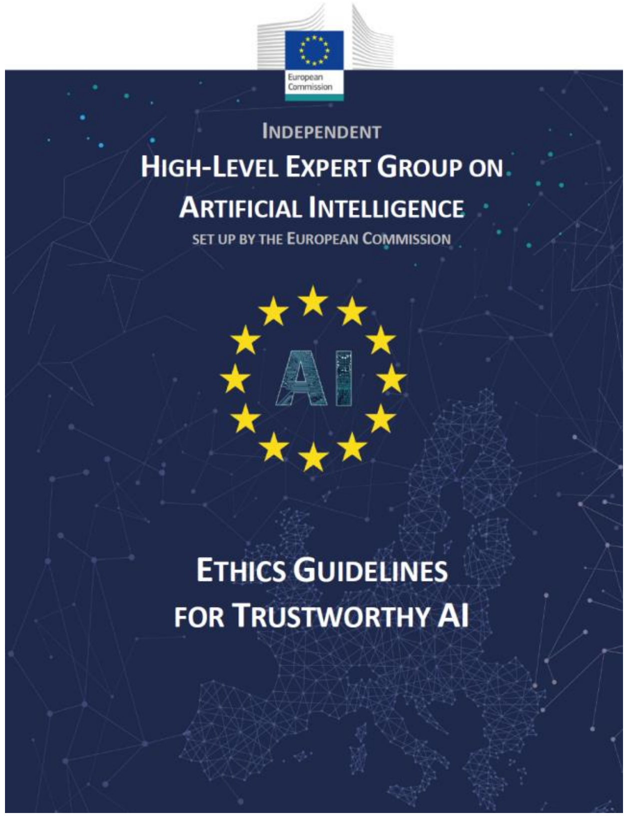
Human-centric approach: AI as a means, not an end

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)



https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

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

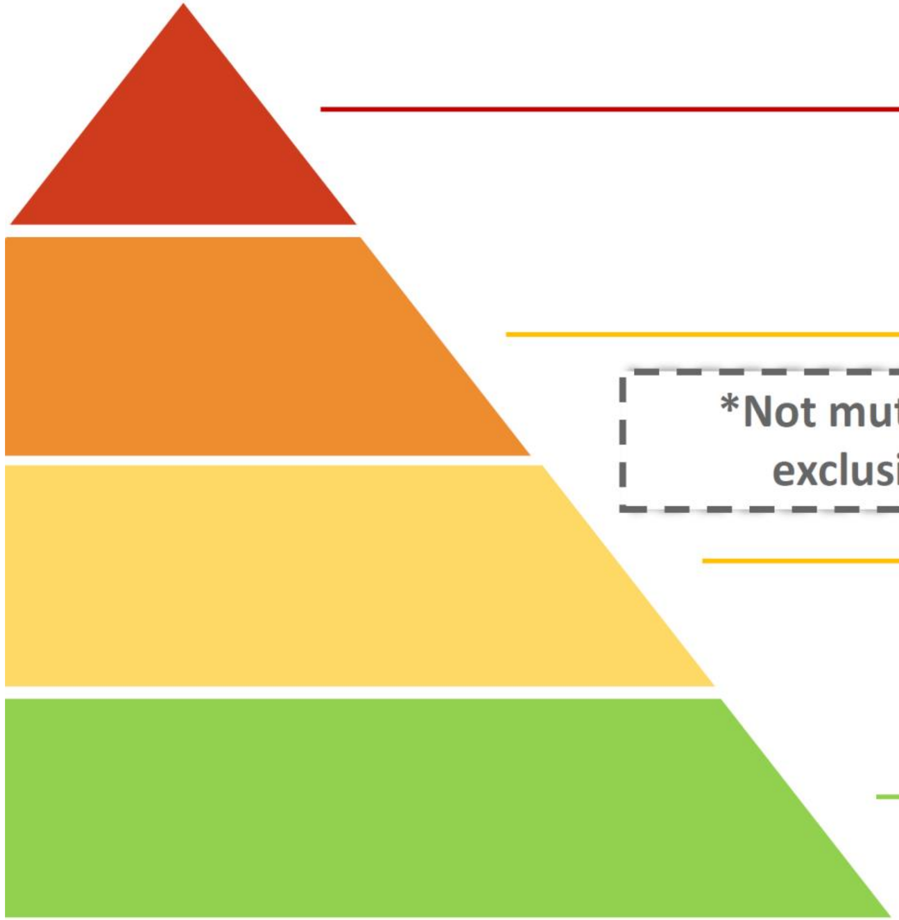
Equal and just distribution of benefits and costs.

Explicability

Transparent, open with capabilities and purposes, explanations

https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

A risk-based approach



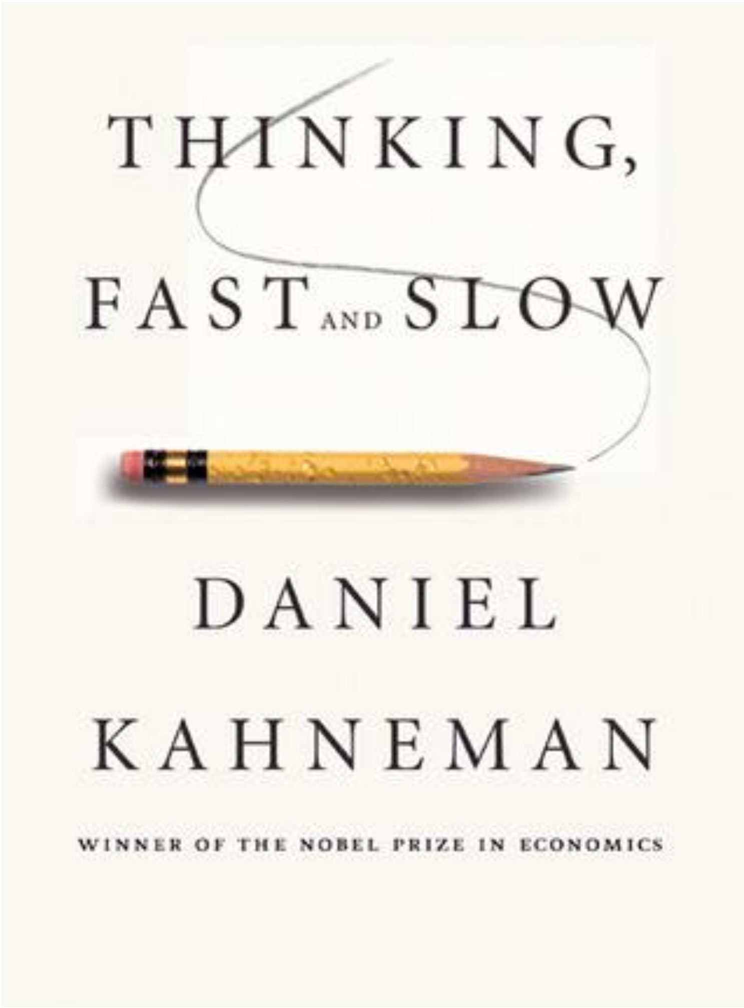
*Not mutually exclusive

Unacceptable risk e.g. social scoring	**Prohibited**
High risk e.g. recruitment, medical devices	**Permitted** subject to compliance with AI requirements and ex-ante conformity assessment
‘Transparency’ risk ‘Impersonation’ (bots)	**Permitted** but subject to information/transparency obligations
Minimal or no risk	**Permitted** with no restrictions

Human and Computational Thinking

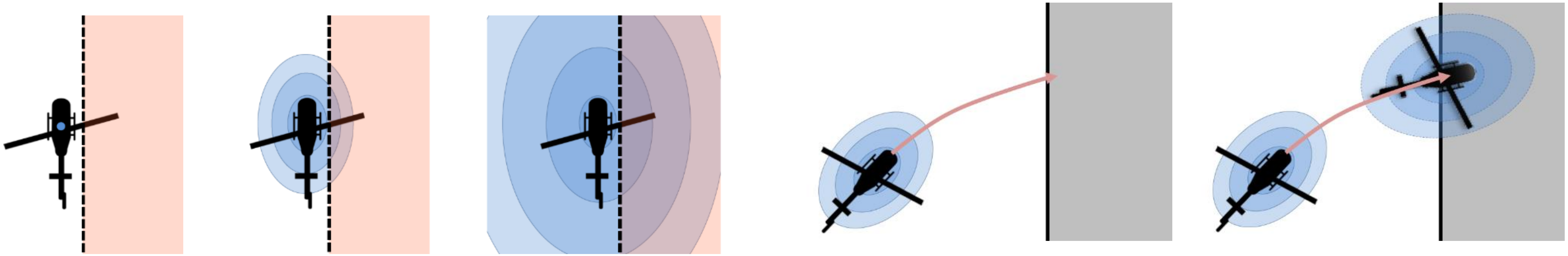
Figure 1: A Comparison of System 1 and System 2 Thinking

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	WITH self-awareness or control
"What you see is all there is."	Logical and skeptical
ROLE Assesses the situation Delivers updates	ROLE Seeks new/missing information Makes decisions



Probabilistic Predictive Stream Reasoning

[Tiger and Heintz TIME 2016, IJAR 2020]



collision: false   Pr(collision) = 0.1   Pr(collision) = 0.4   Pr(collision now) = 0.0...   Pr(collision soon) = 0.5

Reasoning over <u>Uncertainty</u>   **Reasoning over <u>Predictions</u>**

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 (altitude$_0$ > 3)

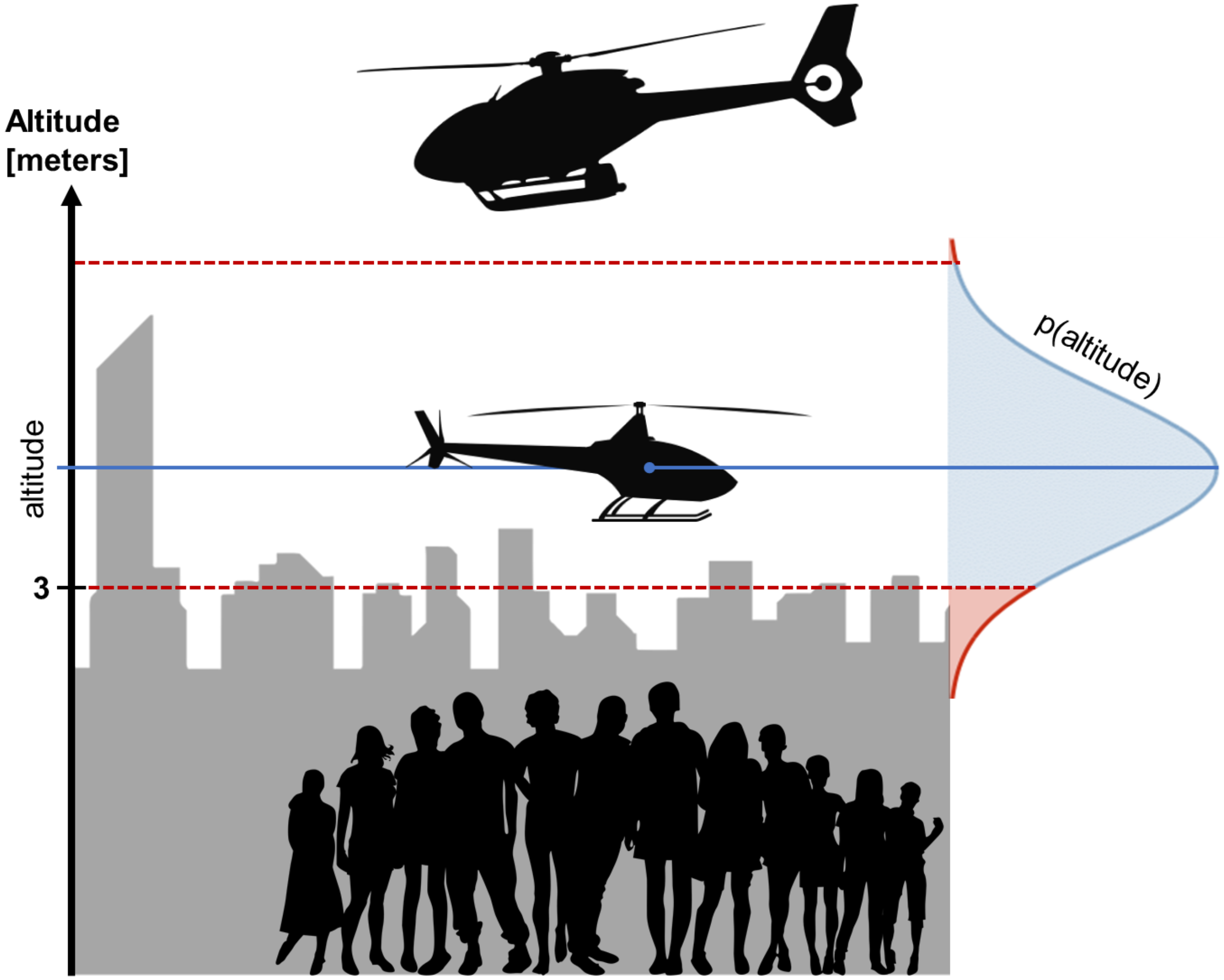
true

always (Pr(altitude$_{0|0}$ > 3) ≥ 0.99)

false

always (Pr(altitude$_{2|0}$ > 3) ≥ 0.99)

Relative time to estimate   Relative time to estimate from

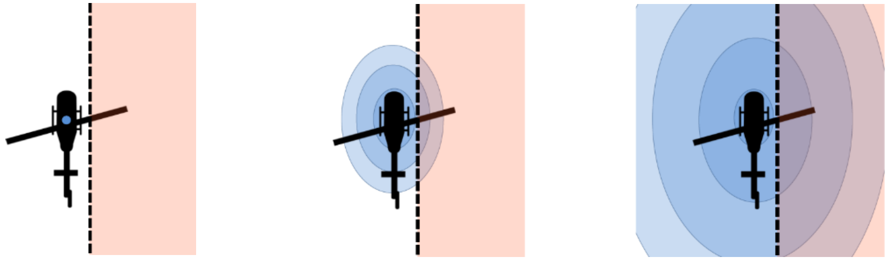


Probabilistic logical reasoning over observed and predicted trajectories

[Tiger and Heintz TIME 2016, IJAR 2020]

- Probabilistic
  - Is the UAV inside the no-fly-zone?

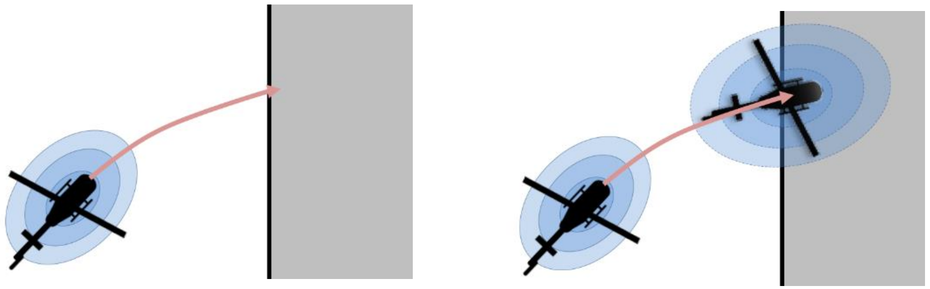
Reasoning over <u>Uncertainty</u>



collision: false   Pr(collision) = 0.1   Pr(collision) = 0.4

- Anticipatory
  - Will the UAV be colliding in the near future?

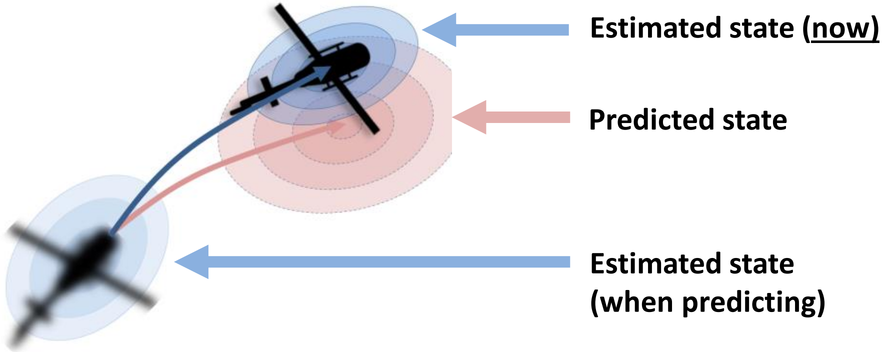
Reasoning over <u>Predictions</u>



Pr(collision now) = 0.0   Pr(collision soon) = 0.5

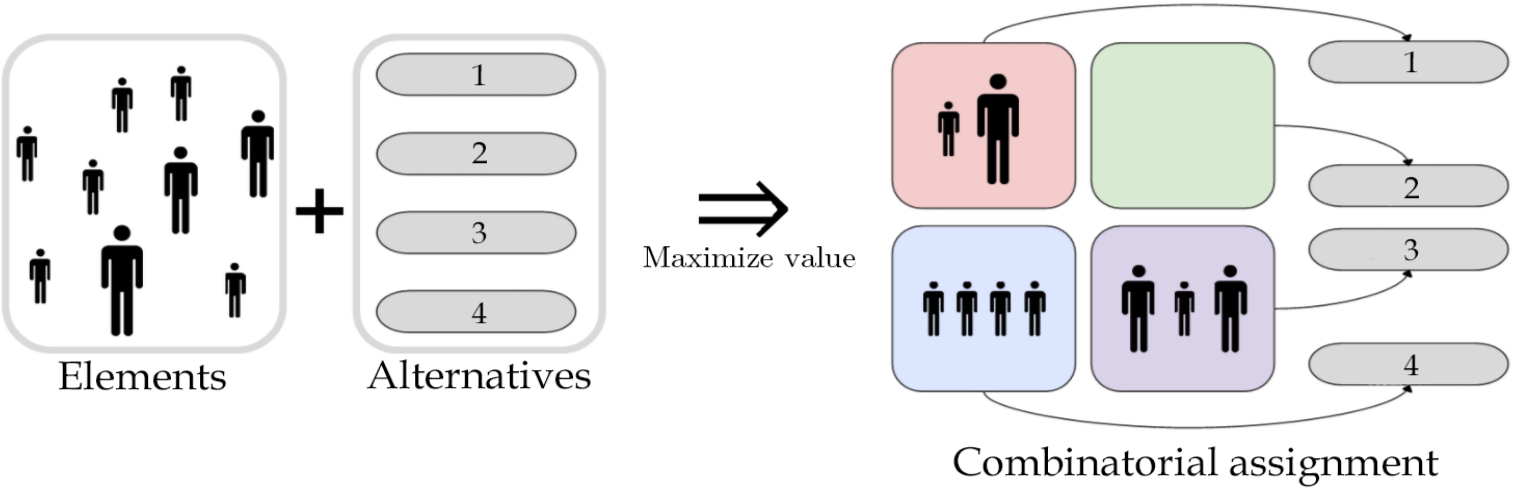
- Introspective
  - Is the prediction similar to the realization?

Reasoning <u>about</u> Predictions

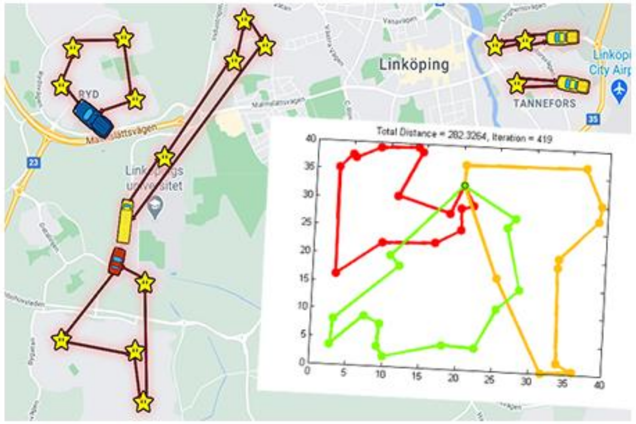


Dividing the Indivisible to Maximize Value

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).

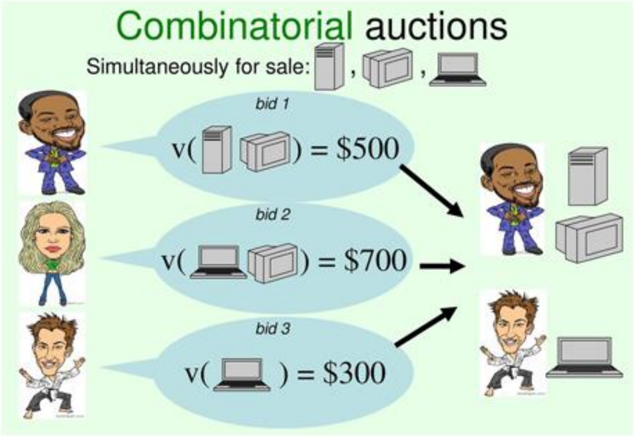


Assigning workers to jobs

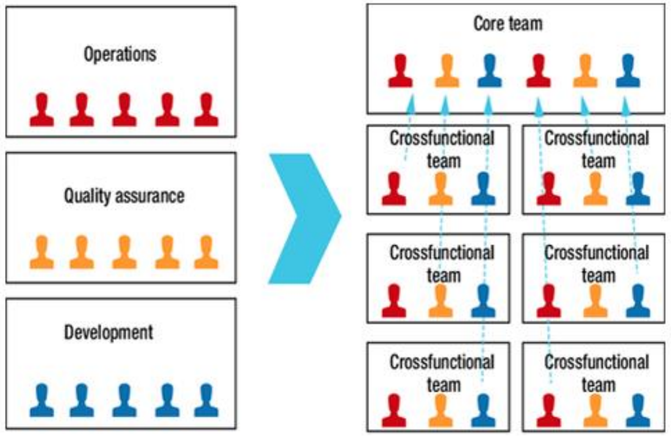


Multi-vehicle routing
(e.g., multiple TSP)

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

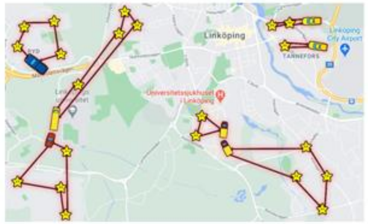
Select References

- Fredrik Präntare and Fredrik Heintz (2020). "An Anytime Algorithm for Optimal Simultaneous Coalition Structure Generation and Assignment". In: *JAAMAS*
- Fredrik Präntare and Fredrik Heintz (2020). "Hybrid Dynamic Programming for Optimal Simultaneous Coalition Structure Generation and Assignment". In: *PRIMA*
- Fredrik Präntare, Herman Appelgren, and Fredrik Heintz (2021). "Anytime Heuristic and Monte Carlo Methods for Large-Scale Simultaneous Coalition Structure Generation and Assignment". In: *AAAI*
- Fredrik Präntare, Mattias Tiger, David Bergström, Herman Appelgren, and Fredrik Heintz (2022). "Learning Heuristics for Combinatorial Assignment by Optimally Solving Subproblems". In: *AAMAS*
- Fredrik Präntare, 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

TAILOR

Foundation of Trustworthy AI:

Integrating Learning, Optimisation and Reasoning

Fredrik Heintz

Dept. of Computer Science, Linköping University

fredrik.heintz@liu.se, @FredrikHeintz

https://tailor-network.eu/

TAILOR is an ICT-48 Network of AI Research Excellence Centers funded by EU Horizon 2020 research and innovation programme GA No 952215

LINKÖPINGS UNIVERSITET

TAILOR – Vision

Develop the scientific foundations for Trustworthy AI

integrating learning, optimisation and reasoning.

This project is funded by the EC under H2020 ICT-48

TAILOR – Unique Selling Point

Actively **bringing together** communities,
especially in **reasoning and learning**,
in an **academic-industrial** network
with the **vision** and **capability**
of developing the **scientific foundations**
for realising the **European vision** of
human-centred **Trustworthy AI**.

This project is funded by the EC under H2020 ICT-48

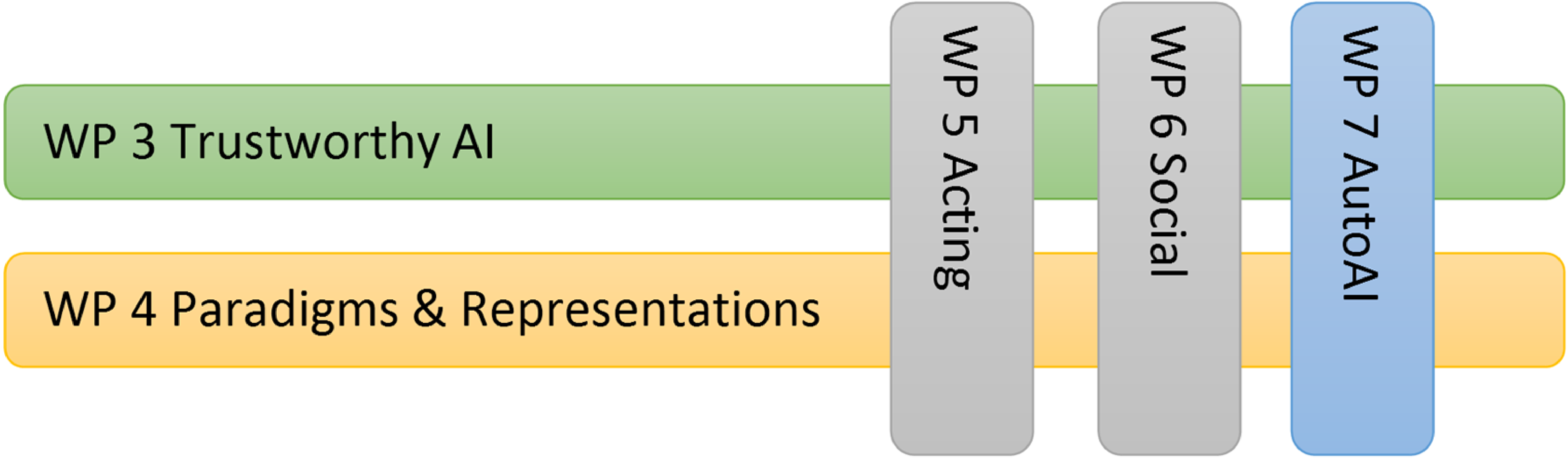
Boosting Capacity to Tackle Major Scientific Challenges

- 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.

This project is funded by the EC under H2020 ICT-48

TAILOR – Basic Research Program

TAILOR
https://tailor-network.eu/

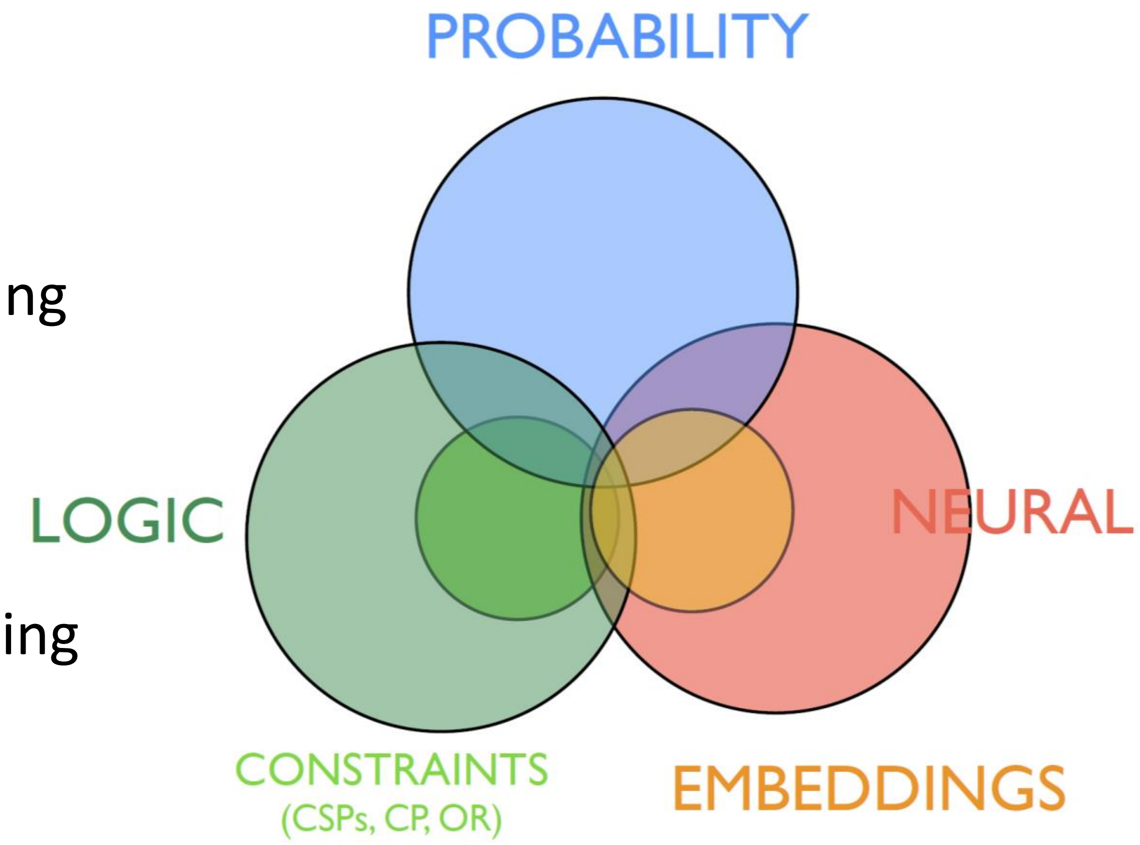


This project is funded by the EC under H2020 ICT-48

LINKÖPINGS UNIVERSITET

Paradigms and Representations

- Goals:
  - Integrate these paradigms
  - Integrate the involved communities
  - Covers five core different communities including
    - Deep & Probabilistic Learning
    - Neuro-Symbolic Computation (NeSy)
    - Statistical Relational AI (StarAI)
    - Constraint Programming & Machine Learning
    - Knowledge graphs for reasoning
    - And apply … in e.g. computer vision



This project is funded by the EC
under H2020 ICT-48

Neuro Symbolic AI

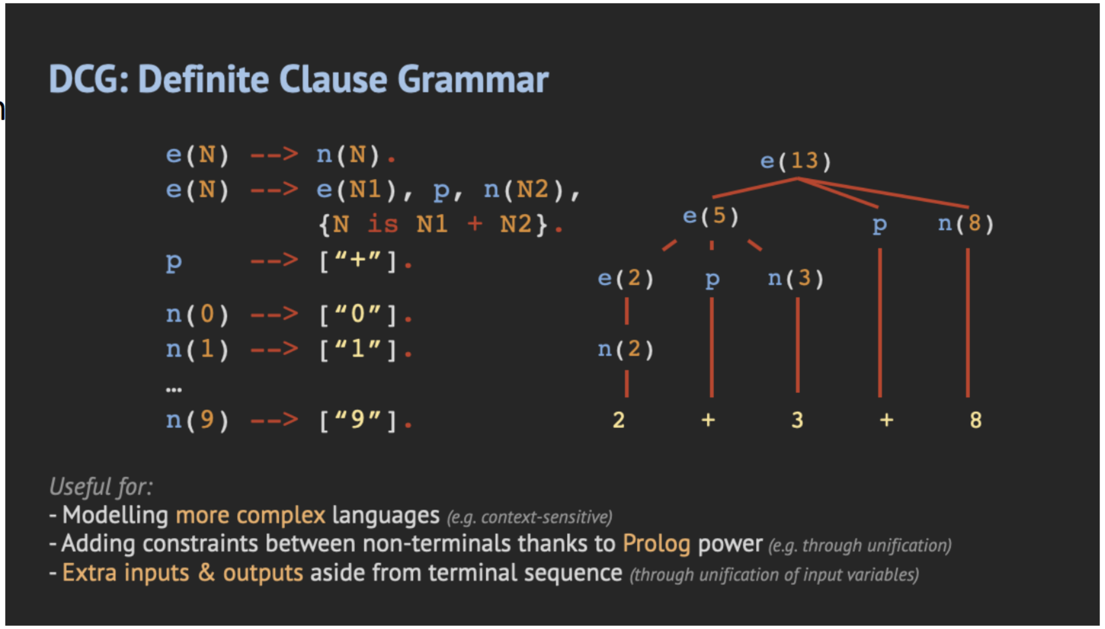
Neural Symbolic AI
one idea :

Take a symbolic (logic / rule based) representation
Turn the 0/1 orTrue/False in Fuzzy or Probabilistic Interpretation
Interpret logical predicates/functions/rules as neural networks

For instance:
map an MNIST image to a number

m() = 2
m as a neural network

mp(,2) =0.93 as a neural predicate
(with a fuzzy/prob. interpretation)



DeepStochLog (Winters et al AAAI 22)

Neuro Symbolic AI

Neural Symbolic AI
one idea :

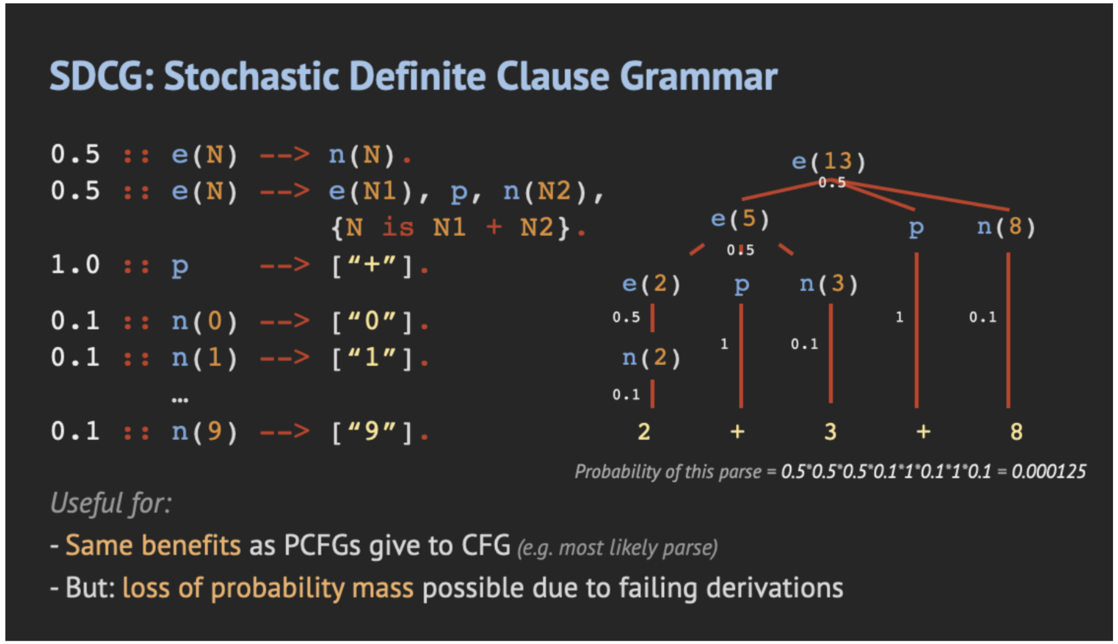
Take a symbolic (logic / rule based) representation
Turn the 0/1 True/False in Fuzzy or Probabilistic Interpretation
Interpret logical predicates/functions/rules as neural networks

For instance:
map an MNIST image to a number

m() = 2

m as a neural network

mp(,2) =0.93 as a neural predicate
(with a fuzzy/prob. interpretation)



DeepStochLog (Winters et al AAAI 22)

Neuro Symbolic AI

Neural Symbolic AI
one idea :

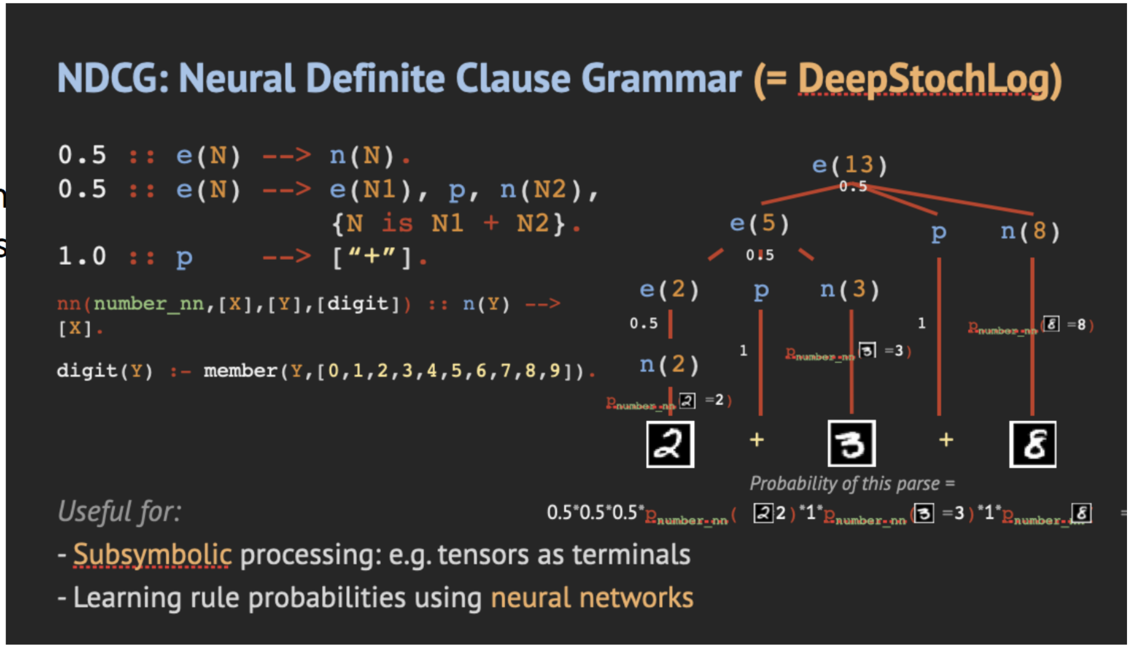
Take a symbolic (logic / rule based) representation
Turn the 0/1 True/False in Fuzzy or Probabilistic Interpretation
Interpret logical predicates/functions/rules as neural networks

For instance:
map an MNIST image to a number

m() = 2

m as a neural network

mp(,2) =0.93 as a neural predicate
(with a fuzzy/prob. interpretation)



DeepStochLog (Winters et al AAAI 22)

This project is funded by the EC
under H2020 ICT-48

LINKÖPINGS UNIVERSITET

Learning and Optimization

Empirical Model Learning (introduced by the UniBo group, 2012, Milano and Lombardi)

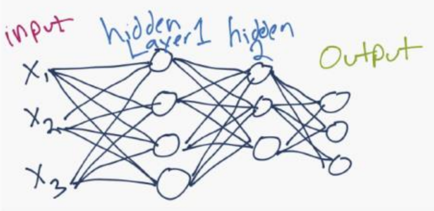
Goal: deal with optimization problems defined over complex systems, and having non-trivial constraints

Step 1: define the core combinatorial structure

$$\min f(x,\ y,\ z)$$
$$x,\ y,\ z \in F$$

- Any cost function
- Any kind of constraint
- ...Just use a suitable solver

Step 2: obtain a ML model for the complex system



$$z = h(x)$$

Step 3: convert the ML model into constraints/predicates

$$\min f(x,\ y,\ z)$$
$$x,\ y,\ z \in F$$
$$z = h(x)$$

- Merge the two models
- ...And solve as before

Currently:

- Support for Neural Networks and Decision Trees
- Support for Constraint Programming, SMT, and Mathematical Programming
- Training done once, prior to search

Also related techniques such as Smart Predict & Optimise

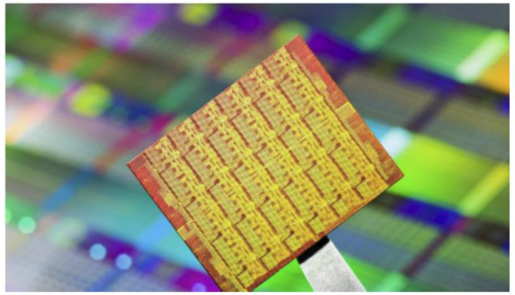
Learning and Optimization

Empirical Model Learning (introduced by the UniBo group, 2012, Milano and Lombardi)
Goal: deal with optimization problems defined over complex systems, and having non-trivial constraints

Thermal Aware Job Allocation

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

- Many-core CPU (Intel SCC, 2009, 48 cores, Xeon Phi precursor)
- Dispatch jobs
- Load balancing constraints
- **Objective:** avoid thermal hot-spots (efficiency loss)



A Case Study: Traffic Light Placement

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

- Add/remove traffic lights in a city
- Traffic lights can be connected (green wave)
- Every operation has a cost
- Budget limit
- **Objective:** improve traffic flow

This project is funded by the EC under H2020 ICT-48

Trustworthy AI – TAILOR Perspective

TAILOR
https://tailor-network.eu/

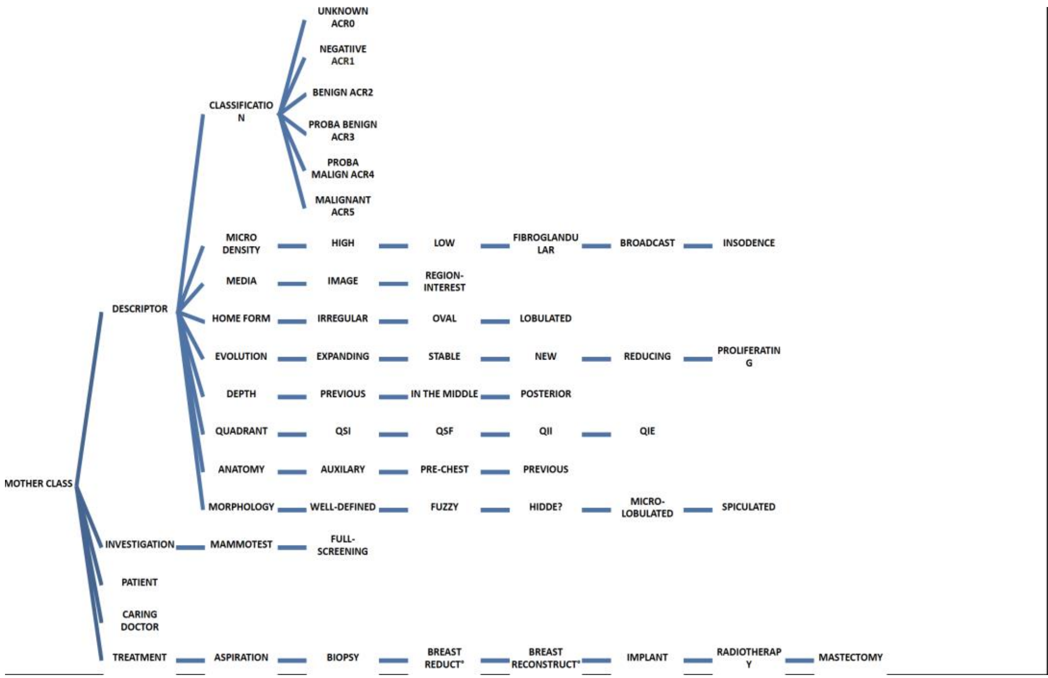
- Goal
  - establish a continuous interdisciplinary dialogue for investigating methods and methodologies
  - ***"To create AI systems that incorporate trustworthiness by design"***
- Organized along the 6 dimensions of Trustworthy AI:
  - Explainability,
  - Safety and Robustness,
  - Fairness,
  - Accountability,
  - Privacy, and
  - Sustainability
- One transversal task that links the 6 dimensions among and ensures coherence and coordination across the activities.

This project is funded by the EC under H2020 ICT-48

Fuzzy Reasoning and Learning: Mammography with BI-RADS attributes

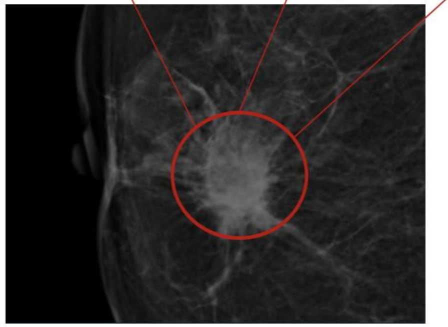
- Hybrid approach
  - Fuzzy Reasoner + ML for interpretable and explainable
- Input:
  - Mammography Ontology and Data
  - Target class: MalignantTumor
- Learned Output: e.g,

"A mammography region of old woman, whose density mass is high, whose margin is spiculated, and whose shape is irregular is a malignant tumour"



(hasAge some hasAgeHigh) and (hasDensity some hasDensityHigh) and (hasMargin some spiculated) and (hasShape some irregular) SubClassOf MalignantTumour

0.98 0.84 0.78



In red: fuzzy concepts

Integration with Image/Text Classifiers (mammography/anamnesis) ongoing

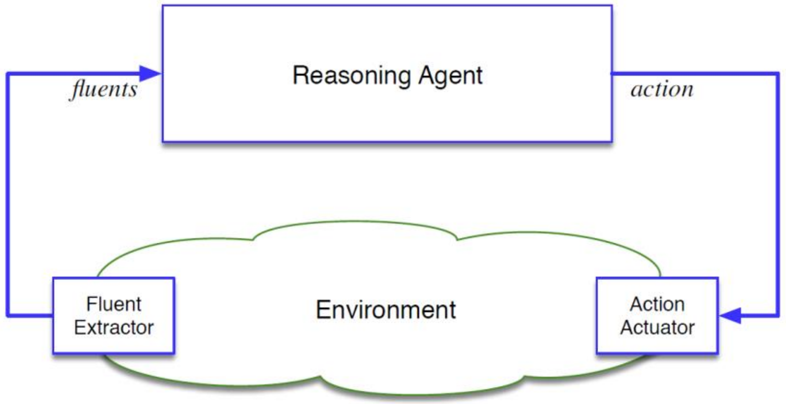
Franco Alberto Cardillo and Umberto Straccia (2021).
Fuzzy OWL-BOOST: Learning Fuzzy Concept Inclusions via Real-Valued Boosting.
In *Fuzzy Sets and Systems,* Elsevier. DOI: https://doi.org/10.1016/j.fss.2021.07.002

LINKÖPINGS UNIVERSITET

Reasoning Agents and Learning Agents

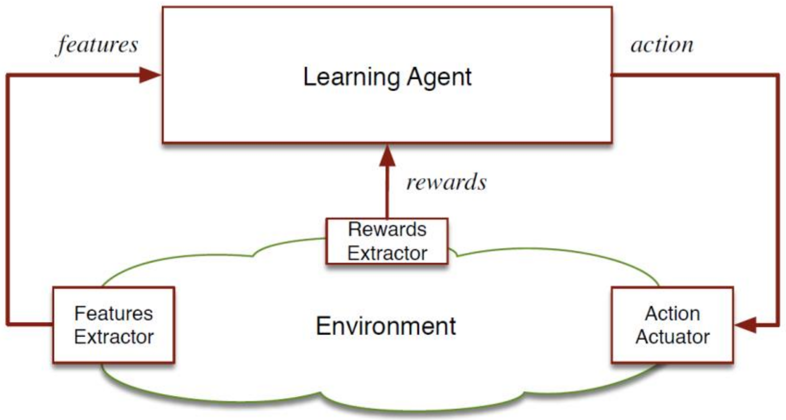
Reasoning agent:

- Senses and acts on the environment
- Has model of its environment and task
- Does Planning



Learning agent:

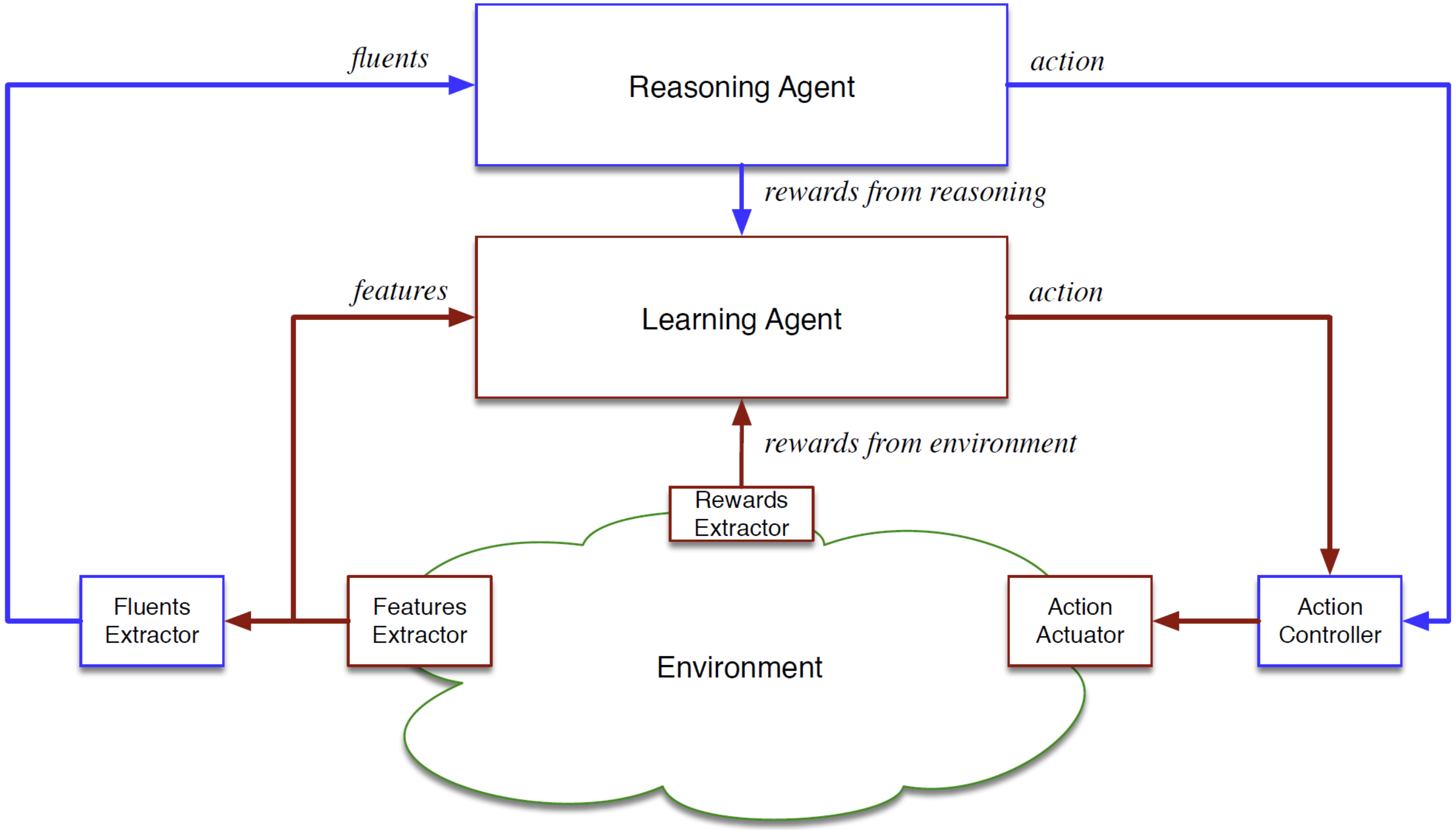
- Senses and acts on the environment
- Gets rewards when right
- Does Reinforcement Learning



Reasoning and Learning Agents

Merging:

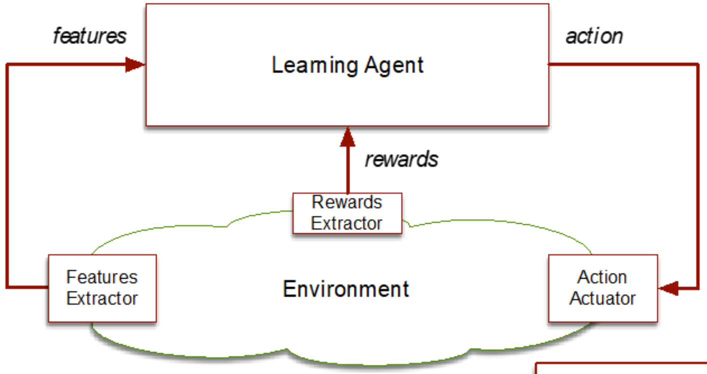
- Reasoning agent
  - E.g. reasoning in temporal logics
- Learning agent
  - E.g. doing reinforcement learning



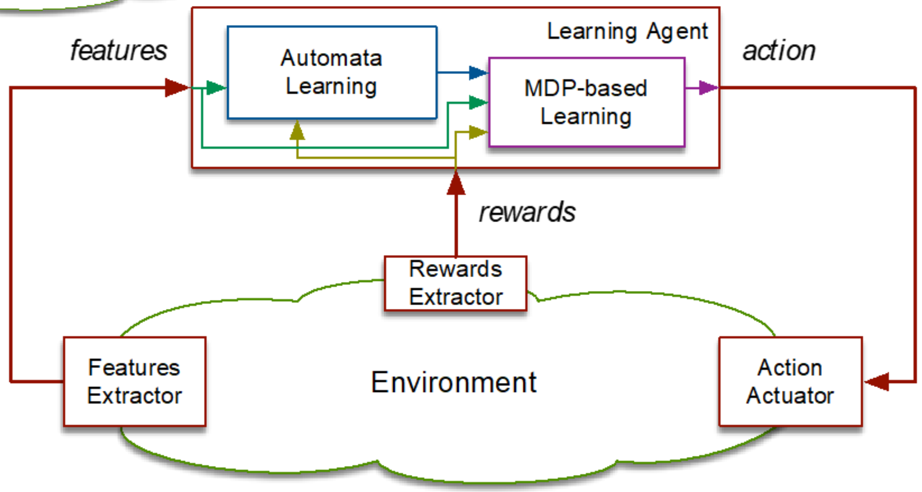
This project is funded by the EC under H2020 ICT-48

Challenge: RL in non-Markovian Domains

- Reinforcement Learning in non-Markovian Domains
- Based on Regular Decision Processes (RDP) instead of MDPs
- Handle non-Markovian dynamics (i.e., depending on the history) without postulating a priori existence of hidden variable, as in POMDPs!
- RL on RDPs requires simultaneously learning an automaton for the dynamics and an optimal policy wrt rewards:
  - Polynomial PAC-learnability
  - With no prior knowledge

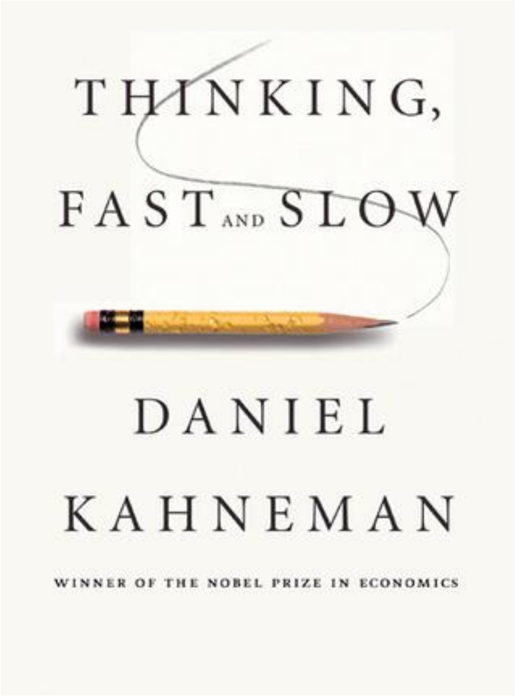


E. Abadi, R. Brafman. Learning and Solving Regular Decision Processes. IJCAI 2020



A. Ronca, G. De Giacomo. Efficient PAC Reinforcement Learning in Regular Decision Processes. IJCAI 2021

The Way Forward



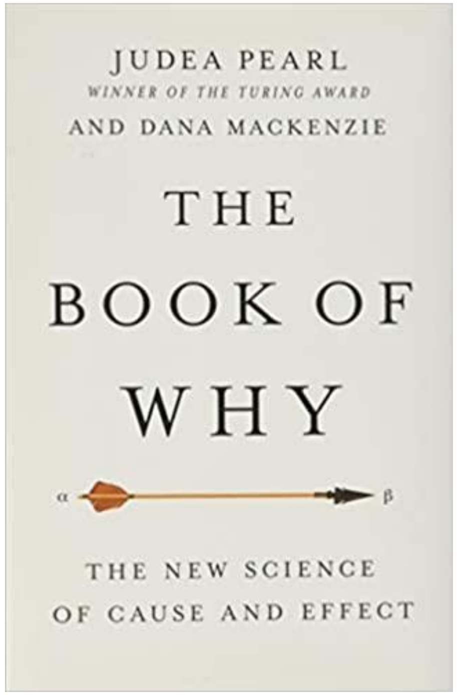
Data

Knowledge/
Assumptions

Causal Models
Combining reasoning
and learning

Explanations

Predictions



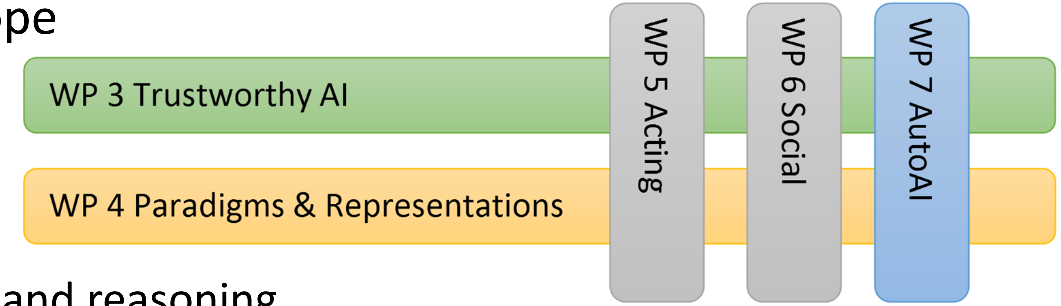
This project is funded by the EC
under H2020 ICT-48

TAILOR ICT-48 Network

https://tailor-network.eu/

*TAILOR brings together 54 leading AI research centres from **learning**, **optimisation** and **reasoning** together with major European companies representing important industry sectors into a single scientific network addressing the **scientific foundations of Trustworthy AI** to reduce the fragmentation, boost the collaboration, and increase the AI research capacity of Europe as well as attracting and retaining talents in Europe.*

- 54 research excellence centres from 20 countries across Europe coordinated by Fredrik Heintz, Linköping University, Sweden
- Four instruments
  - An ambitious research and innovation roadmap
  - Five basic research programs integrating learning, optimisation and reasoning in key areas for providing the scientific foundations for Trustworthy AI
  - A connectivity fund for active dissemination to the larger AI community
  - Network collaboration promoting research exchanges, training materials and events, and joint PhD supervision

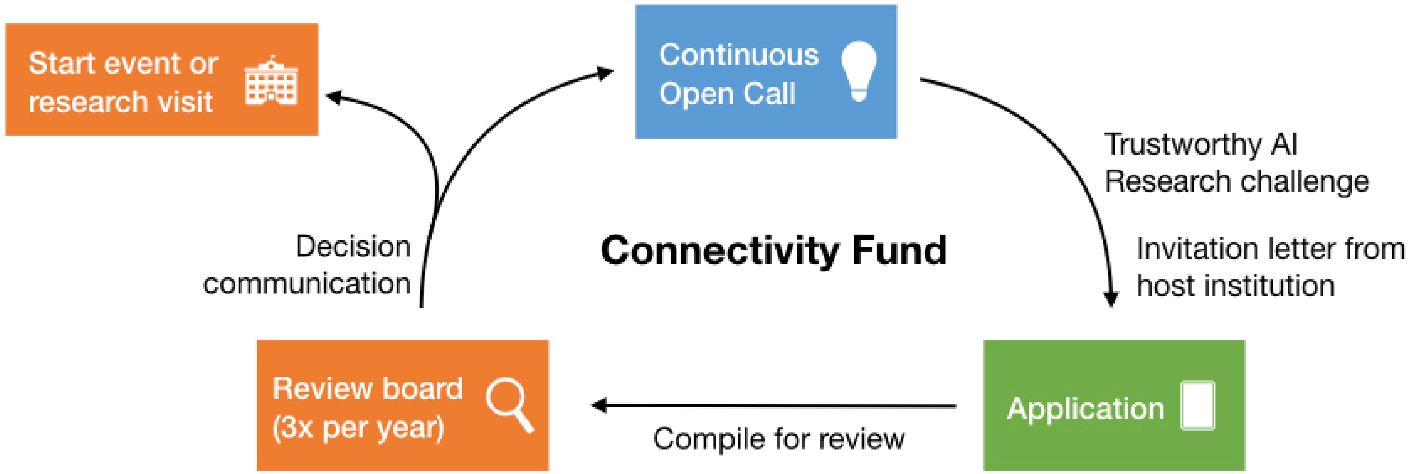


This project is funded by the EC under H2020 ICT-48

Connectivity Fund

Call 4 closes Mar 15!

- 1.5 million EUR fund, third-party funding (guest or host is non-TAILOR)
- Open call, reviewed every 4 months (March, July, November)
  - Submitted by non-TAILOR host or guest
  - Max. 60.000 EUR per visit/workshop, covers travel, housing, and sustenance
- https://tailor-eu.github.io/connectivity-fund/



Research Visits

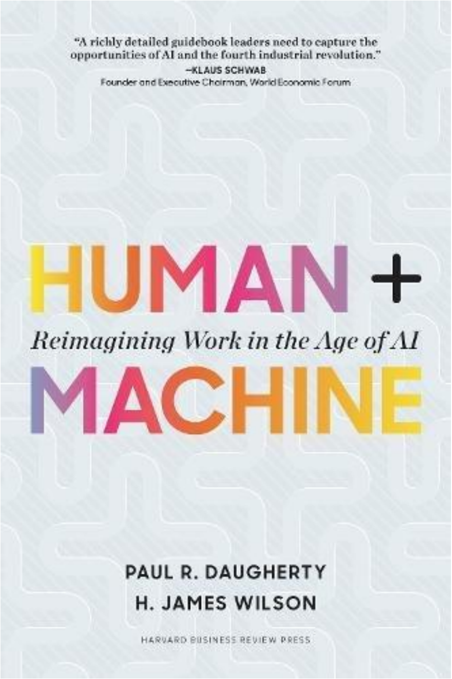
We support research visits between 1 and 12 months. We will pick up the bills so that you can focus on doing excellent AI. You must either be from a non-TAILOR lab visiting a TAILOR lab, or vice versa.

Workshops

We support workshops that bring people all across Europe together to solve hard problems in an open atmosphere. Workshops should explicitly bring TAILOR and Non-TAILOR researchers together.

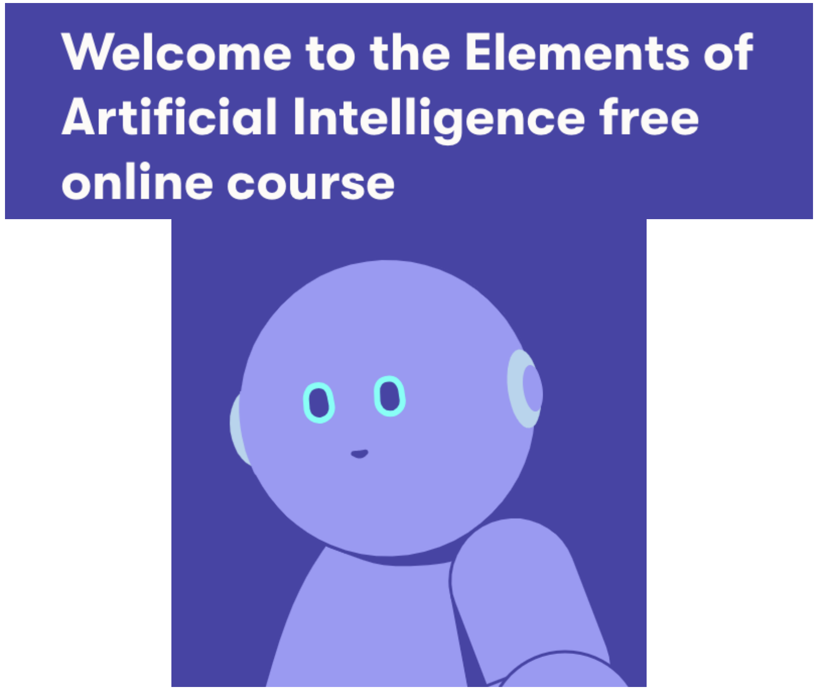
Other Components to Achieve Trustworthy AI

Humans + AI



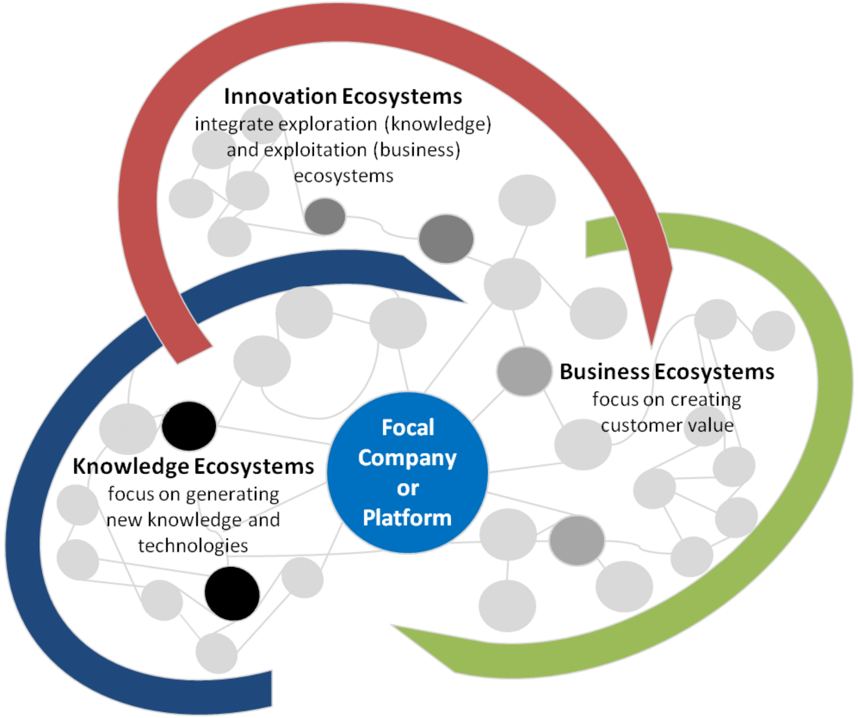
https://knowledge.wharton.upenn.edu/article/reimagining-work-age-ai/

Education



https://elementsofai.se

Ecosystems



https://timreview.ca/article/919

AI Innovation, Competence and Research Ecosystem

AI SUSTAINABILITY CENTER

AI SWEDEN

AI INNOVATION of Sweden

Elements of AI

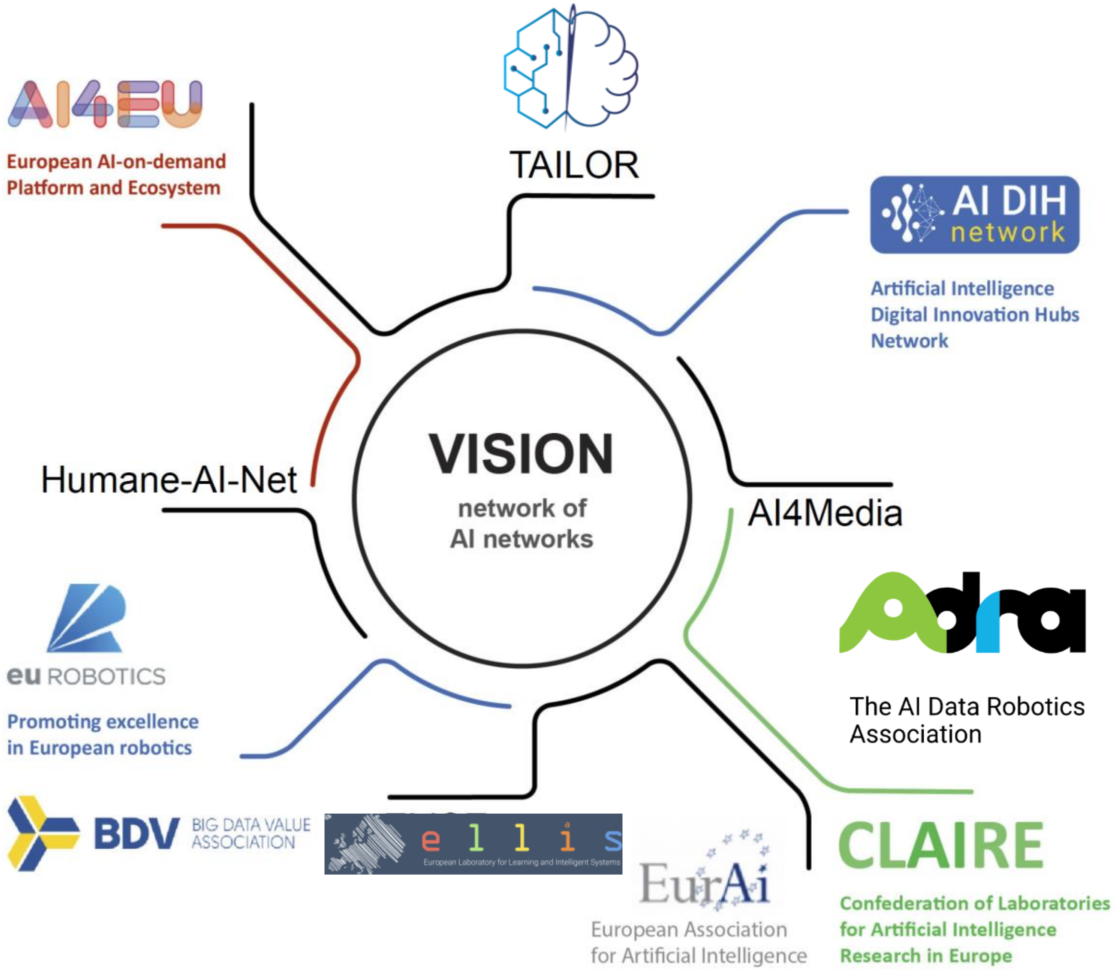
AI Competence of Sweden

WASP-ED

WASP—HS

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

CHALMERS · KTH VETENSKAP OCH KONST · LINKÖPINGS UNIVERSITET · LUNDS UNIVERSITET · UMEÅ UNIVERSITET · UPPSALA UNIVERSITET · ÖREBRO UNIVERSITY



Take Away Message

- AI is about understanding intelligence and develop systems that exhibit intelligent behavior.
- AI will affect all aspects of our society. **Trust is essential!**
- To be **trustworthy** an **AI-system** should be **legal**, **ethical** and **robust**.
- Approaches to address the challenges include
  - Human + AI
  - Education
  - Research
  - Ecosystems
- Very active and interdisciplinary research problems that are still mostly unsolved.
- Europe has **many initiatives** in the area, but **more** is needed.
- **The TAILOR project is committed to develop the scientific foundations for Trustworthy AI**
- **Will most likely require integrating model-free data-driven learning approaches with model-based knowledge-driven reasoning approaches**