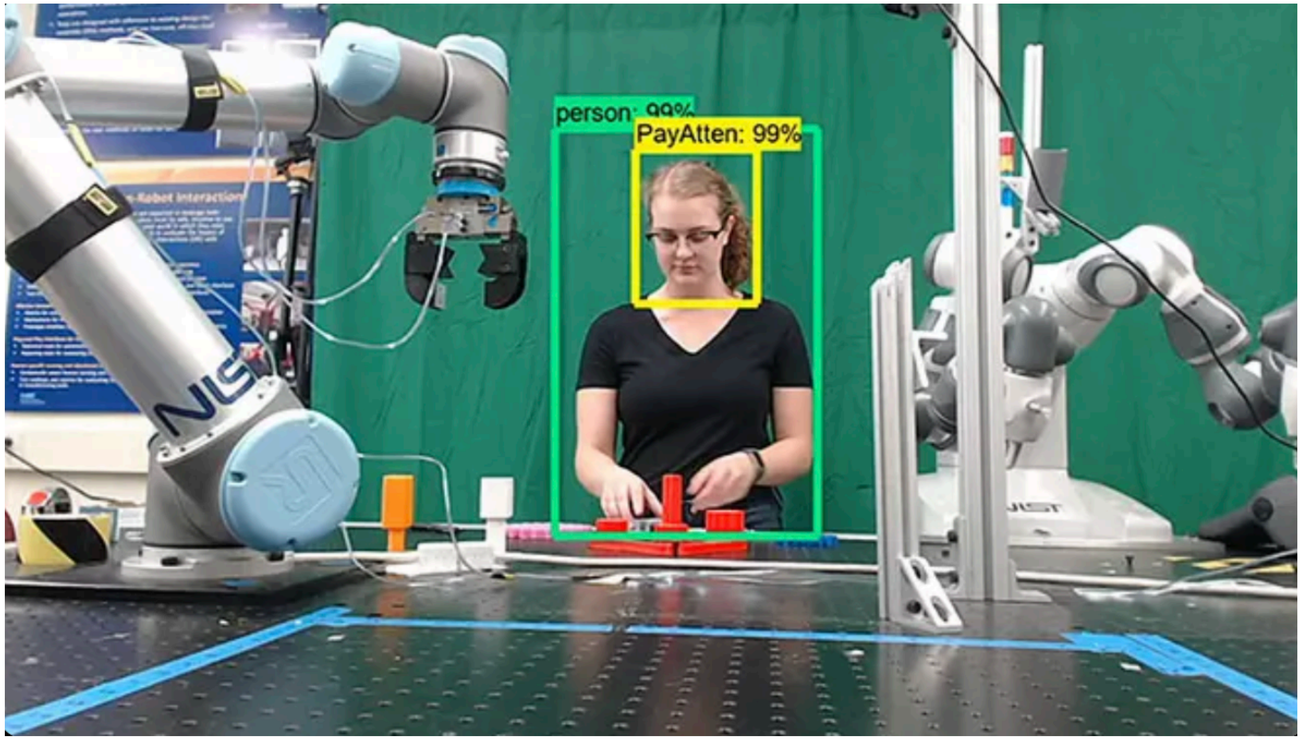
AI and the impact on work

A EU R&I perspective



Dr T.Timan, 13 December, 2023
mail@tjerktiman.nl

Speaker background

- Dr. Tjerk Timan
- EU project manager and principal consultant
- Former lecturer and researcher in Responsible AI and public sector digitisation
- PhD in Science, Technology and Policy, University of Twente, The Netherlands (2013)
- 10+ years of experience in research and teaching in ethics and regulation of technologies
- Published on topics of privacy, surveillance, digital identity and data & AI, with a focus on the public sector (as regulator, developer and user)

A reflection in 3 parts

AI - risks, promises and opportunities

AI and innovation agendas

How to research AI?

Intermezzo on AI oversight

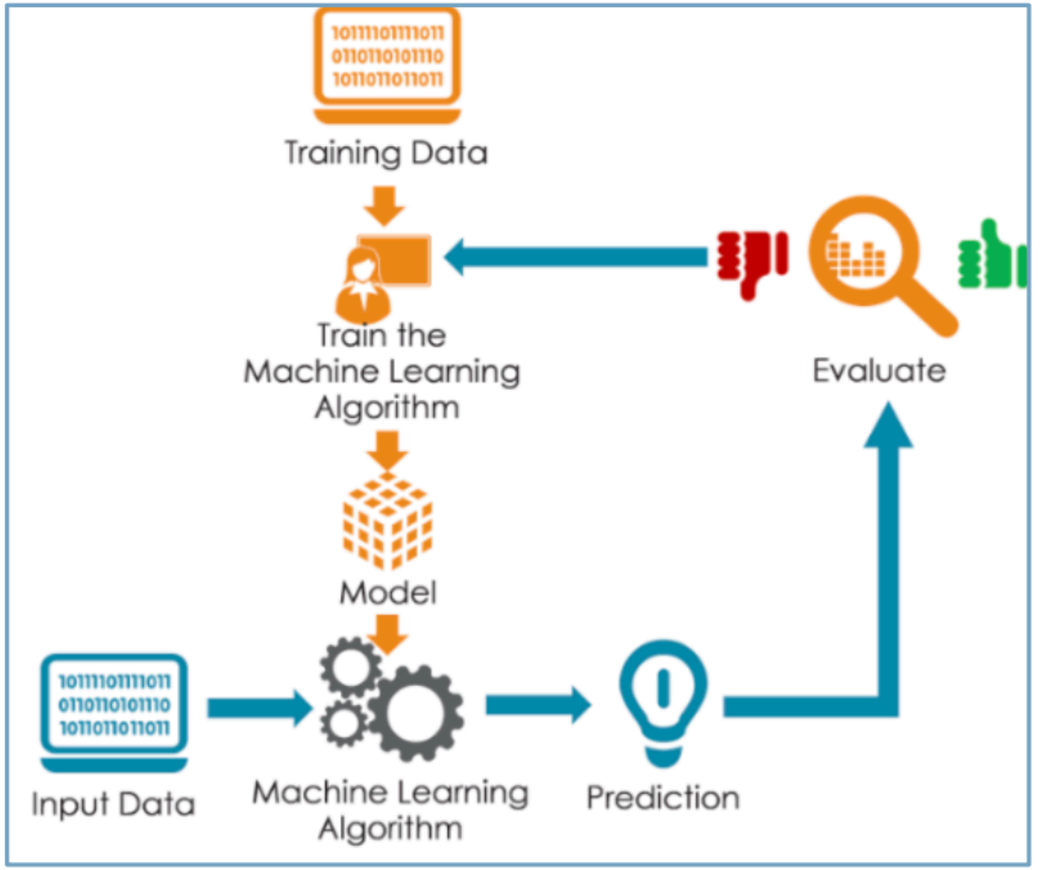
Part 1: AI - risks, promises and opportunities

- What is AI, and how to look at this concept?
- Risk, Harms and Ethics of AI
- Some projects in practice
- Current response; AI regulation and oversight

1.1 What is AI, and how to look at this concept?

“An AI system is a machine-based system that is capable of influencing the environment by producing an output (predictions, recommendations or decisions) for a given set of objectives.”

- Within Machine Learning , there are roughly **two distinctions** in types of algorithms.
- On the one hand, there are **rule-based algorithms**, meaning humans created rules based on which the algorithm makes distinctions.
- On the other hand, there are algorithms that are not based on rules, which are often called **black-box algorithms**.
- These black-box algorithms have two types: **supervised** learning and **unsupervised** learning.



How machine learning works. Source: 7wdata.be

- Is intelligence **only human**?
- Or can it **only be defined** by humans?
- Does that matter?
- What is the role of embodiment and the human scale?
- What **other forms of learning** can we teach a computer (hybrid AI/ open norms?)
- How about **projection** / anthropomorphism of intelligence?

(Cheatsheet:[Intelligence – Wikipedia])

Simone Giertz the Inventor Of Useless Robots. - YouTube

1.2 Risk, Harms and Ethics of AI

"One of the most trivial types of weakness of ML revolves around classification. ML or deep learning can correctly classify a large set of images. However, after just a single pixel is changed, a large number of those images is classified both wrongly and, at the same time, with a high degree of confidence (Wang et al., 2021, *paraphrased*). They continue to assert that "*no matter how much data have been supplied; a neural network needs a "symbolic engine" – such as those in calculators that can deal with basic arithmetic". This illustrates the danger of the big data dogma – the belief that enlarging the training set will solve all learning challenges of a neural network*" (OECD handbook on AI and the productivity of science)

Risks and opportunities

public

negative

positive

private

Why did the A-level algorithm say no?

Sean Coughlan
Education correspondent

14 August 2020

Exam results 2020

Tech Policy / AI Ethics

AI is sending people to jail —and getting it wrong

Using historical data to train risk assessment tools could mean that machines are copying the mistakes of the past.

by Karen Hao Jan 21, 2019

Artificial Intelligence / Machine Learning

Training a single AI model can emit as much carbon as five cars in their lifetimes

Deep learning has a terrible carbon footprint.

Hao Jun 6, 2019

Startup Life/Hiring & Workforce/News/

Fired by robots — Uber faces legal challenge for algorithmic dismissals

The dismissed drivers believe they've been penalised for picking and choosing when to work.

BY TIM SMITH IN BARCELONA 26 OCTOBER 2020

EDT | 1,091 views

Emerging AI Will Drive The Next Wave Of Big Tech Monopolies

Ed Stacey Contributor
Big Data
I write about disruptive technology, AI and computer science.

Follow

Listen to article 8 minutes

Pepper robot to work in Belgian hospitals

14 June 2016

Using AI to find disease-causing genes

Feb 1, 2019, 08:01am EST | 2,797 views

How AI Is Now Being Trained To 'Detoxify' Social Media

Zak Doffman Contributor
Cybersecurity
I write about security and surveillance.

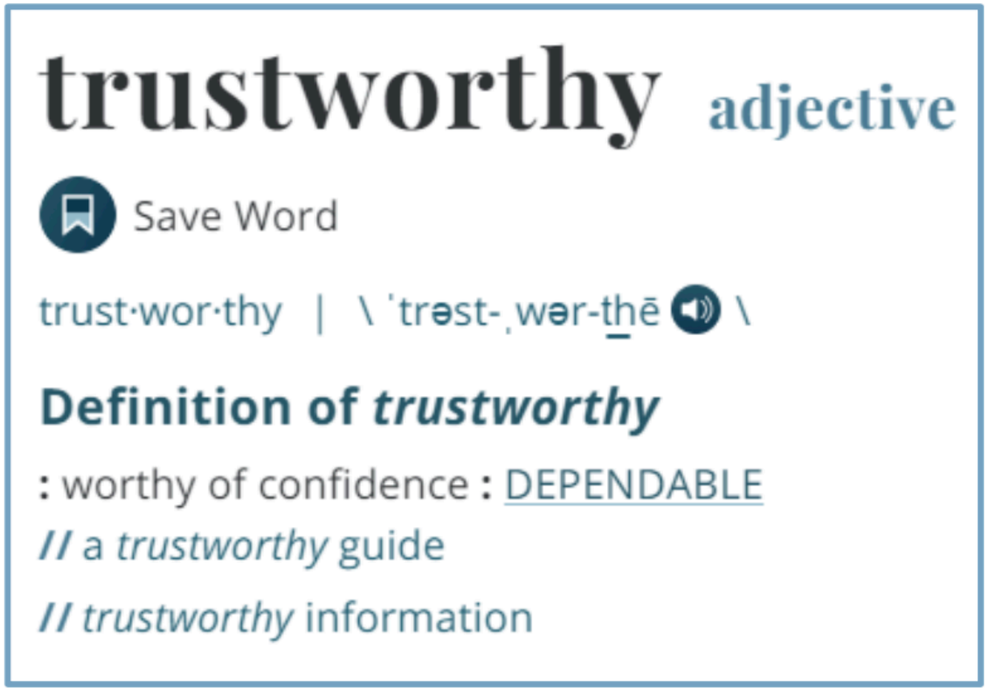
This article is more than 2 years old.

Trust in AI

WHERE AND HOW IS TRUST BEING SHAPED?

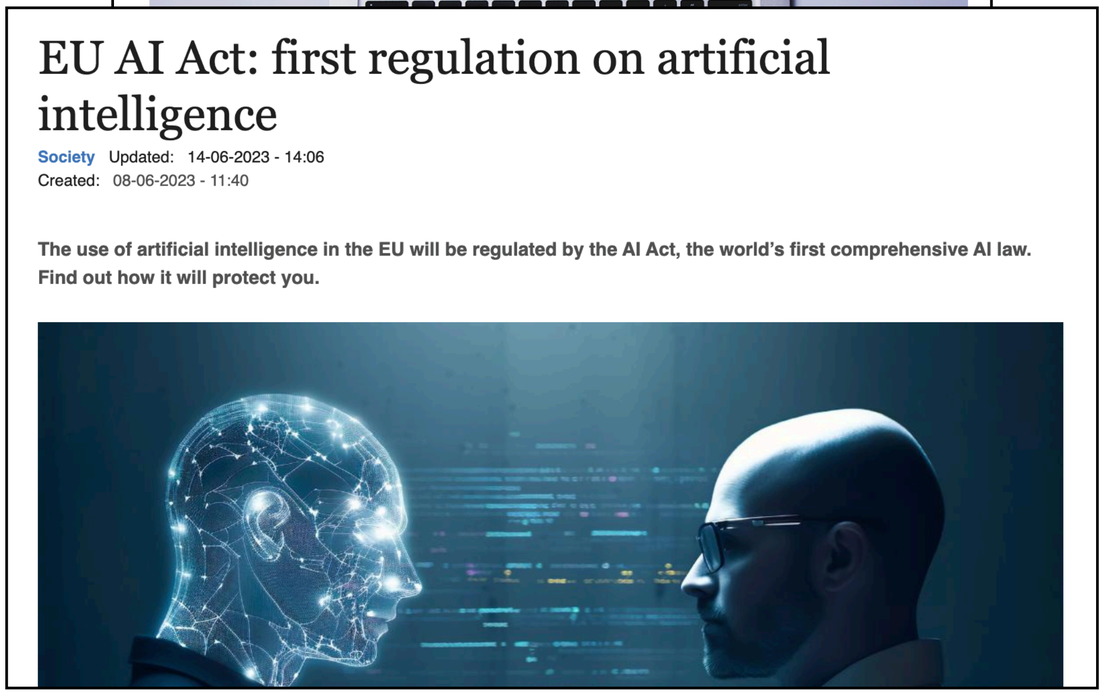
There are many **high-level ethical principles** for AI and many guidelines. But how do they fare in practice? And what does that mean for policy and regulation of AI? Different viewpoints exist:

- **Technical**: standardisation and certification
- **Social**: education, innovation in applications
- **Legal**: sectoral and general, context-dependent frameworks
- **Political**: investments and dependencies on external data and infrastructure
- **Ecological**: AI for good vs AI as novel power-hungry paradigm – added value beyond economical?
  - Trust is brittle – and is linked to **institutional reputation**
  - Built through **data and model availability** and communication



1.3 Current responses: AI regulation and oversight

- From AI Ethics to AI governance
- Increasing awareness through media attention
- Oversight and regulation
- Self-regulation and professional ethics
- Government investment programs
- VC hypes
- …



1.4 Project examples

Project
Fairness in AI

Approach
AI oversight lab - co creation with tech, policy and social science

Client
Local governments, BZK

Outcomes
Methodology, evaluation, paper

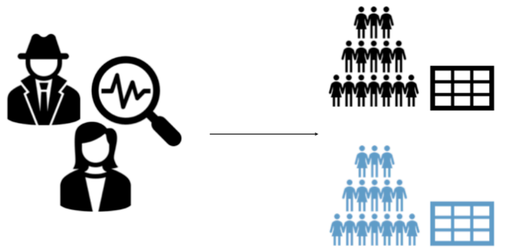
FLAGGING AND PREDICTING: AI AND FRAUD DETECTION

AI algorithm detects fraud by “predicting” cases of potential fraud and puts “flags” for these cases – citizens.

WHICH GROUPS DO WE AIM TO DISTINGUISH?

“The ground truth”: conceptualization on how fraudulent and non-fraudulent citizens are distinguished?

- Definition of fraudulent
- Approximation with the available data?
- Any model is an abstract representation



Experiences:

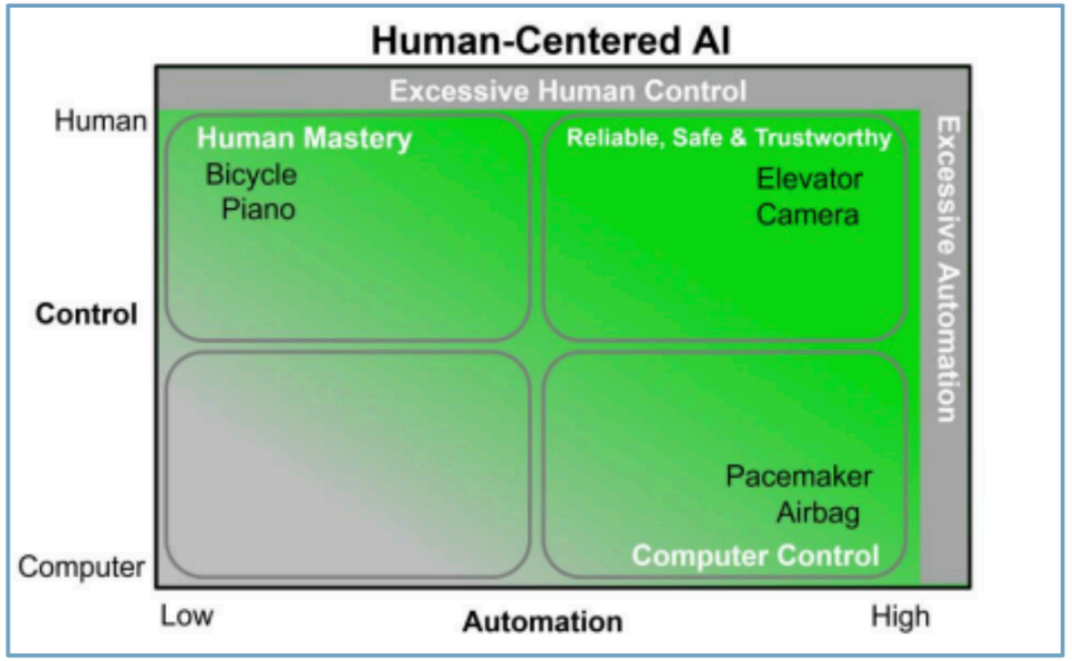
- Availability dominates.
- Room for challenging the ground truth?

Project
From AI ethics to practice

Approach
Analyzing over 44 EU ethical frameworks and connect that to ‘by design’ approaches

Client
TNO

Outcome
- Scientific paper
- Guiding framework for developers



source: Shneiderman, 2021

Home > AI and Ethics > Article

From AI ethics principles to data science practice: a reflection and a gap analysis based on recent frameworks and practical experience

Original Research | Published: 12 January 2022 | **2**, 697–711 (2022)

Project
Social& Economic Impact Assessment of AI in the public sector

Approach
Workshops, database analyses, desk research, expert interviews

Client
JRC

Outcome
- Reports
- Roadmap for EU

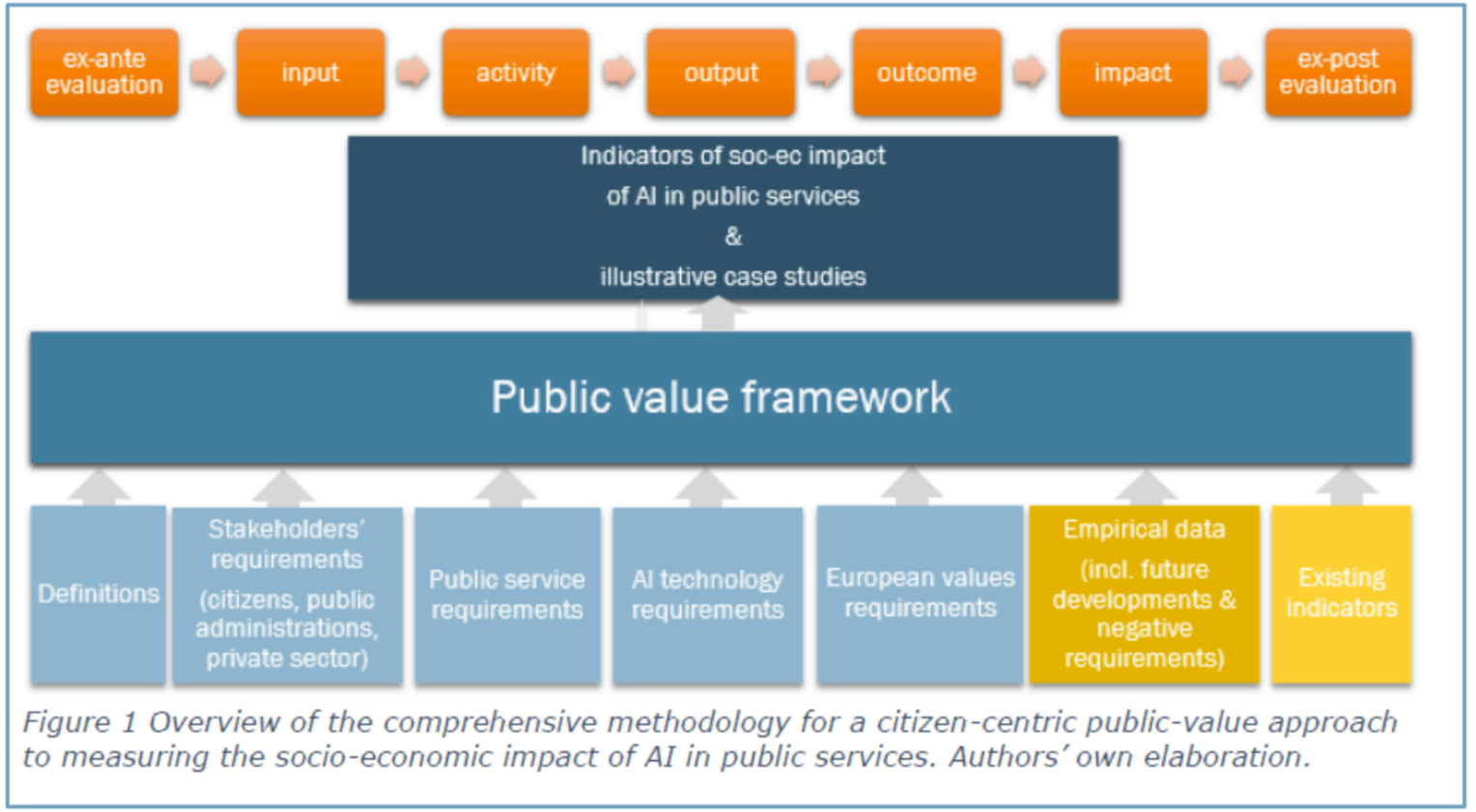


Figure 1 Overview of the comprehensive methodology for a citizen-centric public-value approach to measuring the socio-economic impact of AI in public services. Authors' own elaboration.

source: Bodea et al

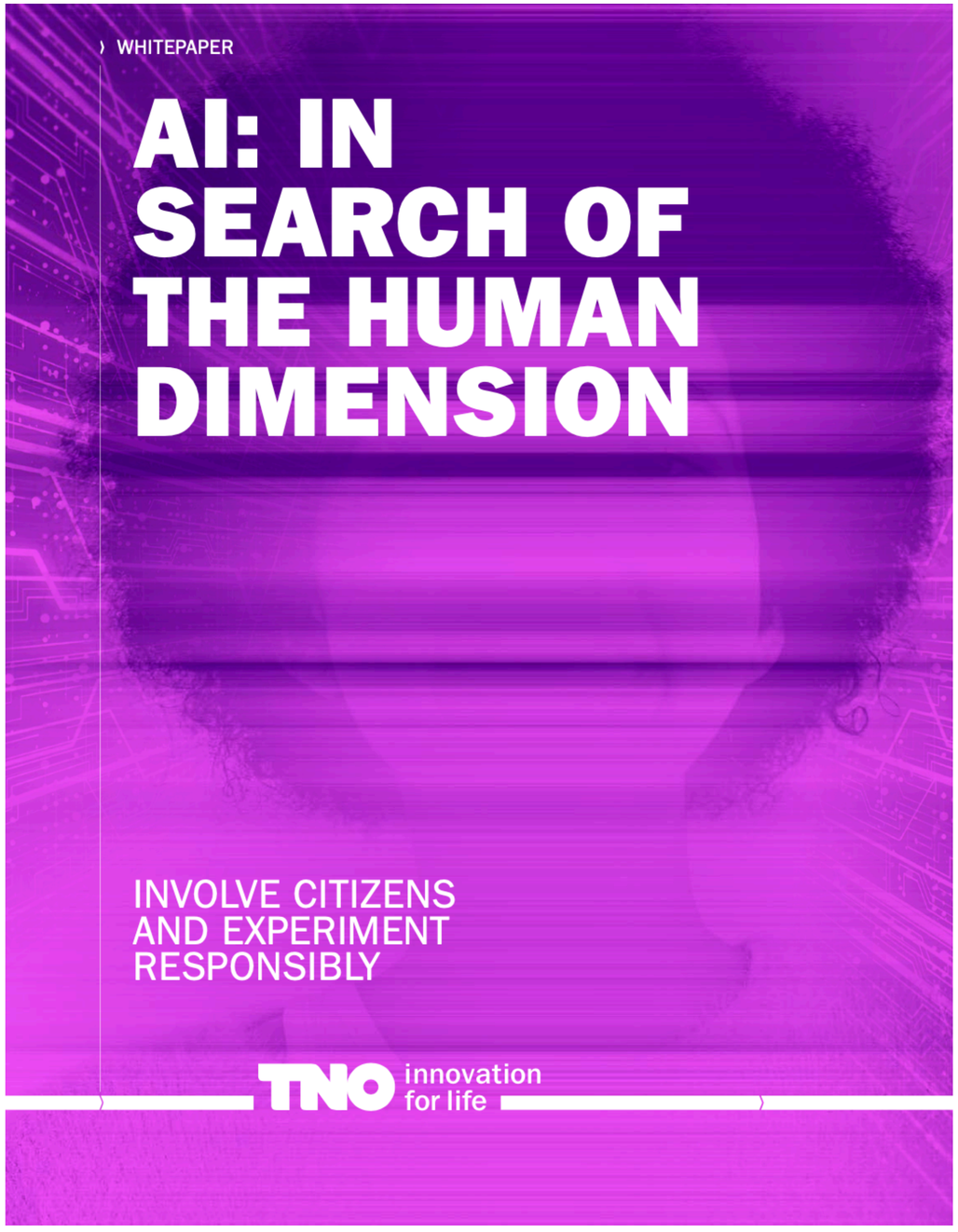
TNO innovation for life

Project
Methodology Development Human Centric AI

Approach
Interviews, Desk Research, Sprints

Client
TNO

Outcomes
Position paper
Podcast



Project
VISION4AI CSA

Approach
Ecosystem mapping, strategic agenda development

Client
EC

Outcomes
Conferences, workshops, SRAs etc

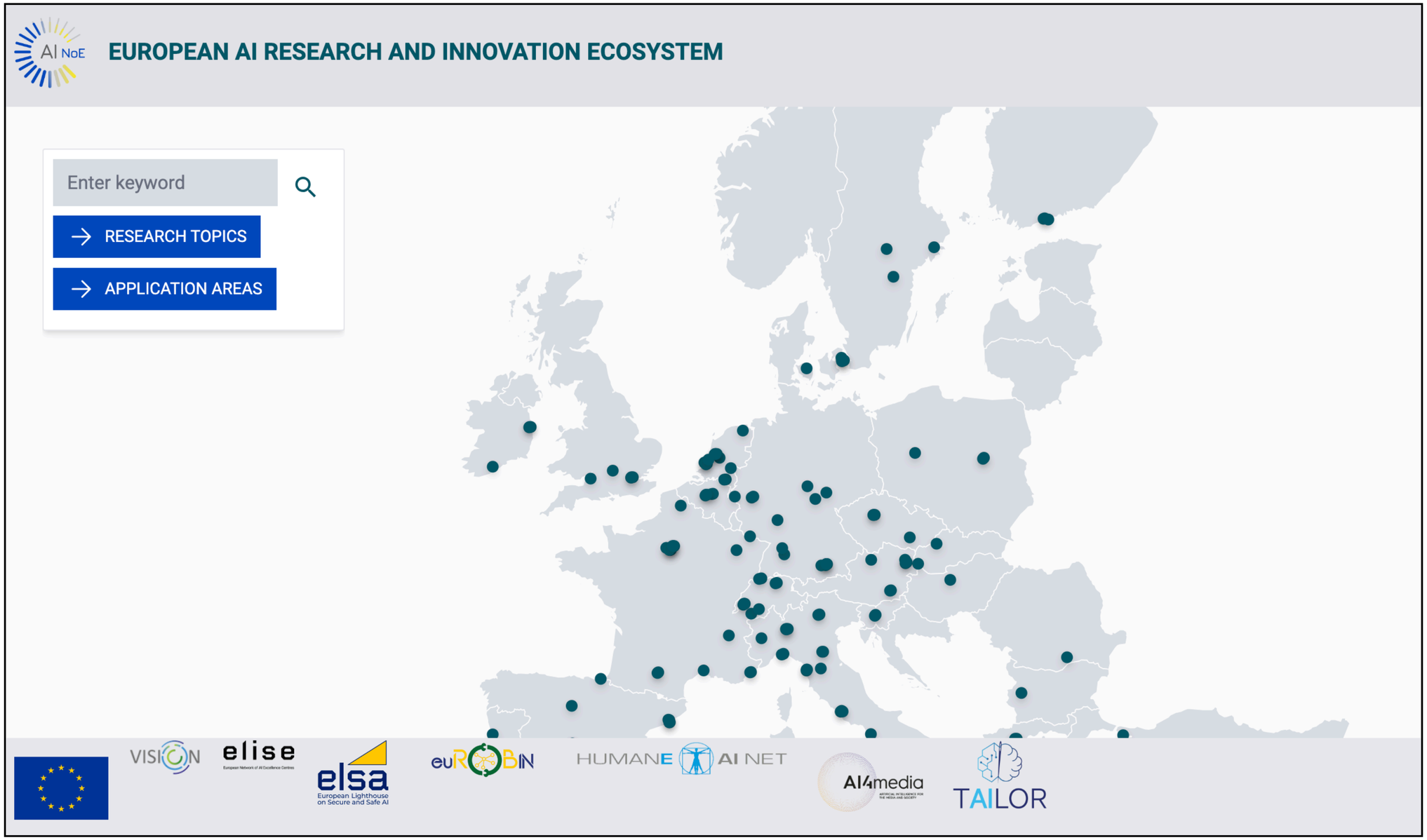
Ecosystem mapping - research topics & application areas

The ecosystem map requires a structured list of topics. The NoEs agreed to combine our initial structure the AAAI-23 keywords and topics provided by the NoEs, performing several reviews. Each topic has several subtopics (the amount of subtopics is shown after each topic). The topics are alphabetically sorted.

Research topics	
AI Hardware & High-Performance Computing	4
Computer Vision & Audition	36
Ethical, Legal, Social Aspects	34
Human- Machine Interaction & Cognition	32
Knowledge Representation & Management	17
Machine Learning	60
Multi-Agent Systems & Agent-Based Modelling	10
Planning, Routing & Scheduling	15
Reasoning	31
Robotics	19
Search & Optimisation	26
Speech & Natural Language Processing	27

×

Application areas	
Agriculture & fisheries	9
Economy & financial markets/services	9
Environment, energy & sustainability	18
Health & wellbeing	12
ICT infrastructure	4
Industry	11
Learning & education	10
Media, communication, web & entertainment	16
Mobility & transportation	18
Public sector & citizen services	6
Safety & Security	8
Scientific research, design & engineering	6



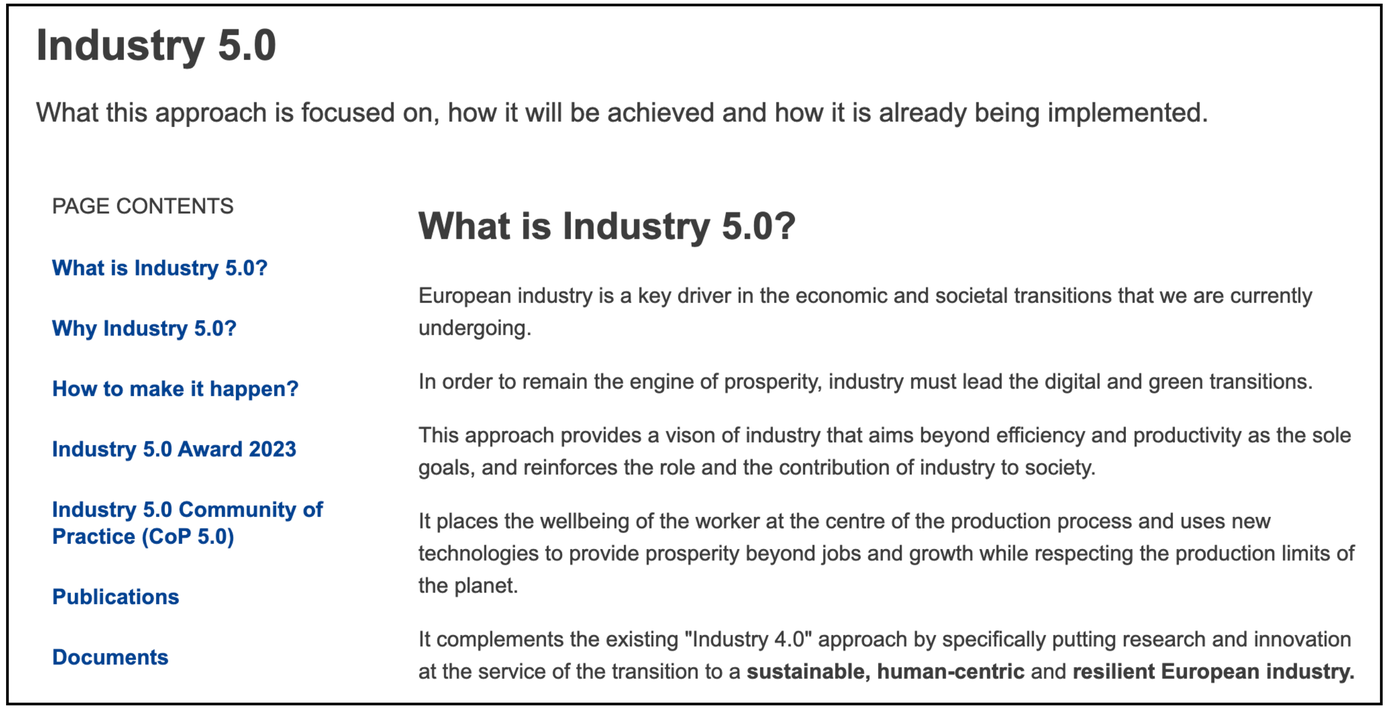
https://eu-ai-ecosystem.tnods.nl/

Project
Industry 5.0 - human centric innovation on the shop floor

Approach
Stakeholder mapping, agenda development, analysis of scientific challenges

Client
EC- DG RTD

Outcomes
In process



Industry 5.0

What this approach is focused on, how it will be achieved and how it is already being implemented.

PAGE CONTENTS

What is Industry 5.0?

European industry is a key driver in the economic and societal transitions that we are currently undergoing.

In order to remain the engine of prosperity, industry must lead the digital and green transitions.

This approach provides a vison of industry that aims beyond efficiency and productivity as the sole goals, and reinforces the role and the contribution of industry to society.

It places the wellbeing of the worker at the centre of the production process and uses new technologies to provide prosperity beyond jobs and growth while respecting the production limits of the planet.

It complements the existing "Industry 4.0" approach by specifically putting research and innovation at the service of the transition to a **sustainable, human-centric** and **resilient European industry.**

Project
RAAIT program strategic agenda

Approach
Mapping and combining SME challenges in the 3rd sector to education and field labs

Client
HRotterdam

Outcomes
Running programs, including new master (‘AI translator’)

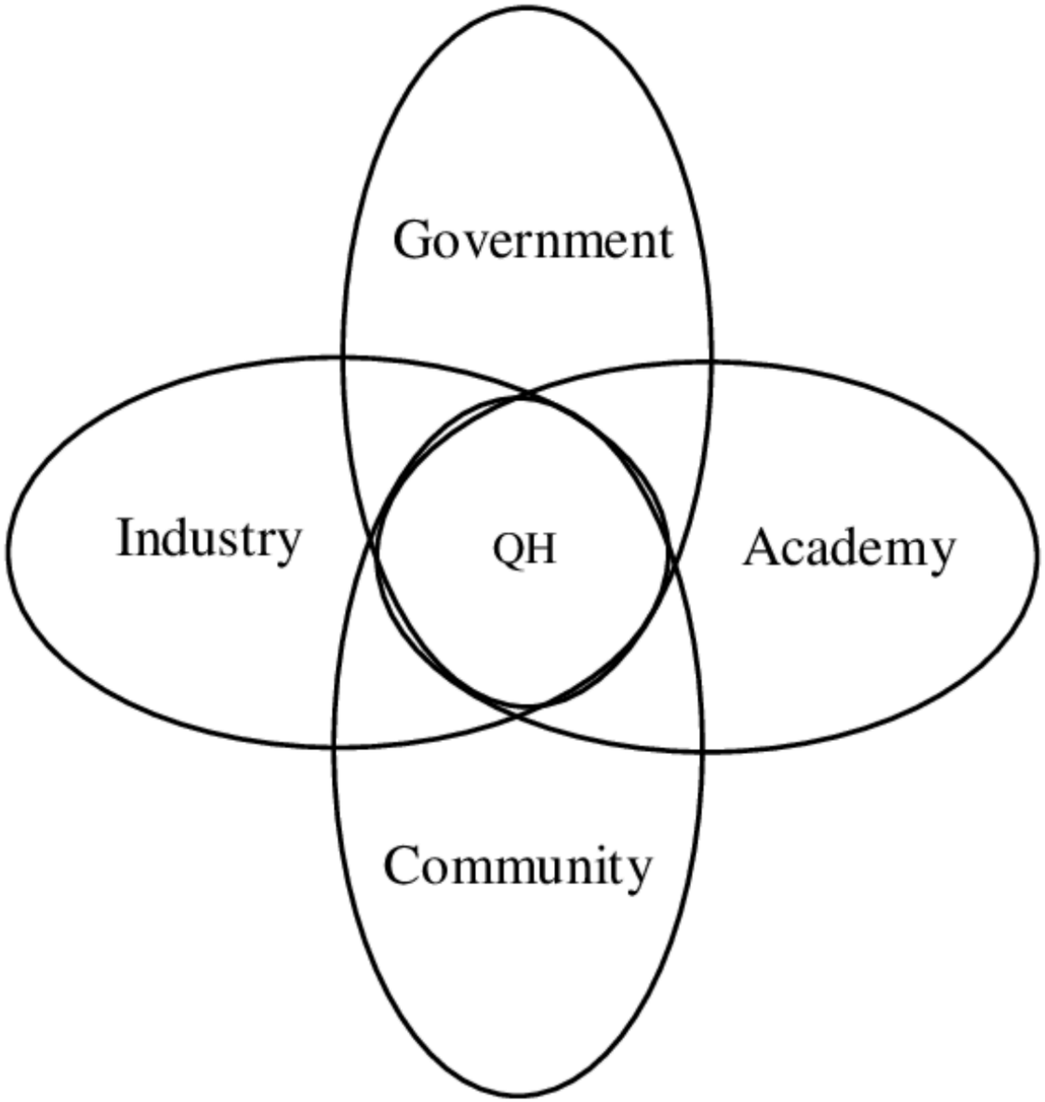
Part 2: AI and innovation agendas

- How are innovation agendas formed and by whom?
- What is the role of public actors in this formation? (EU focus)
- A bit of theory on power, technology and politics (sphere transgressions)
- AI and the (incomplete) quadrupel helix (and a bit of Lessig and Doctorow)

https://www.wsj.com/tech/ai/artificial-intelligence-united-states-future-76c0082e?mod=hp_listb_pos1

2.1 How are innovation agendas formed and by whom?

- Since WWII increasing awareness that innovation is not linear
- Leaving uni and industry develop things alone was see as too risky
- From triple helix (gov, industry, academia), more public engagement for large and high risk programs (see nano, nuclear etc)
- Innovation strategies and agendas co-shaped, both nationally and internationally (EU, OECD, UN, IMF, WHO etc)

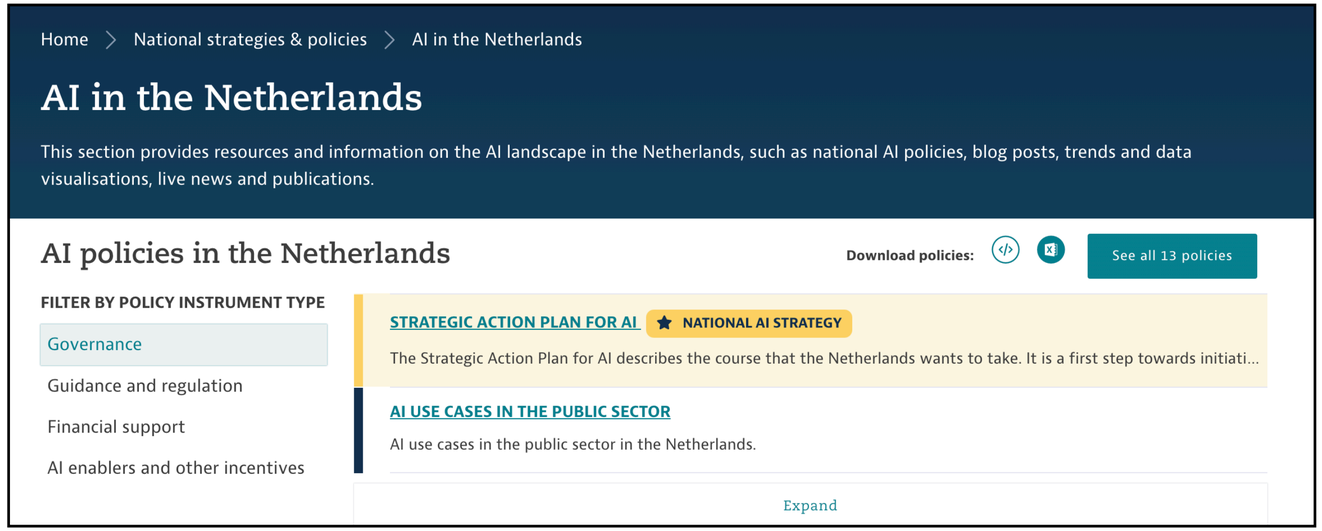


The role of academia

- Progressing and improving AI
- Applying AI to scientific disciplines
- Also risks of AI in science - replicability black-box, falsifiability, hypothesis-free science
- Data, talent and infra; risk of doing bad science with superficial AI
- Co-shaping AI through critical reflection

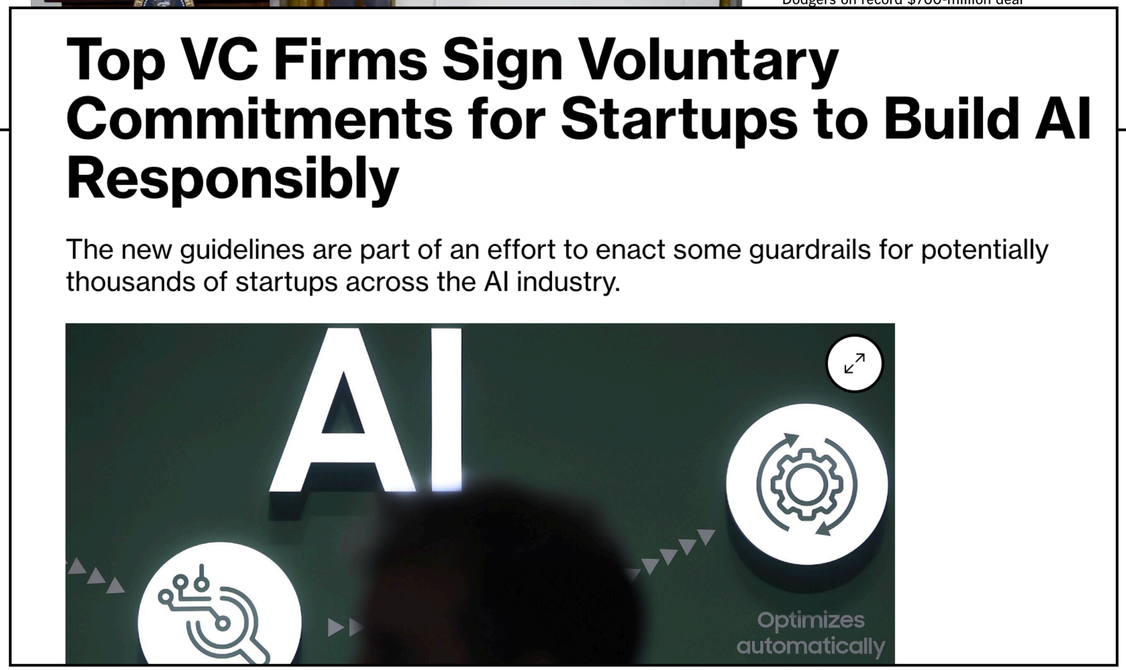
The role of governments

- Both user and regulator of AI
- Beyond, or before applications, also key questions on infrastructure, sovereignty, investments
- Organising technology development through laws, standards, oversight bodies, stimulation programs
- Needs to develop an national- and international strategy



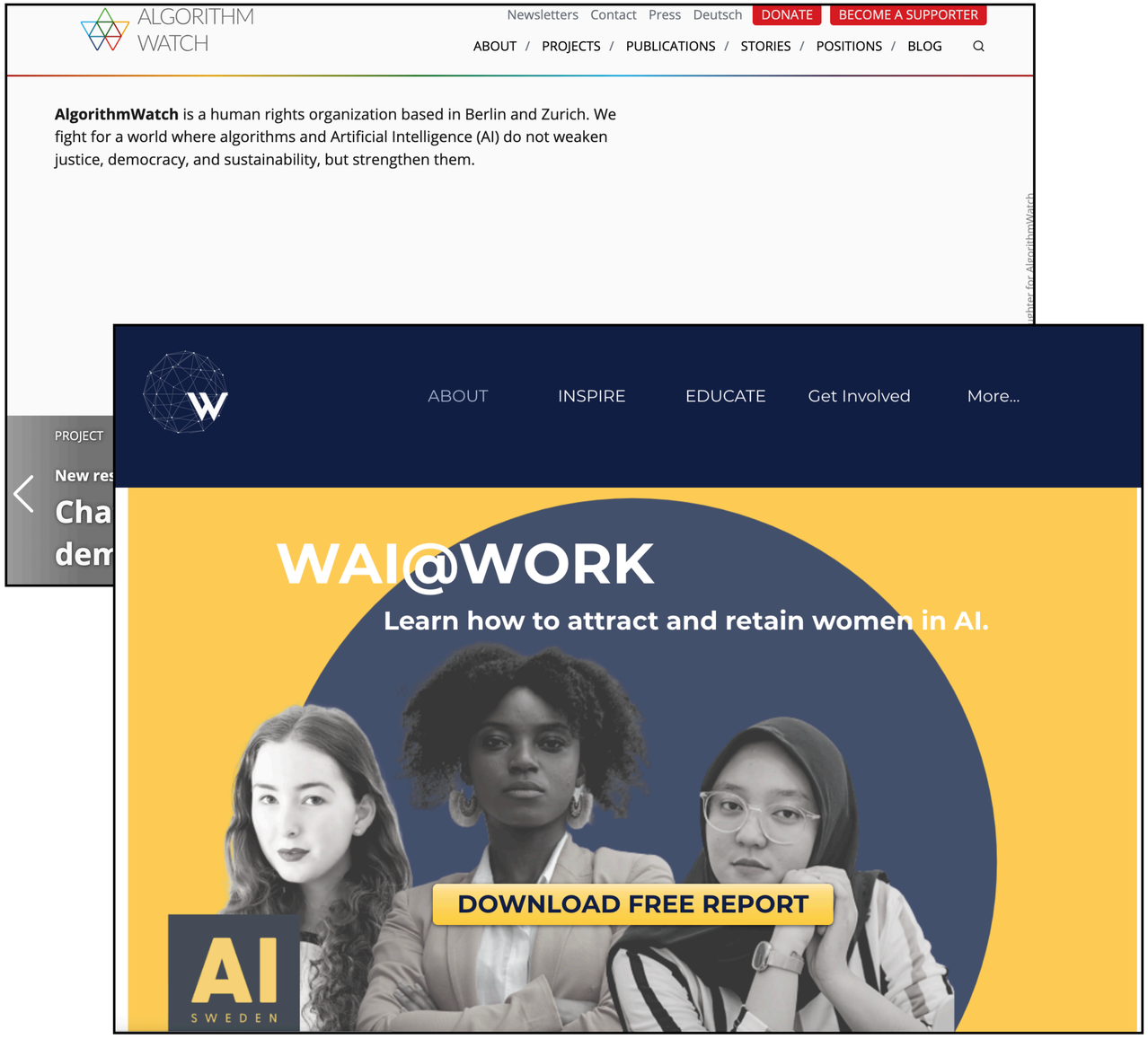
The role of companies

- The ones developing AI and need to find added value
- But also in risk of losing business, becoming dependent of horizontal AI actors
- Role of worker, role in sector, role in larger economy
- Waiting games, doubts and dependencies



The role of NGOs

- Representing specific **public** interests
- National or inter-national
- Can put pressure on governments and companies
- Fragile yet key in a democracy
- In AI - voicing the **marginalised**, the subjects



https://www.aiethicist.org/ai-organizations

2.2 What is the role of public actors in this formation? (EU focus)

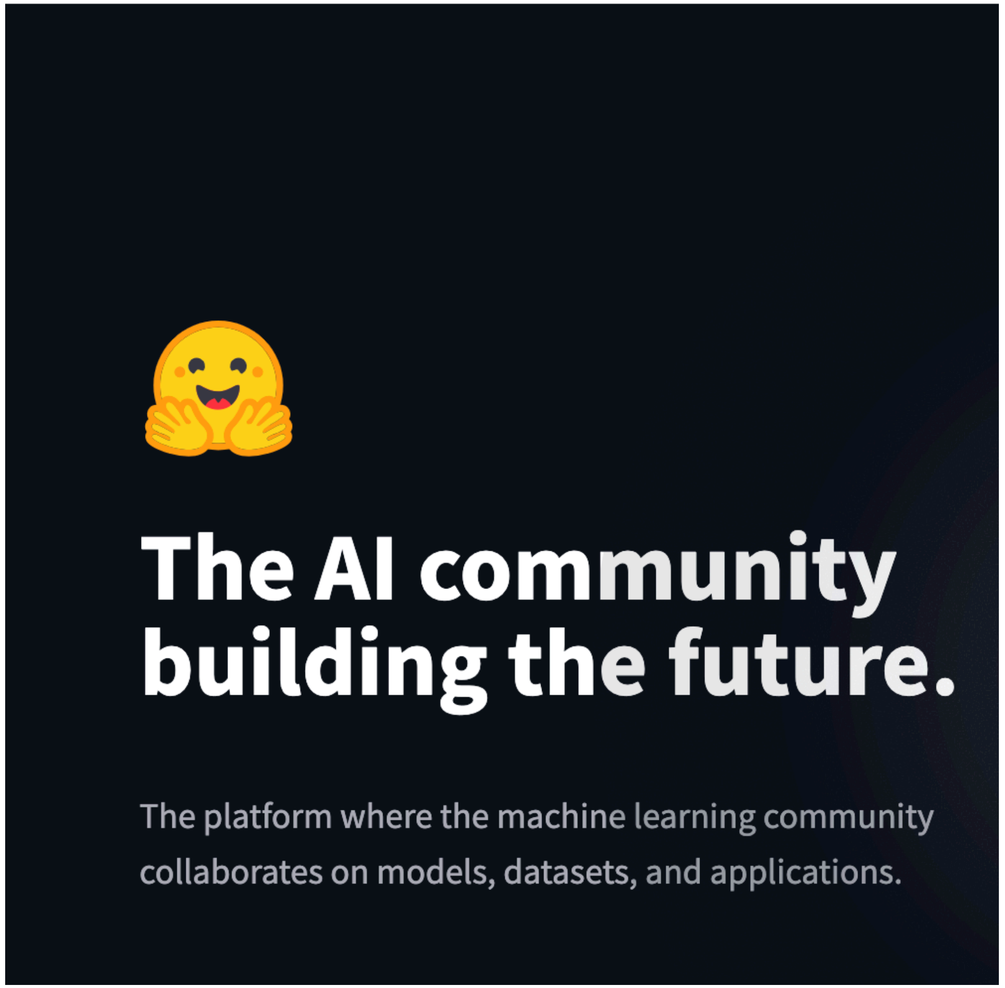
- Quadrupel helix ideal; formation of gov via voting, active role in science via opening up of unis, influence on working conditions via unions and supporting NGOs and journalism

- Current situation; active breaking down of unions, larger gap between academia and public, journalism has marginal role, governments have become more stupid (technocratic neoliberal) - large ICTB companies have a free range (power of palo alto)

- Role should increase by using power of institutions and through scrutiny and oversight

2.3 Interactive; a bit of theory on power, technology and politics

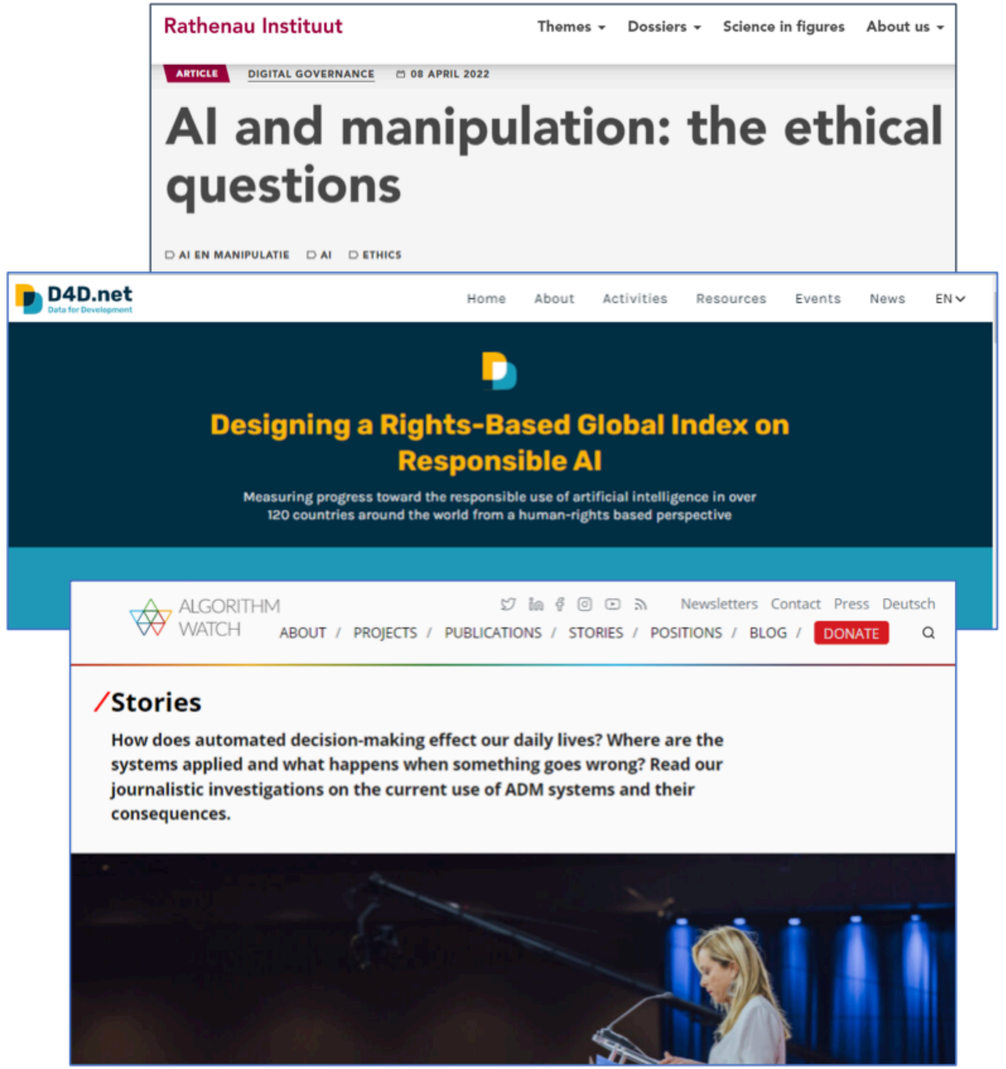
- Doctorow’s article
- Key questions on how owns a technology and why that matters
- Freedom of expressions vs echo chambers and surveillance
- Tinkering with tech and the knowledge gap
- Pros and Cons of open source…



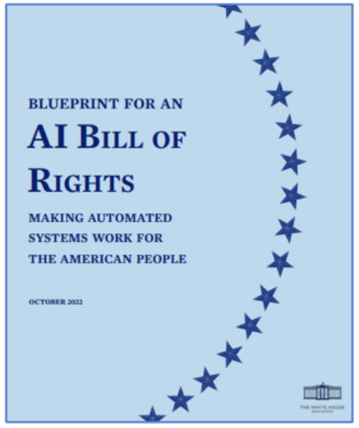
Intermezzo: AI ethics, risks and oversight

What to monitor, how to monitor, and who is to monitor?

- Governments as **both developer, user** and **regulator** of AI
- AI-indices and readiness levels exist: but **very few Trustworthy or Responsible AI monitors**
- From **voluntary** monitoring to official **audits**
- Not clear yet **who** will do the monitoring (self-regulation, supra-national Acts, NGO's and watchdogs, Ombudsmans etc)
- Not clear yet **what** to monitor (AI, data, application, impact?)
- How to deal with **long-term & second-order effects** (from model drift to changing values to energy depend to co-learning or deskilling etc)?
- Different effects on organisations, companies, users, and **implicated actors** (AI-subjects)



Supra-national, national and regional examples

	supranational	*national*	*regional/local*
Ethics & impact **ON** public sector (gov as user/ buyer of AI systems)			
Ethics & impact assessment **BY** the public sector (gov as regulator)			

What to monitor?

from ethical principles and guidelines to public values

↓

from public values to policy & law

↓

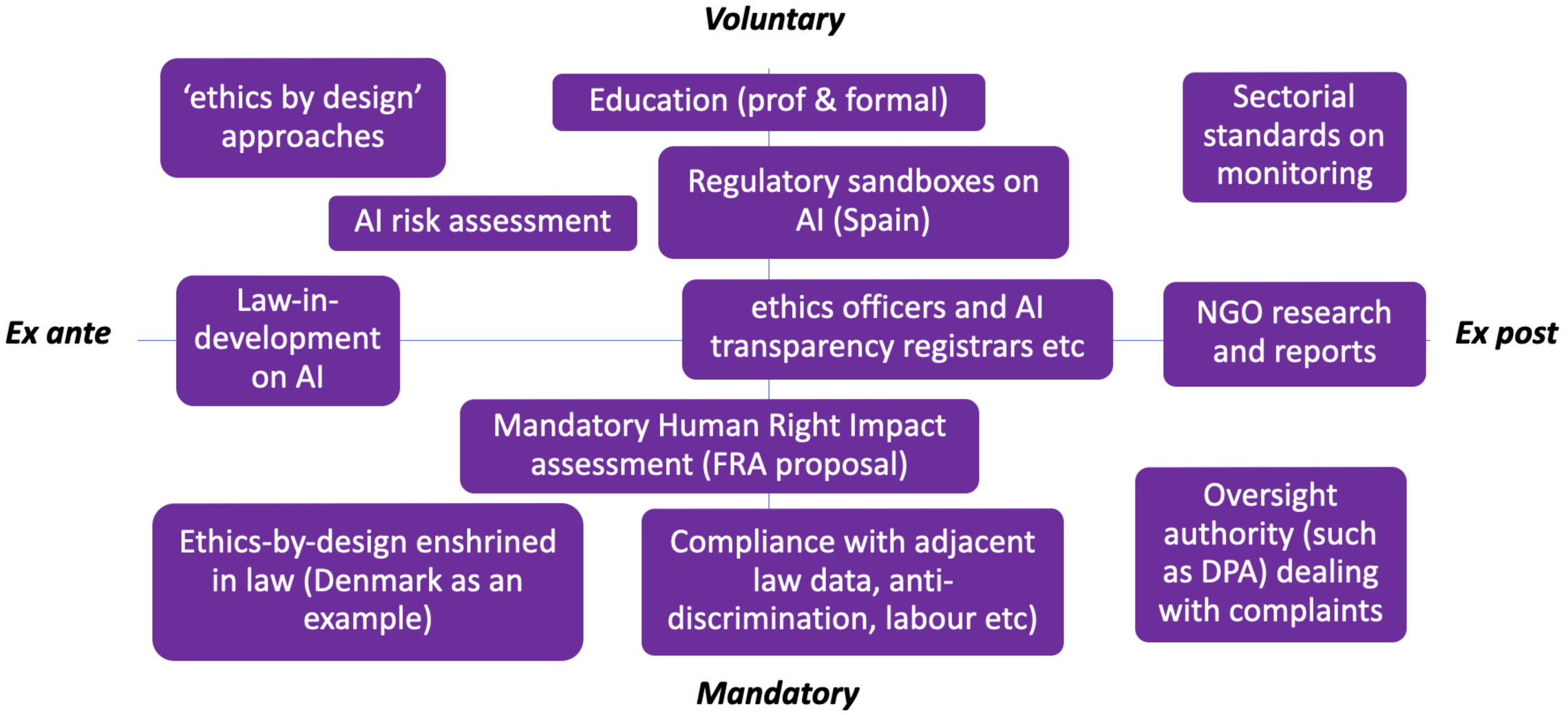
from policy & law to practice

↓

from practice to long-term effects feeding back into ethical principles etc.

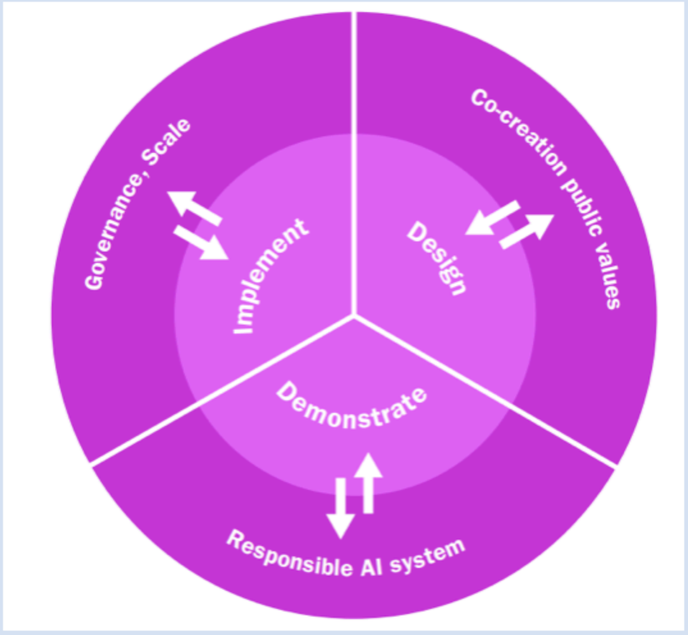
Requirements of Trustworthy AI | FUTURIUM | European Commission (europa.eu).
See also Herrmann, T., Pfeiffer, S. Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence. *AI & Soc* (2022). https://doi.org/10.1007/s00146-022-01391-5

How to monitor ethics irt to AI?



Three foci

AI & inclusive design approaches

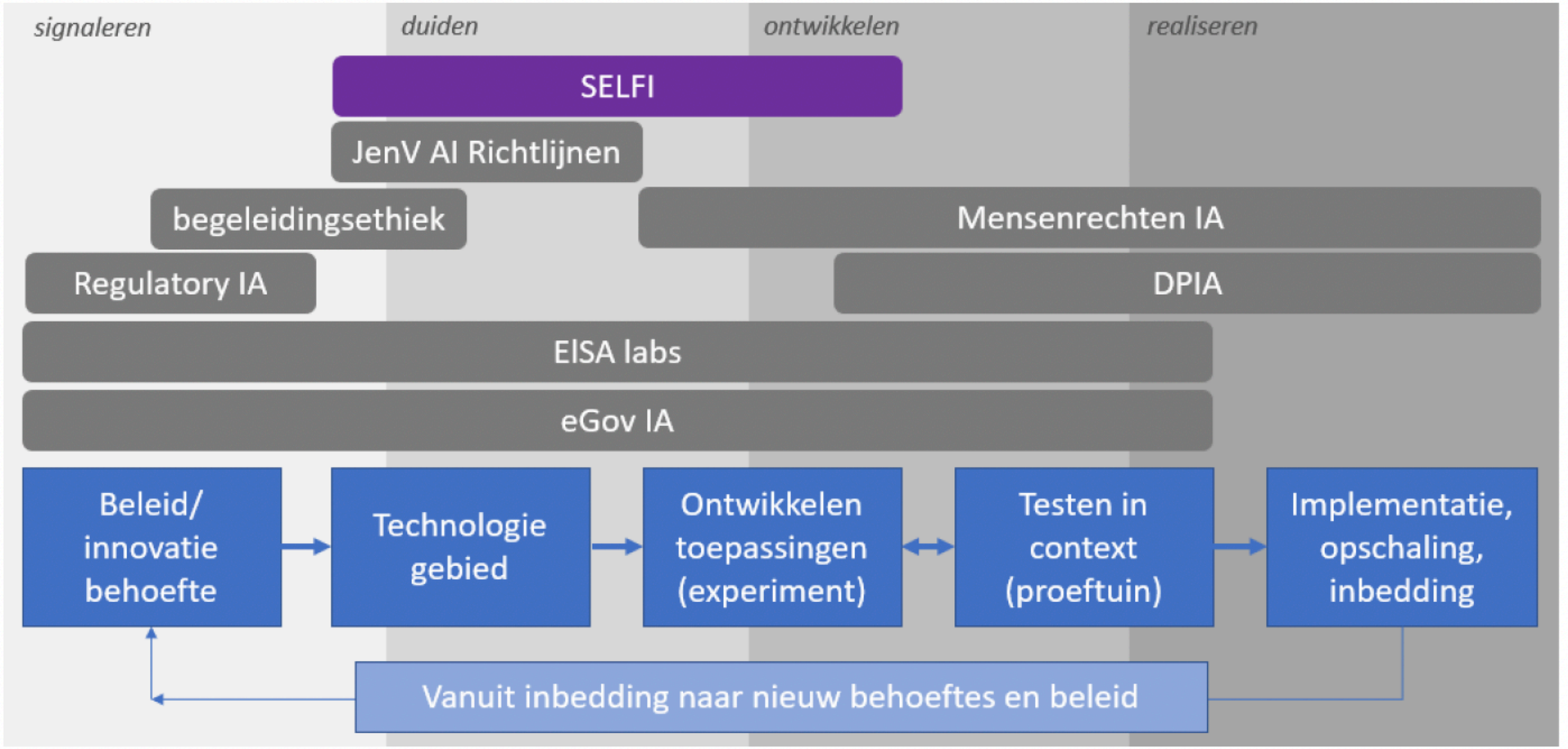


AI & governance

AI & democratic oversight

End of intermezzo

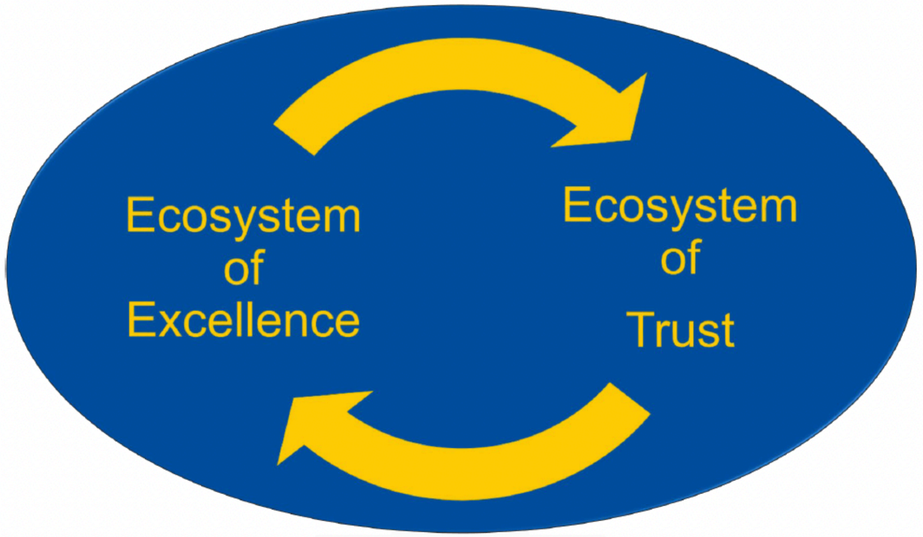
2.1 How are innovation agendas formed and by whom (2)?



Figuur 3 Plaatsing van SELFI t.o.v. andere effectenbeoordelingsmethoden

The EU AI R&I landscape

Ecosystem of Excellence in AI



Our Unit's mission:

Develop **ECOSYSTEM OF EXCELLENCE IN AI**

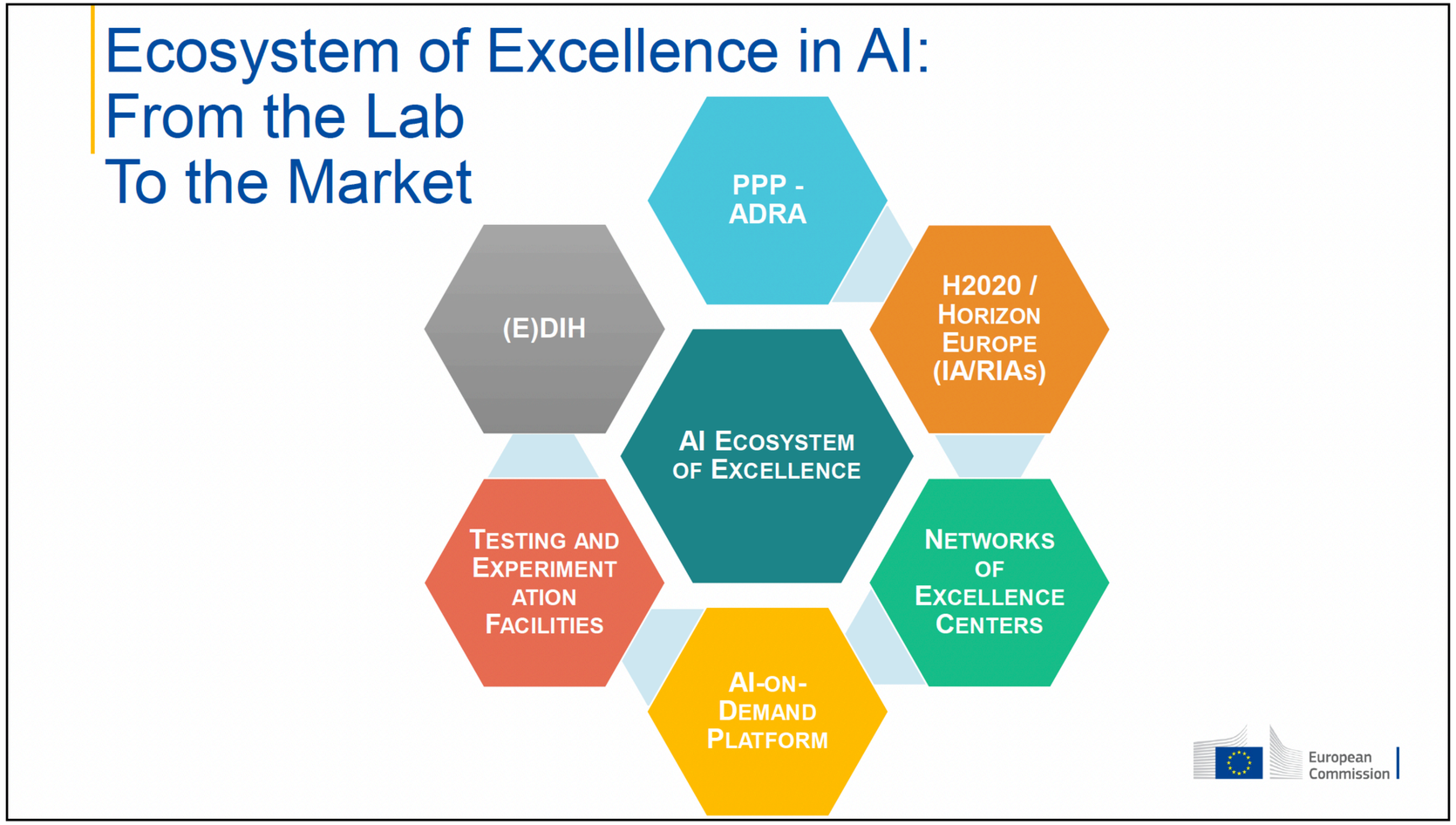
- **EXCELLENCE**, from research to adoption
- **AI** ➔ AI, Data and Robotics

Rich landscape

- Multidisciplinary, Multi-stakeholders

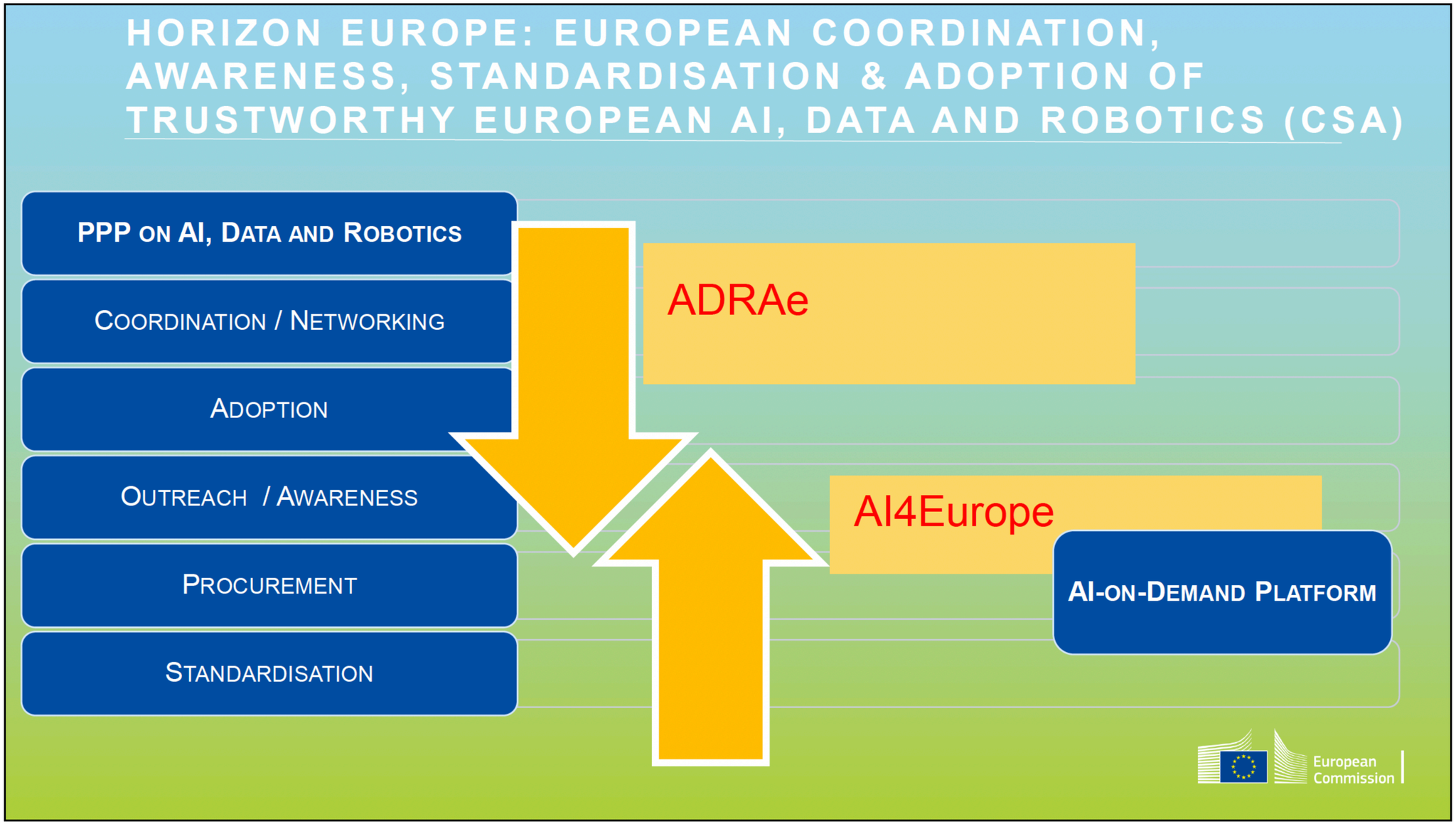
European Commission

The EU AI R&I landscape - 2



Source: ADRA partnership - https://adra-e.eu/sites/default/files/2022-10/Launch%20event%20-%20EC%20CNECT%20A1_Cécile%20HUET.pdf

The EU AI R&I landscape - 3



Looking deeper into HEU AI

Application	Description	Examples of use cases
Computer vision	Extracting information from visual data.	Object recognition, image classification, facial recognition.
Natural Language Processing (NLP):	Processing and understanding human language.	Sentiment analysis, text classification, machine translation.
Computer audition	Processing and analysing audio signals.	Speech to text and text to speech, sound classification, music analysis
Predictive analytics	Forecasting future outcomes based on historical data.	Epidemiology, risk assessment and prognosis of diseases, forecasting traffic conditions
Big data mining	Extracting useful information and patterns from large datasets.	Market analysis, customer segmentation, trend analysis, Anomaly Detection etc.
Generative Models	Creating new data based on existing patterns.	Image generation, text generation, music composition.
Advanced robotics and control	Designing and controlling physical robots	Autonomous robots Industrial automation, robotic surgery, humanoid robots.
Optimization and Planning	Solving complex problems and optimising processes	Resource allocation, production planning, route optimization.
Autonomous systems	Performing tasks without human intervention.	Self-driving cars, drones, automated machinery.

Looking deeper into HEU AI - 2

Clusters	Area of Intervention	Share of projects using AI among total projects
CL1: Health	Tools, Technologies and Digital Solutions for Health and Care, including personalised medicine	59%
	Health throughout the Life Course	48%
CL2: Culture, creativity, and inclusive society	Cultural Heritage	21%
	Democracy and Governance	13%
CL3: Civil Security for Society	Cybersecurity	70%
	Protection and Security	43%
	Disaster-Resilient Societies	30%
CL4: Digital, Industry and Space	Artificial Intelligence and Robotics	100%
	Advanced Computing and Big Data	72%
	JU SNS	57%
	JU KDT	52%
	JU EUROHPC	50%
	Next Generation Internet	44%
	Manufacturing Technologies	38%
	Space, including Earth Observation	27%
	Circular Industries	24%
	Key Digital Technologies	24%
	Emerging enabling technologies	18%
	Advanced Materials	18%
CL5:Climate, Energy and Mobility	Energy Storage	31%
	Buildings and Industrial Facilities in Energy Transition	24%
	Clean, Safe and Accessible Transport and Mobility	19%
	Industrial Competitiveness in Transport	17%
	Smart Mobility	15%
	Energy Systems and Grids	9%
	Climate Science and Solutions	6%
	Energy Supply	5%
CL6: Food, Bioeconomy Natural Resources, Agriculture and Environment	Environmental Observation	33%
	Seas, Oceans and Inland Waters	11%
	Agriculture, Forestry and Rural Areas	11%
	Food Systems	10%
	Circular Systems	9%
	Bio-based Innovation Systems in the EU Bioeconomy	7%

Questions around EU innovation agenda on AI

- Who decides? For now RTOs and small circle of scientists
- How does it fair against national programs in the EU?
- Is it good enough internationally?
- Directionality, efficiency, effectiveness questions in the recent eval.

Part 3: How to research AI?

- Between ignorance and expertise - (how to) be smart and humble
- Descriptive vs action research in AI
- Research methods, ethics and the question of normativity
- Goals on motivations - what impact are you aiming for?

3.1 Between ignorance and expertise

- Knowing the field, key debates, key actors
- Learn about technology, but do not always show you know
- Positive, open inquiry (try to remain non-normative in data collection - you can be normative in your analysis and writing!)
- Focus and limit of scope - boundaries and details
- AI and work - large questions, but incremental implementations

3.2 Descriptive vs action research in AI

- Pros and cons of action research
- Pros and cons of descriptive and conceptual research
- AI is both sexy and boring - embrace the daily, boring, repetitive (bowker & star)
- It is also a fast-moving field; juggling trends and long-term impacts is hard
- Interdisciplinarity is key, but not evident

AN *EXTENDED* ERROR MATRIX

LET'S TALK ABOUT IT

The error matrix from multiple perspectives:

- Data science
- Legal and administration
- Civil servants and citizens

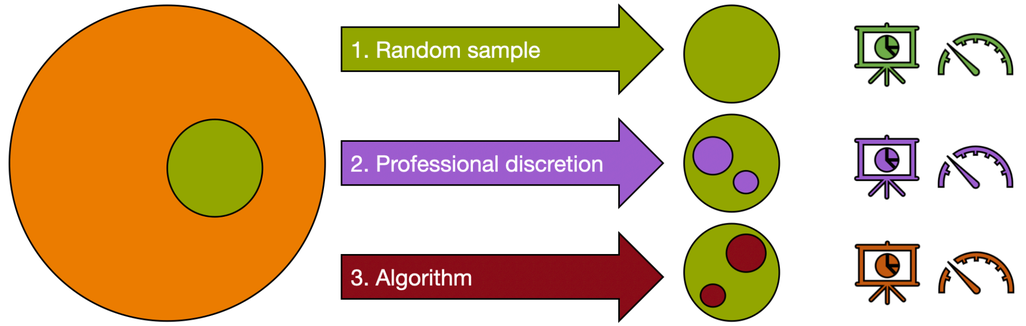
Main finding:

The categories are too simplified;

Does it fit the purpose of the AI to group them together?

	*(B) Would **qualify**, if rules were more just*		
*(A) Made **errors**, unintentionally*	**True positives**	**False positives**	*(C) **Bias** towards group; discrimination*
*(F) **Investigate,** e.g., via sampling, to find fraud*	**False negatives**	**True negatives**	*(D) Companies that **avoid** paying tax*
		*(E) Receive benefit, but **do not need it***	

A PROPOSAL FOR A BETTER APPROACH

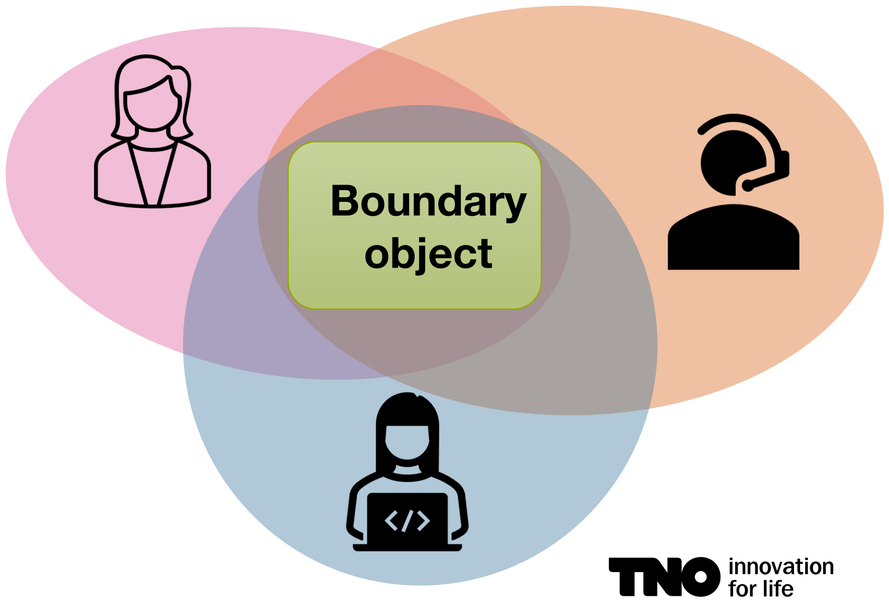


An example of a specific methodology

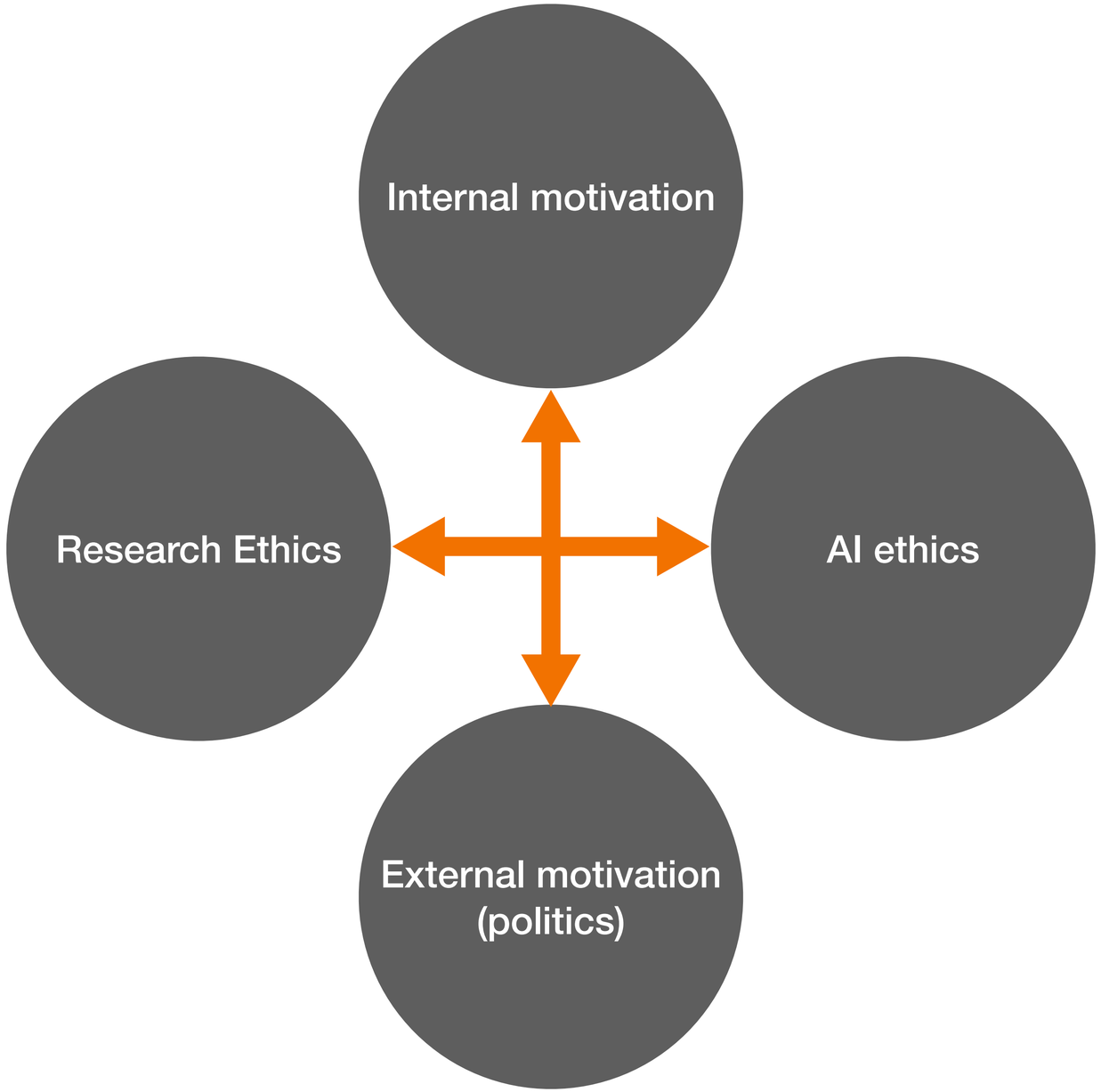
BOUNDARY OBJECTS

- Used in Science-& Technology Studies to describe ‘bridging’-artifacts between science and policy
- As an analytical tool to see if and how convergence of terms, concepts or practices takes place
- To avoid “depends on the situation”, Let’s state the situation and discuss actionable considerations towards social justice

“an object that people can point at, project meanings on, have conversations about, and collaborate around.”



3.3 Research methods, ethics and the question of normativity



Discussion