

Transdisciplinary approaches towards Inclusive, Trustworthy & Responsible AI

Nitin Sawhney

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

*Department of
Computer Science*

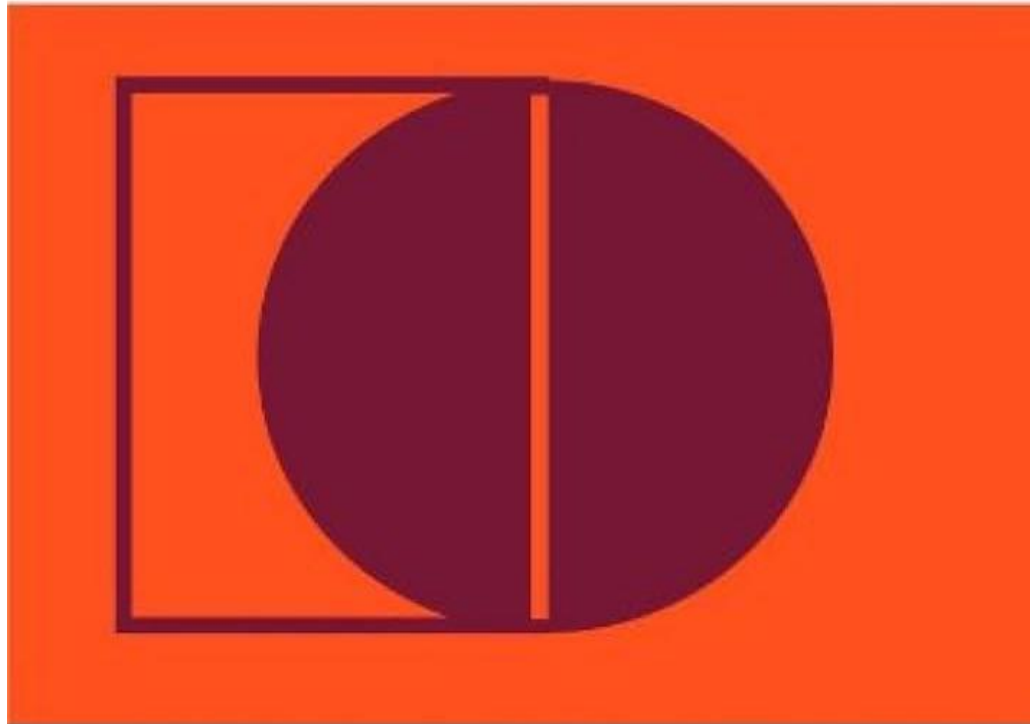


Aalto University
School of Science

Pioneer Center for AI
University of Copenhagen
September 29, 2023

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

MANIFESTO of TRANSDISCIPLINARITY



BASARAB NICOLESCU
TRANSLATED by KAREN-CLAIRE VOSS

Published in 2002



“Today there are hundreds of disciplines. **How can a theoretical particle physicist truly dialogue with a neurophysiologist**, a mathematician with a poet, a biologist with an economist, a politician with a computer programmer, beyond mouthing more or less trivial generalities?

Yet, **a true decision-maker must be able to dialogue with all of them** at once. Disciplinary language is an apparently insurmountable barrier for a neophyte, and each of us is a neophyte in some area. Is a modern tower of Babel inevitable?”

- Basarab Nicolescu, Manifesto of Transdisciplinarity (2002)

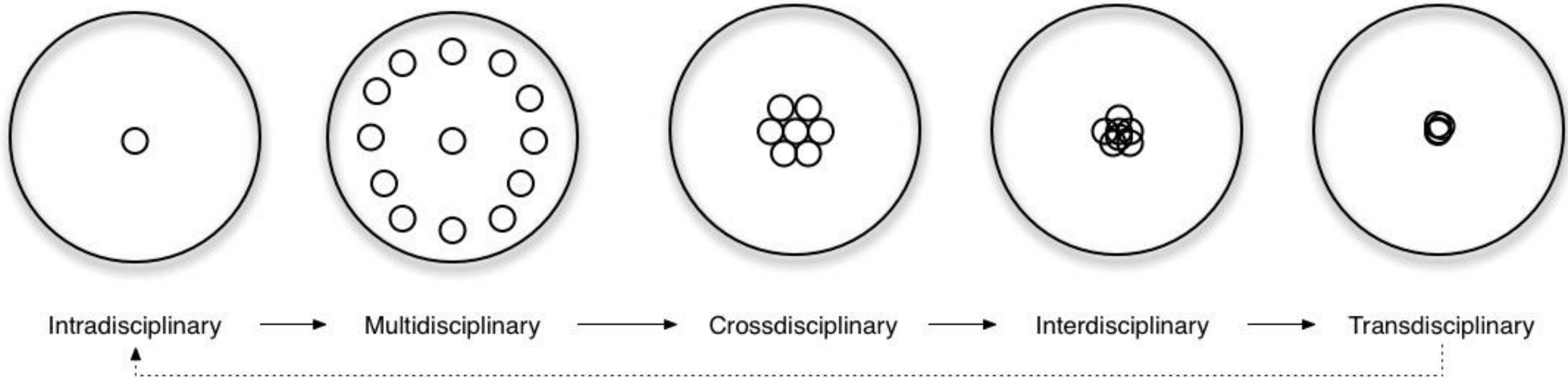


Giovanni Pico della Mirandola
Renaissance humanism
(1463 – 1494)

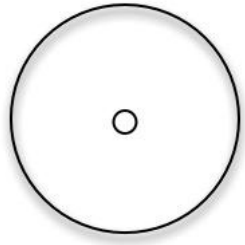
“Perhaps a Pico della Mirandola in our time could be conceivable if he took the form of a **supercomputer** into which one could load all the known data which has been generated by all existing disciplines. This supercomputer would be **capable of knowing everything while understanding nothing**. Its user would be no better off than the supercomputer itself. The user would have immediate access to any results from any discipline, but would be incapable of understanding their meanings, still less of making connections between the results of different disciplines.”

- Basarab Nicolescu, Manifesto of Transdisciplinarity (2002)

Pathways to Transdisciplinarity

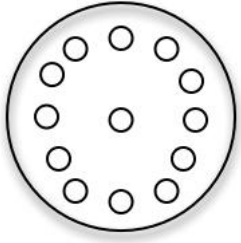


Jensenius, A.R. (2012). Disciplinaryities: intra, cross, multi, inter, trans”, Available at: www.arj.no/.2012/03/12/disciplinaryities-2



INTRADISCIPLINARY — WITHIN

Latin *intra*: “within, inside, on the inside” — **working within the frame of a single, recognized discipline.**

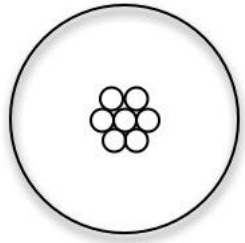


MULTIDISCIPLINARY — ADD, MULTIPLY

Latin *multus*: “much, many” — looking at one problem by adding **multiple perspectives and disciplines** to the mix.

In this process, a root discipline may involve other disciplines to solve a problem. Participants **exchange knowledge and compare results**, but stop short of integrating them.

The **disciplines maintain their distinctiveness** and the results remain grounded in the framework of the root discipline.

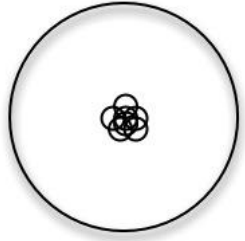


CROSSDISCIPLINARY — INTERSECT, TRANSFER

Latin *crux*: “intersecting, lying athwart each other” — **intersecting knowledge from two or more disciplines**, viewing one discipline from another perspective.

Researchers collaborate with the goal of **transferring knowledge from one discipline to another**.

Understanding the complex dynamics of environmental problems in a socioecological context is a typical cross-disciplinary approach.



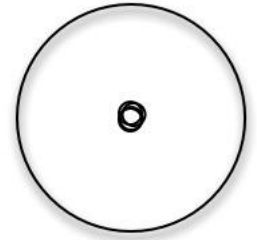
INTERDISCIPLINARY — INTEGRATE, SYNTHESE

Latin *inter*: “among, between, betwixt, in the midst” — **integrating knowledge and methods** from different disciplines using a **synthesis of approaches**.

Interdisciplinary concerns **collaborations between contrasting academic disciplines** or research methods for new applications, new analyses, or the creation of entirely new disciplines.

For example, interdisciplinary research on information systems and biomedical research has given rise to the field of bioinformatics.

TRANSDISCIPLINARY – TRANSCEND, WHOLE



Latin *trans*: “across, over, beyond” – emergence of a new discipline **transcending** the boundaries of disciplinary perspective.

Transdisciplinarity combines interdisciplinarity with a **participatory** approach.

The research paradigms involve **non-academic participants** as (equal) participants in the process for a common goal – towards a **transformational condition in society** (*not necessarily a solution*).

Culmination of interdisciplinary efforts, relating all disciplines into a **coherent whole**.

PEACEBUILDING AND THE PROMISE OF TRANSDISCIPLINARITY

Richard Lappin

Ph.D. Candidate
Centre for Peace Research and Strategic Studies
Catholic University of Leuven
Belgium

Richard Lappin is a Ph.D. candidate at the Centre for Peace Research and Strategic Studies at the Catholic University of Leuven in Belgium under the supervision of Prof. Dr. Luc Reyckler. Richard has participated in several post-conflict peacebuilding missions with a variety of organizations.

A transdisciplinary approach to peacebuilding is compelling. Undeniably there is a semantic appeal to its claims of rising above disciplinary limitations and forging new ways of thinking. But what do we really understand by this concept? What are its possibilities and limitations? And, moreover, how can we harness its promise of creative and sustainable solutions in the service of peacebuilding?

THE CHALLENGE OF PEACE-BUILDING

Peacebuilding is a wide-ranging process which depends upon several interdependent approaches. From the establishment of peace agreements and new institutions of government to the promotion of economic reconstruction and social reconciliation, peacebuilding is a highly complex, and often daunting, task. It is also a long-term and multi-faceted process and, if a lasting and sustainable peace is to emerge, peacebuilding has to address not only the symptoms of a conflict—manifest, physical violence—but also the broader, underlying causes of the conflict. As such, peacebuilding not only involves addressing multiple problems, but also multiple interpretations of the origins of these problems and the reasons why

Theory, Culture & Society
Volume 32, Issue 5-6, September-November 2015, Pages 159-182
© The Author(s) 2015, Article Reuse Guidelines
<https://doi.org/10.1177/0263276415594238>



Article - II. Transdisciplines



Contradiction of Terms: Feminist Theory, Philosophy and Transdisciplinarity

Stella Sandford

Abstract

This article addresses the question of the relation between disciplines and transdisciplinary practices and concepts through a discussion of the relationship between philosophy and the emblematically transdisciplinary practice of feminist theory, via a discussion of interdisciplinarity and related terms in gender studies. It argues that the tendency of philosophy to reject feminist theory, as alien to it as a discipline, is in a sense correct, to the extent that the two defining features of feminist theory – its constitutive tie to a political agenda for social change and the transdisciplinary character of many of its central concepts – are indeed at odds with, and pose a threat to, the traditional insularity of the discipline of philosophy. If feminist philosophy incorporates feminist theory, its transdisciplinary aspects thus open it up to an unavoidable contradiction. Nonetheless, I will argue, this is a contradiction that can and must be endured and made productive.

Keywords

critical theory, feminism, gender, philosophy, transdisciplinarity

Kingston University London

Corresponding author(s):

Stella Sandford. Email: S.Sandford@kingston.ac.uk

What happens when well-defined disciplines meet or are confronted with transdisciplinary discourses

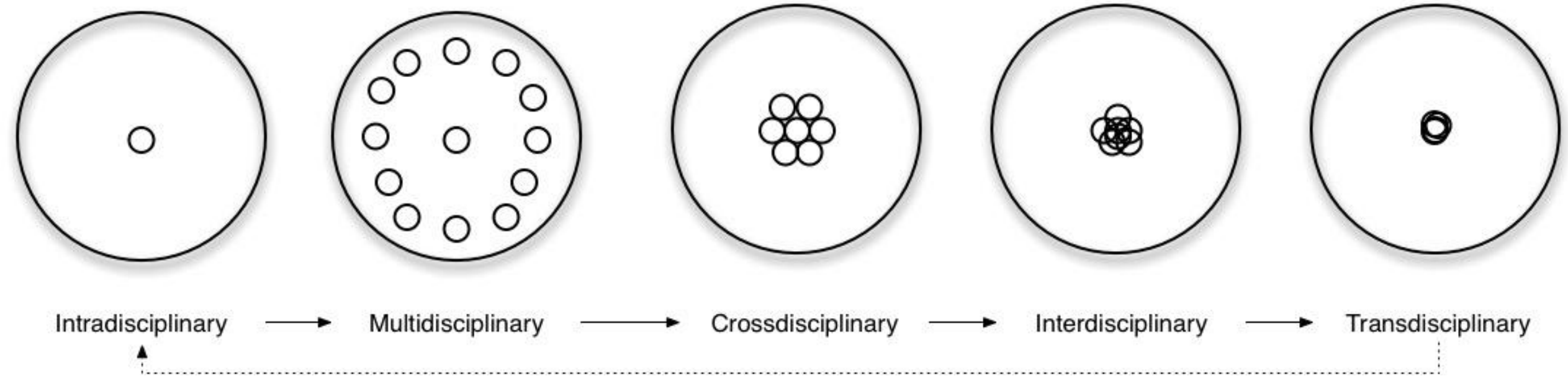
“transdisciplinarity concerns that which is at once between the disciplines, across the different disciplines, and beyond all discipline. Its goal is the understanding of the present world , of which one of the imperatives is the unity of knowledge.” ...

“Disciplinarity, multidisciplinary, interdisciplinarity and transdisciplinarity are like **four arrows shot from but a single bow: knowledge .**

As in the case of disciplinarity, **transdisciplinary research is not antagonistic but complementary** to multidisciplinary and interdisciplinarity research.”

- Basarab Nicolescu, Manifesto of Transdisciplinarity (2002)

Developing a Transdisciplinary Orientation or Practice?



Jensenius, A.R. (2012). "Disciplinarity: intra, cross, multi, inter, trans", Available at: www.arj.no/.2012/03/12/disciplinarity-2

*Do we need transdisciplinary approaches towards
Inclusive, Trustworthy & Responsible AI?*

*How can researchers across disciplines from
machine learning, human-centered design,
linguistics, law, and sociology collaborate at the
intersections of AI and society?*

FCAI

Finnish Center for Artificial Intelligence

Search



[PHD AND POSTDOC POSITIONS: APPLY NOW](#)

[REGISTER NOW FOR AI DAY →](#)



[: GENERATIVE AI, CREATIVITY AND FOUNDATION MODELS →](#)

Real AI made by real people: Meet the humans of FCAI



For Kevin Luck, a postdoc at FCAI was an important step to a faculty position.

[Read More →](#)



Patrick Rinke's pioneering expertise in finding sustainable and climate-friendly materials with machine learning methodology has arguably never been more in demand.

[Read More →](#)



The impact of creative AI is unfolding before our eyes, yet we struggle to understand it. It's the perfect time to ask researchers what they see and think.

[Read More →](#)



FCAI postdoc Ulpu Remes develops ELFI, the open-source software for likelihood-free inference.

[Read More →](#)



Aalto University visiting professor and member of ELLIS Unit Helsinki, Guoying Zhao, on emotion AI, face analysis and visual intelligence

[Read More →](#)



FCAI professor Vikas Garg's research applies deep learning to the fields of quantum computing and drug discovery

[Read More →](#)

Finnish Center for Artificial Intelligence (FCAI)

Fundamental AI Research

Research Programs

Agile probabilistic
Prof. Aki Vehtari

Simulator-based
Prof. Jukka Corander

Deep learning
Prof. Arno Solin

Privacy & security
Prof. Antti Honkela

Interactive AI
Prof. Antti Oulasvirta

Autonomous AI
Prof. Ville Kyrki

AI in society
Prof. Petri Ylikoski

Joint Methodological Goal (FCAI Teams)

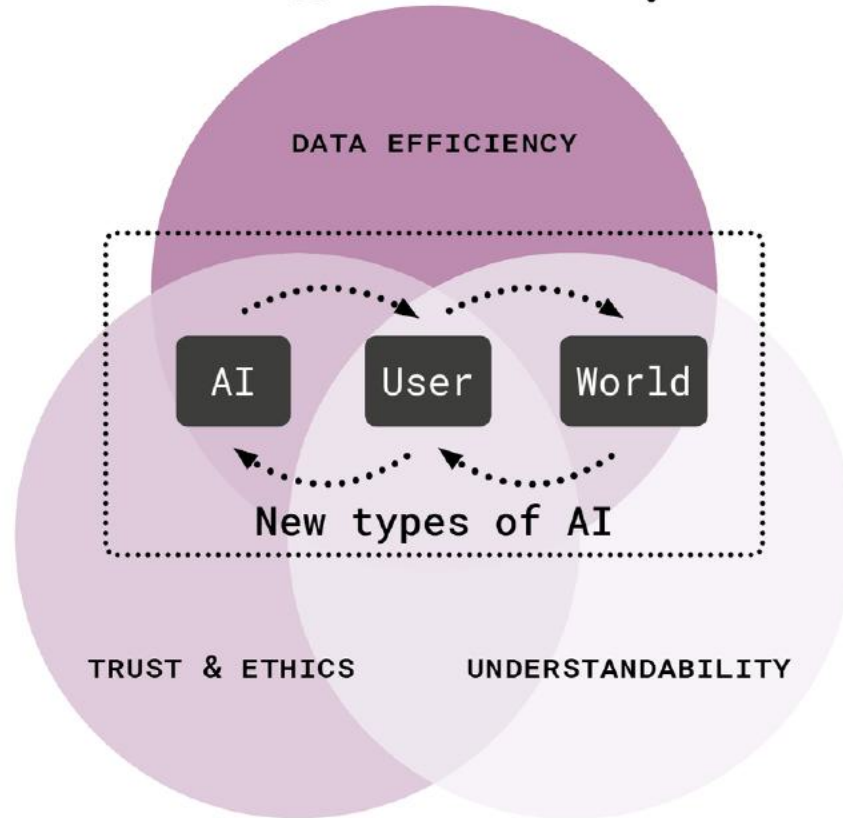
AI-assisted decision-making, design and modeling

Prof. Samuel Kaski

Joint Objectives

Advances in methodology

Transformative AI tools



AI Across Fields

SIGs: communities with shared interest

Health

Computer vision

Modeling tools
Prof. Arto Klami

AI for public health
Prof. Pekka Marttinen

Atmospheric AI
Prof. Kai Puolamäki

Highlight Programs

AI and materials
Prof. Patrick Rinke

AI for sustainability
Prof. Laura Ruotsalainen

+ *new Highlights*

Edge AI

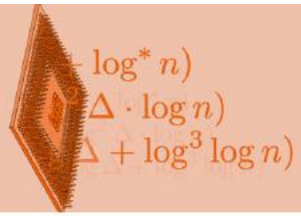
Language

Neuro-science



Department of Computer Science





Algorithms and Theoretical Computer Science

Fundamental methods and mathematics of computation.

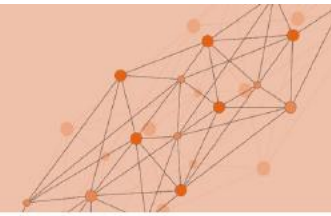
Department of Computer Science



Artificial Intelligence and Machine Learning

Fundamentals and practical impact of AI

Department of Computer Science



Complex Systems

Network science, stochastic processes and nonlinear dynamics.

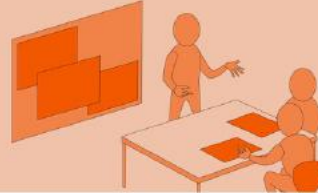
Department of Computer Science



Computational Life Sciences

Modelling, analysis and design of biological systems.

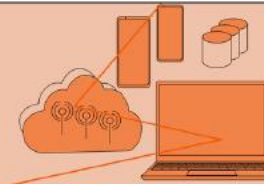
Department of Computer Science



Computing education research and educational technology

Psychology and education.

Department of Computer Science



Computing Systems

The study, design and development of modern computing systems.

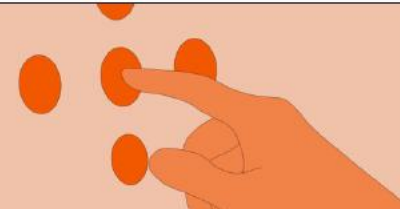
Department of Computer Science



Digital Ethics, Society and Policy



Engineering Psychology



Human-Computer Interaction and Design

Digital Ethics, Society and Policy

Engaging the societal impact of technologies through transdisciplinary inquiry and ethical practices



Digital Ethics, Society and Policy (Digital-ESP) brings together diverse researchers, scholars and practitioners engaging at the intersection of technology, social science and policy concerns in a societal context. From a Computer Science perspective this is a crucial area for critically interrogating the ethics and values embedded in technology, the interplay with society, and the role of policy-making to better understand and mediate positive or adverse impacts.

Increasingly, contemporary concerns with emerging technologies such as AI, surveillance, cybersecurity, robotics, drones and autonomous systems in society have been much debated. However, issues of gender and racial equity, Global South inclusion, political participation, democracy, protest, civic engagement, crisis, and sustainability in technological contexts, are also critical topics of emphasis for this area.

Digital-ESP is focused on understanding the societal impact of technologies in a transdisciplinary manner (i.e. drawing on research conducted at the intersection of different fields and engaging diverse stakeholders), while proposing critical guidance and policy interventions.



'For many people, it's always been a crisis' – a new Aalto collective engages intersecting dimensions of crises

The transdisciplinary Crisis Interrogatives collective

7.4.2021 | News



How to achieve digital equality in smart cities

Who we design digital devices and apps for can unintentionally exclude some users, a process referred to as 'digital inequality'

28.11.2019 | News



Will AI make us better humans? Why we must be cautious

It's not news that digitalisation

27.4.2021 | News



'We managed to turn a major risk into a great opportunity'

To make sure the Fusion Grid

1.4.2020 | News



A 'lifelong interest in coding' is not a requirement for seeking a career in technology

29.10.2019 | News

A
Aalto University
School of Science



Critical AI and Crisis Interrogatives

The CRITICAL AI and Crisis Interrogatives (CRAI-CIS) research group at Aalto University explores the impact of technology in critical societal contexts, in particular for ethical AI, civic agency and crisis narratives, working at the intersection of computational and social sciences engaging HCI and participatory design.

crai-cis.aalto.fi

Projects



Crisis Narratives

Crisis Narratives is a multidisciplinary joint project between Aalto University and the Department of Health and Welfare (THL), which examines the construction of stories about the crisis on various communication platforms and in public debate.



Civic Agency in AI

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.



Trust-M

The Trust-M project aims to create trustworthy digital public services for improved integration of migrants in Finland, hence strengthening Finnish society through increased inclusion, resilience of the labor market, and economic vibrance.



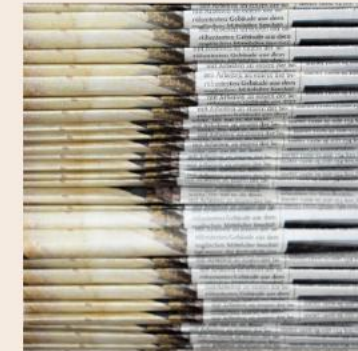
Algorithmic Literacy

The project explores how new forms of digital pedagogy and participatory engagement with algorithmic and data-centric concepts and technologies can promote critical, playful, and inclusive digital citizenship and algorithmic literacy among young learners.



AI x Music and Creativity

A joint project between Aalto and the Google Brain team which looks at developing AI models for digitally facilitating co-creation processes between human musicians and creative AI agents.



Countering AI-infused Disinformation

A joint research project between Aalto University and University of Helsinki aims to examine the increasing emergence of AI-infused disinformation and the challenges faced by news media practitioners and fact-checking organizations.

People



Nitin Sawhney
Professor of Practice



Minttu Tikka
Postdoctoral Researcher



Viivi Eskelinen
Postdoctoral researcher



Henna Paakki
Doctoral Researcher



Kaisla Kajava
Doctoral Researcher



Karolina Drobotowicz
Doctoral Researcher



Antti Rannisto
Doctoral Researcher



Ana Paula Gonzalez Torres
Doctoral Researcher



Uttishta Varanasi
Doctoral Researcher



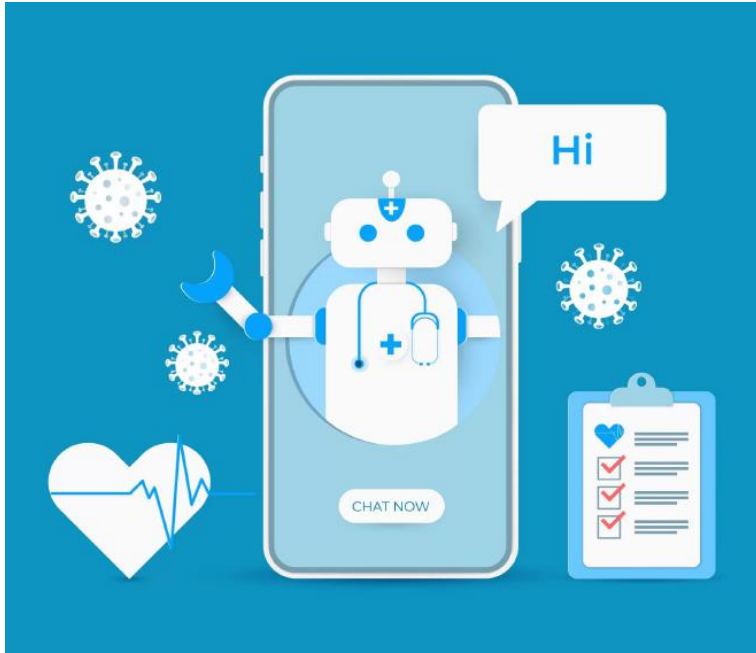
Rūta Šerpytė
Design Researcher



Sophie Truong
Research Assistant



Fatima Sounny
Research Assistant



Can Algorithms
be Biased, Fair
or Racist?



Fostering Inclusive
Trustworthy and
Responsible AI in
the Public Sector



Piloting AI
Regulatory
Sandboxes
in Finland

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

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

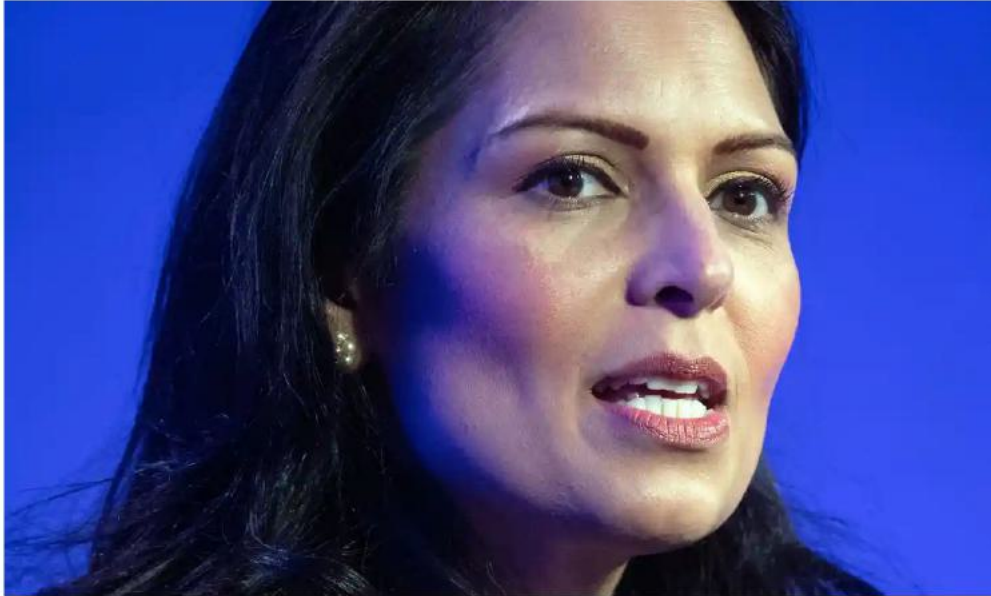


“Machine-learning translations are not yet in a place to be trusted completely without human review.”

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

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

“Racist Algorithms” would (should) never happen in Finland!

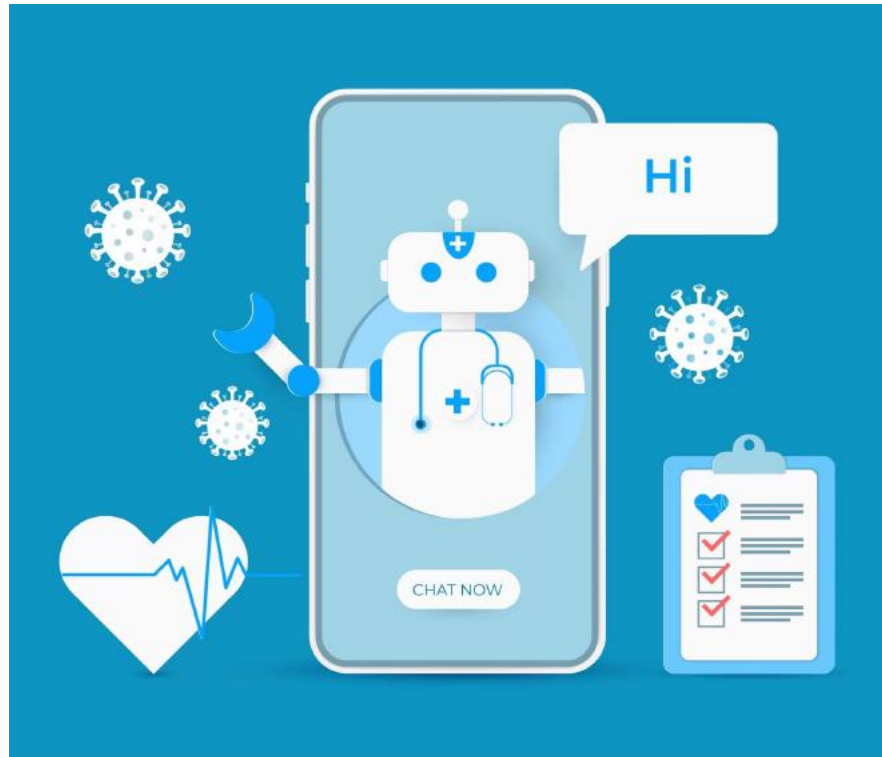


**ASYLUM, INTEGRATION, AND
THE POLITICS OF DIFFERENTIATION**
IN THE 2023 GOVERNMENT PROGRAMME

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

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

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



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

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



Creating trustworthy and accessible digital public services for migrants



- 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.



TRUST
TECH



DESIGN



POLICY



IMPACT

Track #1

Project management and coordination



Track #2

Reconceptualizing
trust in public
services



Track #3

Legal, ethical,
policy
considerations



Track #4

Participatory
socio-technical
design of services



Track #5

Developing
conversational
systems

Track #6

Interaction, piloting, impact assessment



Participatory Research & Collaborative Design



21.7.2023 / Collaborations

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

By Rūta Šerpytė – Design researcher at Trust M

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

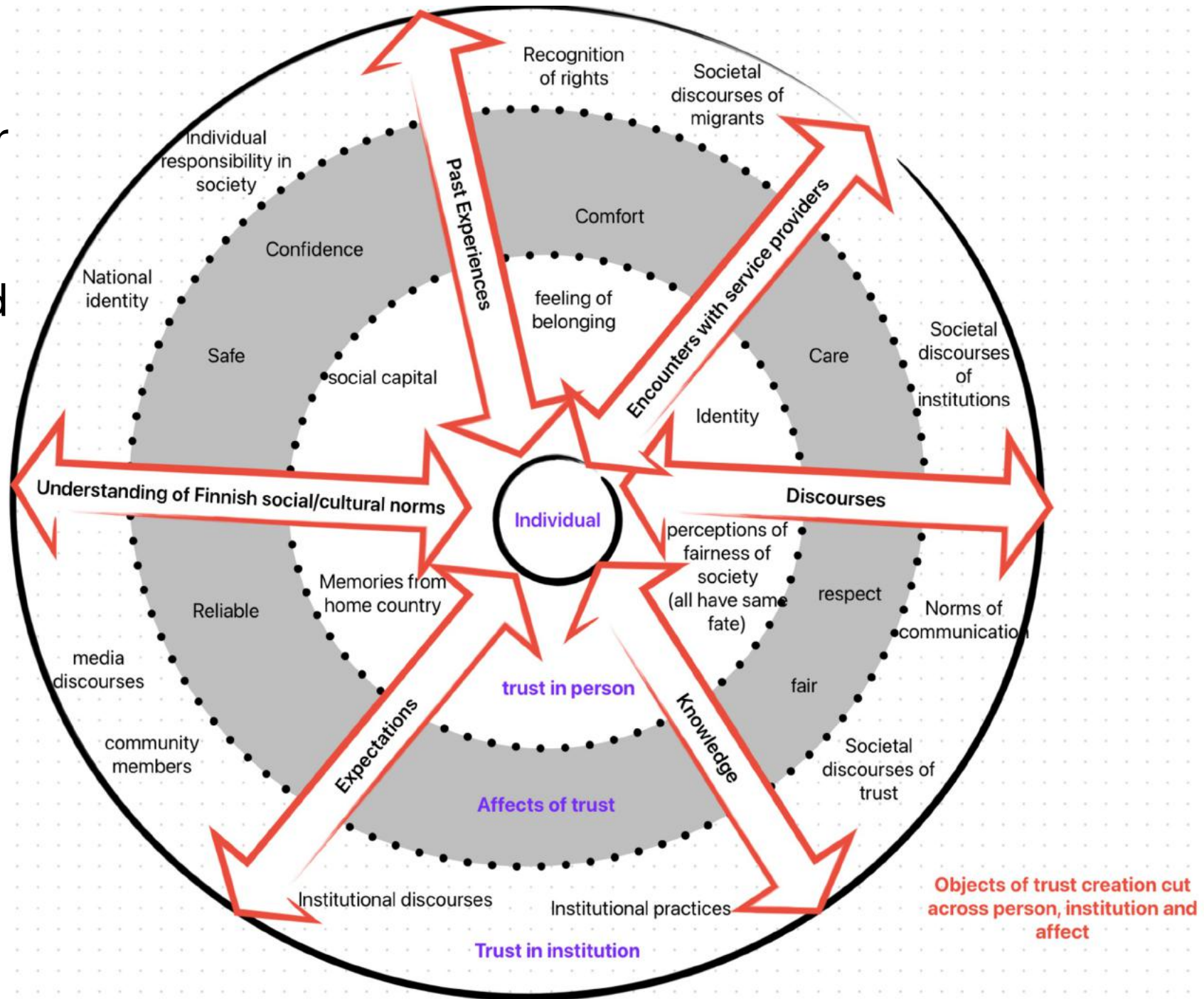
Rahim Ahsanullah and Lucy Truong, Aalto University



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

- Lucy Truong, Thesis Research (2023)

Preliminary framework for concepts of Trust by Avanti Chajed



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

USE OF AI-BASED SYSTEMS IN THE PUBLIC SECTOR

Opportunities

(Barker *et al.*, 2021; Manzoni *et al.*, 2022)

- Support context-specific public values:
 - Operational
 - Political
 - Social
- Foster citizen trust & participation
- Improve efficiency & decision making
- Provide innovative digital services
- Personalisation of public services

Challenges

(Pechtor & Basl, 2022; Pūraitė *et al.*, 2020)

- 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.

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.

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.



Case Study of a Public AI-Enabled Service

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.

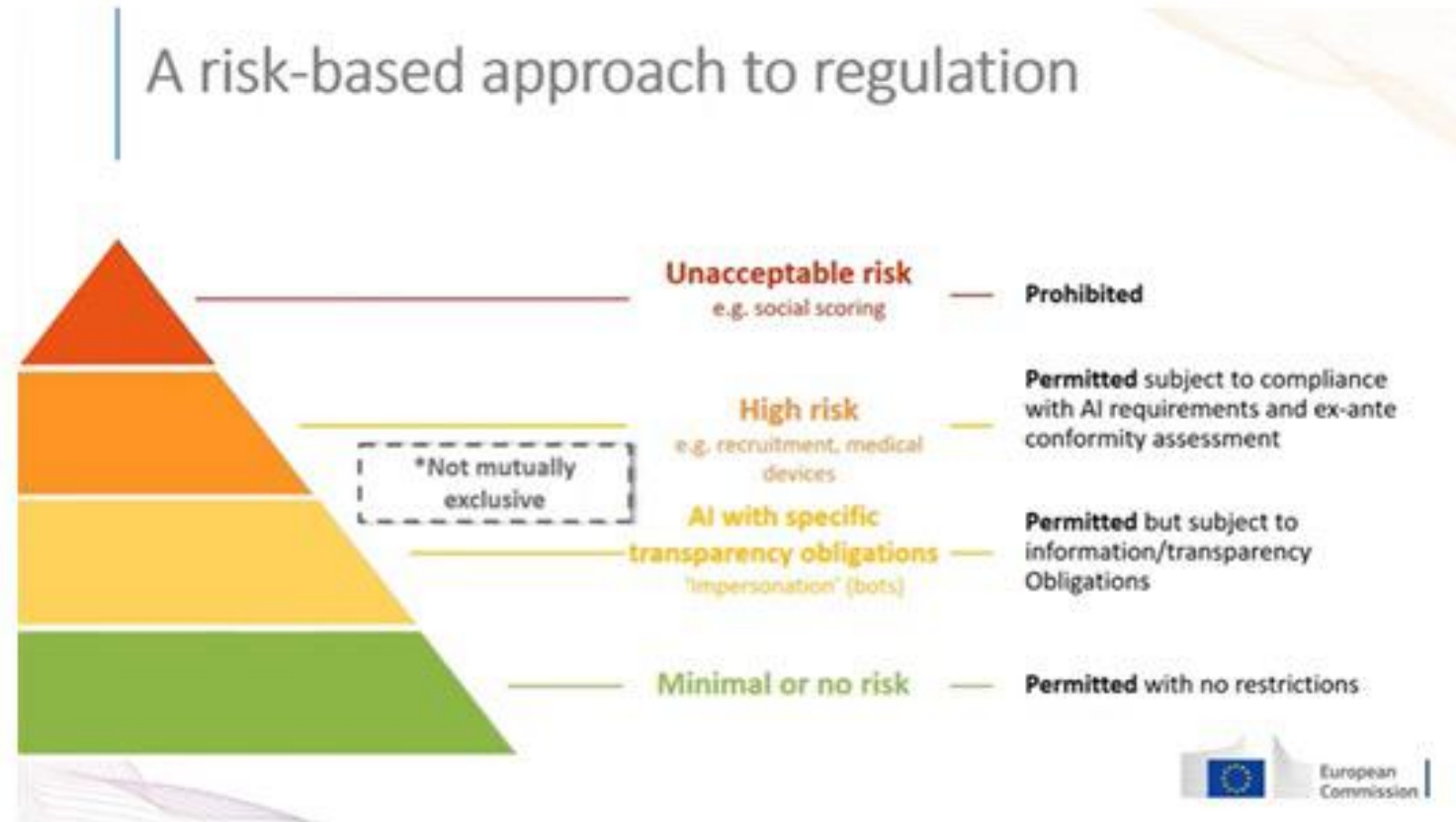
Uses deep interviews as the main data collection. Might also use ethnography and design workshop with public sector actors in the future.

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.

AI Act proposed by 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

A''

Aalto University
School of Science



Kaisla Kajava
Doctoral Researcher
Computational Linguistics

Justifying AI and its Regulation: Examining Multi-Stakeholder Responses to the AI Act

Kaisla Kajava, Ana Paula Gonzalez Torres, Antti Rannisto

kaisla.kajava@aalto.fi, CRAI-CIS Research Group, Department of Computer Science

This work is part of the Civic Agency in AI (CAAI) project funded by the Kone Foundation and the Research Council of Finland.

Reactions to the Proposed AI Act

The rapid uptake of algorithmic systems across sectors reflects in discourse around AI. The AI Act (AIA) regulation proposal, introduced by the European Commission in 2021 (EC, 2021), is one **narrative turn, introducing a legal dimension to the debate** and prompting reactions from diverse stakeholders. The AI discourse reflected in stakeholder responses is structured around the **justification of perspectives and arguments**. How do stakeholders from different sectors justify their views on the AIA and on the use of AI? How do they frame key issues around the regulation, development, and use of AI? Additionally, we examine perspectives on general-purpose and generative models, which were already discussed by stakeholders prior to being considered by regulators.

Data & Methods: Justifications Analysis & Natural Language Inference

DATA 128 English-language feedback documents: 15 academic or research institutions, 57 companies or business organizations, 25 non-profit and non-governmental organizations, 28 business associations, and 3 public sector stakeholders.

METHOD

- Theory of justifications by Boltanski & Thévenot [2] and later additions [3][1][4]: nine *worlds* of justification: *civic, domestic, fame, green, industrial, inspired, market, network, vital*.
- Empirical operationalization **Justifications Analysis (JA)** [6].
- Assistive zero-shot **Natural Language Inference (NLI)** [5] to aid in analyzing more documents.

GOAL Examine the feedback submitted on the first draft of the AIA, focusing on justifications of views on **a) the regulation of AI** and **b) the use and uptake of AI**.

Justifications for and against the AI Act

“Parts of AI research, development, and deployment will require government oversight and regulation. That oversight should **encourage innovation** [inspired] in systems that **enhance human quality of life** [civic].” (OpenAI, 2021)

“Rather than just dictating that data sets be ‘free of errors and complete,’ the AI Act should leave space for the **collaborative development** [network] of **standards and best practices** [industrial] that can help establish measures for high-quality data sets.” (Facebook, 2021)

“[...] must also be covered by the AIA, to prevent states and private sector actors from fast tracking the deployment of high-risk AI systems to **avoid compliance requirements** [industrial]” (Amnesty International, 2021)

Industry: **responsibility, complexity, efficiency, connectedness/agility**

Academic/research: **validity, societal impact**

Nonprofits/NGOs: **reliability, accessibility, deliberation**

Public sector: **expertise, accessibility, responsibility**

NLI Model Evaluation

The performance of the NLI model was evaluated against a manually coded set of 20 documents. **Figure 2** shows the recall for each justification world as well as the total recall across all justification worlds. The recall metric represents the number of justifications which were meaningfully captured using the model (true positives). The total precision score was 78.7%, indicating the number of true positives out of all captured sentences to detect the amount of noise introduced by the model.



Kaisla Kajava
Doctoral Researcher
Computational Linguistics

Justifying AI and its Regulation: Examining Multi-Stakeholder Responses to the AI Act

The most common justifications across stakeholders were based on the *civic* and *industrial* worlds, followed by *market*, *inspired*, and *network* (Figure 1). Stakeholders agree that some uses of AI technologies are adverse to human rights and should be prohibited. As for general employment applications, safety and security are highlighted as a **human-a** across domains proposed to specify **cretized as po** ways both to

Amendments to the AIA & Future Work

In our ongoing research, we build on the analysis of justifications by introducing a policy perspective to **trace the changes to the AIA proposal** following the multi-stakeholder feedback. Our aim is to examine which of the arguments for and against sections of the regulatory text have subsequently been implemented in the form of amendments. The process of regulation-making is complex and both custom- and rule-driven as well as an ongoing deliberation between diverse actors. Thus, we consider the relationship between the multi-stakeholder feedback analyzed in this work and the amended version of the regulation to be more correlative rather than causative. Nonetheless, many modifications to the AIA text have likely been influenced by stakeholder feedback. In this regard, we consider the feedback to the first draft of the AIA a **discursive indicator of the regulatory debate**.



By **Laura Lamberti**

Laura Lamberti is a junior reporter at The Parliament Magazine

25 Jan 2023

[@LauraLamberti10](#)

READ NEXT:



Opinion

OLAF report shows yet again how Frontex systematically ignores human rights

by [Tineke Strik](#)

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

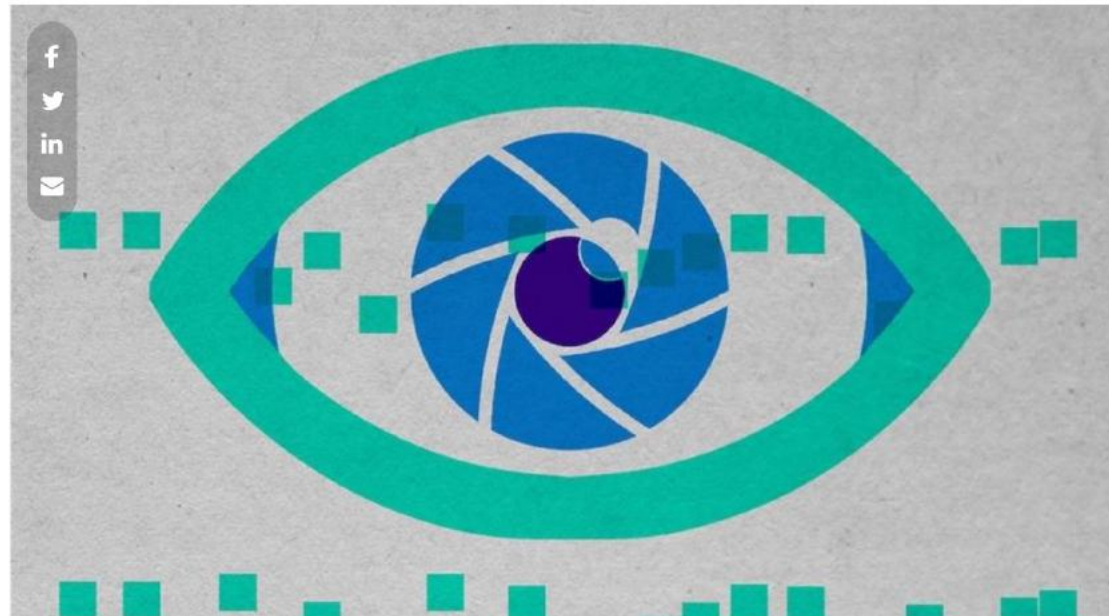


Illustration by Joe Magee

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.

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...

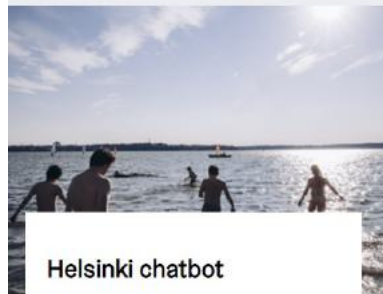
> Read more

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

AI REGULATORY SANDBOXES



EXPERIMENTING WITH
AI INNOVATIONS IN
THE PUBLIC SECTOR



Ana Paula Gonzalez Torres
Doctoral Researcher
Law, Policy & Technology

WHY AI REGULATORY SANDBOXES?

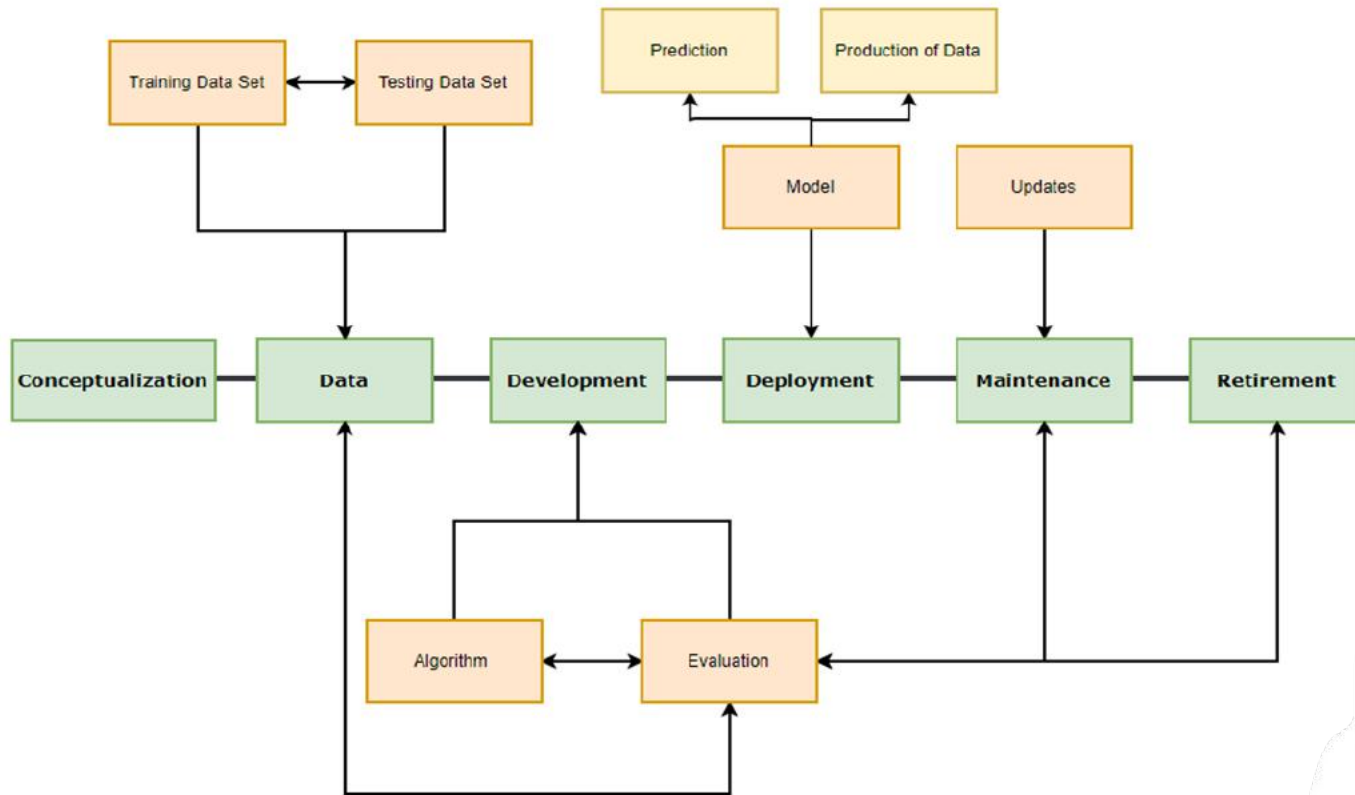
- In **high-risk domains** (e.g., financial sector), regulatory sandboxes used to explore possibilities and implications of algorithmic systems before wider deployment (Manzoni *et al.*, 2022).
- Allow for **experimentation** and critical exploration of both technical and regulatory implications of AI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

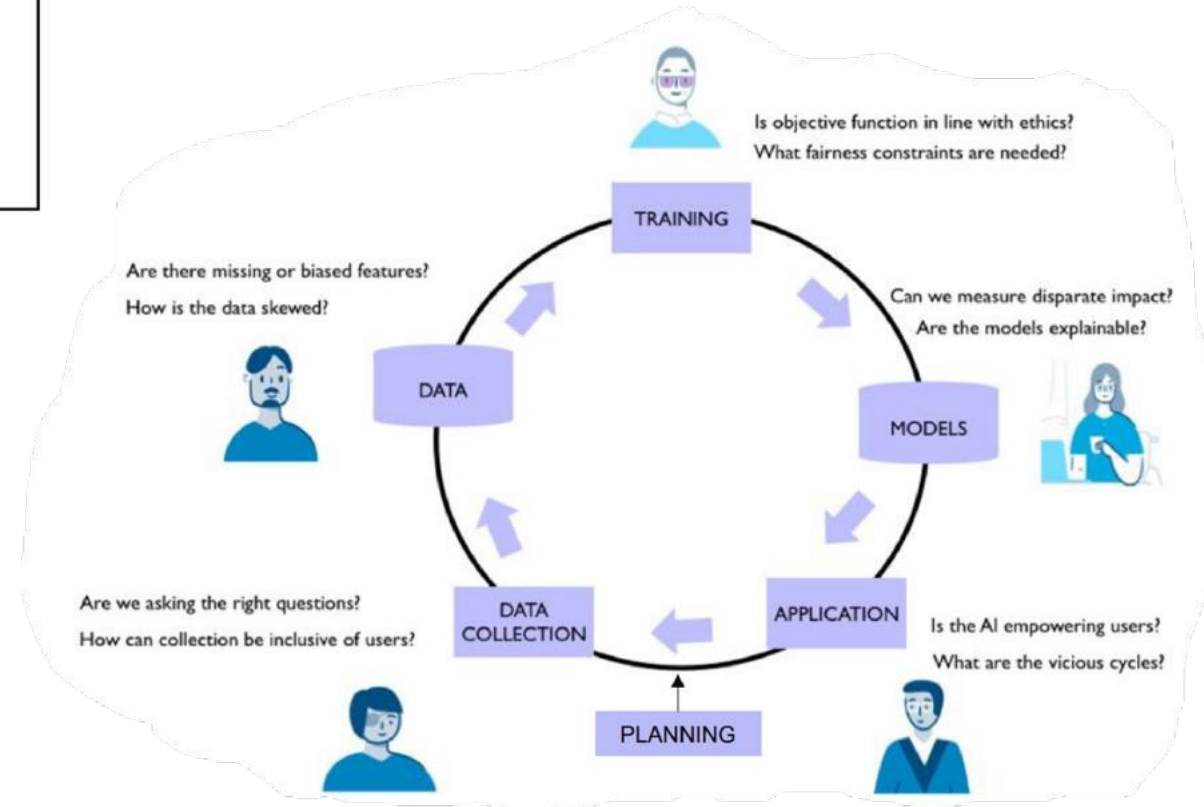
STAGES OF THE AI LIFECYCLE



- 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:

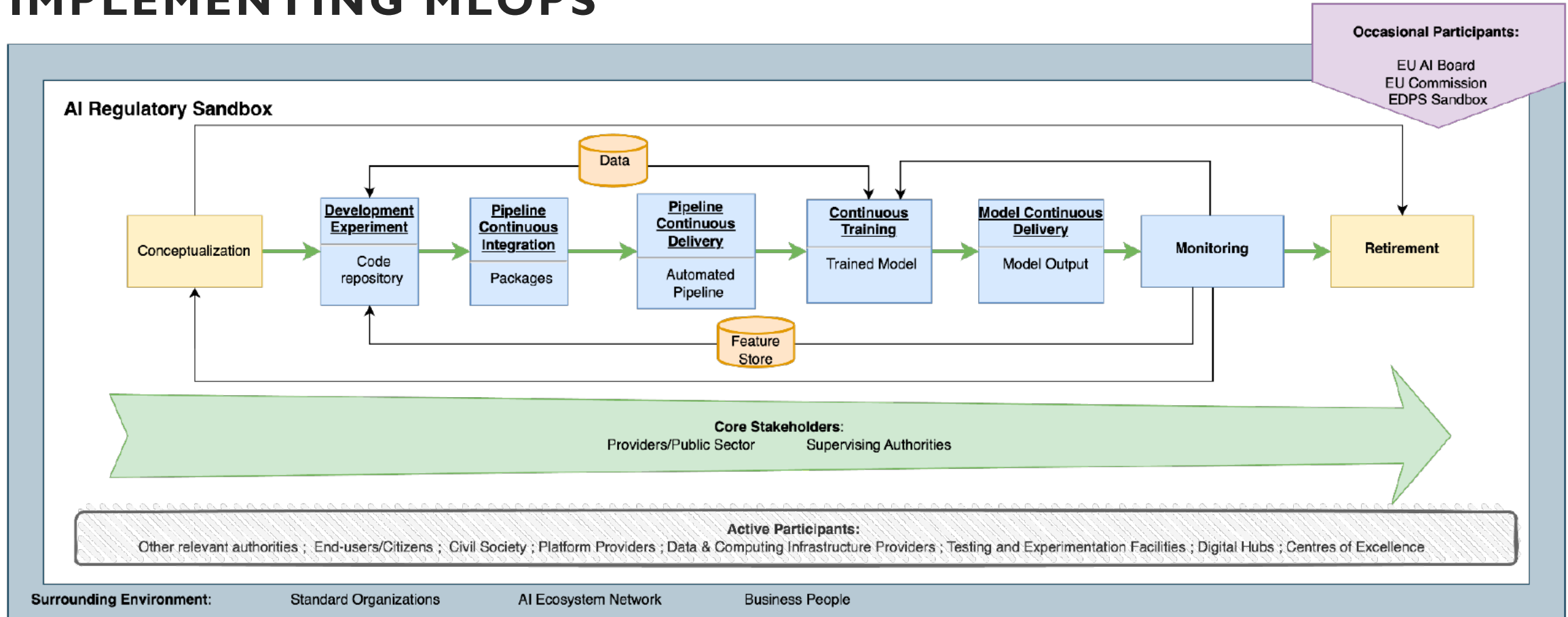
1. **Design:** why an AI-based approach?
2. **Training:** are the datasets biased?
3. **Development:** how outcomes are validated?
4. **Deployment:** what harmful impacts may emerge?
5. **Maintenance:** are there discriminatory feedback loops?
6. **Retirement:** what happens if system recalled?



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 & 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



PILOTING AI REGULATORY SANDBOXES IN FINLAND?

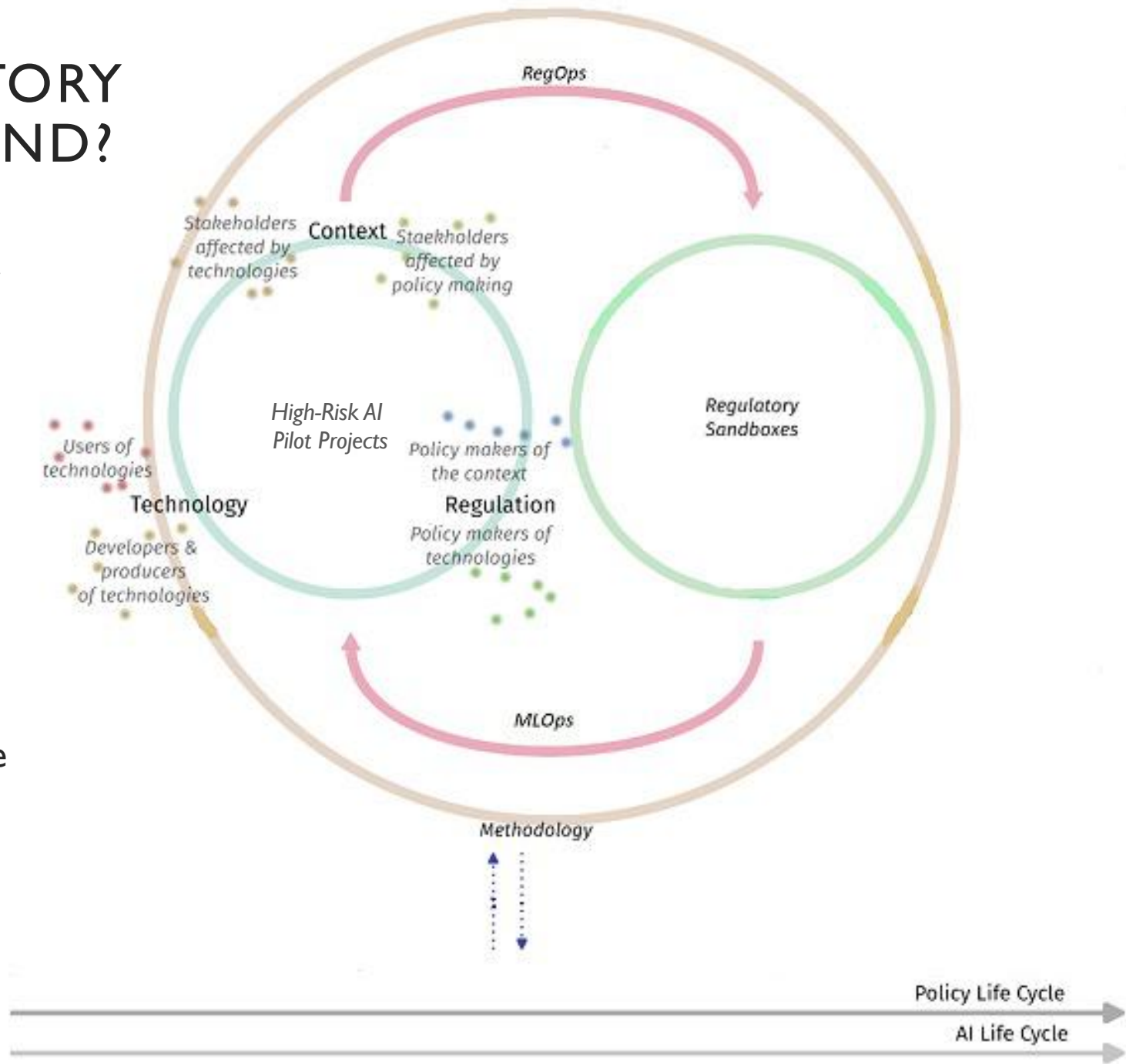
Conduct **pilots with Finnish Public Sector** organizations using AI Regulatory Sandboxes and integrative frameworks

Engage with multiple stakeholders:

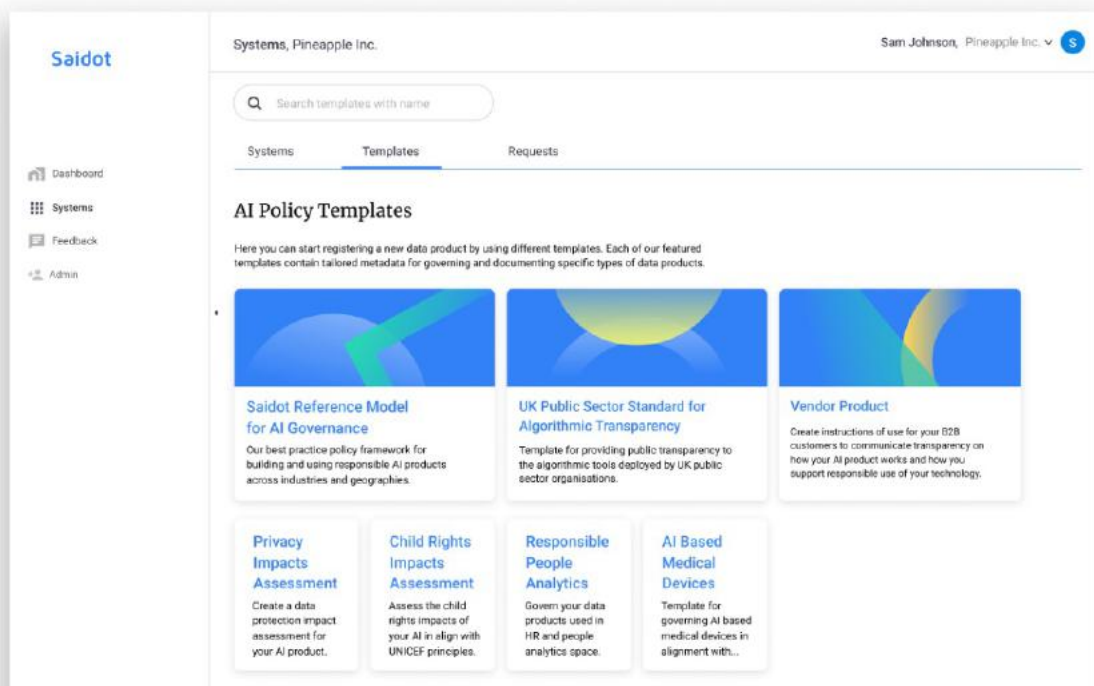
- Regulators, public administration, infrastructure providers, developers, auditing/compliance facilitators

Expected outcomes:

- Provide the space for interactions and mutual collaboration through AI lifecycle
- Examine the limits and possibilities of legislation and technological innovation for public sector AI



COLLABORATING WITH INDUSTRY PARTNERS



Platform for AI governance & transparency

Our platform translates your responsible AI principles into practice and helps you adopt systematic AI governance methodology in alignment with ethical standards and regulatory requirements.

We connect AI teams with internal and external stakeholders for building a better world with fair, accountable and transparent AI.

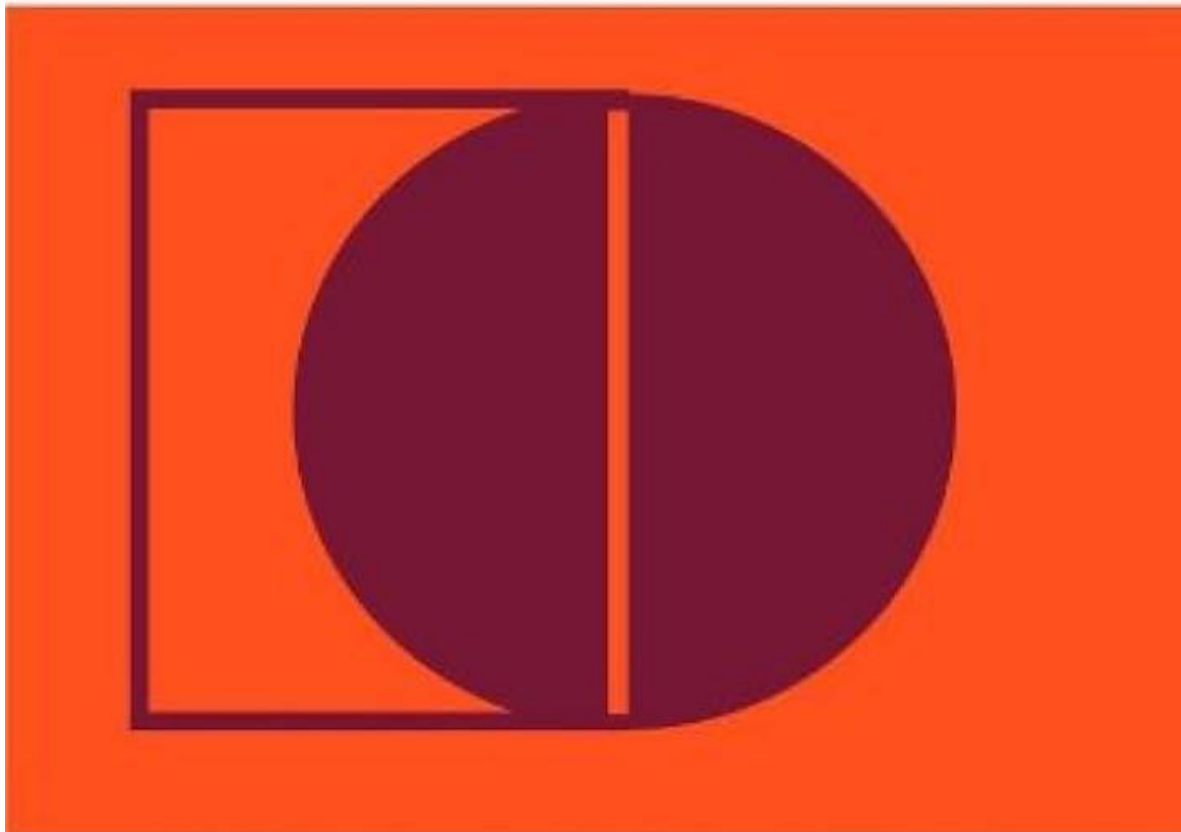
<https://www.saidot.ai>

RECENT 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.



MANIFESTO of TRANSDISCIPLINARITY



for AI & Society?



Transgressing Disciplines Through Critical Pedagogy

Critical AI & Data
Justice in Society

Designing Voice &
Auditory Interaction

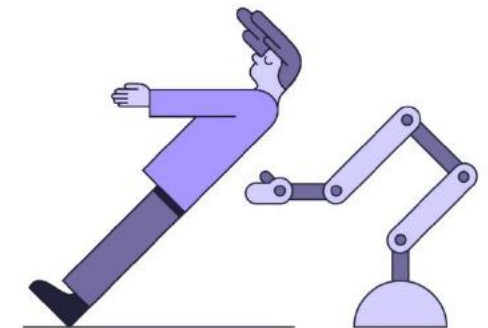
Designing
Trustworthy AI
Systems & Practices



2021



2022



2023

CRAI-CIS Public Seminars

Interdisciplinary Dialogues

CRAI-CIS
seminar

Sarah Pink
Life and Tech at the Edge:
Refiguring Possible Futures

16:00 - 17:30 | 20.09.2023
Room T1,
Computer Science Building,
Aalto University



Life and Tech at the Edge:
Refiguring Possible
Futures

Speaker: Sarah Pink

20.9.2023 16:00–17:30 | Events

CRAI-CIS
seminar

Michel Beaudouin-Lafon
Theory-driven HCI

14:00 - 15:30
11.05.2023

Room T1,
Computer Science Building



Theory-driven HCI

Speaker: Michel Beaudouin-
Lafon

Professor of Computer

11.5.2023 14:00–15:30 | Events

CRAI-CIS
seminar

Wendy Mackay
Applying Generative Theory to
Human-Computer Partnerships

16:00 - 17:30
10.05.2023

Room T2,
Computer Science Building



Applying Generative
Theory to Human-
Computer Partnerships

Speaker: Wendy Mackay

10.5.2023 16:00–17:30 | Events

CRAI-CIS
seminar

Arathi Sethumadhavan
Demystifying Trust and
Reliance on AI

16:00 - 17:30
19.04.2023

Room T1,
Computer Science Building



Demystifying Trust and
Reliance on AI

Speaker: Arathi
Sethumadhavan


19.4.2023 16:00–17:30 | Events

CRAI-CIS
seminar

Ivan Habernal
Towards Privacy-Preserving Natural
Language Processing

16:00 - 17:30
15.03.2023

Room T2,
Computer Science Building



Towards Privacy-
Preserving Natural
Language Processing

Speaker: Ivan Habernal

15.3.2023 16:00–17:30 | Events

CRAI-CIS
seminar

07.12.2022

Massimo Airoidi
Machine habitus: toward
a sociology of algorithms



Machine Habitus: Toward
a Sociology of Algorithms

Speaker: Massimo Airoidi
Assistant Professor of

7.12.2022 16:00–17:30 | Events

CRAI-CIS
seminar

26.10.2022

Noura Howell
Toward an affirmative biopolitics:
Reimagining biodata with feeling
and fabulation



Toward an affirmative
biopolitics: Reimagining

CRAI-CIS
seminar

5.10.2022

Virginia Dignum
Responsible AI: from
principles to practice



Responsible AI: from
principles to practice

CRAI-CIS
seminar

14.9.2022

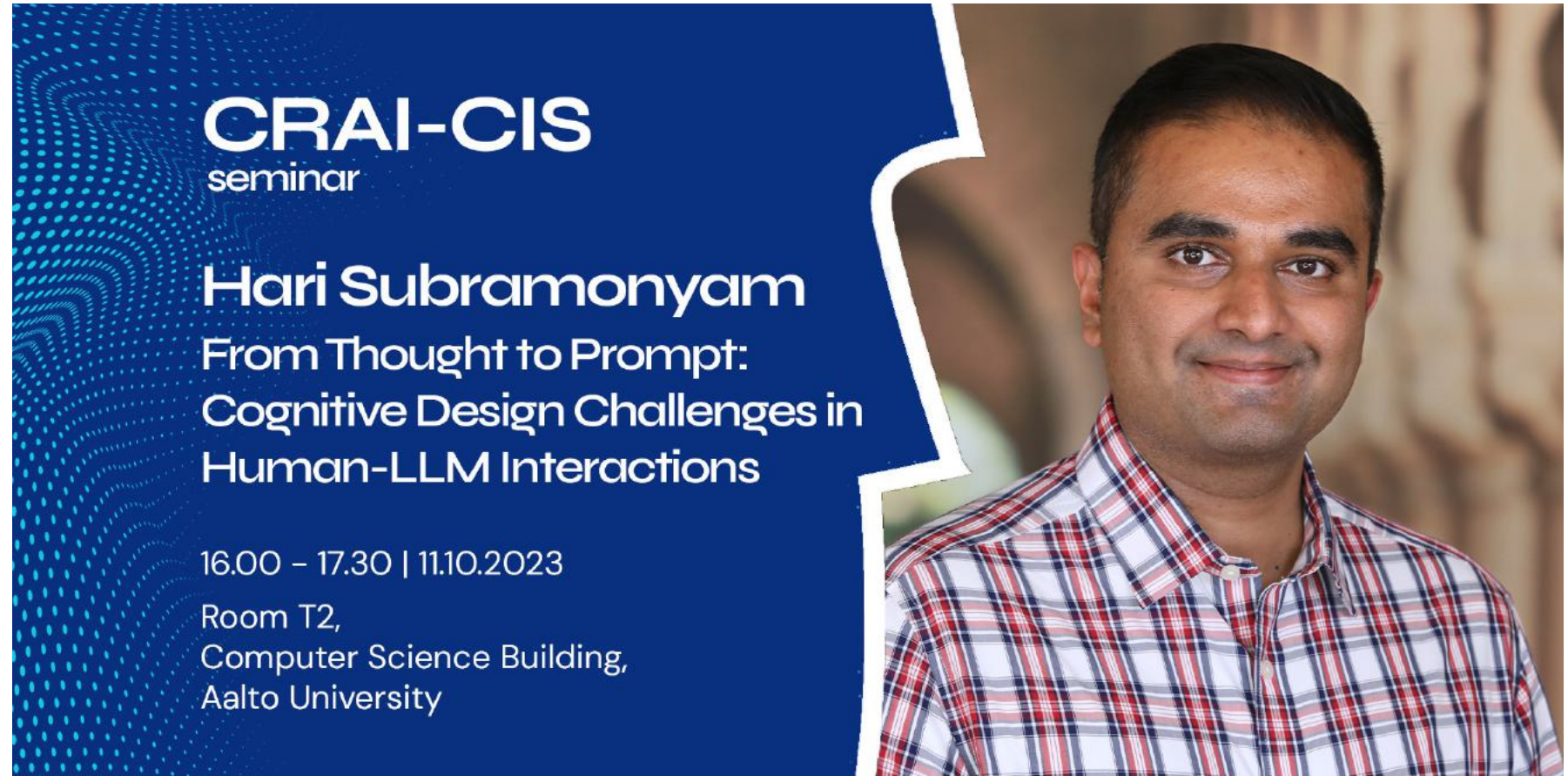
Katja Velaskivi
Imagining communities with
“intelligent” machines



Imagining communities
with “intelligent” machines

CRAI-CIS Public Seminars

Hosting Visiting Researchers to Foster Collaborations



CRAI-CIS
seminar

Hari Subramonyam
From Thought to Prompt:
Cognitive Design Challenges in
Human-LLM Interactions

16.00 – 17.30 | 11.10.2023
Room T2,
Computer Science Building,
Aalto University

ELLIS Distinguished Lecture & CRAI-CIS Seminar – Serge Belongie

Wednesday, November 1, 2023
16:00 – 17:30

1 Otakaari, Espoo, Uusimaa, 02150, Finland (map)

Google Calendar · ICS



ELLIS DISTINGUISHED LECTURE & CRAI-CIS SEMINAR

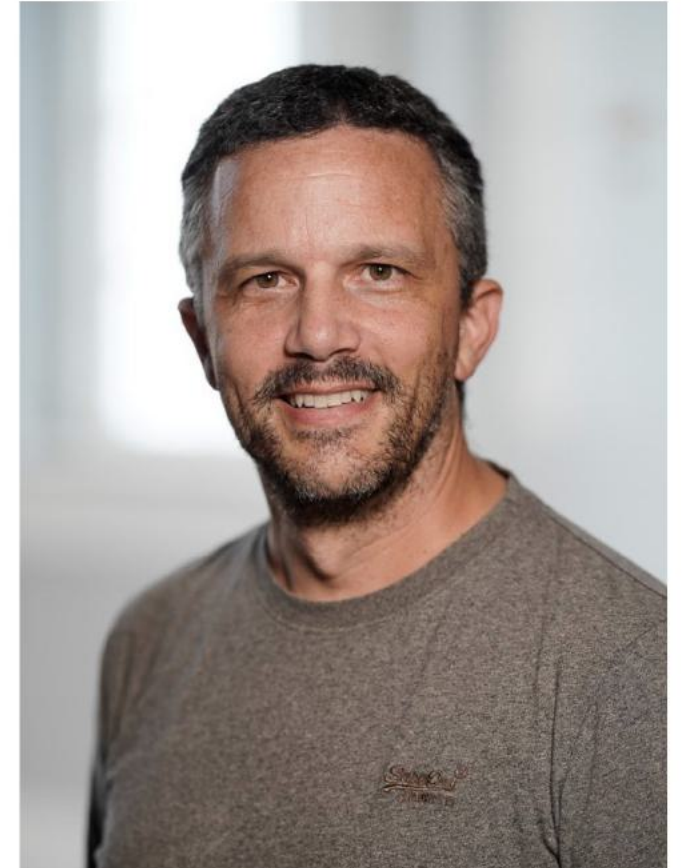
SERGE BELONGIE: SEARCHING FOR STRUCTURE IN UNFALSIFIABLE CLAIMS

Time: November 1, 2023, 16:00-17:30 EET

Venue: Y203a, B Hall (Undergraduate Centre,
Otakaari 1, Espoo) / Zoom (link TBC)

Abstract:

While advances in automated fact-checking are critical in the fight against the spread of misinformation in social media, we argue that more attention is needed in the domain of unfalsifiable claims. In this talk, we outline some promising directions for identifying the prevailing narratives in shared content (image & text) and explore how the associated learned representations can be used to identify misinformation campaigns and sources of polarization.



<https://fcai.fi/calendar/2023/ellis-distinguished-lecture-belongie>