Responsible AI Practices in the Public Sector
Piloting AI Regulatory Sandboxes

Nitin Sawhney
Professor of Practice

Critical AI & Crisis Interrogatives
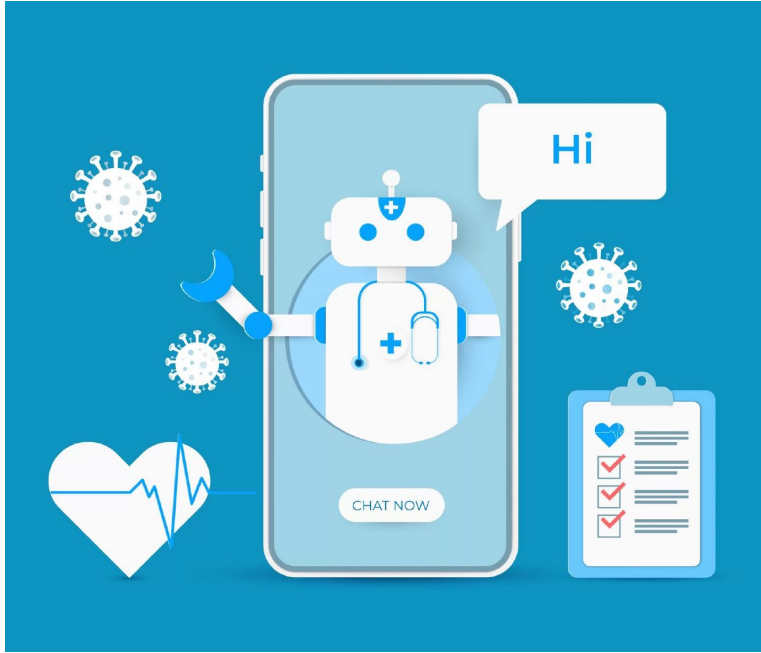*CRAI-CIS Research Group*

Department of
Computer Science

nitin.sawhney@aalto.fi
http://crai-cis.aalto.fi

September 1, 2023

A"
AaltoUniversity
School of Science

Can Algorithms be Biased, Fair or Racist?

Fostering Inclusive Trustworthy and Responsible AI in the Public Sector

Piloting AI Regulatory Sandboxes in Finland

Aalto University
School of Science

Coping with Humanitarian Crises

## Public Sector & Civil Society on the Frontlines!

Afghan refugees arrive at Dulles International Airport on Aug. 27, 2021, after being evacuated from Kabul following the Taliban takeover of Afghanistan.

Olivier Douliery/AFP via Getty Images

A"
Aalto University
School of Science

rĕst ŏf wŏrld
Reporting Global Tech Stories

AI translation is jeopardizing Afghan asylum claims

Cost-cutting translations are introducing errors and putting refugees at risk.
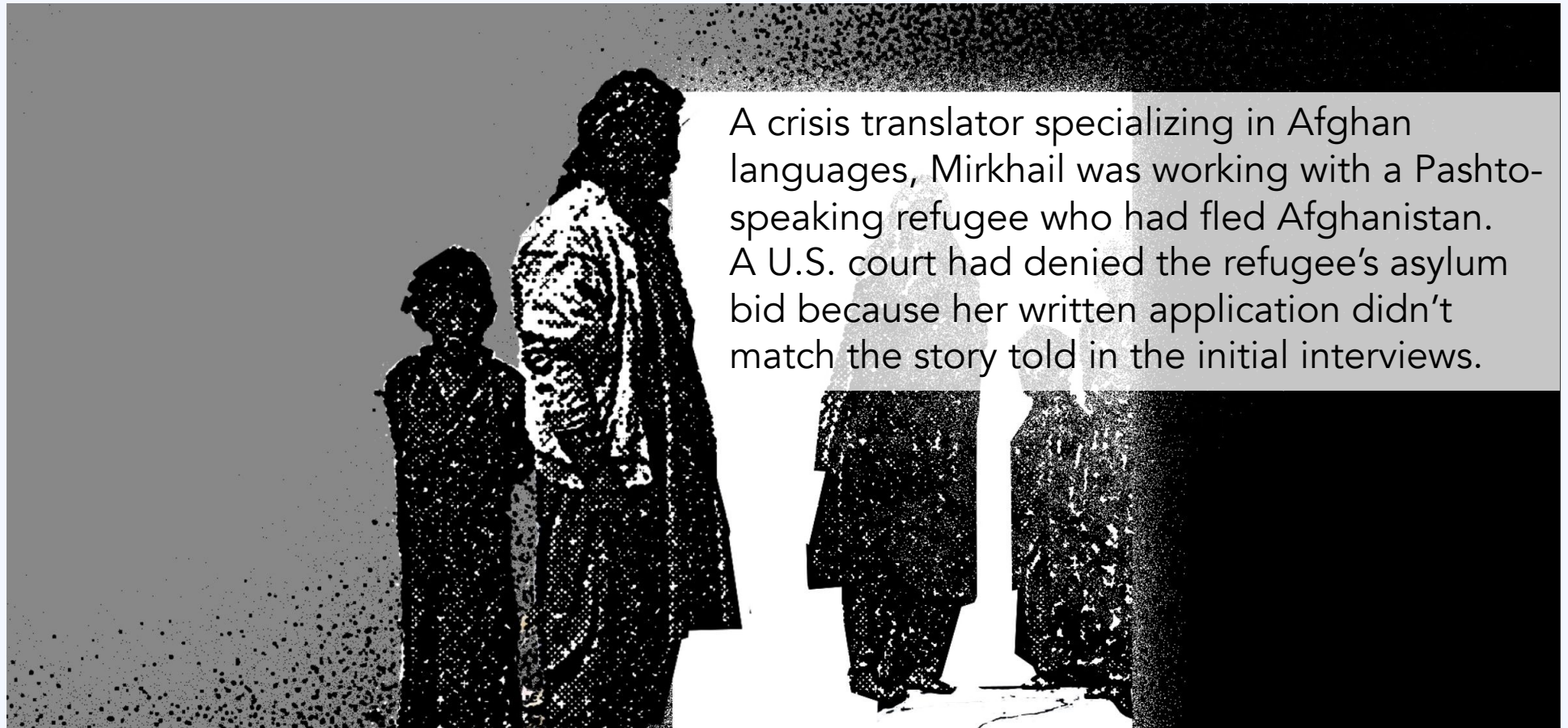
A crisis translator specializing in Afghan languages, Mirkhail was working with a Pashto-speaking refugee who had fled Afghanistan. A U.S. court had denied the refugee's asylum bid because her written application didn't match the story told in the initial interviews.

A"
Aalto University
School of Science

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!

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.

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

# "Racist Algorithms" would (should) never happen in Finland!

ASYLUM, INTEGRATION, AND THE POLITICS OF DIFFERENTIATION IN THE 2023 GOVERNMENT PROGRAMME

Erna Bodström, DSSc, Swedish School of Social Science University of Helsinki 25 Aug 2023

Photo credit: © Lauri Heikkinen, Prime Minister's Office

Aalto University School of Science

Dystopian Digital Futures: Automated Denial of Medical Services in the Future?
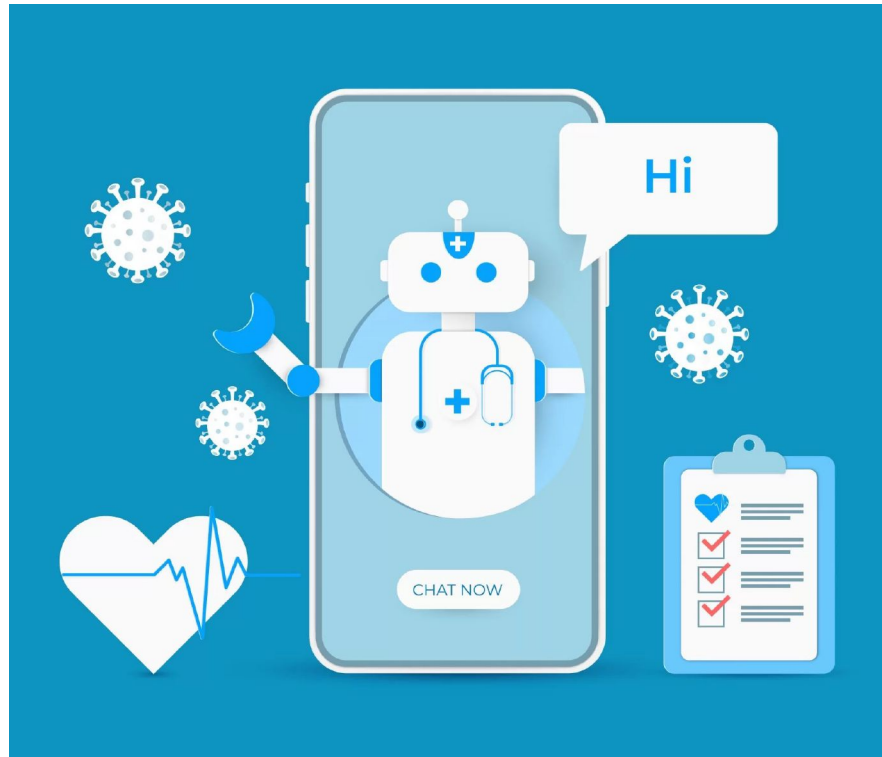
Rethinking "Algorithms" as a political & social-technical system (not just code)!

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

Way Forward? Fostering Inclusive Trustworthy and Responsible AI in the Public Sector

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)

A"
Aalto University
School of Science

Civic Agency in AI: Democratizing Algorithmic Services in the Public Sector

Supported by Kone Foundation & Research Council of Finland (2022 – 2027)

Ana Paula Gonzalez Torres, Antti Rannisto, Nitin Sawhney, Kaisla Kajava & Karolina Drobotowicz

Photo Credit: Matti Ahlgren / Aalto University

CAAI: Civic Agency in AI Project

The CAAI (Civic Agency in AI) project aims to understand citizens' algorithmic literacy, agency, and participation in the design and development of AI services in the Finnish public sector in order to advance more democratic and citizen-centric digital infrastructures. In the first stage of the project, we engage with a public sector to study AI-enabled services in plan, design or development using the following lenses:

Social Lens:

How (civic) values are incorporated and manifested in concrete practices of building AI-based services at Kela. How different values are recognized, explicated, deliberated, and negotiated within these practices.

Uses qualitative/ethnographic methods: participant observation, interviews, and analysis of related documents and objects; possibly also workshops and focus groups.

Interaction Lens:

How the interaction between citizens and AI-enabled services is being designed and developed, with a special focus on trust, transparency, empowerment and inclusion of the interface.

Uses deep interviews as the main data collection. Might also use ethnography and design workshop with public sector actors in the future.

Case Study of a Public AI-Enabled Service

Legal Lens:

What are the rights, risks and responsibilities of different actors in civil society (including industry and public sector) for algorithmic services in light of the "AI Act". How can we promote multi-stakeholder participation throughout the AI lifecycle to realize opportunities, mitigate risks, and ensure compliance?

Uses interviews, document analysis, and accountability theories as the approach to develop responsible AI governance frameworks.

Linguistic Lens:

How the AI-enabled service is communicated, described, conceptualized, and documented within the organization, to citizens, and to other stakeholders involved in its development.

Uses textual documents as the data: public or internal texts related to the planning and development of the service, e.g. development plans, service descriptions, user stories, ethical assessments, technical documentation, leaflets, and announcements.

USE OF AI-BASED SYSTEMS IN THE PUBLIC SECTOR

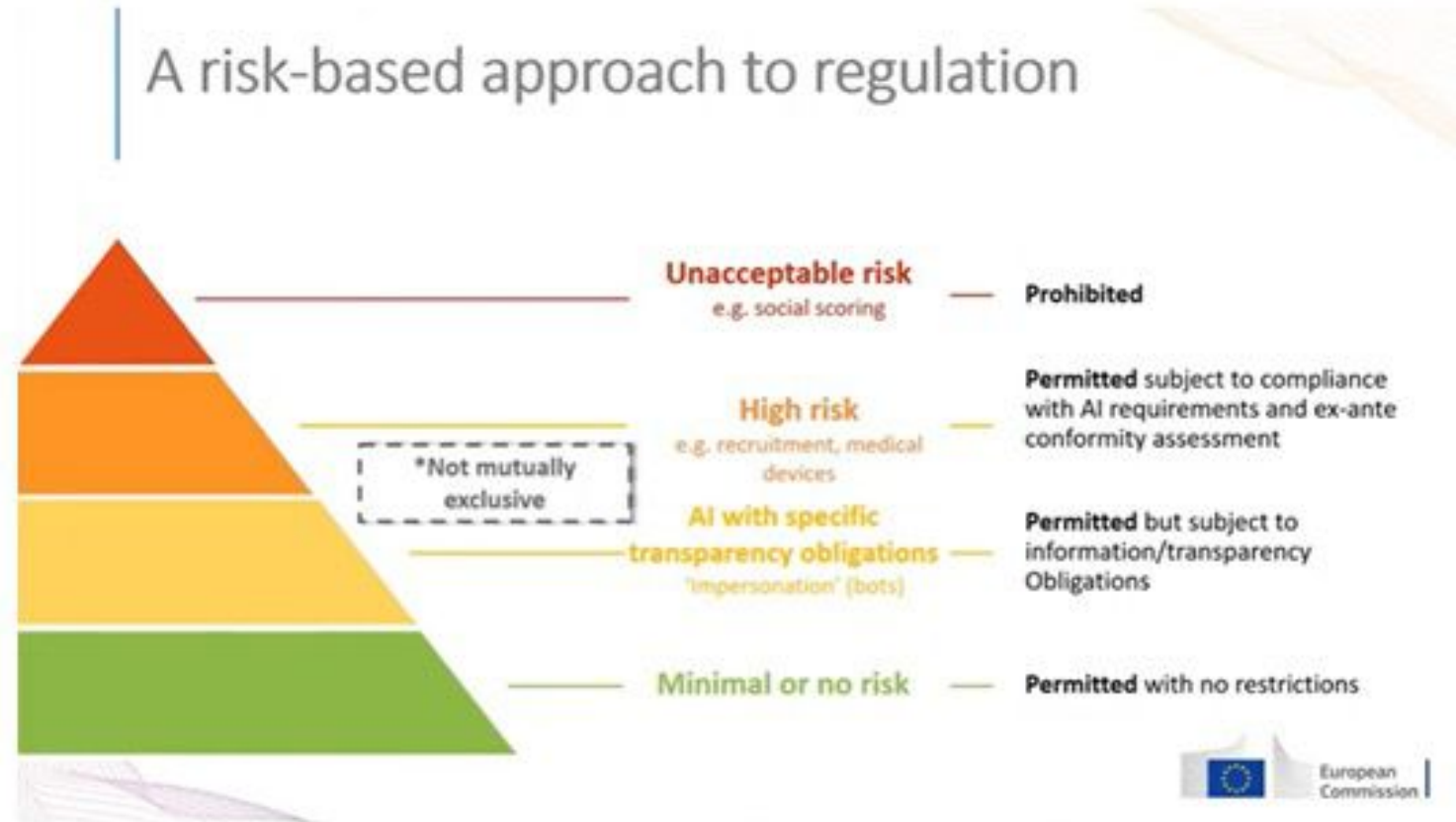| **Opportunities** (Barker *et al.,* 2021; Manzoni *et al.,* 2022)	**Challenges** (Pechtor & Basl, 2022; Pūraitė *et al.,* 2020)
- Support context-specific public values: - Operational - Political - Social - Foster citizen trust & participation - Improve efficiency & decision making - Provide innovative 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.

A"
Aalto University
School of Science

AI Act proposed by European Commission

A risk-based approach to regulation

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

*Not mutually exclusive

AI with specific transparency obligations
'Impersonation' (bots)
— **Permitted** but subject to information/transparency Obligations

Minimal or no risk
— **Permitted** with no restrictions

European Commission

"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

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

By **Laura Lamberti**

Laura Lamberti is a junior reporter at The Parliament Magazine

25 Jan 2023

🐦 @LauraLamberti10

READ NEXT:

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

T he 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.

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 Finnish 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.

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*).

Culture and leisure

Intelligent material management system

IMMS (Intelligent Material Management System) is an intelligent material management system for the entire library collection. The City Library's collection contains approximately 1.8 million items. An intelligent material management system was acquired while the city library moved away...

> Read more

Social services and health care

Health center chatbot

The chatbot provides health and illness-related advice easily without queuing. Chatbot directs the users to the right digital health services and advises on questions related to dental, mental health, substance abuse and social services. The service answers the most frequently asked...
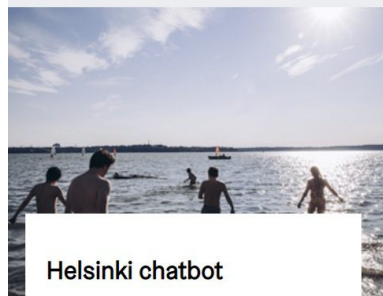
> Read more

Urban Environment

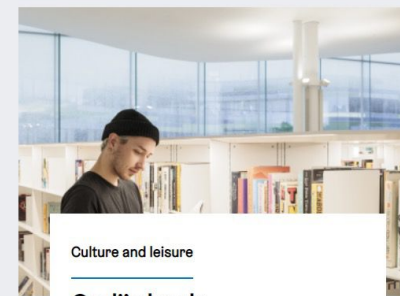**The rental apartment search chatbot**

The rental apartment search chatbot is a 24-hour customer service channel of the City of Helsinki housing services aimed at improving the accessibility of customer service and the customer experience as well as increasing the interactivity of the self-service. The service provides...
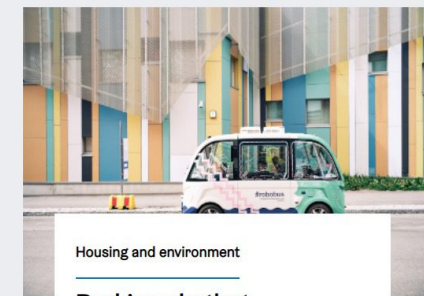
> Read more

Helsinki chatbot

Helsinki chatbot is a 24-hour customer service channel of the Helsinki City Information aimed at improving the accessibility of customer service and the customer experience and increasing the interactivity of the self-

Culture and leisure

Oodi's book recommendation service...

Obotti is Central Library Oodi's recommendation chatbot. The service recommends books from Oodi's

Housing and environment

Parking chatbot

The parking chatbot is a customer service channel of city's parking services. Service provides automated answers to the parking-related

City of Helsinki AI Register – setup by Saidot.ai

AI REGULATORY SANDBOXES

EXPERIMENTING WITH AI INNOVATIONS IN THE PUBLIC SECTOR

WHY AI REGULATORY SANDBOXES?

- In **high-risk domains** (e.g., financial sector), regulatory sandboxes used to explore possibilities and implications of algorithmic systems before wider deployment (Manzoni *et al.,* 2022).

- Allow for **experimentation** and critical exploration of both technical and regulatory implications of AI systems with diverse stakeholders.

- The proposed **AI Act** in title V, 'measures in support of innovation', establishes **sandboxes**.

  '[A] controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan.' Article 53(1).

- 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.
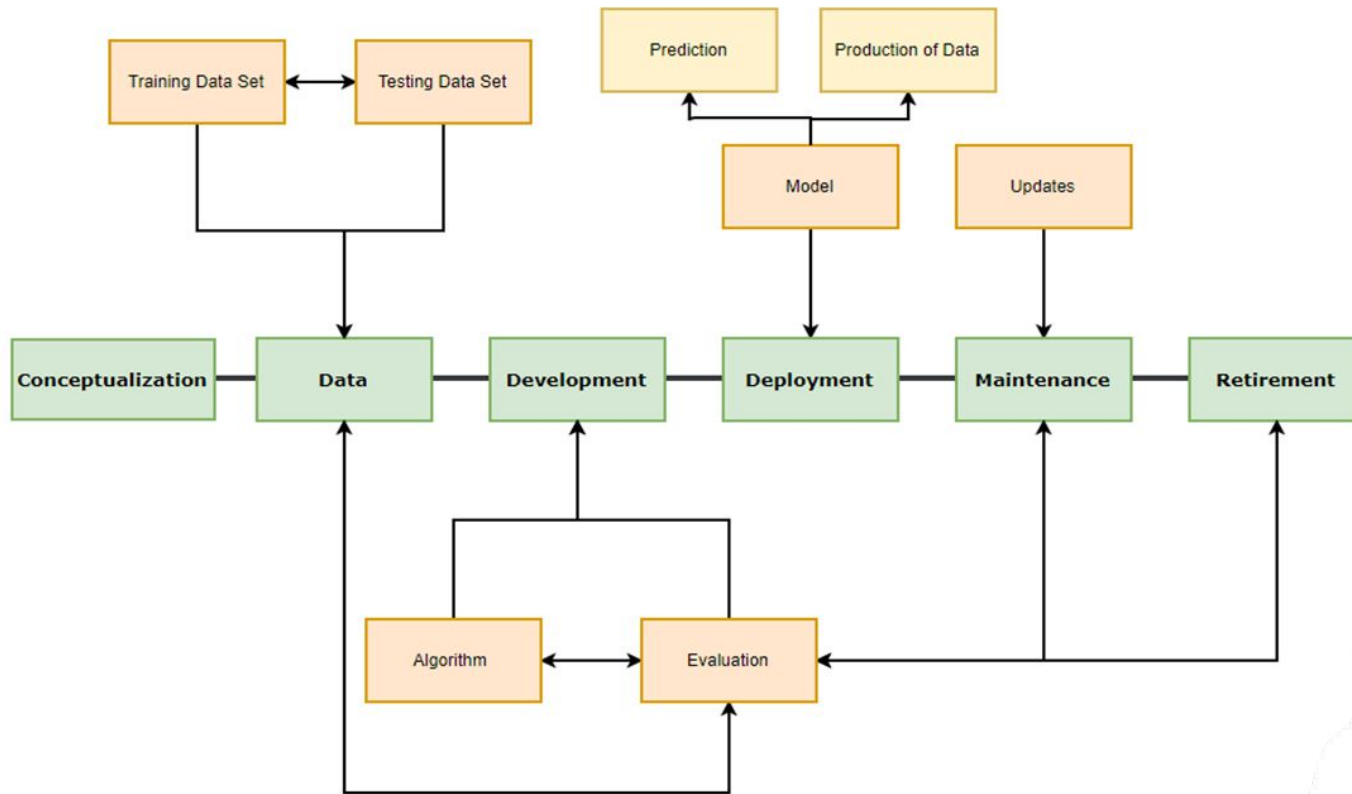
AI LIFECYCLE APPROACH

- 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


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?

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
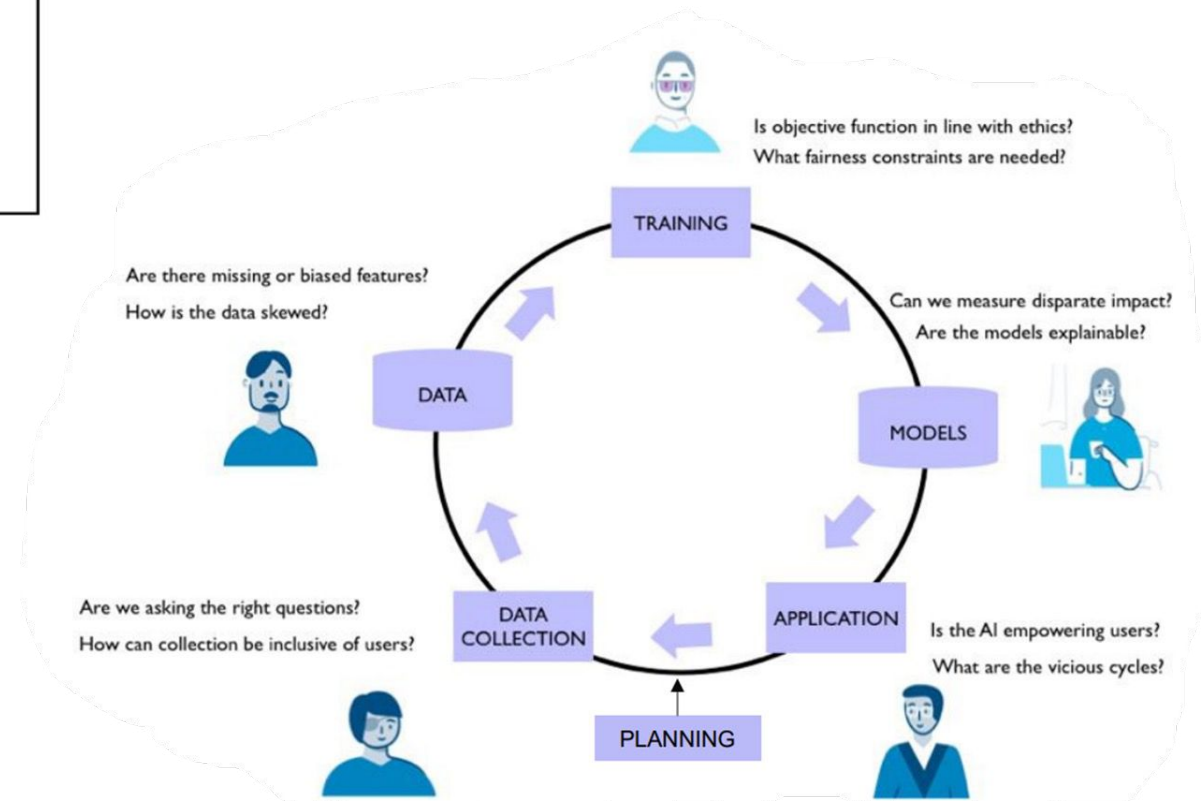
INTEGRATIVE FRAMEWORKS: FROM MLOPS TO REGOPS

- **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 & Kuranananda, 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
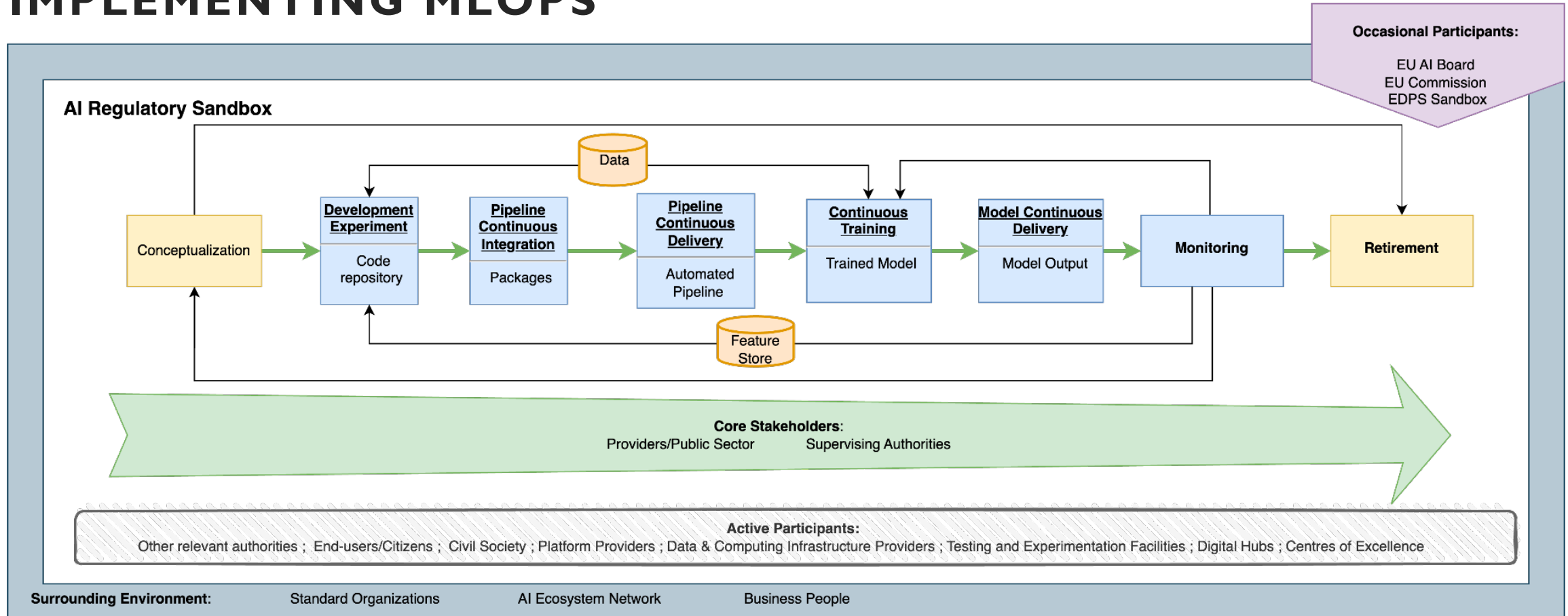
RELEVANT PUBLICATIONS

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

2. Kajava, K. & Sawhney, N. 2023. **Language of Algorithms: Agency, Metaphors, and Deliberations in AI Discourses.** Forthcoming in *Lindgren, S. Handbook of Critical Studies of Artificial Intelligence. Edward Elgar Publishers.*

3. 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.

4. 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.

5. Drobotowicz, K., Sekar, B., & Truong, L. 2023. **Engaging civil society in designing public sector AI: What participatory methods can we use?** Workshop on Designing the City in *Communities & Technologies Conference.*

6. Varanasi, U., Šerpytytė, R., & Sawhney, N. 2023. **Re-evaluating Evaluation: Looking for Value-based Metrics in Public Service Design.** Workshop on Designing the City in *Communities & Technologies Conference.*

7. Sawhney, N. & Gonzalez Torres, A. P., 2022. **Devising Regulatory Sandboxes and Responsible Practices for Designing AI-based Services in the Finnish Public Sector.** *WAICOM Workshop at the International Conference on Legal Knowledge and Information Systems (JURIX 2022),* Saarland University, Saarbrücken, Germany.

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.*

Aalto University
School of Science

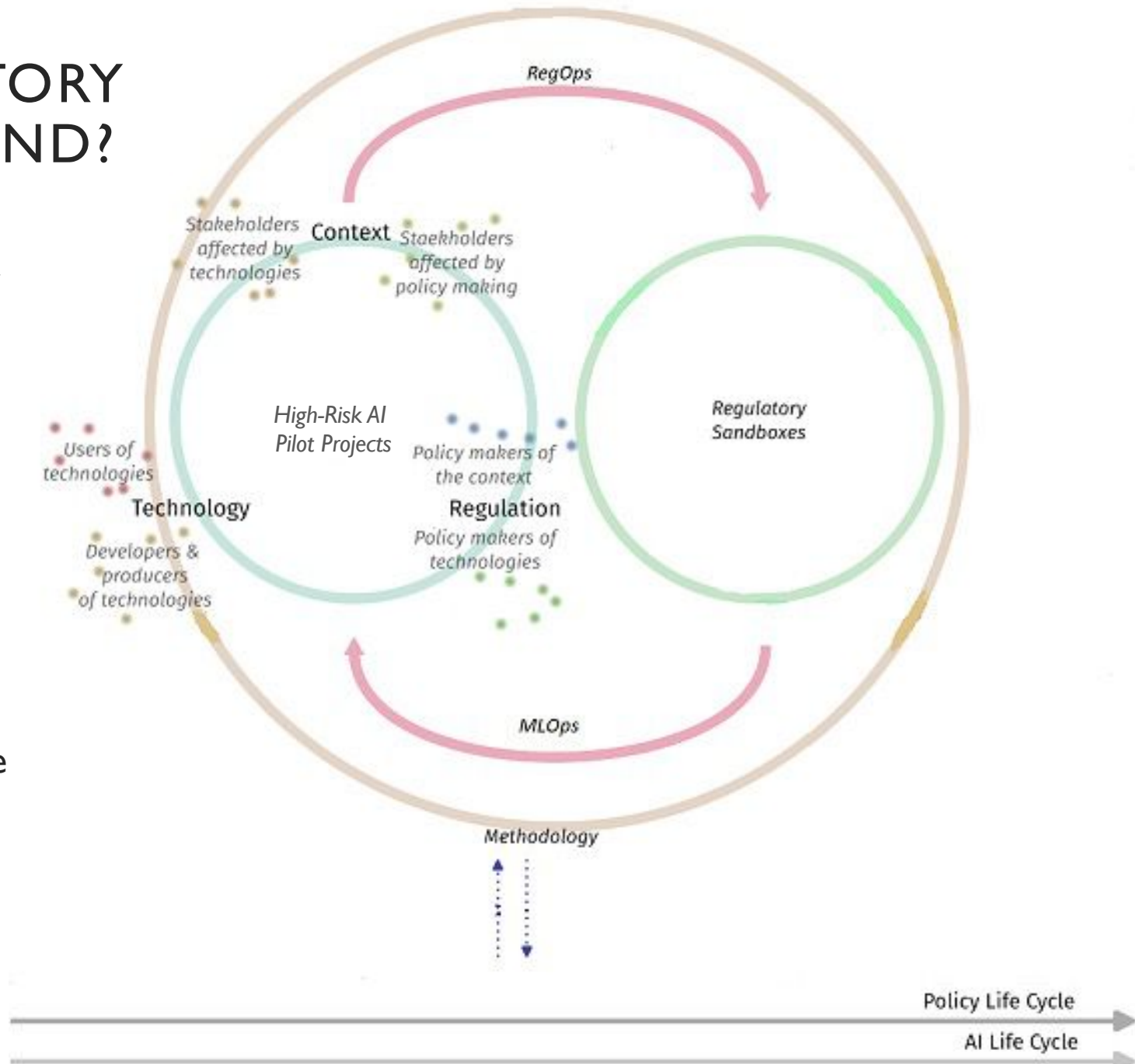# PILOTING AI REGULATORY SANDBOXES IN FINLAND?

Conduct **pilots with Finnish Public Sector** organizations using AI Regulatory Sandboxes and integrative frameworks
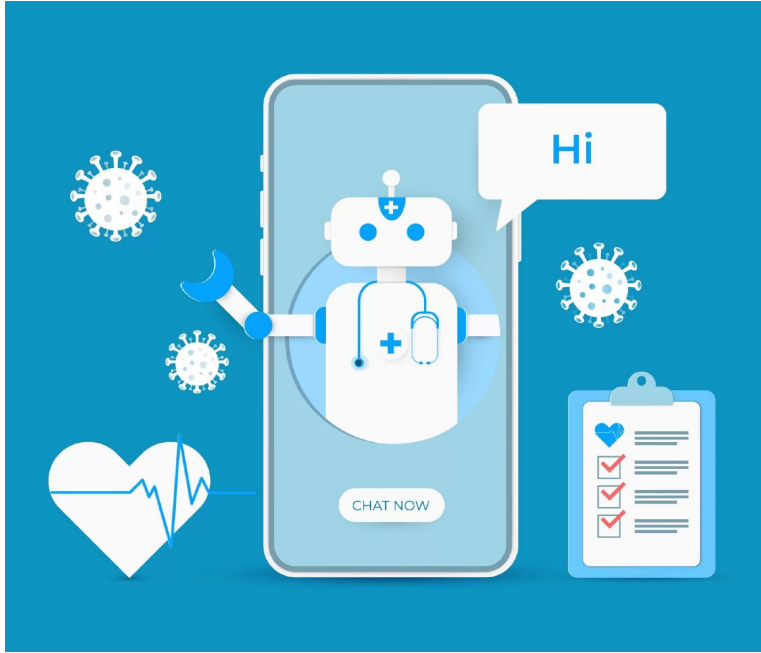
Engage with multiple stakeholders:

- Regulators, public administration, infrastructure providers, developers, auditing/compliance facilitators

Expected outcomes:

- Provide the space for interactions and mutual collaboration though AI lifecycle

- Examine the limits and possibilities of legislation and technological innovation for public sector AI

Can Algorithms be Biased, Fair or Racist?

Fostering Inclusive Trustworthy and Responsible AI in the Public Sector

Piloting AI Regulatory Sandboxes in Finland

Aalto University
School of Science

Compassionate Algorithms?

Fostering Inclusive Trustworthy and Responsible AI in the Public Sector

A"
AaltoUniversity
School of Science

Nitin Sawhney
Professor of Practice
nitin.sawhney@aalto.fi
http://crai-cis.aalto.fi

Photo Credit: Jussi Nukari / IS