TAILOR

- Developing the Scientific Foundations for Trustworthy AI

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

LINKÖPINGS UNIVERSITET

Ethics Guidelines for Trustworthy AI – Overview

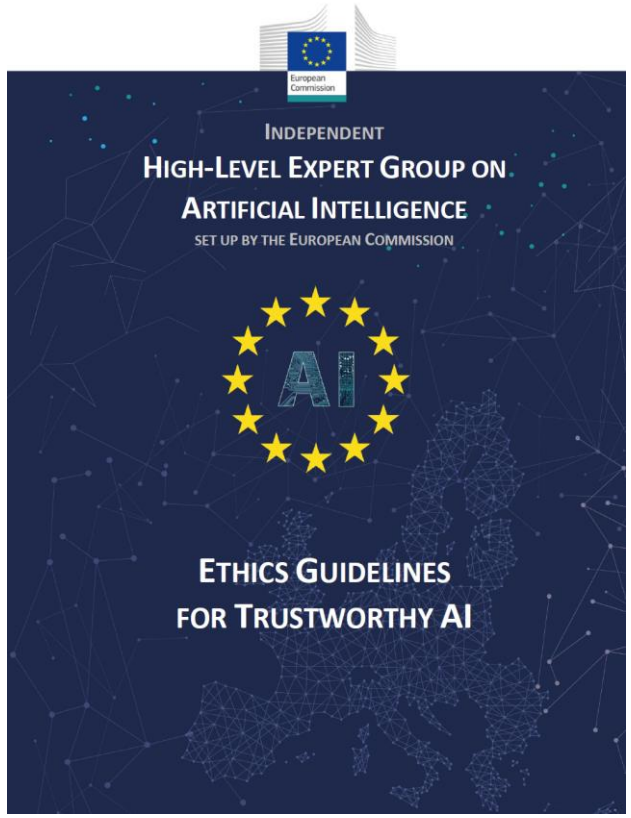
Human-centric approach: AI as a means, not an end

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

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

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.

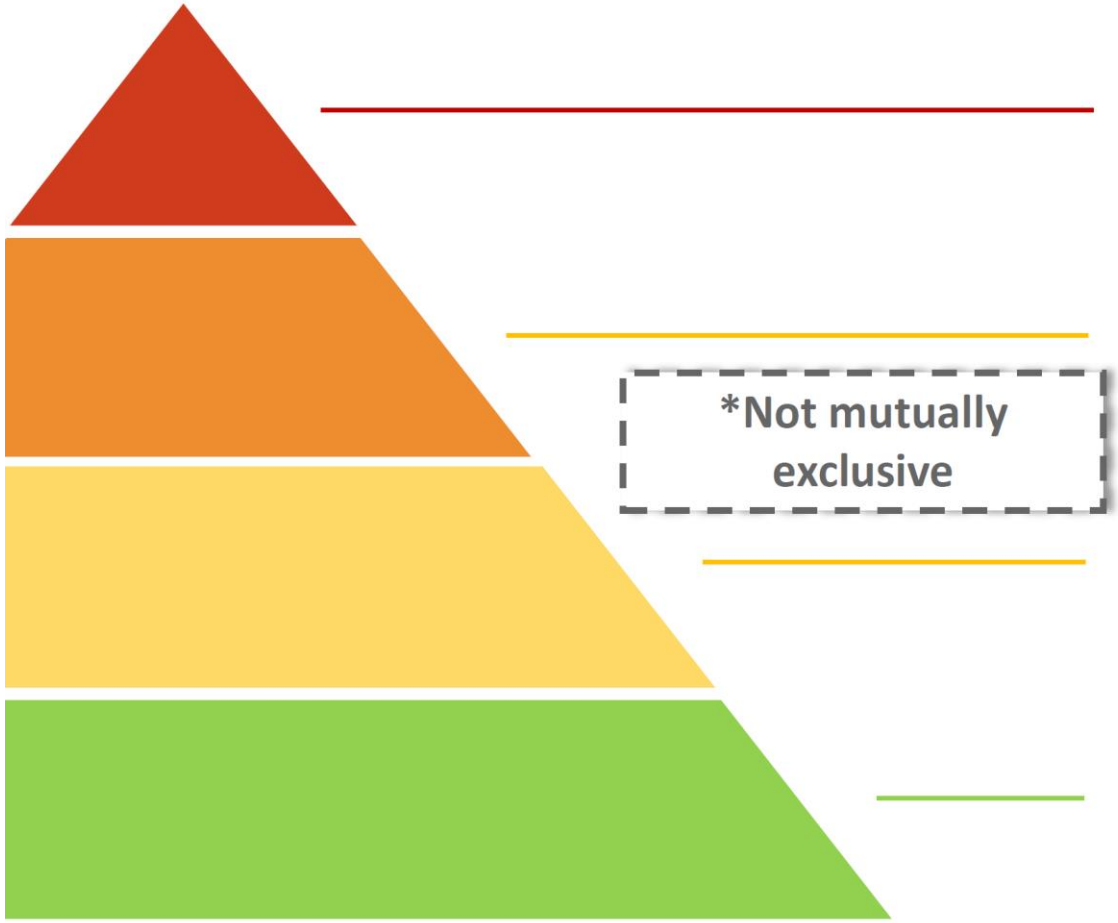
Explicability

Transparent, open with capabilities and purposes, explanations

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

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

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

European Commission

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

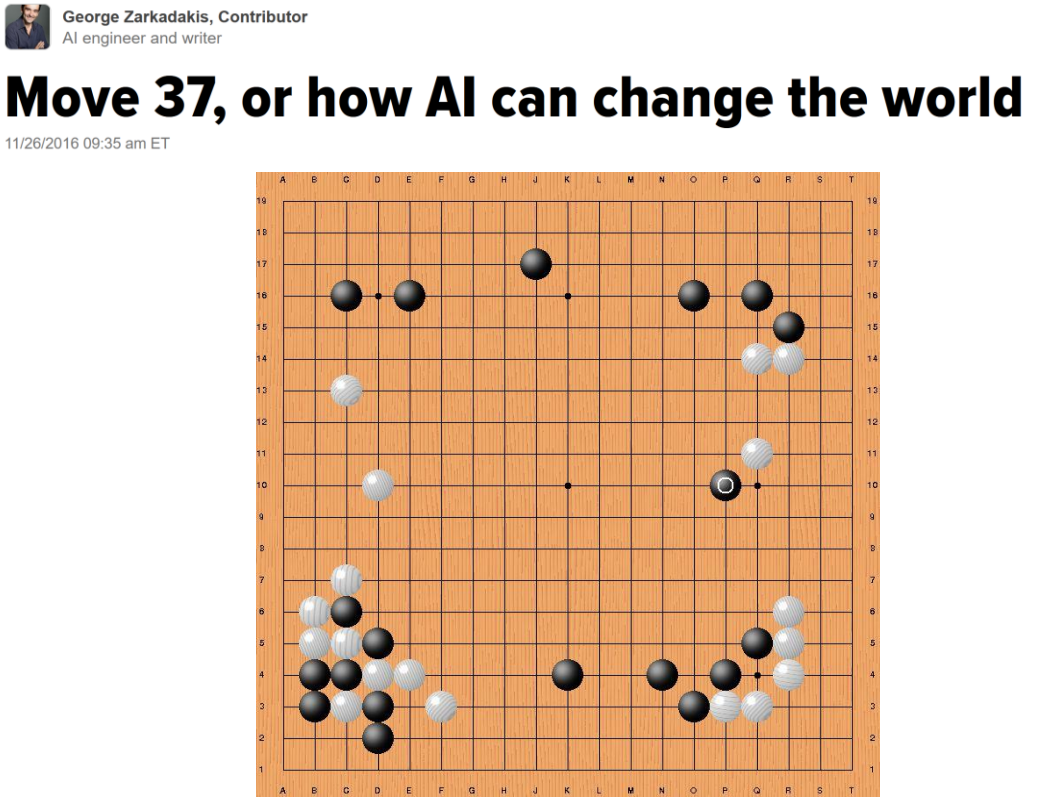
Establish and implement **risk management system**

&

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**

How to Evaluate AI Systems?



https://www.huffpost.com/entry/move-37-or-how-ai-can-change-the-world_b_58399703e4b0a79f7433b675

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

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

TAILOR
https://tailor-network.eu/

TAILOR Consortium

- 54 partners from 18 EU countries (AT, BE x2, CZ x2, DE x8, ES x4, FI, FR x6, GR, IE, IT x8, LU, NL x6, PL, PT, SE x2, SI, SK, UK x4), Israel and Switzerland x2.
- More than 60 network members.
- 23 Core partners (LiU, CNR, INRIA, UCC, KUL, UOR, LEU, IST-UL, UPF, UNIBO, BIU, TUE, CNRS, JSI, TUDA, UNIBRIS, ALU-FR, UOX, UNITN, DFKI, EPFL, FBK, CINI)
- 21 Partners (VUB, CUNI, CEA, CRIL, CVUT, TUD, FhG, TU Graz, IIIA-CSIC, LIRA, UOA, NEO-UMA, PUT, RWTH, slovak.AI, TNO, UniPI, UGA, UNIBAS, UPV, ICL)
- 10 Industry partners (VW, ENG, Tieto, Philips, EDF, ABB, ZF, LIH, CBS, Bosch)

CLAIRE

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

LINKÖPINGS UNIVERSITET

TAILOR
https://tailor-network.eu/

TAILOR Objectives

O1: Establish	O2: Define and maintain	O3: Create	O4: Build	O5: Progress	O6: Increase
O1: Establish a strong pan-European network of research excellence centers on the Foundations of Trustworthy AI	O2: Define and maintain a unified strategic research and innovation roadmap for the Foundations of Trustworthy AI	O3: Create the capacity and critical mass to develop the scientific foundations for Trustworthy AI	O4: Build sustained collaborations with academic, industrial, governmental, and community stakeholders on the Foundations of Trustworthy AI	O5: Progress the Scientific State-of-the-Art for the Foundations of Trustworthy AI	O6: Increase Knowledge and Awareness of the Foundations of Trustworthy AI across Europe

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

LINKÖPINGS UNIVERSITET

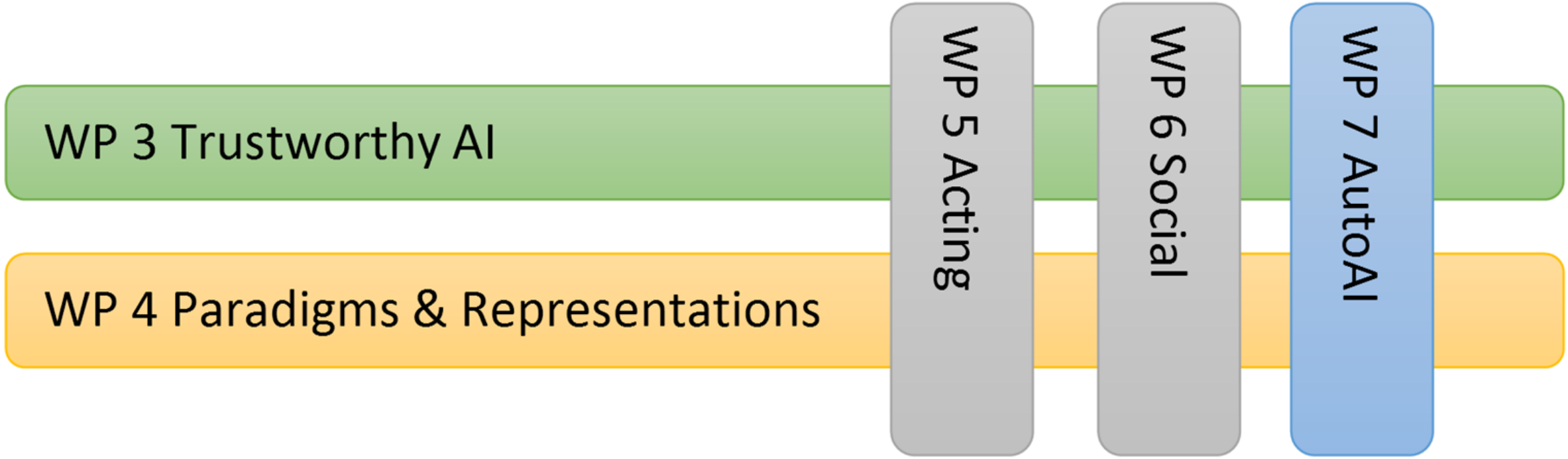
Boosting Capacity to Tackle Major Scientific Challenges

- 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
https://tailor-network.eu/

TAILOR – Basic Research Program



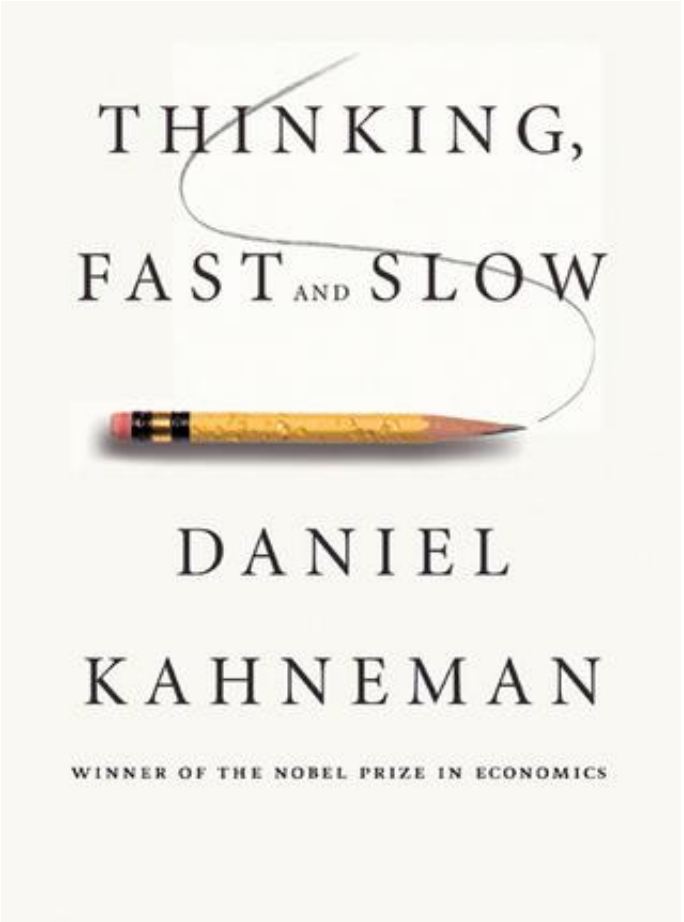
This project is funded by the EC
under H2020 ICT-48

LINKÖPINGS UNIVERSITET

Human and Computational Thinking

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

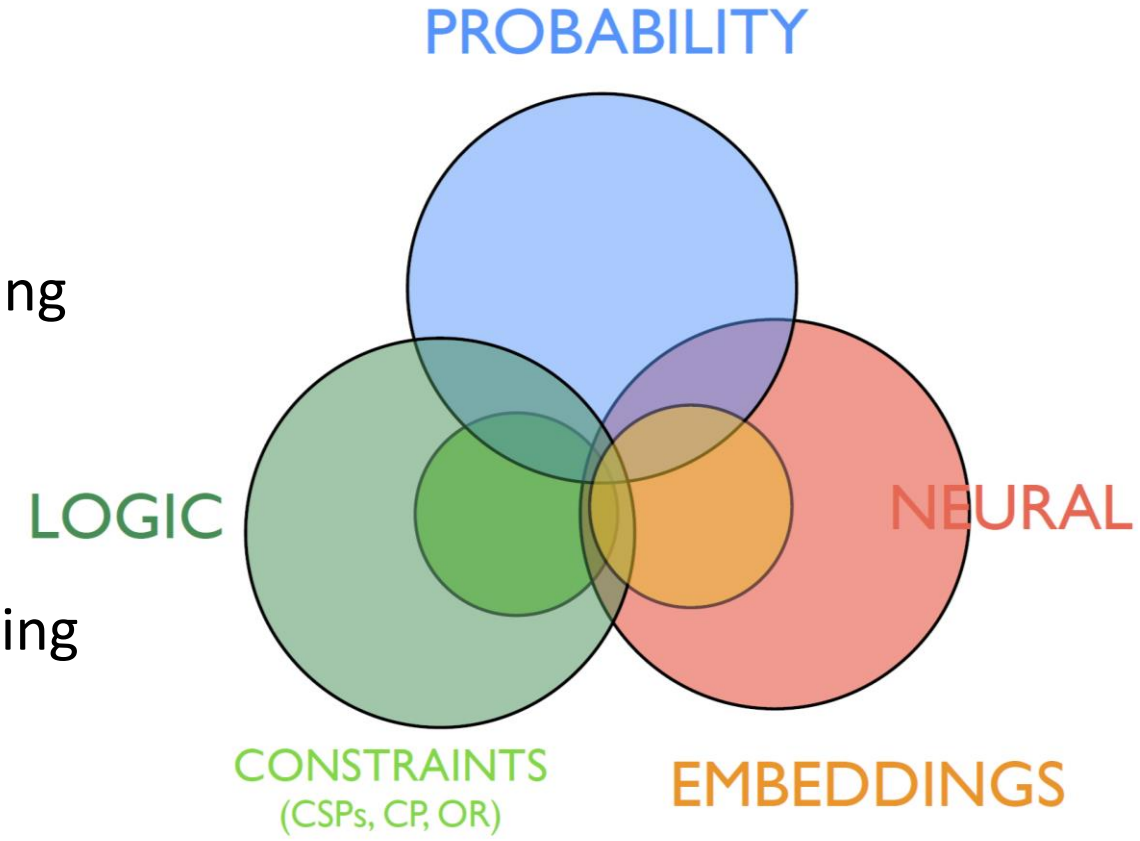
System 1 "Fast"	System 2 "Slow"
DEFINING CHARACTERISTICS Unconscious Effortless Automatic	DEFINING CHARACTERISTICS Deliberate and conscious Effortful Controlled mental process
WITHOUT self-awareness or control	WITH self-awareness or control
"What you see is all there is."	Logical and skeptical
ROLE Assesses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



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

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

Learning and Optimization

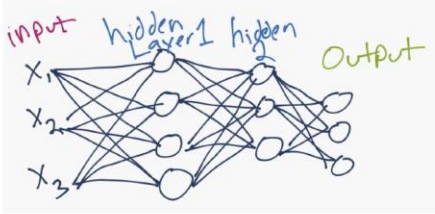
Empirical Model Learning (introduced by the UniBo group, 2012, Milano and Lombardi)
Goal: deal with optimization problems defined over complex systems, and having non-trivial constraints

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:

Also related techniques such as Smart Predict & Optimise

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

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

TAILOR
https://tailor-network.eu/

Trustworthy AI – TAILOR Perspective

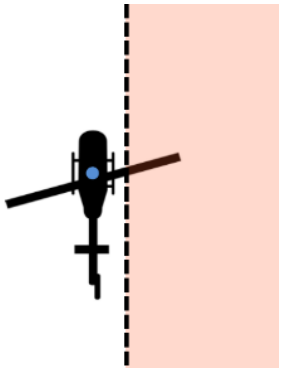
- Goal
  - establish a continuous interdisciplinary dialogue for investigating methods and methodologies
  - ***"To create AI systems that incorporate trustworthiness by design"***
- Organized along the 6 dimensions of Trustworthy AI:
  - Explainability,
  - Safety and Robustness,
  - Fairness,
  - Accountability,
  - 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

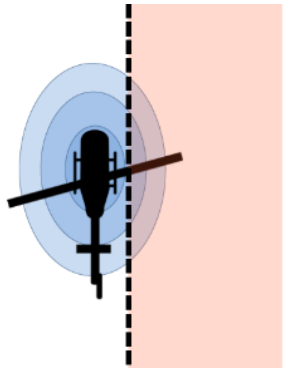
LINKÖPINGS UNIVERSITET

Probabilistic logical reasoning over observed and predicted trajectories

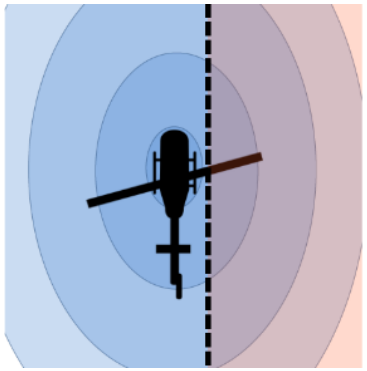
[Tiger and Heintz TIME 2016, IJAR 2020]



collision: false

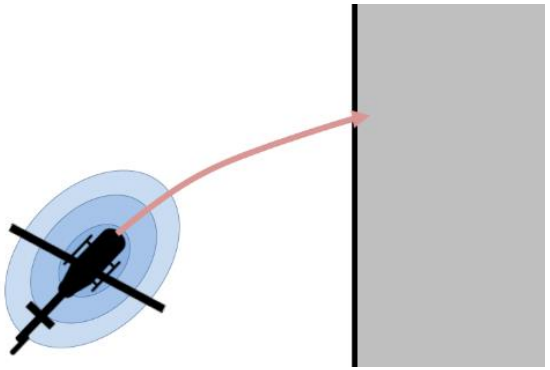


Pr(collision) = 0.1

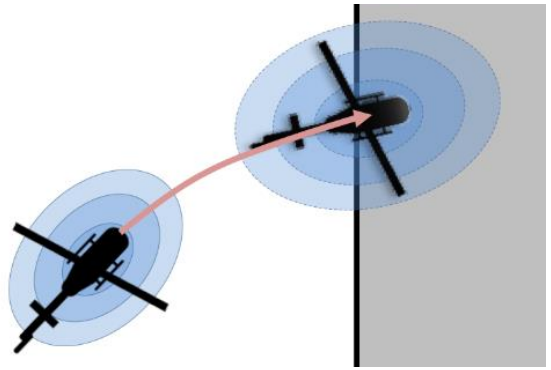


Pr(collision) = 0.4

Reasoning over Uncertainty



Pr(collision now) = 0.0...



Pr(collision soon) = 0.5

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.

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

Probabilistic logical reasoning over observed and predicted trajectories

[Tiger and Heintz TIME 2016, IJAR 2020]

always ($altitude_0 > 3$)

true

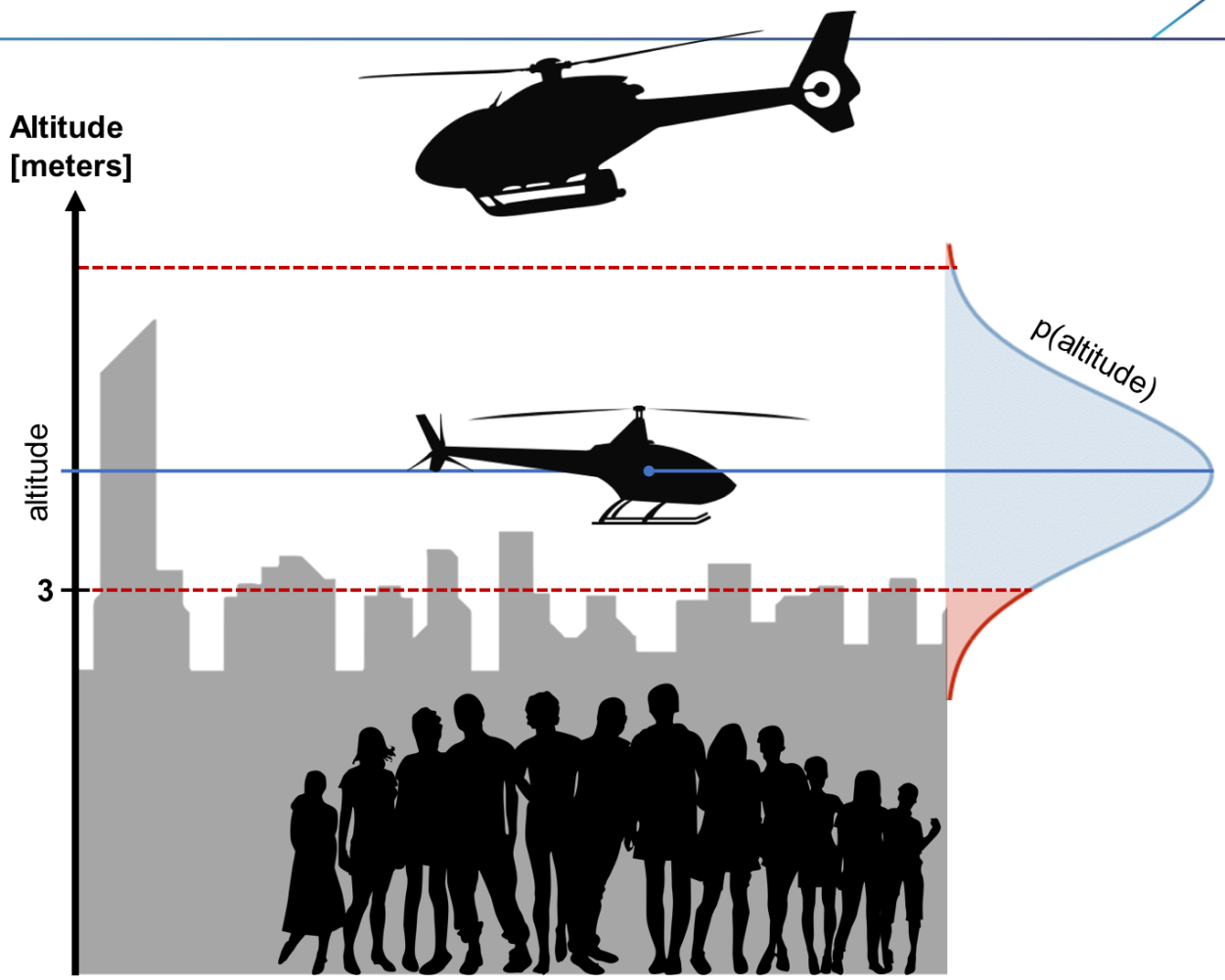
always ($Pr(altitude_{0|0} > 3) \geq 0.99$)

false

always ($Pr(altitude_{2|0} > 3) \geq 0.99$)

Relative time to estimate

Relative time to estimate from



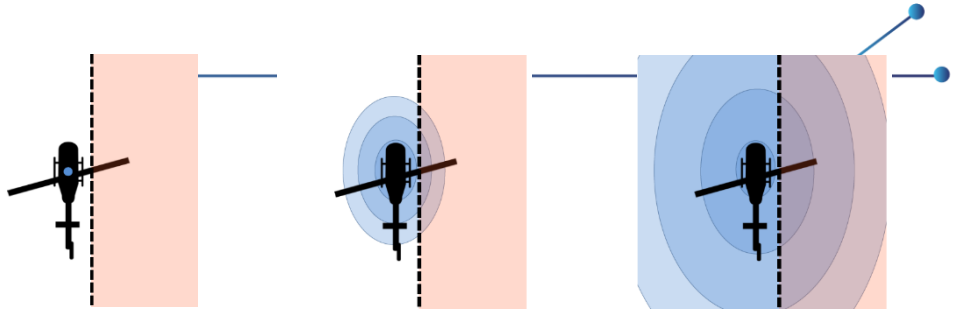
This project is funded by the EC
under H2020 ICT-48

Probabilistic logical reasoning over observed and predicted trajectories

[Tiger and Heintz TIME 2016, IJAR 2020]

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

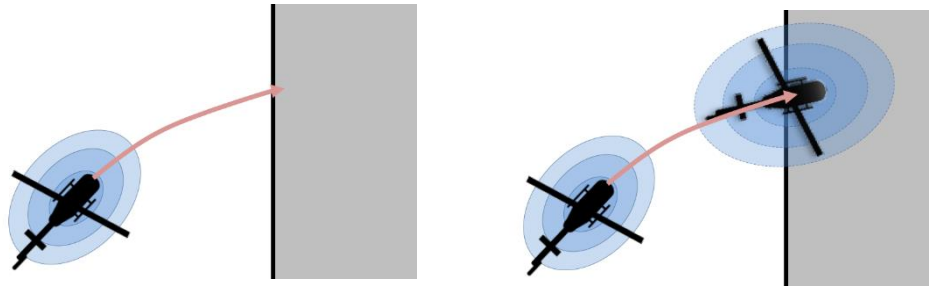
Reasoning over Uncertainty



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

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

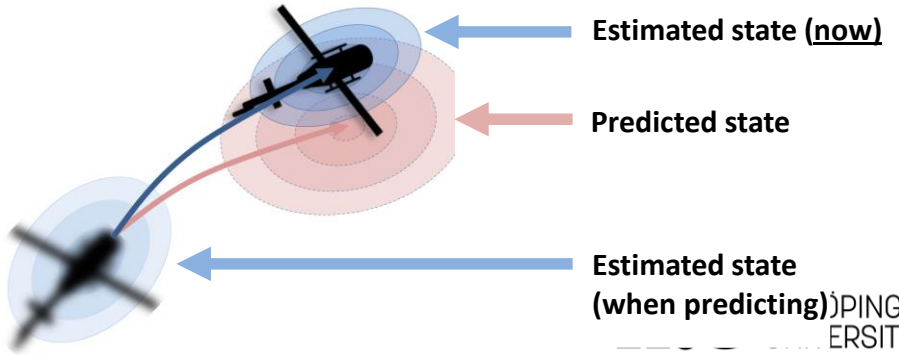
Reasoning over Predictions



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

- Introspective
  - Is the prediction similar to the realization?

Reasoning about Predictions



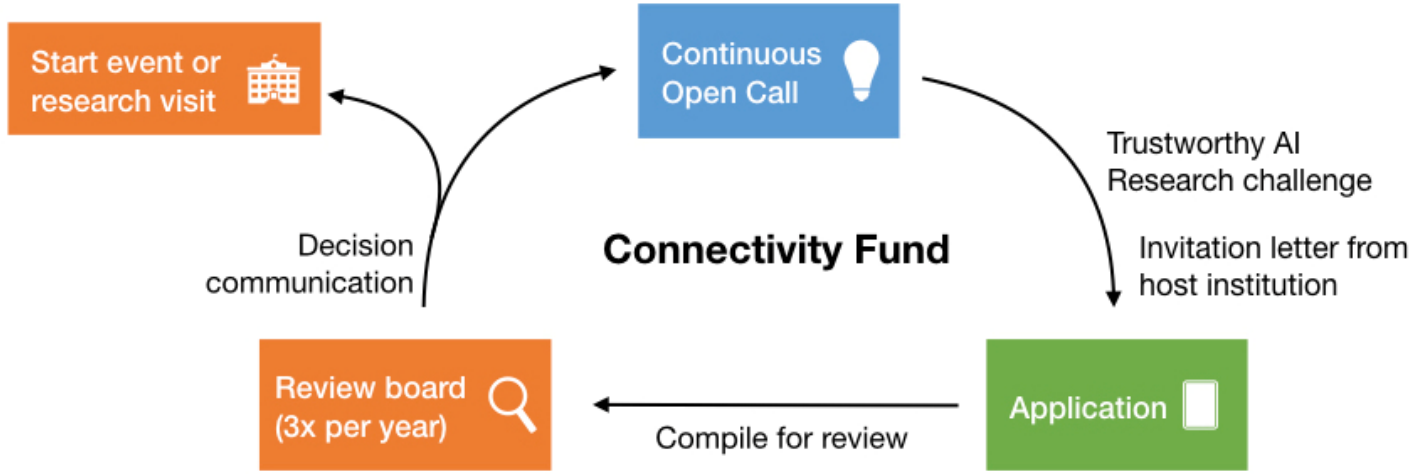
This project is funded by the EC under H2020 ICT-48

TAILOR
https://tailor-network.eu/

Connectivity Fund

Call 6 closes Nov 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.

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

LINKÖPINGS UNIVERSITET

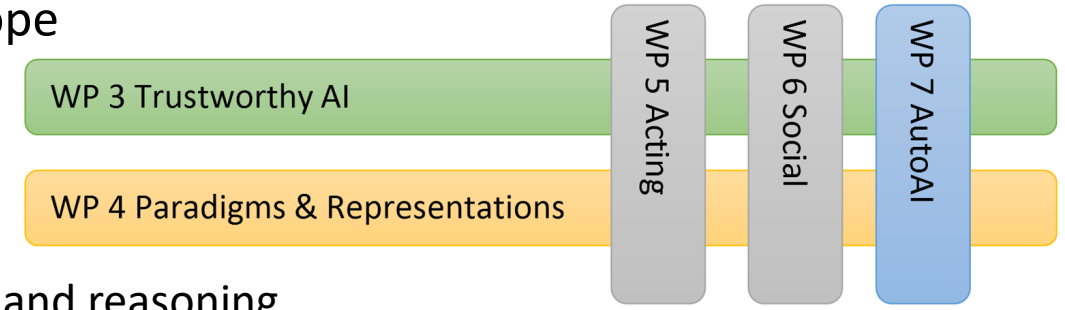
TAILOR ICT-48 Network

TAILOR https://tailor-network.eu/

CLAIRE

*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.*

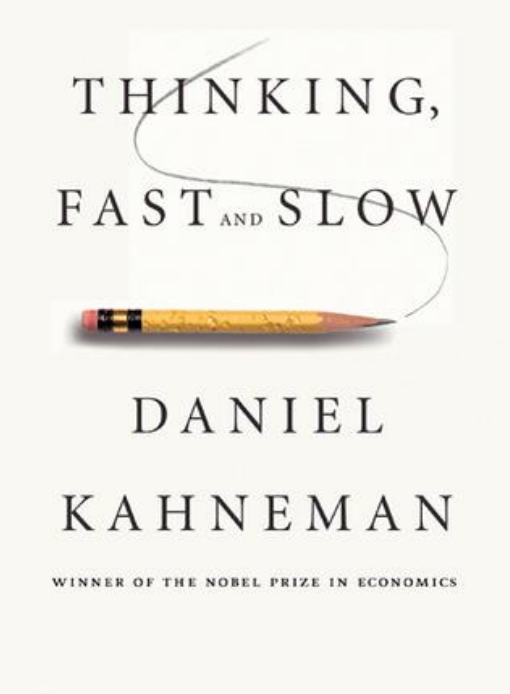
- 54 research excellence centres from 20 countries across Europe coordinated by Fredrik Heintz, Linköping University, Sweden
- Four instruments
  - An ambitious research and innovation roadmap
  - Five basic research programs integrating learning, optimisation and reasoning in key areas for providing 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

LINKÖPINGS UNIVERSITET

The Way Forward



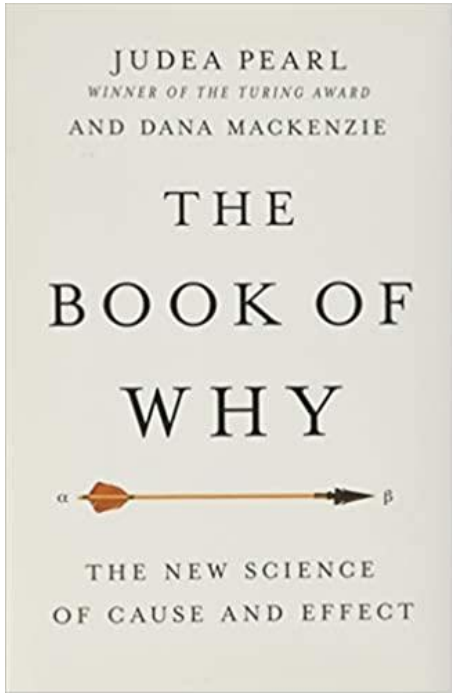
Data

Knowledge/
Assumptions

Causal Models
Combining reasoning
and learning

Explanations

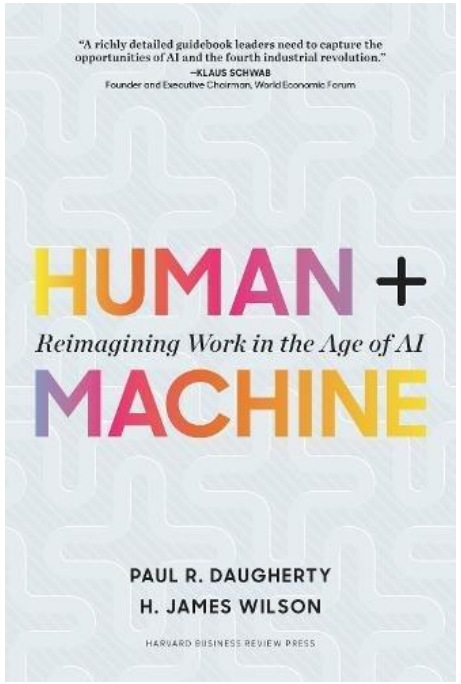
Predictions



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

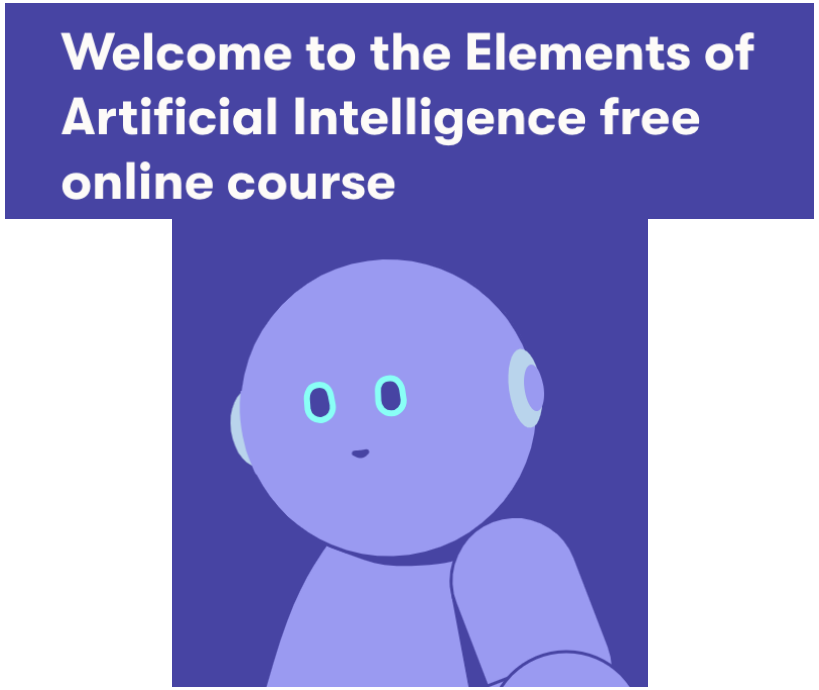
Other Components to Achieve Trustworthy AI

Humans + AI



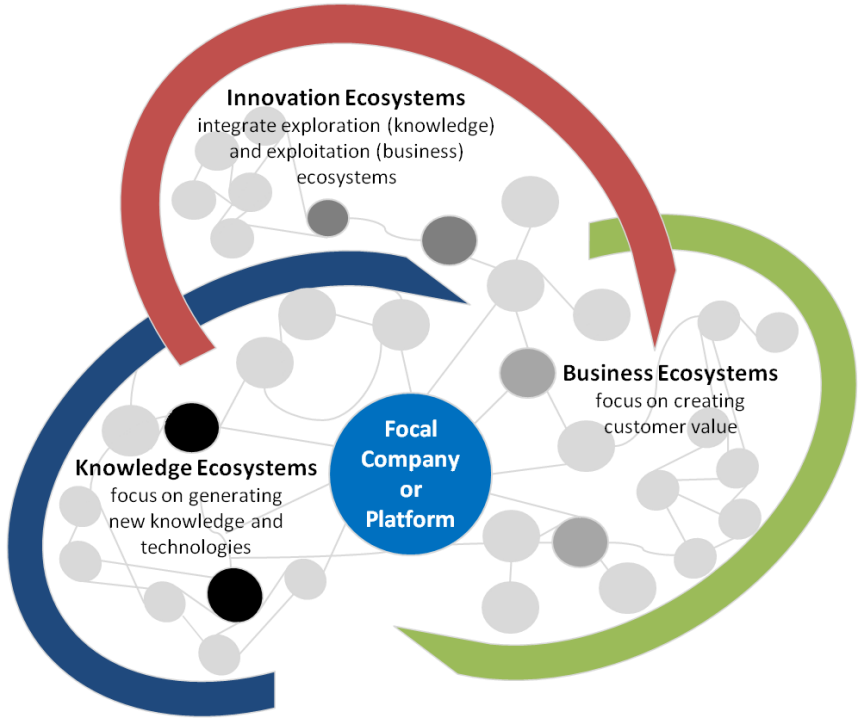
https://knowledge.wharton.upenn.edu/article/reimagining-work-age-ai/

Education



https://elementsofai.se

Ecosystems



https://timreview.ca/article/919

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