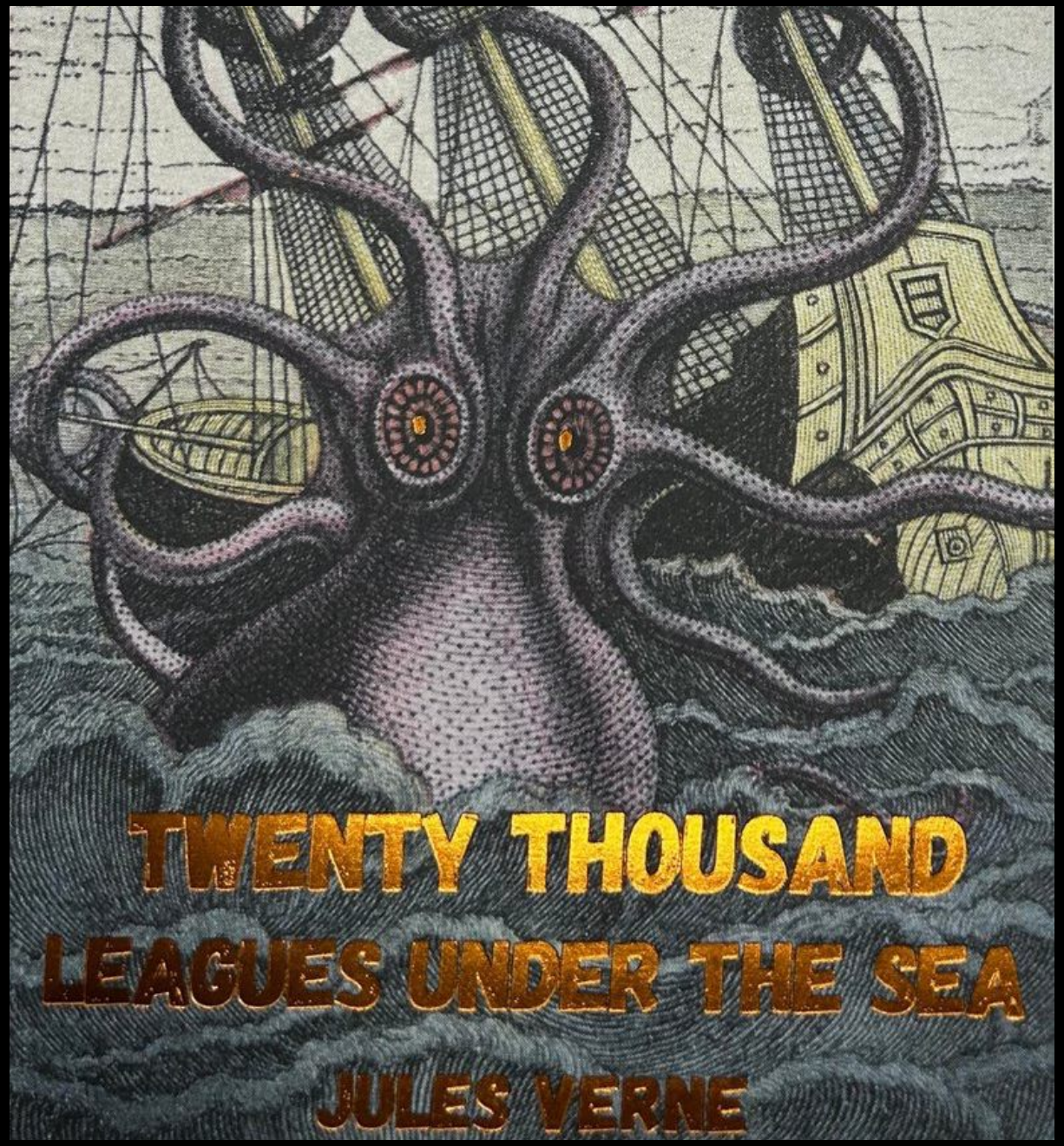
SANDBOXING RISK AND PARTICIPATION FOR RESPONSIBLE AI IN THE PUBLIC INTEREST?

Invited Research Talk

Sociotechnical AI Systems Lab
Delft University of Technology
(TU Delft)
November 12, 2025

Nitin Sawhney
Visiting Researcher
Uniarts Research Institute
Helsinki, Finland


TWENTY THOUSAND
LEAGUES UNDER THE SEA
JULES VERNE


Vingt Mille Lieues sous les Mers

MAKING SENSE OF RISK?

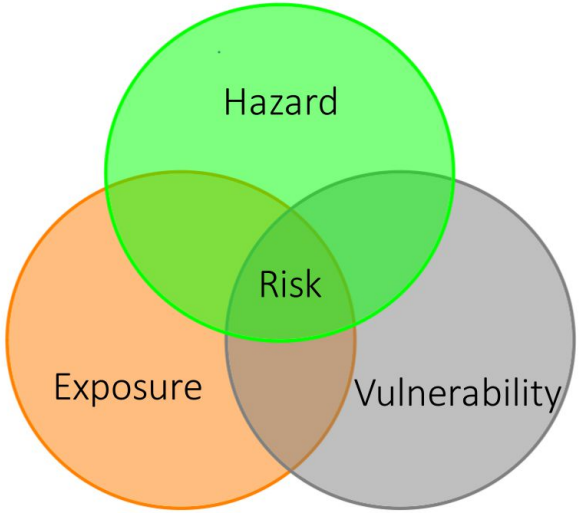
Definitions of Risk from the Oxford dictionary: 1) *a situation involving exposure to danger*; 2) *the possibility that something unpleasant or unwelcome will happen.*

Psychology: *"Psychosocial adversity or event considered a stressor that may hinder normal functioning."* - Risk and Resilience in Child Development (Anne S. Masten, 1994)

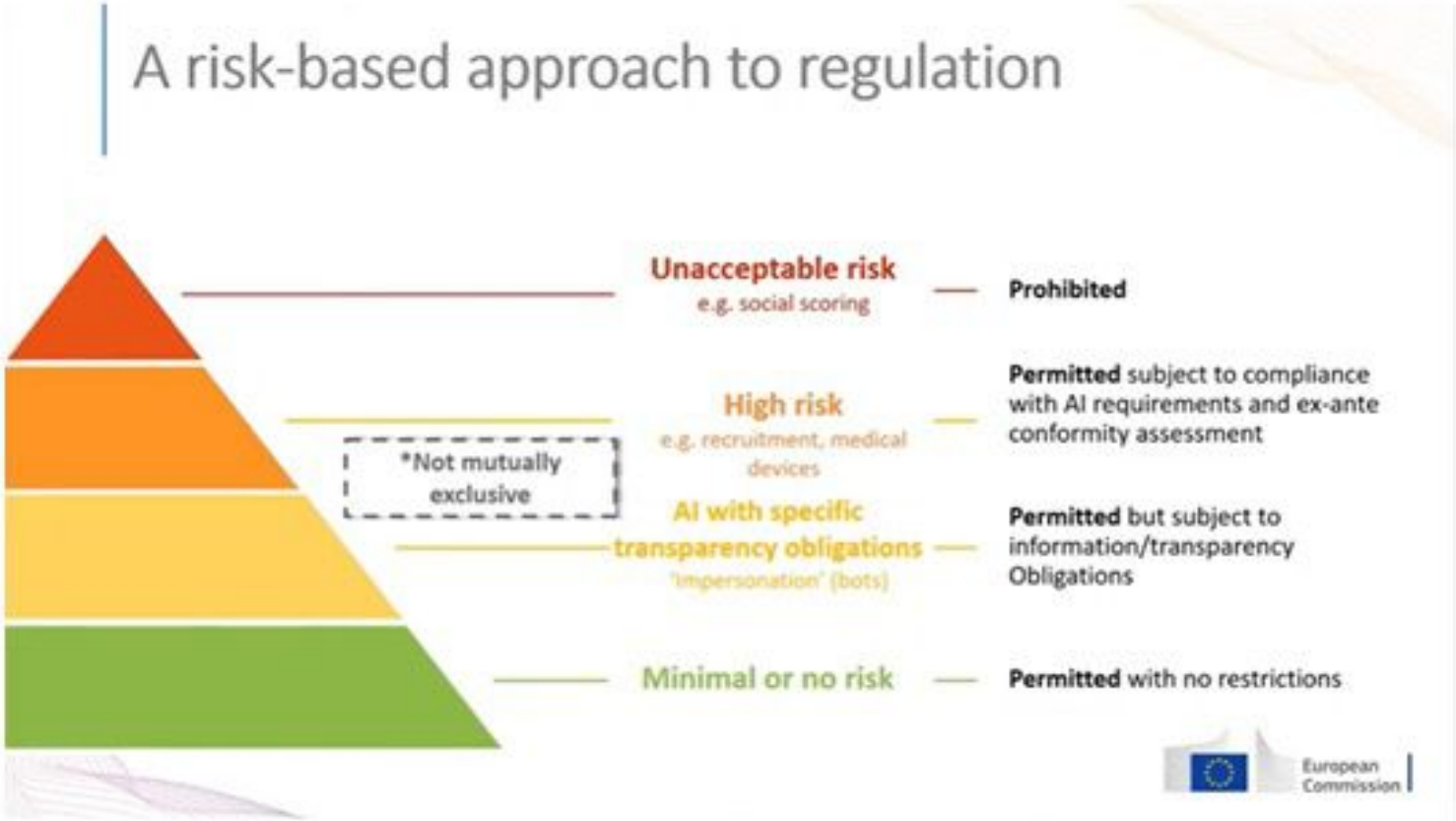
Social Science: *"the lack of perceived controllability", "set of possible negative consequences" and "fear of loss"* - Risk is a construct; Perceptions of risk perception (C. Vlek, 1996).

Engineering: *Risk is the probability of an undesired event multiplied by the consequences* - Risk and Reliability for Engineers, TU Delft (Robert Lanzafame, 2024)

Law: The Role of Law in Managing the Tension between Risk and Innovation, European Journal of Risk Regulation (Maria Weimer & Luisa Marin, 2017)



AI Act proposed by European Commission



Source: Lucilla Sioli, European Commission, European Vision for AI, 22 April 2021

"No one size fits all"

Avoid overregulation

Trustworthy **and** innovative AI

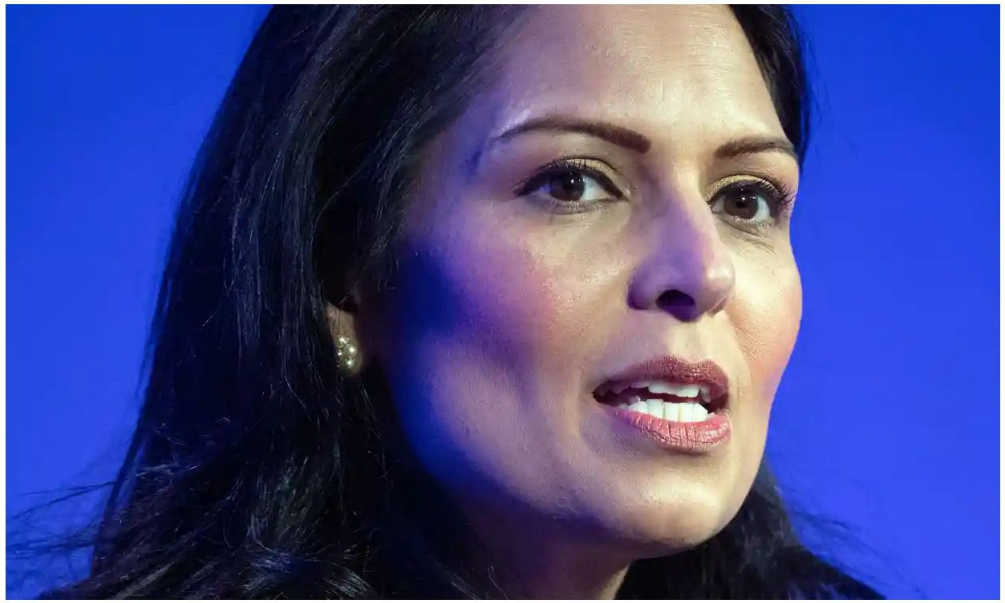
This approach offers a balance between innovation and regulation

Main criticism refers to the definition of high-risk

Subject to the existing legislation without additional legal obligations

Home Office to scrap 'racist algorithm' for UK visa applicants

Tool criticised for creating hostile environment for migrants and 'speedy boarding for white people'



Home Office solicitors confirmed that the home secretary, Priti Patel (pictured), 'has decided that she will discontinue the use of the streaming tool to assess visa applications, pending a substitute review of its operation'. Photograph: Dominic Lipinski/PA

The Home Office is to scrap a controversial decision-making algorithm that migrants' rights campaigners claim created a "hostile environment" for people applying for UK visas.

The "streaming algorithm", which campaigners have described as racist, has been used since 2015 to process visa applications to the UK. It will be abandoned from Friday, according to a letter from Home Office solicitors seen by the Guardian.

The decision to scrap it comes ahead of a judicial review from the Joint Council for the Welfare of Immigrants (JCWI), which was to challenge the Home Office's artificial intelligence system that filters UK visa applications.

SHARE

< NEWS

October 25, 2021

Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms

The Dutch government risks exacerbating racial discrimination through the continued use of unregulated algorithms in the public sector, Amnesty International said in a damning new analysis of the country's childcare benefit scandal.

The report *Xenophobic Machines* exposes how racial profiling was baked into the design of the algorithmic system used to determine whether claims for childcare benefit were flagged as incorrect and potentially fraudulent. Tens of thousands of parents and caregivers from mostly low-income families were falsely accused of fraud by the Dutch tax authorities as a result, with people from ethnic minorities disproportionately impacted. While the scandal brought down the Dutch government in January, sufficient lessons have not been learnt despite multiple investigations.

> "Governments around the world are rushing to automate the delivery of public services, but it is the most marginalized in society that are paying the highest price.
>
> Merel Koning, Senior Advisor on Technology and Human Rights

Recently added

- Senegal: The State must move from commitment to strong action to protect talibé children
- FIFA misleading world on remedy for migrant workers
- Op-ed: A flicker of hope for human rights in South Asia
- Write for Rights 2022: Championing activists in a year of global protest
- Write for Rights: World's biggest human rights event returns for Human Rights Day 2022

Amnesty International News

AI translation is jeopardizing Afghan asylum claims

Cost-cutting translations are introducing errors and putting refugees at risk.

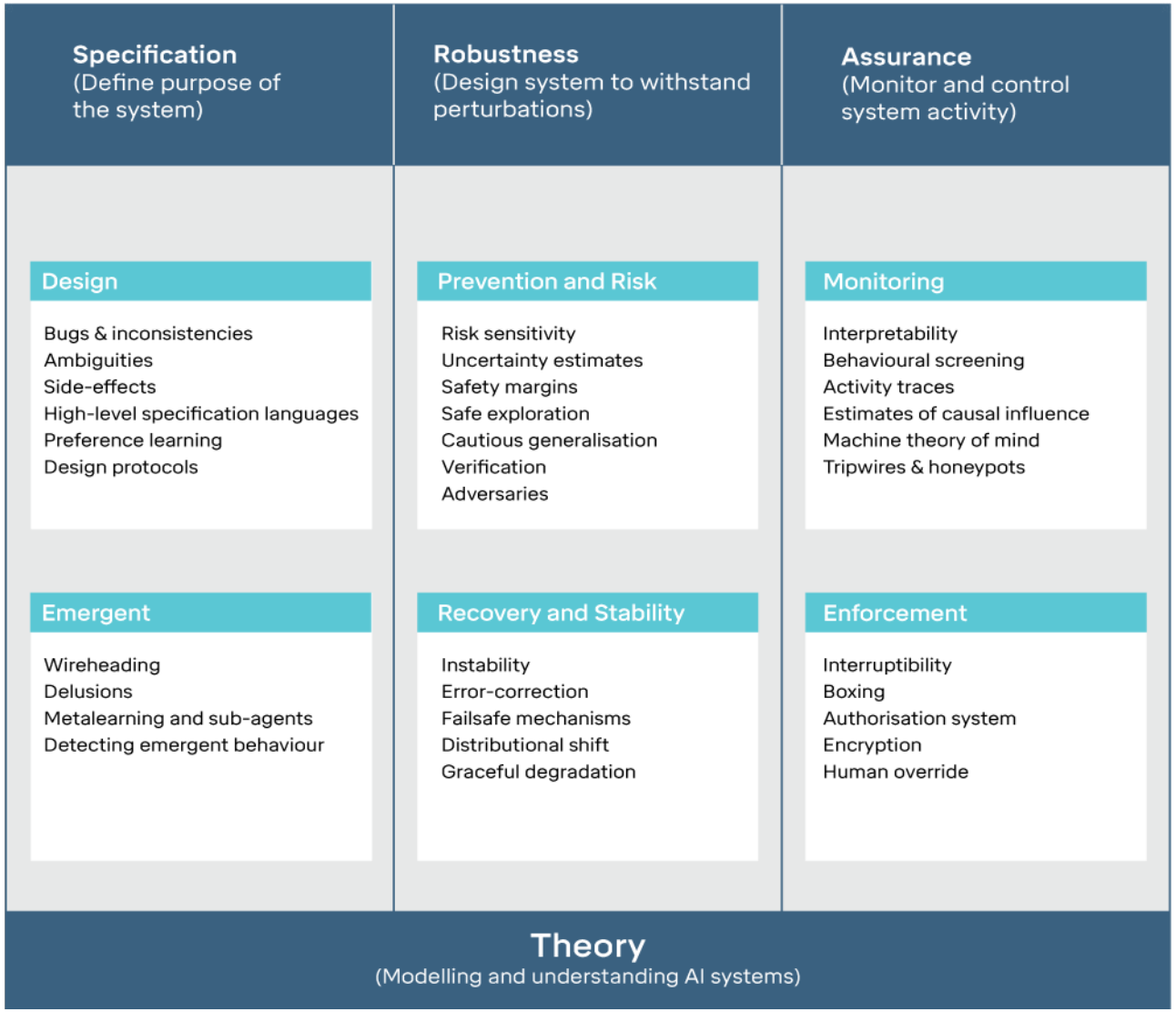
A crisis translator specializing in Afghan languages, Mirkhail was working with a Pashto-speaking refugee who had fled Afghanistan. A U.S. court had denied the refugee's asylum bid because her written application didn't match the story told in the initial interviews.

Lost in Translation: Algorithmic Discrimination

> "Machine-learning translations are not yet in a place to be trusted completely without human review."

The challenge of "low-resource" languages like Pashto and increasing use of machine translations in immigration services!

TECHNICAL CHALLENGES OF AI SAFETY



Pedro A. Ortega, Vishal Maini, and the DeepMind safety team, 2018

UNSOLVED (TECHNICAL) PROBLEMS IN AI SAFETY

Unsolved Problems in ML Safety

Dan Hendrycks UC Berkeley
Nicholas Carlini Google
John Schulman OpenAI
Jacob Steinhardt UC Berkeley

Robustness Create models that are resilient to adversaries, unusual situations, and Black Swan events.

Monitoring Detect malicious use, monitor predictions, and discover unexpected model functionality.

Alignment Build models that represent and safely optimize hard-to-specify human values.

Systemic Safety Use ML to address broader risks to how ML systems are handled, such as cyberattacks.

Position Paper, 2021

AI SAFETY ON WHOSE TERMS?

Why safety? One could view the shift to safety with cynicism. Big Tech, weary from bad publicity, is seizing the chance to be viewed as saviors from algorithmic harms, not perpetrators of them.

Sociotechnical approaches recognize and reject "safety-washing"— giving lip service to safe AI systems, without requisite commitments and practices to ensure this is the case — and call for transparency and accountability to keep companies honest.

Seth Lazar and Alondra Nelson

Editorial in Science, July 13, 2023

SOCIOTECHNICAL & INTERDISCIPLINARY APPROACHES?

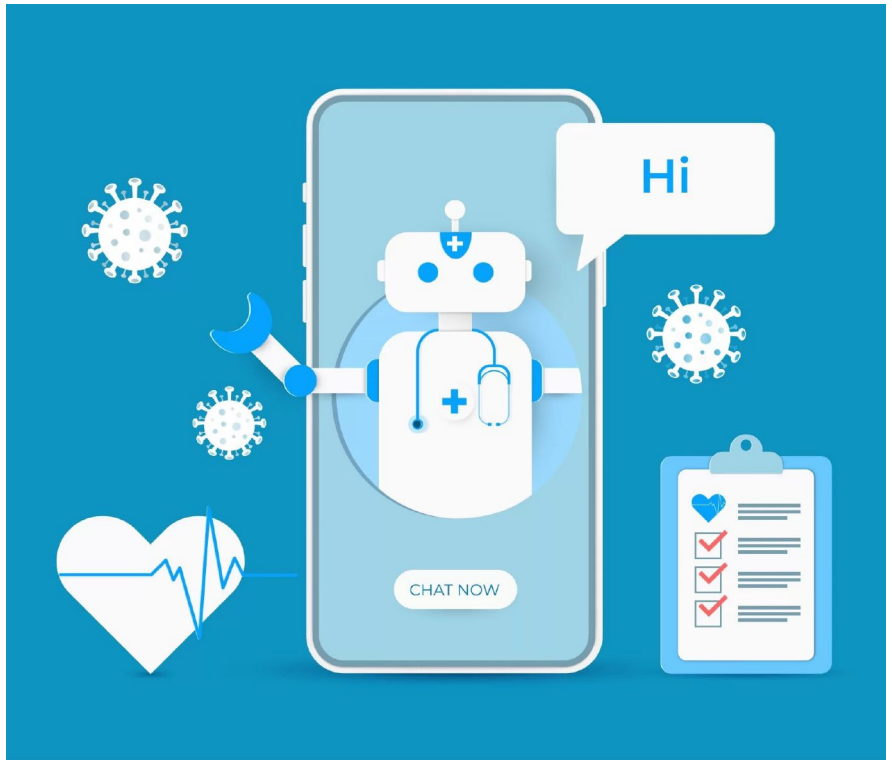
Applying Sociotechnical approaches to critically examine the risks and harms of AI "systems".

Addressing not just the algorithms, data & methods in AI, but the wider socio-political context of its design, deployment and use in society!

How do we make AI systems trustworthy, responsible and inclusive throughout their lifecycle and those affected?

Taking an interdisciplinary approach at the intersection of machine learning, design research, social sciences, law and society.

Dystopian Digital Futures: Automated Denial of Medical Services in the Future?



Rethinking “Algorithms” as a political & social-technical system (not just code)!

yle Uutiset Areena Urheilu Menu

Finnish Medical Association: Government's plans to deny undocumented migrants non-urgent care "irresponsible"

The organisation representing doctors across Finland said the move will not lead to savings, but will instead deepen inequality.

File photo. Image: Timo Metsäjoki / Yle

“In a statement, the association said that the proposal goes against a doctor's duty-of-care as well as the medical profession's code of ethics.” August 30, 2023

crai cis
critical ai & crisis interrogatives

Nitin Sawhney
Professor of Practice
Department of Computer Science, Aalto University

AI & Society Program
Finnish Center for Artificial Intelligence (FCAI)

Steering Committee
Global AI Policy Research Network

Aalto University
School of Science

Principles & Practices

- Explore the impact of technology in critical societal contexts for responsible AI, inclusion and civic agency.
- Interdisciplinary group working at the intersection of computational & social sciences engaging HCI, participatory design, and policy.
- Collaboration with societal partners, especially the public sector in Finland, EU and India.

Public Engagement

From Thought to Prompt: Cognitive Design Challenges in Human-LLM Interactions
11.10.2023 16:00–17:30 | Events

Life and Tech at the Edge: Refiguring Possible Futures
Speaker: Sarah Pink
20.9.2023 16:00–17:30 | Events

Theory-driven HCI
Speaker: Michel Beaudouin-Lafon
Professor of Computer
11.5.2023 14:00–15:30 | Events

Projects

Crisis Narratives

civic agency in artificial intelligence

DiME
ai—infused disinformation in media communications

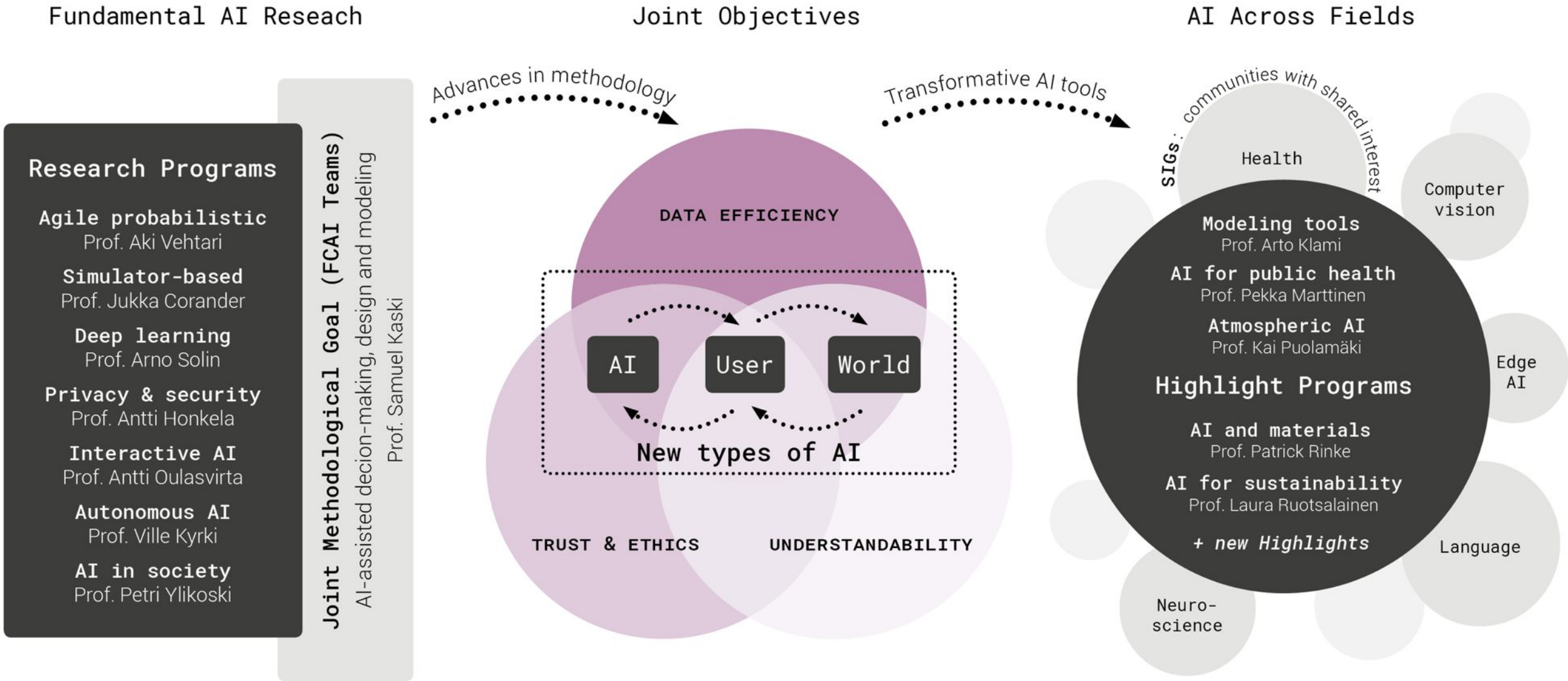
trust-M: Designing Inclusive & Trustworthy Digital Public Services for Migrants

crai-cis.aalto.fi

Department of Computer Science

https://www.aalto.fi/en/department-of-computer-science

Finnish Center for Artificial Intelligence (FCAI)



fcai.fi/research

RESPONSIBLE AI IN THE PUBLIC SECTOR?

1. Understanding the challenges of innovations in public sector AI from **ethical and regulatory compliance** to fostering **experimentation**.

2. Facilitating participation of **diverse stakeholders** throughout the **AI lifecycle** of designing, deploying, and assessing public sector AI services.

3. Aligning the **values and practices of public sector** organisations with how AI-based services are envisioned & deployed.

4. **Piloting AI Regulatory Sandboxes** to explore novel AI services, facilitating technological innovation with regulatory compliance.

USE OF AI-BASED SYSTEMS IN THE PUBLIC SECTOR

Opportunities (Barker *et al.*, 2021; Manzoni *et al.*, 2022)	**Challenges** (Pechtor & Basl, 2022; Pūraitė *et al.*, 2020)

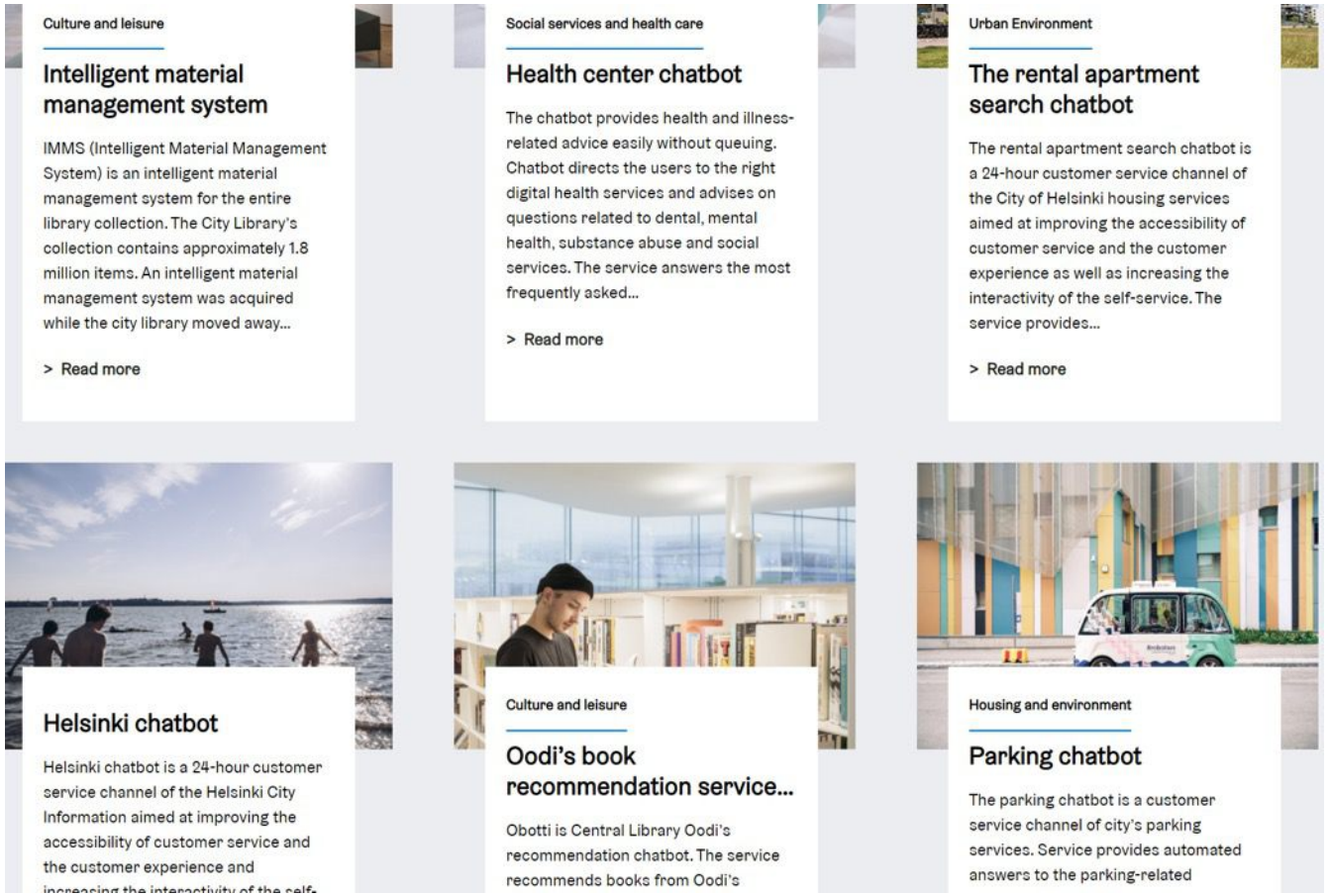
• Support context-specific public values:
  o Operational
  o Political
  o Social
• Foster citizen trust & participation
• Improve efficiency & decision making
• Provide innovative digital services
• Personalisation of public services
	• The rule of law
	• Complex ecosystem, multi-stakeholders involved throughout the AI lifecycle
	• Different values and incentives
	• Balancing benefits & risks of AI-based systems
	• Demonstrate innovation over short time
	• Public administration bodies work in silos

- Need for tools, platforms and practices that facilitate experimentation with AI-based systems.

- Ensuring technologically innovative, ethically responsible, and legally compliant systems.

AI REGISTRIES: CITY OF HELSINKI

- Rule-based chatbots & information services developed for residents of Helsinki. Aims:
  - Leverage advanced analytics such as ML, dynamic optimisation, and predictive models to improve city operations and use of public resources.
  - Adoption of AI-based services according to participatory approaches that fosters trust, accountability and human oversight.
- Incorporate high-level (abstract) ethical AI principles into innovation strategies but cannot easily translating them into concrete measures.
- AI Registries document different aspects of AI services but lack dynamic versioning (*what*), auditability (*where*), & chain of accountability (*who*).



City of Helsinki AI Register – setup by Saidot.ai

Way Forward?

Fostering Inclusive Trustworthy and Responsible AI in the Public Sector

Examining discourses of shared and contested values embedded in the EU AI Act and Public AI Services

civic agency in artificial intelligence

- Gonzalez Torres, A. P., Kajava, K., & Sawhney, N. 2023. Emerging AI Discourses and Policies in the EU: Implications for Evolving AI Governance. *Communications in Computer and Information Science, Springer.*
- Kajava, K. and Sawhney, N. 2023. Language of Algorithms: Agency, Metaphors, and Deliberations in AI Discourses. Lindgren, S. *Handbook of Critical Studies of Artificial Intelligence. Edward Elgar Publishers*.

Creating trustworthy and accessible digital public services for migrants

trust m

- Interdisciplinary team of 30+ people
- Funded for 3 (+3) years
- Looking for hybrid services, possibly based on conversational AI and/or speech-based interaction
- One of the main migrant groups of interest is migrant women
- Piloting services in the City of Espoo.

strategic RESEARCH

Aalto University — TRUST TECH

Tampere University — DESIGN

HELSINGIN YLIOPISTO — POLICY

ESPOO ESBO — IMPACT

trustmproject.aalto.fi

Participatory Research & Collaborative Design

21.7.2023 / Collaborations

Stretching our empathy towards migrant challenges: collaborative workshops with the city of Espoo

By Rūta Šerpytytė – Design researcher at Trust M

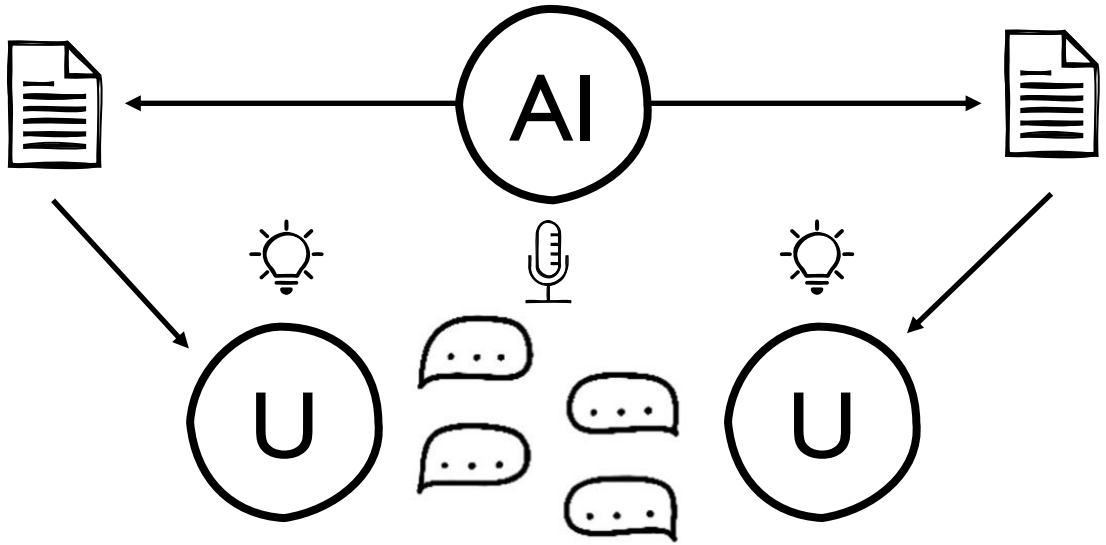
Bhuvana Sekar, Aalto University and Irena Bakic, City of Espoo

Rahim Ahsanullah and Lucy Truong, Aalto University

Enhancing Conversations in Migrant Counseling Services: Designing for Trustworthy Human-AI Collaboration

- Lucy Truong, Thesis Research (2023)

ENHANCING CONVERSATIONS IN MIGRANT COUNSELING SERVICES: DESIGNING FOR TRUSTWORTHY HUMAN-AI COLLABORATION



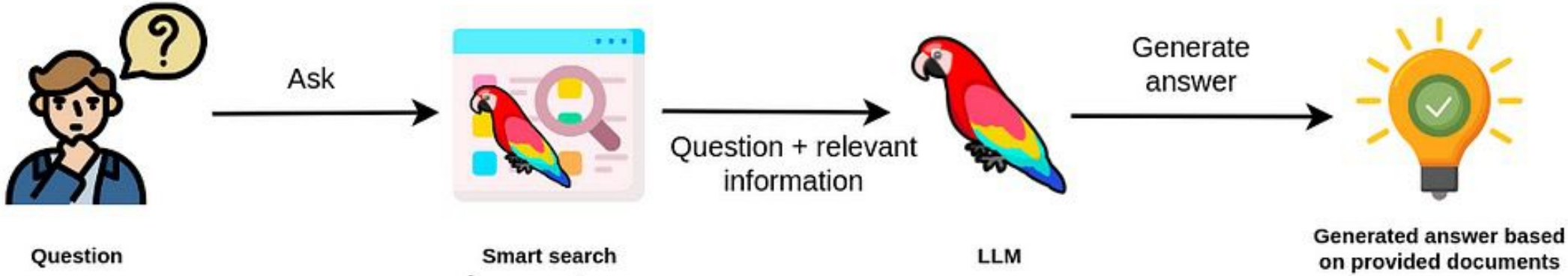
Examining existing values and practices of city counselors from conversations with migrant users. Role of distributed knowledge (& values) emerging in human-AI interaction?

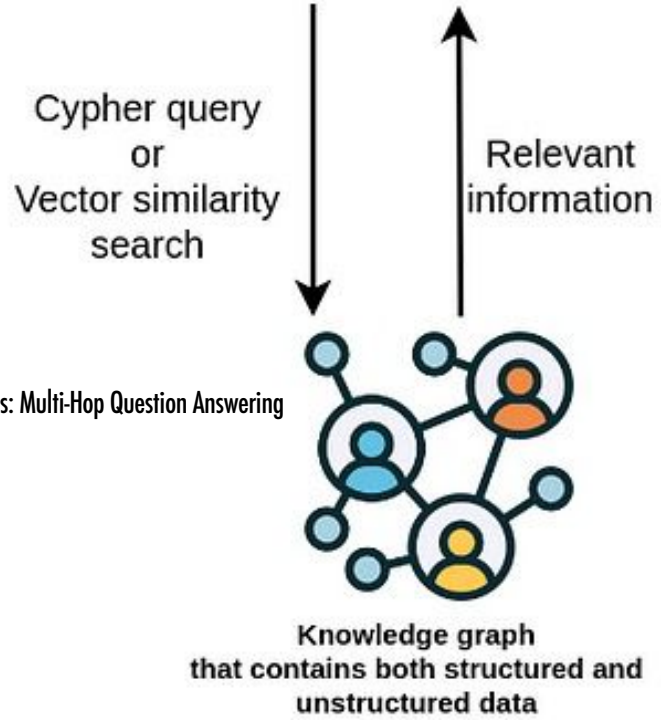
Design research & prototyping at Aalto University, Lucy Truong's thesis research with the City of Espoo (2023)

* Truong, L., Lee, S., & Sawhney, N. 2024. Enhancing Conversations for Migrant Counseling: Designing for Trustworthy AI-mediated Collaboration between Migrants and Service Advisors. *ACM Conference on Computer-Supported Cooperative Work and Social Computing*, San José, Costa Rica, November 9-13, 2024.

GENERATING CONVERSATION SUMMARIES

Embedding Public Values in Knowledge Graphs?





Bratanic, T. (2023) Knowledge Graphs & LLMs: Multi-Hop Question Answering

Advantages:

- logical reasoning and answering multi-hop questions
- use reliable knowledge curated by the area experts
- the answers can be traced and linked to source
- can be scaled to wider knowledge without pretraining

Challenges / Opportunities:

- requires building domain-specific Knowledge Graph from a range of data sources (documents, interviews, conversations)
- slow & deliberative requiring human + automated process
- need to explicitly examine, amplify and validate the nature of diverse or contested values embedded

* Datta, P., Chizhikova, A., Vitiugin, F., & Sawhney, N. 2024. Construction of Hyper-Relational Knowledge Graph Using Pre-Trained Large Language Models. *AAAI Conference on Artificial Intelligence, Workshop on Graphs and more Complex structures for Learning and Reasoning*, February 20-27, 2024, Vancouver, Canada.

AI REGULATORY SANDBOXES

* Some slides based on presentation by Alex Moltzau, EU AI Office, FARI Workshop on Regulatory & Innovation Sandboxes for Applied AI Technologies, Brussels, Belgium, November 19, 2024.

EUROPEAN **ARTIFICIAL INTELLIGENCE** OFFICE

The EU AI Act

...laying down a **uniform legal framework** in particular for the **development**, the placing on the market, the putting into **service** and the **use** of artificial intelligence systems (**AI systems**) in the Union, in accordance with Union values, to promote the uptake of **human centric** and **trustworthy** artificial intelligence (AI) while ensuring a high level of **protection of health, safety, fundamental rights**...

EUROPEAN **ARTIFICIAL INTELLIGENCE** OFFICE

AI ACT AND GENERAL PURPOSE AI (GPAI)

Regulation 2024/1689 (the "AI Act"), entered into force on August 1st, 2024.

General Purpose AI *(Article 3 (63))*:

- *"a model… trained with a large amount of data*
- *using self-supervision at scale,*
- *displays significant generality and*
- *perform a wide range of distinct tasks*
- *and …*
- *can be* ***integrated into a variety of downstream systems or applications****"*

AI Regulatory Sandboxes for Public Sector Innovation

*A concrete and controlled **framework** set up by a competent authority which offers providers or prospective providers of AI systems the **possibility to develop, train, validate and test**, where appropriate in real world conditions, **an innovative AI system**, pursuant to a sandbox plan for a **limited time under regulatory supervision**.*

EU AI Act: Article 3(55)

** However, EC Proposal article 52(4) indicates that participants in AI regulatory sandboxes would remain **liable** for any harm inflicted on third parties as a result of experimentation in the sandbox environment.*

EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE

Why AI Regulatory Sandboxes?
Article 57(9)

The establishment of AI regulatory sandboxes shall aim to contribute to the following objectives:

(a) improving **legal certainty** to achieve regulatory compliance with this Regulation or, where relevant, other applicable Union and national law;

(b) supporting the sharing of **best practices** through cooperation with the authorities involved in the AI regulatory sandbox;

(c) **fostering innovation** and competitiveness and facilitating the development of an AI ecosystem;

(d) contributing to evidence-based **regulatory learning**;

(e) facilitating and accelerating access to the Union market for AI systems, in particular when provided by **SMEs**, including **start-ups**.

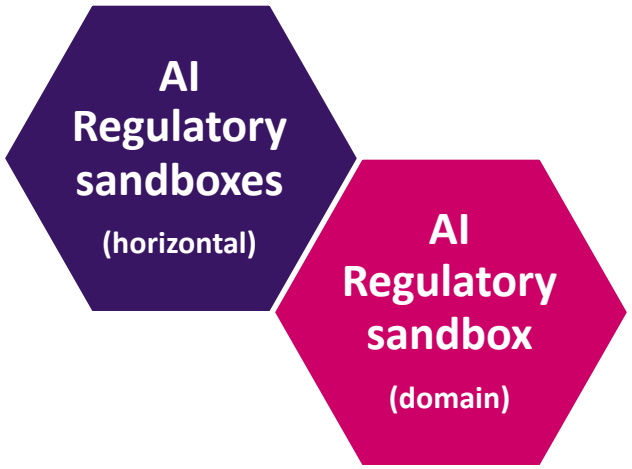
EUROPEAN **ARTIFICIAL INTELLIGENCE** OFFICE

EU role for the AI Regulatory Sandboxes

European AI Office

AI Board of MS

National Competent Authorities



- The **AI Office** together with the **AI Board**, and its sub-group, as framework for Member States, and their **National Competent Authorities** (NCAs), to coordinate their activities and cooperate
- European Commission (EC) to develop detailed arrangements for the establishment, development, implementation, operation and supervision of AI Regulatory Sandboxes (via implementing acts)

DRAFT ILLUSTRATION

EUROPEAN **ARTIFICIAL INTELLIGENCE** OFFICE

EU role for the AI Regulatory Sandboxes

European AI Office

AI Board of MS

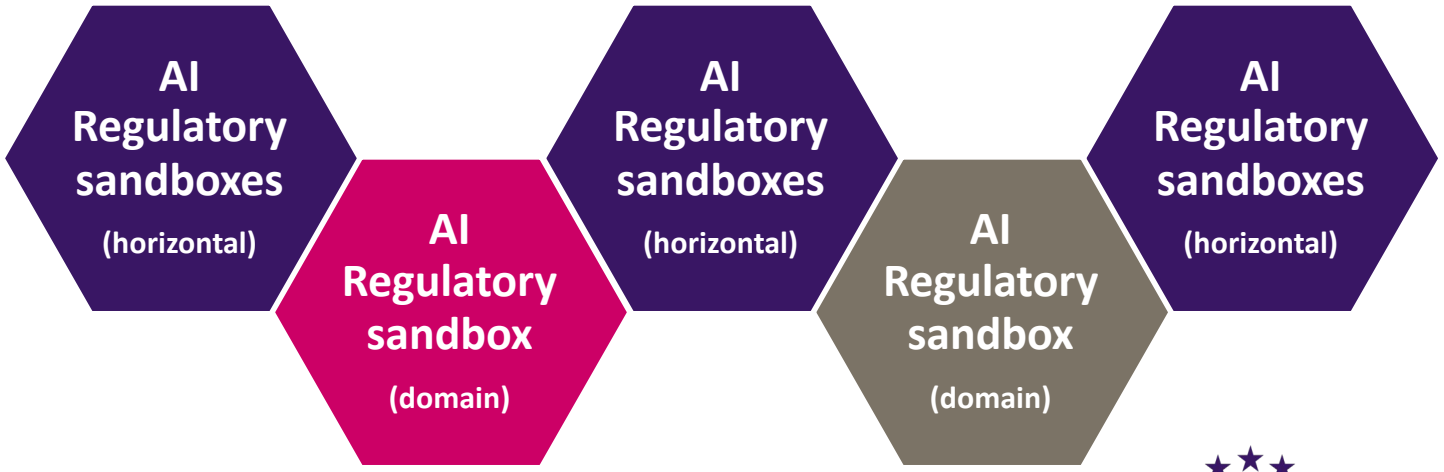
AI Board Subgroup on
AI Regulatory Sandboxes

National Competent Authorities

National Competent Authorities

National Competent Authorities

- EC to coordinate and facilitate cross-border cooperation between NCAs
- EC to develop a single interface containing all relevant information related to sandboxes



DRAFT ILLUSTRATION

EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE

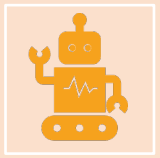
EU AI Regulatory Sandbox Framework

- The **Sandbox** serves as meeting point between Regulators and EU Initiatives providing support to AI providers to comply with EU law and to test, validate, and improve their AI models, i.e. **Regulatory Learning** and **Innovator Learning**
- **Regulators**: AI Office, NCAs, Market Surveillance, Data Protectors, Fundamental Rights Agency, etc supervise and provide guidance guaranteeing compliance to AI EU law
- **Providers**: SMEs and Startups explore AI opportunities
- **EU initiatives**: EDIHs, TEFs and AI factories providing support for testing and development
- **Citizens and civil society stakeholders**: to enjoy safer AI models
- Regulators oversee the process, ensuring safety for consumers and citizens by protecting their rights, as the sandbox generates valuable insights and data, informing regulatory frameworks and fostering innovation and compliance.

EUROPEAN **ARTIFICIAL INTELLIGENCE** OFFICE

AI LIFECYCLE APPROACH

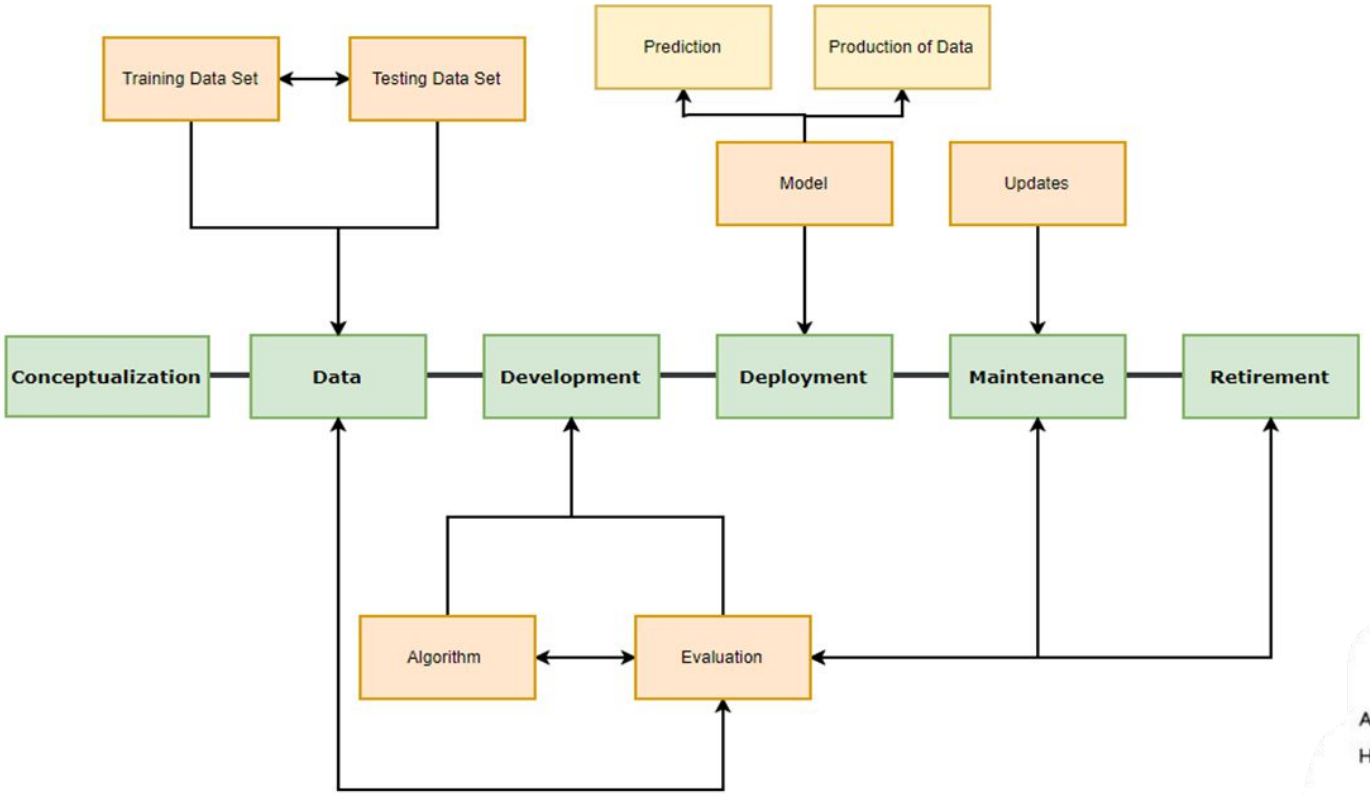
Challenges of adopting and deploying AI-based solutions require engaging responsible & ethical practices with **multiple stakeholders** involved across the entire AI lifecycle
(De Silva & Alahkoon, 2021).



Public sector's organizational logic is based on hierarchy and verticality (Pūraitė *et al.*, 2020), while AI lifecycle approach benefits from **horizontal embedding of roles** and responsible actions from multiple stakeholders across different stages.

Regulatory compliance should be embedded in **different stages** of design, use of data, development, deployment, maintenance and retirement of AI systems.

STAGES OF THE AI LIFECYCLE

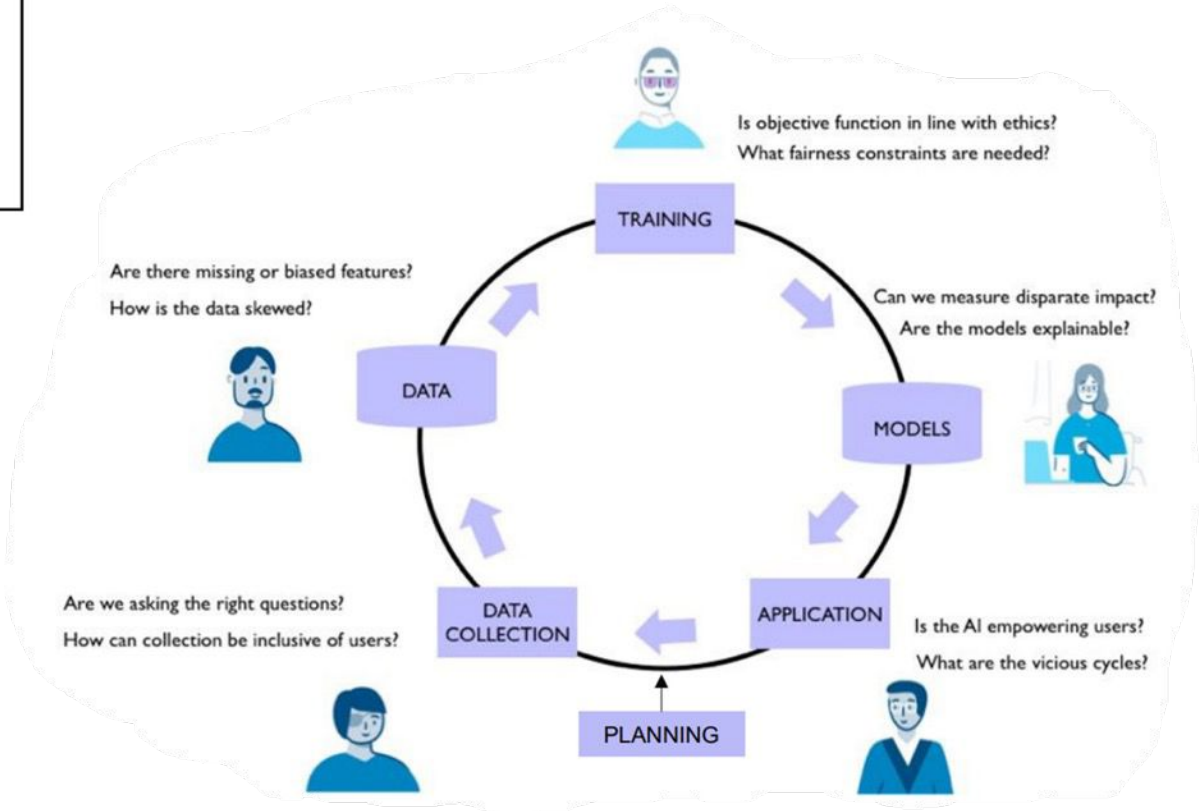


Complying with ethical & regulatory measures during lifecycle:

- Training, testing, and evaluating AI systems with quality data
- Human oversight to prevent or minimise risks
- Mechanisms to address unintended feedback loops constant monitoring through the lifecycle
- Withdrawal or recall of non-compliant high-risk AI systems

Ethical considerations through lifecycle stages:

1. **Design**: why an AI-based approach?
2. **Training**: are the datasets biased?
3. **Development**: how outcomes are validated?
4. **Deployment**: what harmful impacts may emerge?
5. **Maintenance**: are there discriminatory feedback loops?
6. **Retirement**: what happens if system recalled?

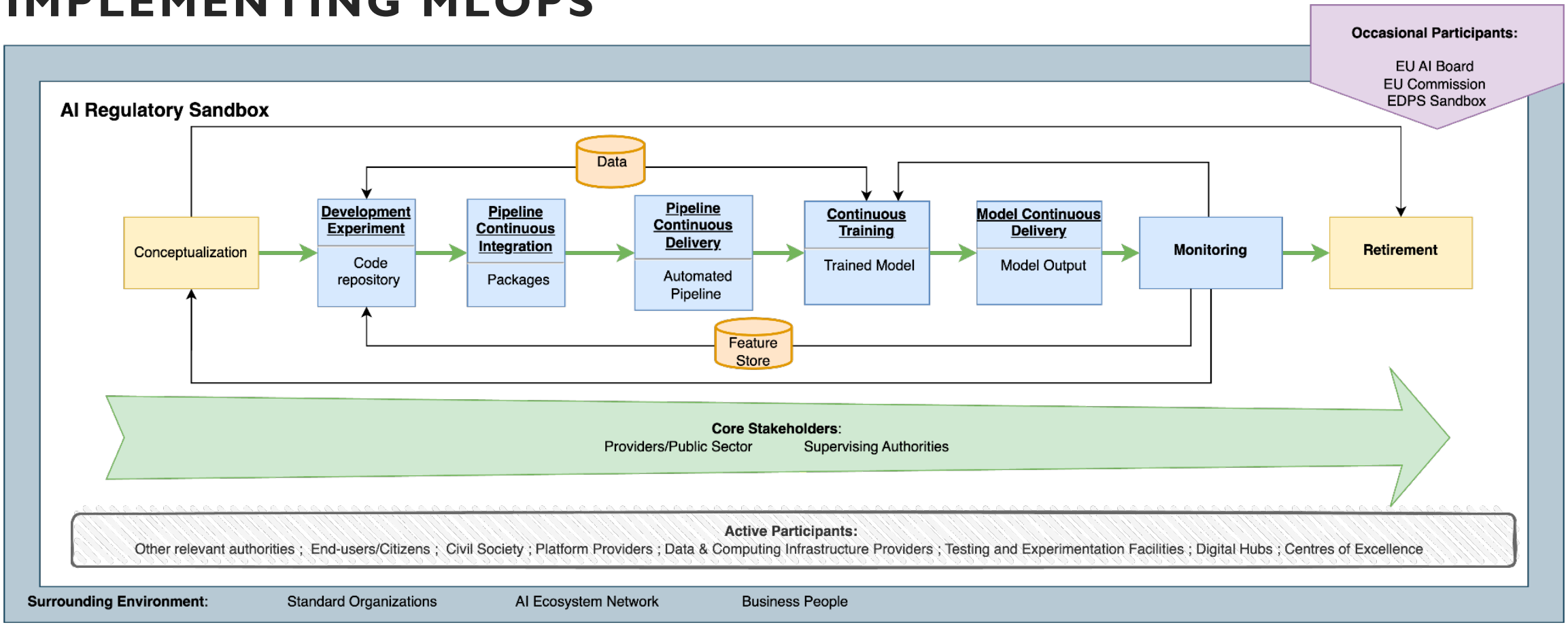


INTEGRATIVE FRAMEWORKS: FROM MLOPS TO REGOPS

- **Machine operations** (**MLOps**) address need for agile and dynamic tools to support technical and responsible adoption of AI-based services across their lifecycle (Pechtor & Basl, 2022).
  - Software framework to support continuous monitoring, versioning, enhanced transparency, auditing & improved usability of resulting AI systems (Ranawana & Kurananda, 2021).
  - Useful for environments with constantly changing needs (like regulatory sandboxes), but automation can pose obstacles to compliance from constant requests by regulatory bodies.

- **Regulatory Operations** (**RegOps**) designed to support regulatory processes e.g., for certification of medical devices and AI-based medical systems.
  - **Continuous monitoring and flagging** of events that can trigger interventions from multiple providers of different aspects of an AI-based system.
  - Facilitate responsible AI lifecycle approach to allow **tracing of impact and liability**

MULTI-STAKEHOLDER AI REGULATORY SANDBOX IMPLEMENTING MLOPS



ENGAGING PUBLIC VALUES IN AI REGULATORY SANDBOXES

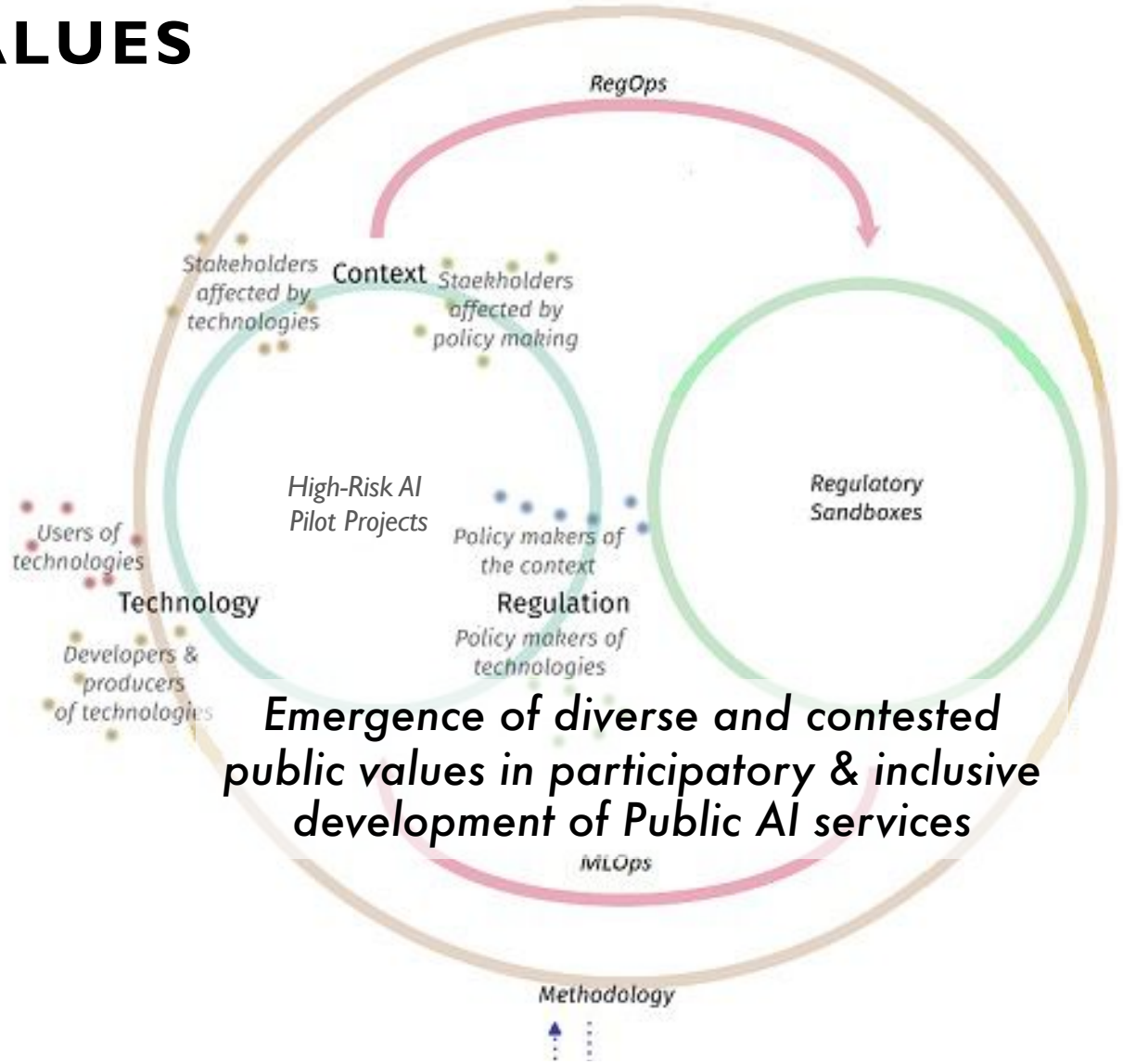
Conduct pilots of AI Public Sector services using AI Regulatory Sandboxes, integrated platforms and participatory infrastructures, in focused domains like public health, municipal services etc.

Engage with multiple stakeholders:

- Regulators, public administration, providers, developers, domain experts, SMEs and citizens and civil society groups.

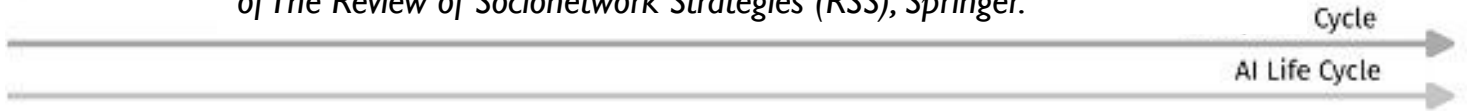
Expected outcomes:

- Developing platforms, guidelines, policies and best practices to support sandboxes
- Provide the space for interactions and mutual collaboration though out the AI lifecycle
- Examine the risks, limitations and possibilities of legislation and technological innovation



Emergence of diverse and contested public values in participatory & inclusive development of Public AI services

* Gonzalez Torres, A. P. and Sawhney, N. 2023. Role of AI Regulatory Sandboxes and MLOps for Finnish Public Sector Services. *Special issue of The Review of Socionetwork Strategies (RSS), Springer.*



civic agency in artificial intelligence

AI & SOCIETY
https://doi.org/10.1007/s00146-022-01502-2

ORIGINAL ARTICLE

Contestations in urban mobility: rights, risks, and responsibilities for Urban AI

Nitin Sawhney[1]

Received: 2 July 2021 / Accepted: 19 May 2022

Abstract

Cities today are dynamic urban ecosystems with evolving physical, socio-cultural, and technological infrastructures. Many contestations arise from the effects of inequitable access and intersecting crises currently faced by cities, which may be amplified by the algorithmic and data-centric infrastructures being introduced in urban contexts. In this article, I argue for a critical lens into how inter-related urban technologies, big data and policies, constituted as Urban AI, offer both challenges and opportunities. I examine scenarios of contestations in *urban mobility*, defined broadly to include equitable access, movement, and liberty to engage with the socio-cultural, political, and urban fabric of cities. I anchor my arguments through a framework of *rights, risks,* and *responsibilities* for critically examining and configuring the roles, values and ethical implications for all stakeholders including human, AI and non-human entities within an urban ecosystem. As a way forward, I examine the European Commission's proposed regulations on AI systems through an illustrative case study of an automated parking control system introduced by the City of Amsterdam. In moving beyond the city to broader urban ecosystems, I highlight the role of engaging Indigenous perspectives for designing and reconciling the implications of equitable and sustainable Urban AI ecosystems in the future.

Keywords Urban AI · Artificial intelligence · Urban · Ethics · Rights · Risks · Regulations · Smart cities · Mobility

Urban & Public Sector AI

Parking Control Systems (PCS) in Amsterdam: Analyzing Ethical & Regulatory Implications

Algorithmic Data Processing
Automated parking control
City of Amsterdam

Dit schema is alleen beschikbaar in het Engels. Als u een Nederlandse vertaling wilt, neem dan contact op met: I.fliert@amsterdam.nl.

City residents and visitors park their cars at the streets.

ANPR-camera scans license plates and translates the images into data.

The license plate is checked with National Parkin Register (NPR) by using function 'Check PsRight'.

Ok

The car has parking rights.
Only metadata is shared with EGIS Parking Service (anonymously).
The scan is deleted within 48hrs.

Not ok

The car has insufficient parking rights.
An Egis Parking Service employee checks the images.

Correct

The scan is correct.
License plate number is sent to municipal tax authorities to issue a fine.

Incorrect

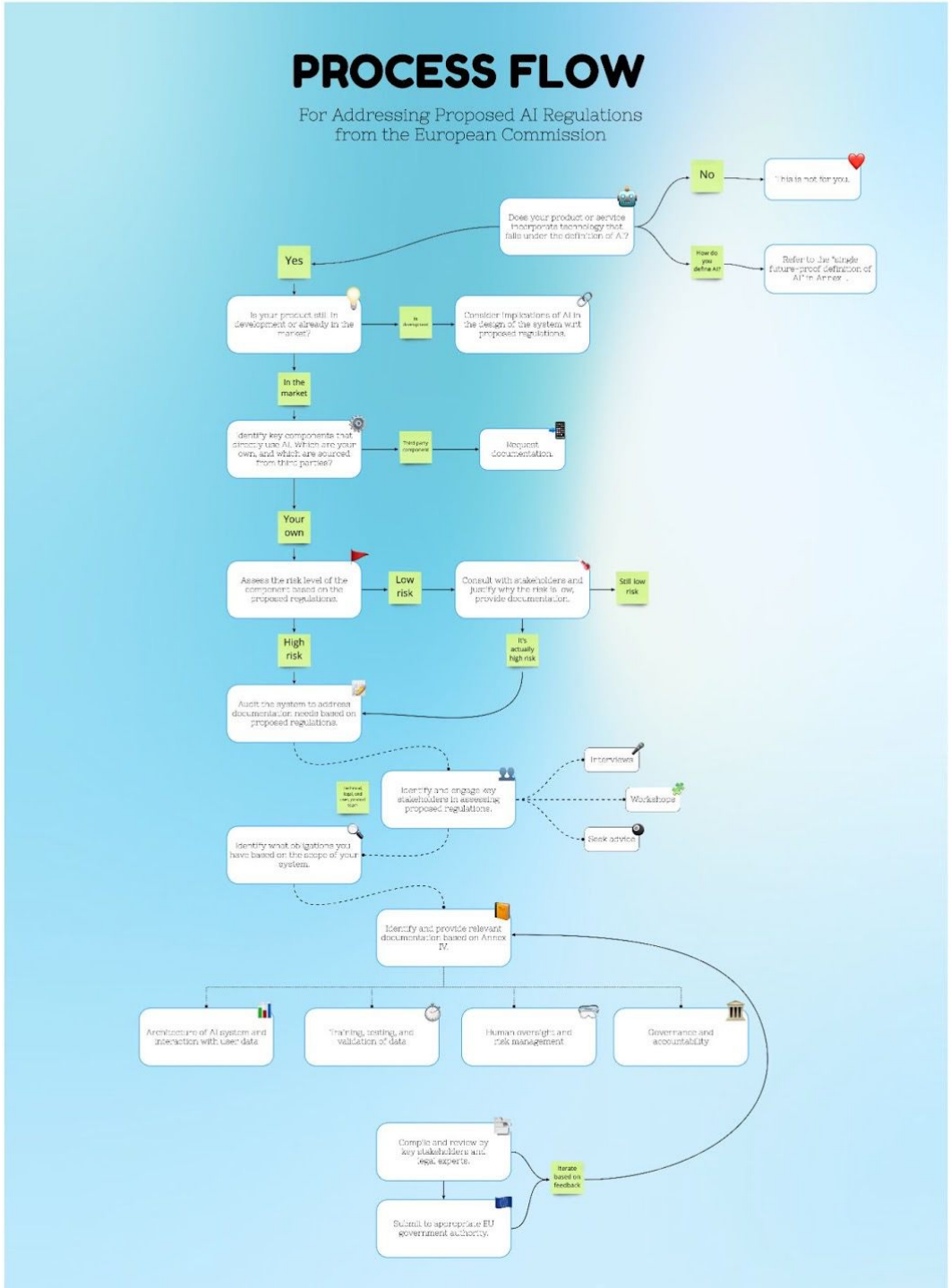
The scan is incorrect.
The license plate data is corrected and re-entered in NPR.

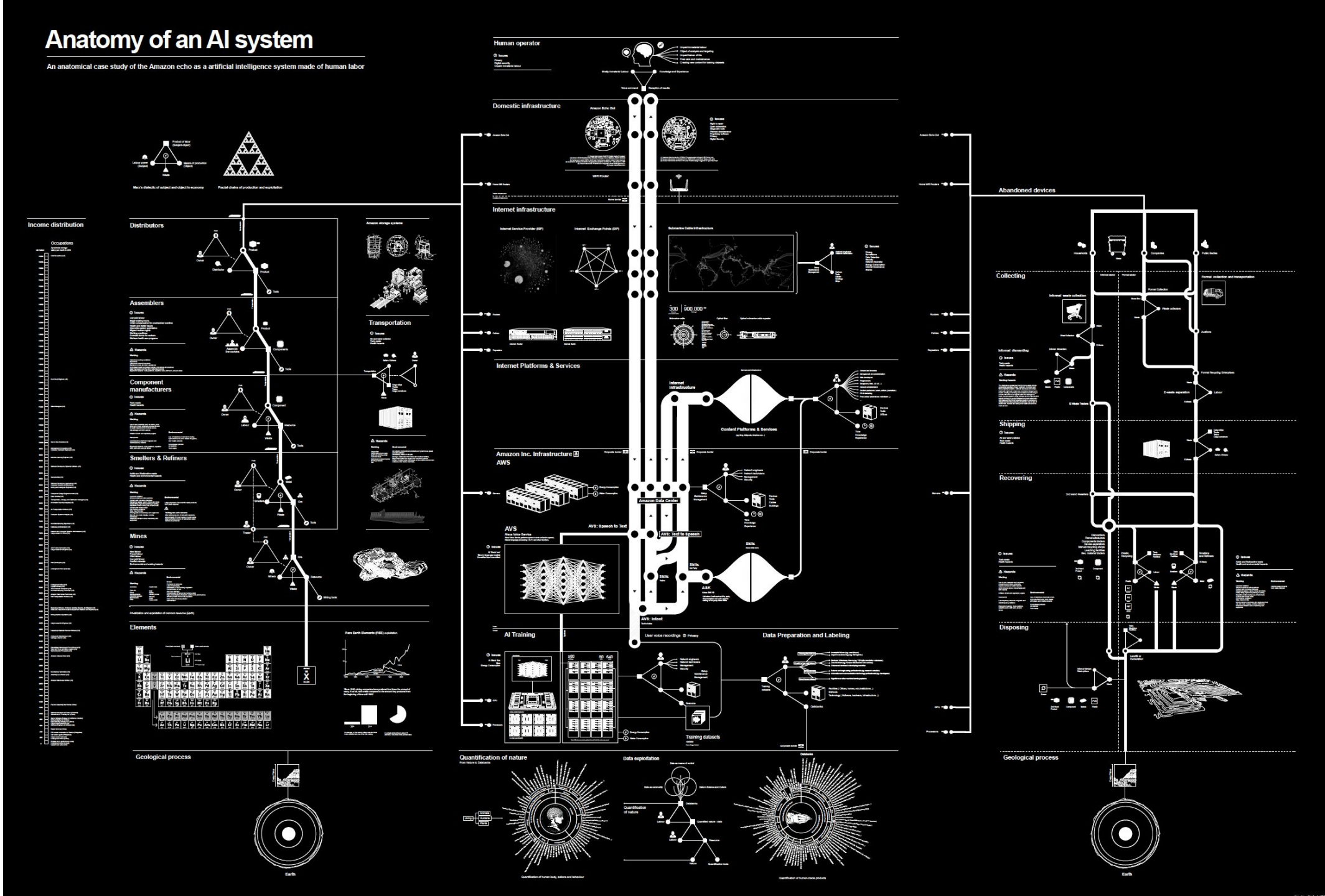
No violation

No violation (e.g. car is waiting at traffic lights).
The scan is deleted within 48hrs.

Legend

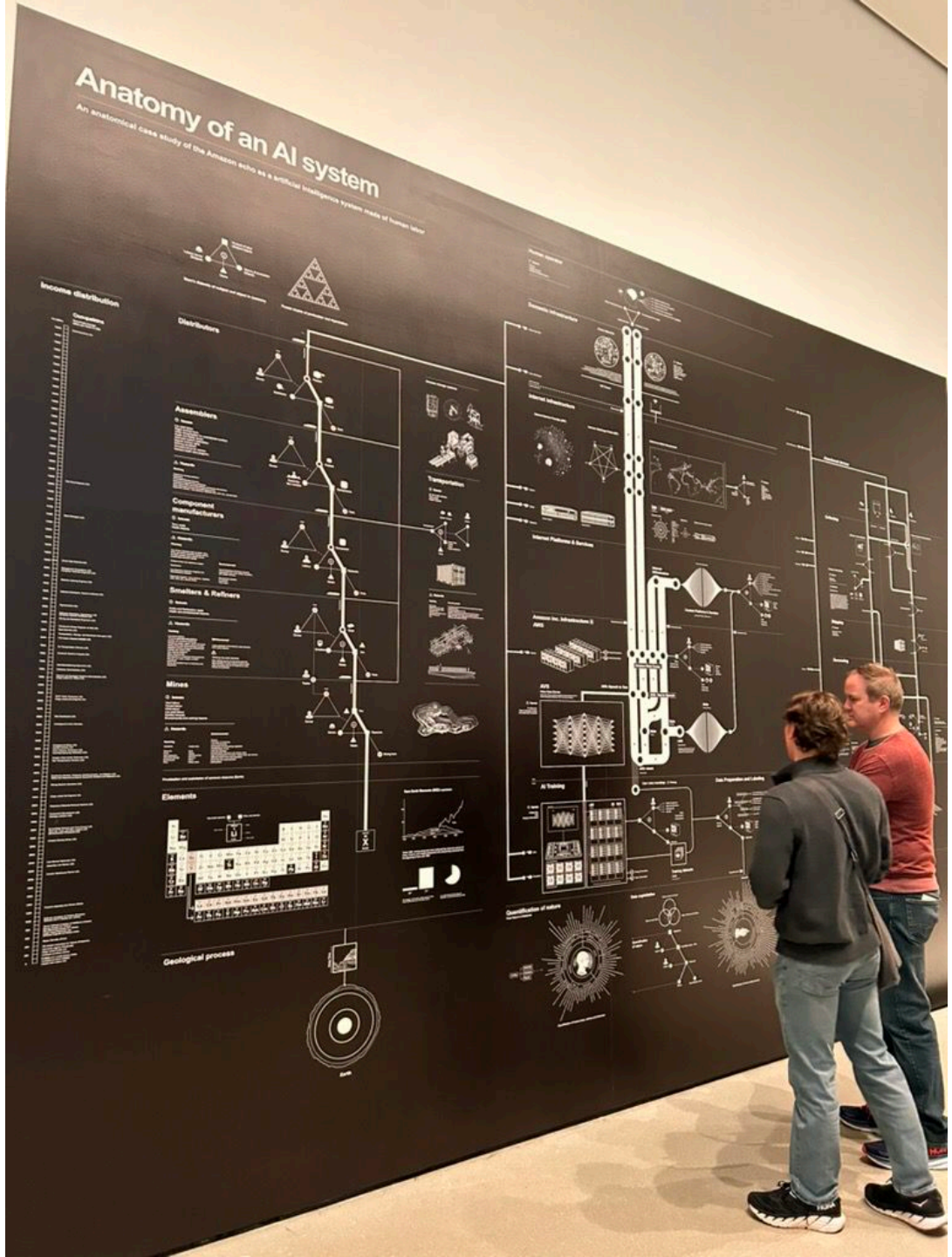
User performed activity | Automated activity | Civil servant activity





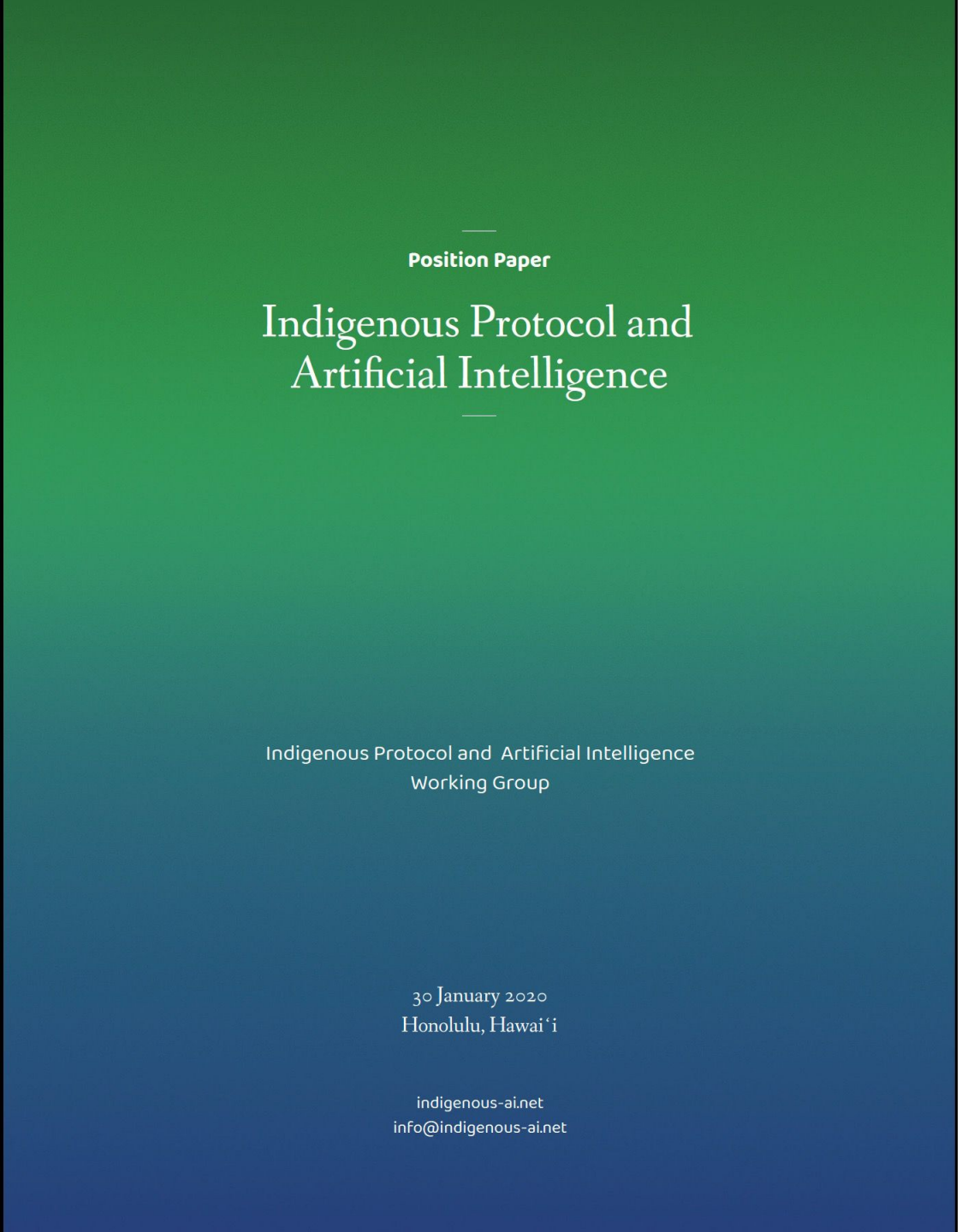
Vladan Joler & Kate Crawford, AI Now Institute, 2018.
www.anatomyof.ai

FOSTERING CRITICAL AI LITERACY & PUBLIC IMAGINATION



Vladan Joler & Kate Crawford, MoMA 2024
www.anatomyof.ai

Aalto University
School of Science



Guidelines for Indigenous-centred AI Design v.1

1. Locality

2. Relationality and Reciprocity

3. Responsibility, Relevance and Accountability

4. Develop Governance Guidelines from Indigenous Protocols

5. Recognize the Cultural Nature of all Computational Technology

6. Apply Ethical Design to the Extended Stack

7. Respect and Support Data Sovereignty

By **Laura Lamberti**

Laura Lamberti is a junior reporter at The Parliament Magazine

25 Jan 2023

@LauraLamberti10

Is the AI Act missing safeguards on migration?

The European Commission's proposed AI Act – the first-ever legal framework on artificial intelligence – includes an exemption that could allow for the use of certain high-risk technologies in migration-related procedures

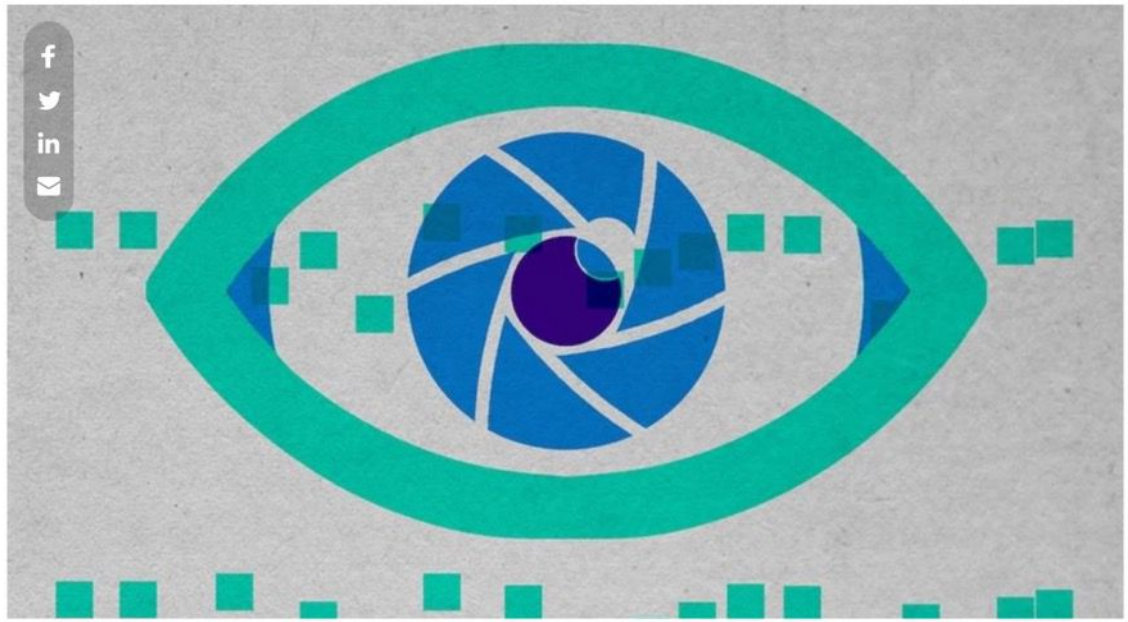


Illustration by Joe Magee

The photo is hard to stomach: it shows a man's back lined with bright red and pink lacerations. The caption reads: "Injuries sustained to the abovementioned respondent's back after expulsion by Croatian authorities."

The image was taken in Vrata, Croatia, in 2019, by affiliates of Border Violence Monitoring Network (BVMN), a coalition of organisations documenting illegal pushbacks and police violence by European Union Member State authorities in the Western Balkans and Greece.

READ NEXT:

Opinion

OLAF report shows yet again how Frontex systematically ignores human rights

by Tineke Strik

ANDREAS SCHÜLLER

JOHANNA VEHKKO

ALEKSI KNUUTILA

PANEL DISCUSSION

International Humanitarian Law, Disinformation and Counter Media Narratives in Warfare

YULIIA DUKACH

SALLA-MAARIA

KRISTIAN HU

TRANSDISCIPLINARY SYMPOSIUM ON
AI, HUMAN RIGHTS AND WARFARE

Contestations.AI

OCTOBER 23, 2024
THE TRAIN FACTORY, HELSINKI
+ LIVESTREAMED

NEIL CHRISTOPHER RENIC

JESSICA DORSEY

PANEL DISCUSSION

Critical perspectives on AI, Big Data and Autonomous Weapons

NITIN SAWHNEY

LAUREN G

UMPULA-NATRI

JOONAS LEPPÄNEN

CONVERSATION

EU and Global Implications for Responsible AI

TEIVO TEIVAINEN

SALLA-MAARIA LAAKSONEN

Global AI Policy Research Summit 2025

Deep Dive Workshop
AI in Warfare:
Actions, Policies and Practices

14 November 09.30 - 12.30
Panorama @Mondai | House of AI

Meet the Hosts & Panellists

Meet the Panellis

Virginia Dignum
Professor in Responsible AI and Director of the AI Policy Lab, Umeå University. Member of the UN High Level Advisory Body on AI

Martine Jaarsma
Doctoral Researcher International Humanitarian Law, Military uses of AI and Cricital Legal Studies, University of Antwerp

Megan Karlshøj-Pederson
Policy Specialist at Airwars

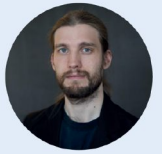
Meet the Panellists

Ilse Verdiesen
Research Fellow, National Defense Academy. Chief of Staff Joint IV Commando (Col)

Taylor Kate Woodcock
PhD Researcher in Public International Law, Asser Institute

Meet the Hosts

Nitin Sawhney
Visiting Researcher, University of the Arts Institue Helskini



Petter Ericson
Staff Scientist, Umeå University

Meet the Panellists

Rainer Rehak
Research Associate, Weizenbaum Institute

Andreas Schüller
Co-Director International Crimes and Accountability Program, European Center for Constitutional Human Rights

mondai
House of AI

TU Delft | AI Initiative

PRINCIPLES OF GLOBAL AI POLICY RESEARCH

Human and Planetary Welfare

Accountability and Transparency

Inclusivity, Diversity and Capacity Building

Ethical Research Practices and Institutional Responsibility

Ethical Governance

Equitable Economic Growth

Roadmap for AI Policy Research, *Global AI Policy Research Network*
Released on January 30, 2025

A VOICE FOR NATURE

The Whanganui River in New Zealand is a legal person. A nearby forest is too. Soon, the government will grant a mountain legal personhood as well. Here's how it happened, and what it may mean.

BY **KENNEDY WARNE**

PHOTOGRAPHS BY **MATHIAS SVOLD**

RELEVANT PUBLICATIONS

1. Truong, L., Lee, S., & Sawhney, N. 2024. **Enhancing Conversations for Migrant Counseling: Designing for Trustworthy AI-mediated Collaboration between Migrants and Service Advisors.** ACM Conference on Computer-Supported Cooperative Work and Social Computing, San José, Costa Rica, November 9-13, 2024.
2. Datta, P., Chizhikova, A., Vitiugin, F., & Sawhney, N. 2024. **Construction of Hyper-Relational Knowledge Graph Using Pre-Trained Large Language Models.** AAAI Conference on Artificial Intelligence, Workshop on Graphs and more Complex structures for Learning and Reasoning, Feb 20-27, 2024, Vancouver, Canada.
3. Gonzalez Torres, A. P. & Sawhney, N. 2023. **Role of AI Regulatory Sandboxes and MLOps for Finnish Public Sector Services.** *The Review of Socionetwork Strategies (RSS), Springer.*
4. Gonzalez Torres, A. P., Kajava, K., & Sawhney, N. 2023. **Emerging AI Discourses and Policies in the EU: Implications for Evolving AI Governance.** Communications in Computer and Information Science, Springer.
5. Kajava, K. & Sawhney, N. 2023. **Language of Algorithms: Agency, Metaphors, and Deliberations in AI Discourses.** *Lindgren, S. Handbook of Critical Studies of Artificial Intelligence. Edward Elgar Publishers.*
6. Truong, L. 2023. **Conversations with Service Advisors: The Role of Trust in Supporting Vulnerable Migrants.** Workshop on Participatory Design for Whom?, *ACM Conference on Conversational User Interaction (CUI)*, Eindhoven, Netherlands.
7. Drobotowicz, K., Truong, L., Gonzalez Torres, A. P., Ylipulli, J., & Sawhney, N. 2023. **Practitioners' Perspectives on Inclusion and Civic Empowerment in Finnish Public Sector AI.** *Proceedings of ACM Communities & Technologies Conference*, Lahti, Finland.
8. Sawhney, N. 2022. **Contestations in Urban Mobility: Rights, Risks & Responsibilities for Urban AI.** *Special Issue on Urban AI, AI & Society, The Journal of Culture, Knowledge and Communication. Springer.*