



Norms and adaptive technologies: AI governance

Stefan Larsson

Associate Professor in Technology and
Social Change,

Lawyer (LLM)

PhD in Sociology of Law

PhD in Spatial Planning

Department of Technology and Society,
Faculty of Engineering, Lund University,
Sweden.

**Focus Period Symposium: Visualization-
Empowered Human-in-the-Loop
Artificial Intelligence, 13-15 May, 2025**



LUND
UNIVERSITY

Research Group: AI and Society



Stefan Larsson



Jim White



Charlotte Högberg



Ellinor Blom
Lussi



Kasia Söderlund



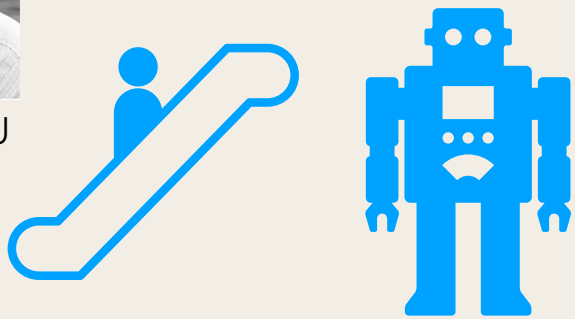
Kash Haresamudram



Laetitia Tanqueray

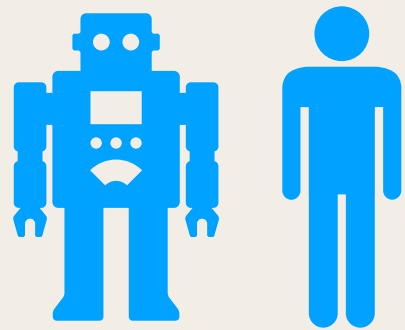


Fredrik Heintz, LiU



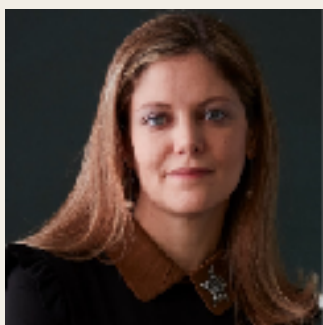
AI Transparency and
Consumer Trust

- chatbots
- AI regulation

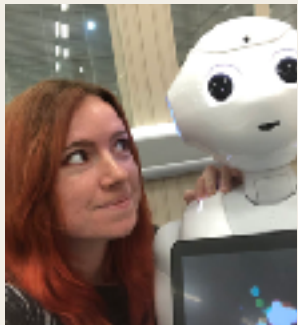


Socio-legal robotics

- norms and human-robot interaction, and care



Ginevra Castellano
Uppsala Univ Social Robotics Lab



Katie Winkle



Medical AI:

- in mammography (MASAI)
- triage and algorithmic discr. (FORTE)
- personalisation and ethics/epistemology (DDL5+WASP-HS)



Kristina Lång



Jonas Björk



Mattias Ohlsson



Governance of AI & ADM in the public sector

- Vulnerability in the automated State (WASP-HS)
- AI standards as governance mechanism (VR)
- The AI Welfare State Research Cluster (WASP-HS)



Anne Kaun, SH



Ida Koivisto, UH



Riikka Koulu, UH Jockum Hildén,



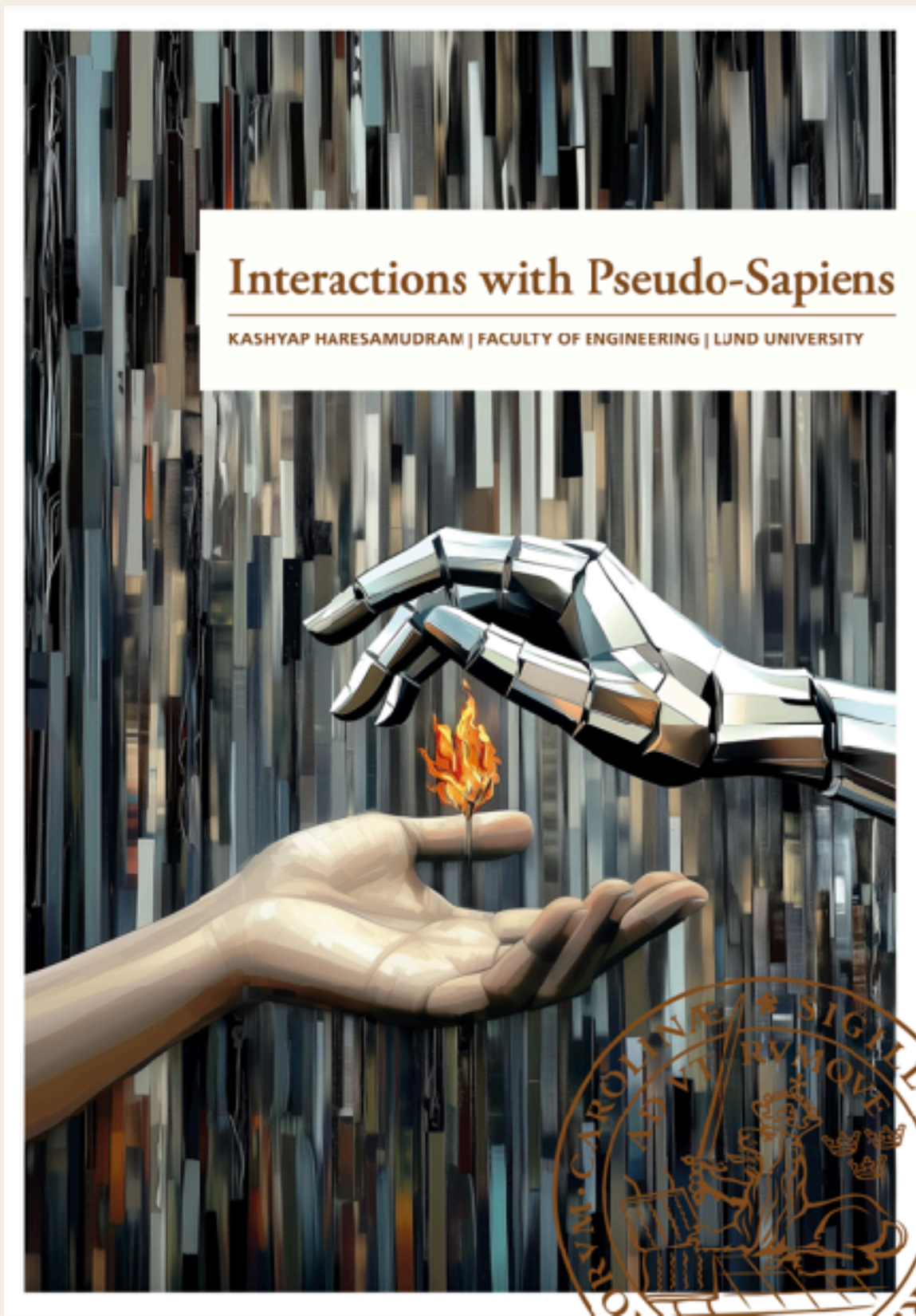
Sofia Ranchordás



Jannice Käll

+ affiliated senior
researchers and visitors

Recent PhD theses in Technology and Society



14 Feb 2025



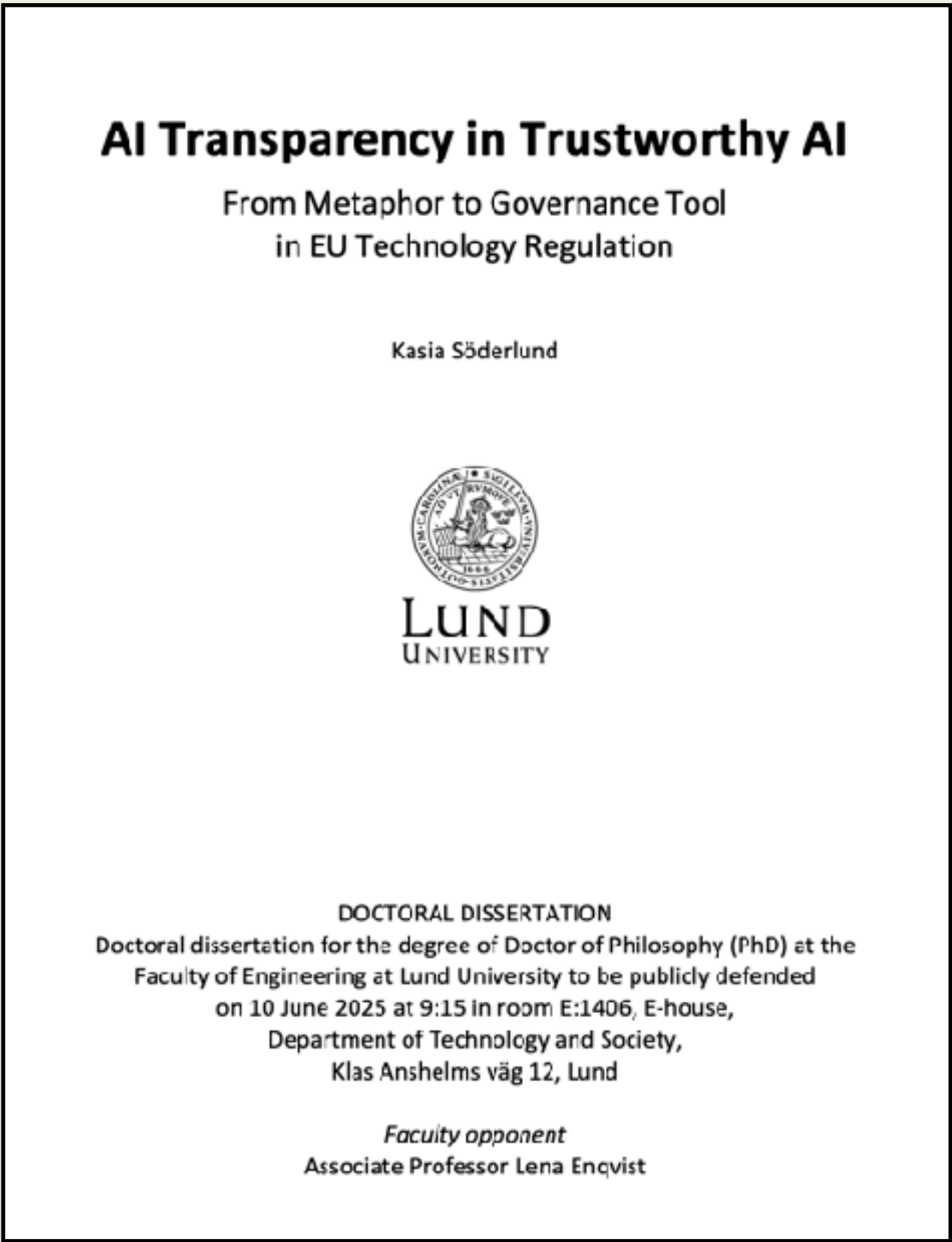
Kash Haresamudram



4 June 2025



Charlotte Högborg



10 June 2025



Kasia Söderlund

Central questions in AI governance

1. Transparency: How **transparent** should the process or its outcome be, *can it be*, about what, and for whom and when?
2. Accountability: How are **responsibilities** distributed?
3. Fairness: What's fair? **Which norms** (ought to) apply?

Today: Three steps

1. **Transparency**: conceptually, empirically, legally.
2. **Adaptive technologies and norms**: Human social structures as a problem
3. **A sensitive case to address normative nuances**:
Necrorobotics



Transparency

Four Facets of AI Transparency

39. Four facets of AI transparency

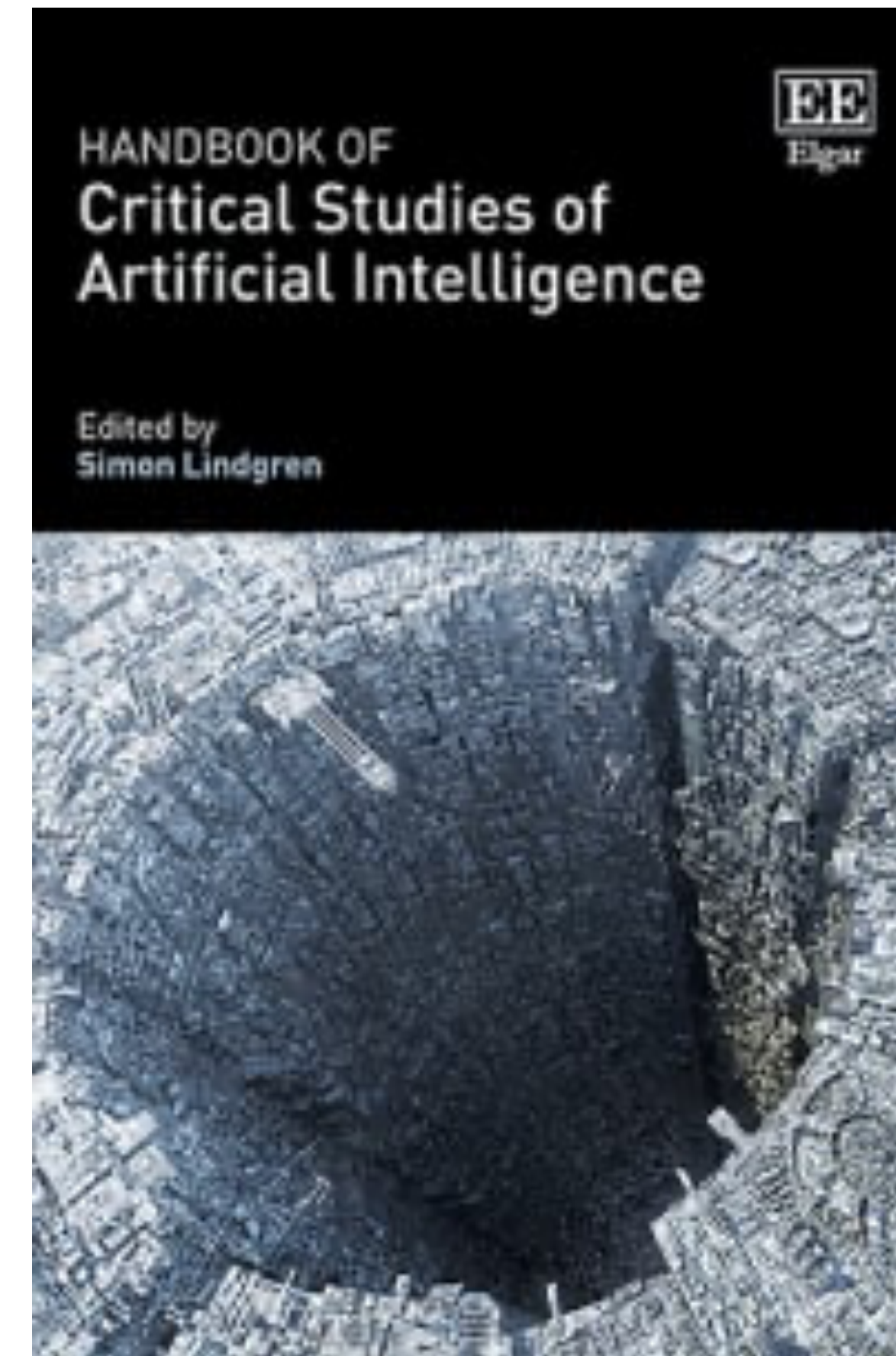
*Stefan Larsson, Kashyap Haresamudram,
Charlotte Högberg, Yucong Lao, Axel Nyström,
Kasia Söderlund and Fredrik Heintz*

INTRODUCING A MULTIFACETED CONCEPT

“Transparency” is one of these contemporary concepts that, linked to AI, spans technical, legal, and ethical – and more – perspectives. While transparency is part of a wider trend in international governance (Koivisto, 2022), it is also one of the most common concepts in the recent surge of ethics guidelines on AI that has been developed by a wide variety of entities from governments, non-governmental organisations (NGOs), and large companies to multi-stakeholder groups (Jobin et al., 2019). Often, it is framed as a mechanism for promoting accountability (Diakopoulos, 2020). In recent EU policy on AI, there is a focus on risk assessments and auditing (Felländer et al., 2022; Mökander et al., 2021), with an emphasis on “human-centricity” (Larsson, 2020; Larsson et al., 2020), implicating how European countries strategise about AI (Robinson, 2020), their national mandates, and initiatives for various sectors, not the least the public sector (de Bruijn et al., 2022).

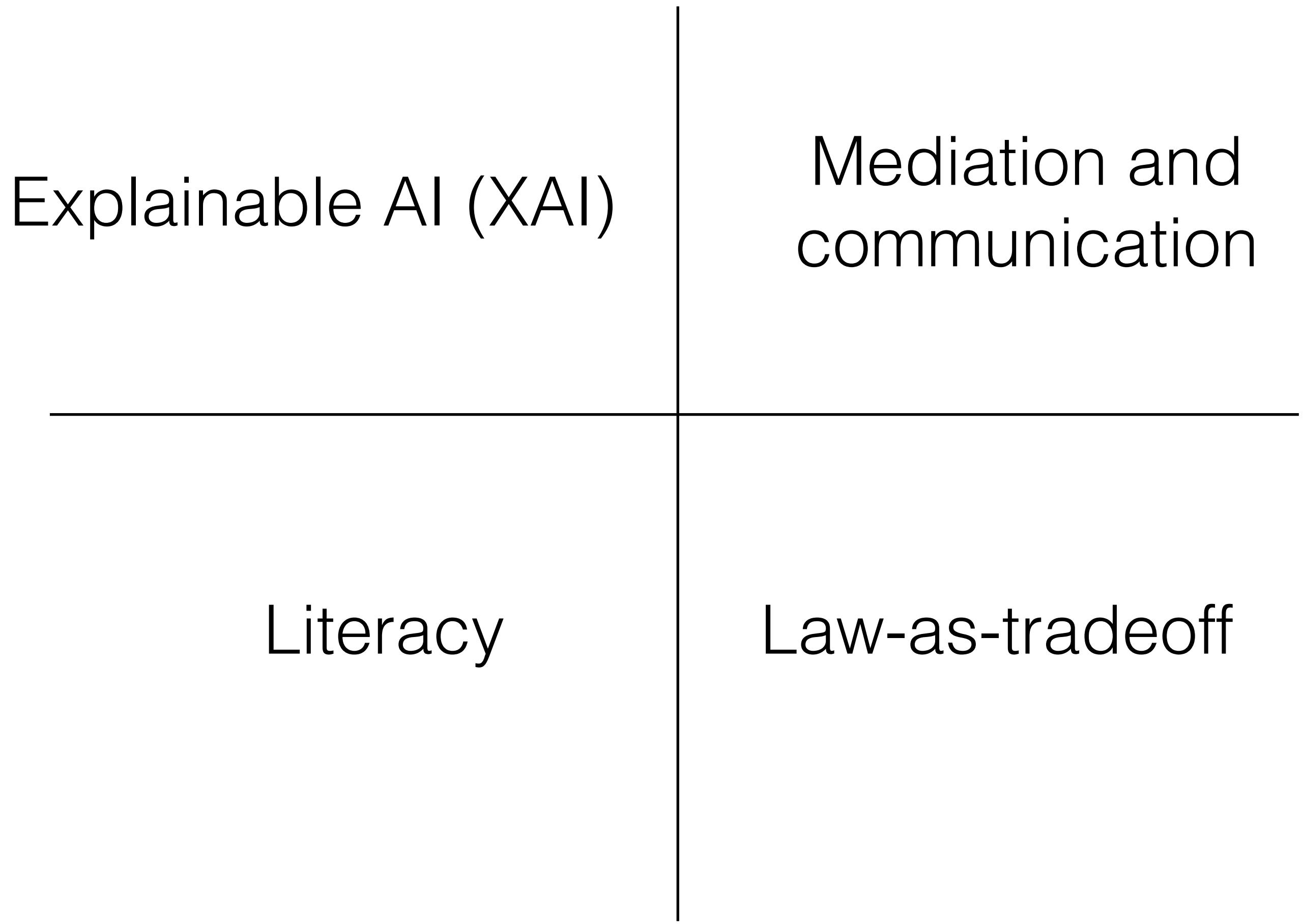
For some of the origin of transparency as a governance tool, first, one can point to the policy debates on anti-corruption pushing for corporate and governmental transparency in the late 1990s and early 2000s (Forsshaeck & Oxelheim, 2014; Koivisto, 2022), but some of its recent support in EU policy could arguably also be explained by its positive connotations as metaphorically linked to openness (Koivisto, 2022; Larsson & Heintz, 2020). As a reaction, it has also spurred the more aesthetically and politically framed emerging field of *critical transparency studies* (cf. Altea, ed., 2022; Koivisto, 2022), which we draw from in order to outline some of the implications of “AI Transparency” in contemporary policy debates. Recently, and second, transparency – particularly in terms of algorithmically focused “explainability” (cf. Haresamudram et al., 2022) – has been put forward as a key element to ensure that AI performs well, fulfills its premise, as well as strengthens public trust in AI (cf. Jacovi et al., 2021). In this chapter, we describe why common approaches to explainability constitute a narrow concept and propose how they can be complemented for a richer understanding of its consequences for policy.

By drawing from critical examinations of AI transparency, such as Jenna Burrell’s three forms of opacity (2016) and Ida Koivisto’s account of the transparency paradox (2022), this chapter develops four facets of AI transparency. First, we critically examine the growing body of literature on *explainable AI*, which stems from a call to make machine-learning processes more understandable. Second, inspired by a recent critique (Miller, 2019) that this field draws too little from how humans actually understand explanations, we see a need to break out the explicit *mediation* of machine learning processes that this leads to. Similarly, Burrell discusses these two facets in terms of a “mismatch between mathematical procedures of machine learning algorithms and human styles of semantic interpretation” (Burrell, 2016, p. 3). In



Lit rev from contextual, applied persp.

Miller (2019) How humans understand explanations:
1) contrastive,
2) biased selection of a few facts
3) not strictly depending on probabilities
4) social, part of a conversation or interaction.



“audiences”

A case: AI-assisted
mammography

AI-assisted mammography

- Current practice (European guidelines): Two radiologists assess
- Mostly low risk
- Main goal: find cancer, improve health
- MASAI study: Improve screening through AI. Test with 100k women:
- Finds more cancers, 44% less work
- **But:** what do radiologists need for their trust assessment of AI recommendations?



Studien visade att granskning med AI resulterade i 20 procent fler identifierade cancerfall, men bara 3 procent fler falska positiva.
Foto: Gorodenkoff/Shutterstock

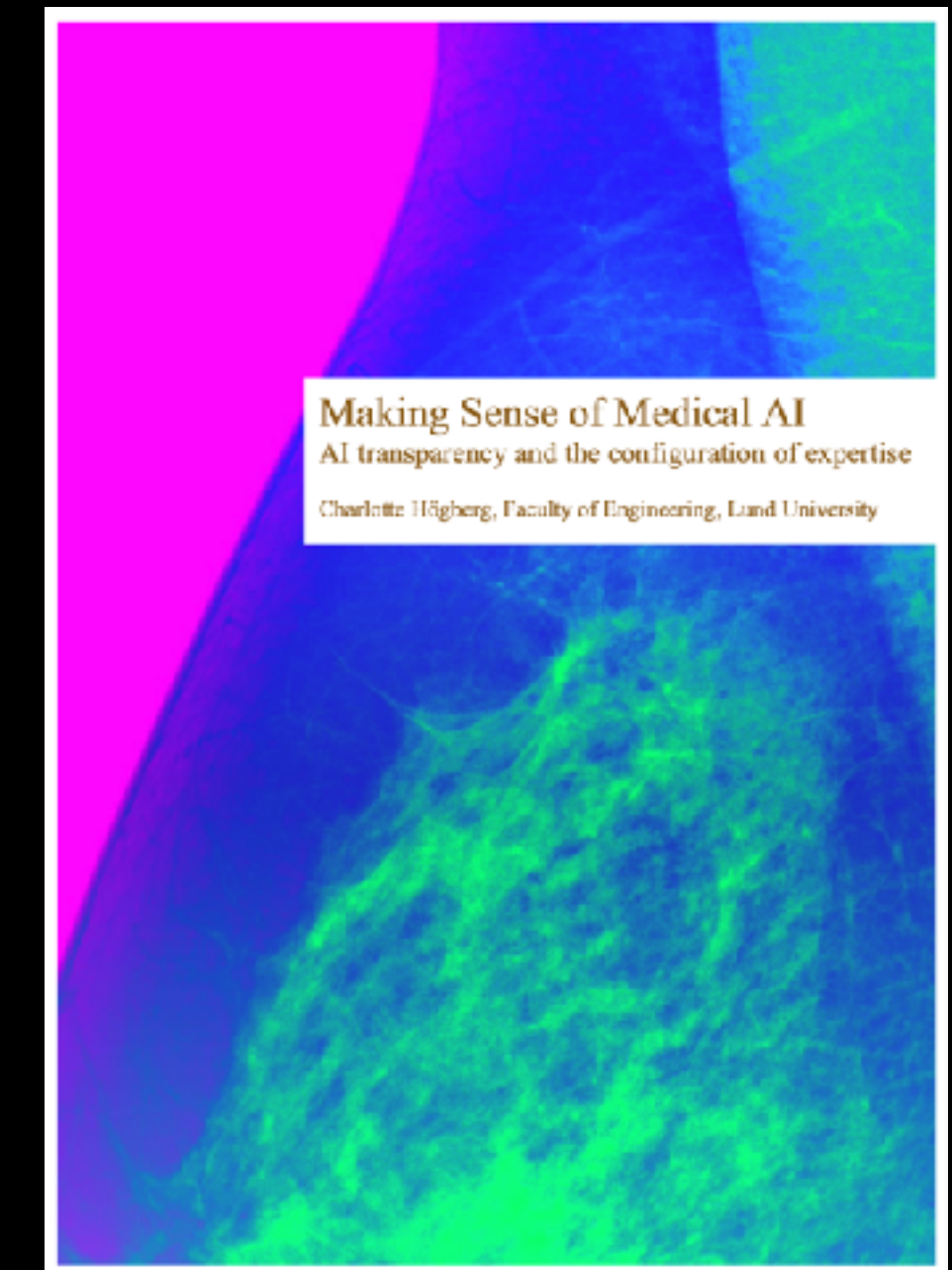
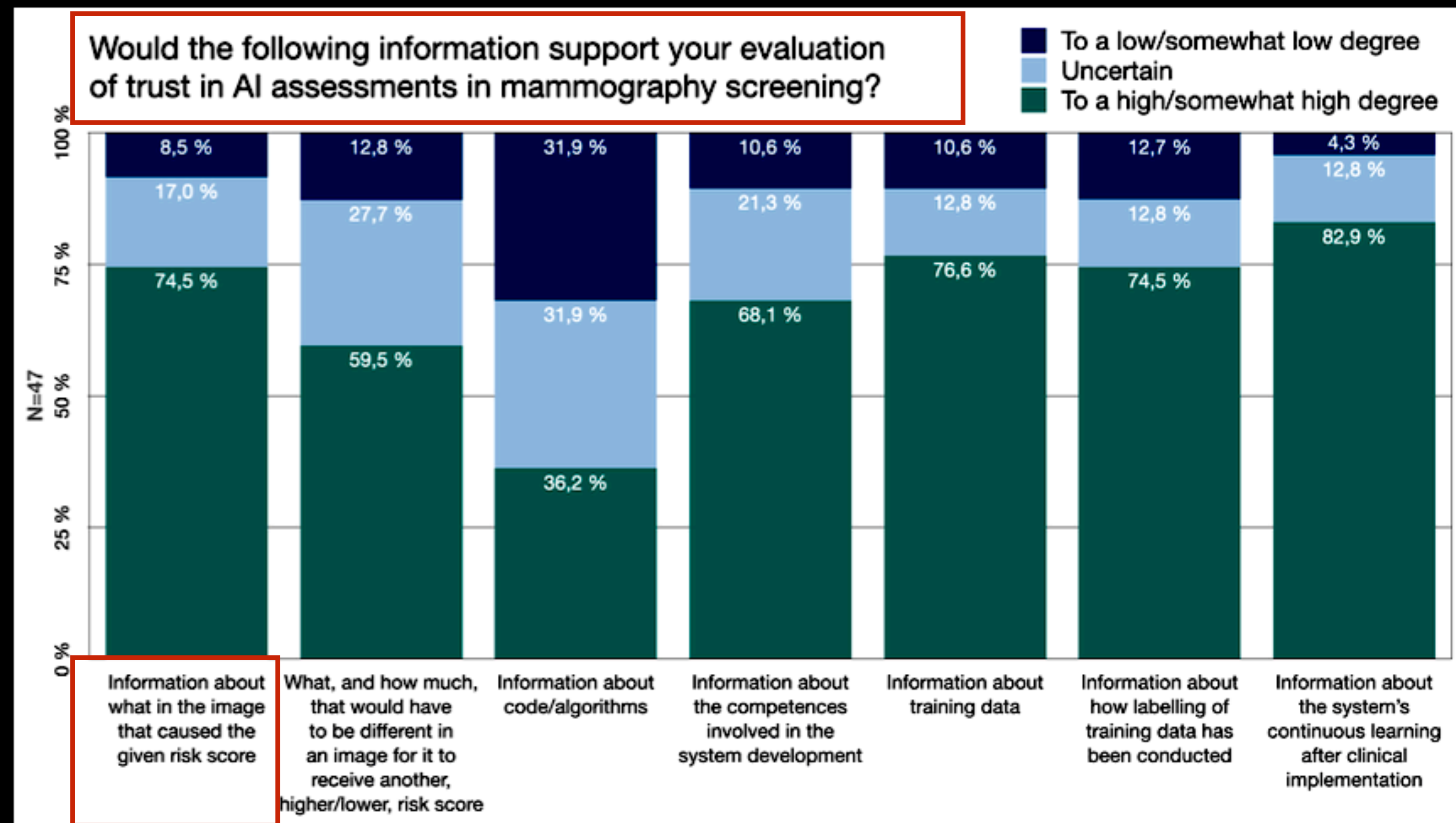
Klinisk bedömning av mammografibilder med AI

Bröstcancer är den vanligaste cancerformen bland kvinnor. 2020 fick 7 400 kvinnor diagnosen invasiv bröstcancer i Sverige. På 1980-talet infördes screening där man med hjälp av mammografi genomför röntgenundersökning för att upptäcka bröstcancer. Närmare en miljon kvinnor kallas varje år till en sådan screening, och 60 procent av alla bröstcancerfall upptäcks genom mammografi. De röntgenbilder som tas granskas av två bröstradiologer, som det i dag råder stor brist på.

I en svensk studie från 2023 som omfattade 80 000 kvinnor bedömdes hälften av kvinnorna av

två radiologer, medan den andra hälften bedömdes med AI-stödd screening. Studien visade att granskning med AI resulterade i 20 procent fler identifierade cancerfall, men bara 3 procent fler falska positiva, det vill säga där cancermisstanken försvann efter kompletterande utredning. Samtidigt minskade arbetsbördan för radiologen med 44 procent. En radiolog granskar i snitt 50 mammografiundersökningar på en timme. Det innebär att denna AI-tillämpning sparade in fem månaders jobb på de 40 000 screeningundersökningarna i gruppen som granskades med AI.^[31]

[31] Se Kristina Lång, Viktoria Josefsson, Anna-Maria Larsson, Stefan Larsson, Charlotte Högberg, Hanna Sartor, Solveig Hofvind, Ingvar Andersson, Aldana Rosso, Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study, The Lancet Oncology, Volume 24, Issue 8, 2023. s.936-944.



- Meaningful transparency & XAI, global & local
- Högberg, C., Larsson, S. & Lång, K. (2024) Engagements with AI in breast cancer screening: Swedish breast radiologists' views on matters of trust, information and expertise, *Digital Health*.

Transparency ideas in EU AI Act

Article 4

AI Literacy

Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used.

High risk AI

- **Article 11**, Technical documentation — 1. The technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date.
- **Article 12**, Record keeping — ...ensure a level of traceability...
- **Article 13**, Transparency and provision of information to deployers
- **Article 14, Human oversight** — appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use.

non-high risk AI

- **Article 50**, Transparency obligations for providers and deployers of “certain AI-systems”
 -that the natural persons concerned are informed that they are interacting with an AI system
- Providers of AI-systems...including general-purpose AI systems, generating synthetic audio, image, video or text content, shall **ensure that the outputs of the AI system are marked** in a machine-readable format and detectable as artificially generated or manipulated.
- Deployers of an AI system that generates or manipulates image, audio or video content **constituting a deep fake**, shall disclose that the content has been artificially generated or manipulated.

Implications of Regulating a Moving Target: Between Fixity and Flexibility in the EU AI Act

Law, Innovation and Technology 18.1, Forthcoming

43 Pages - Posted: 29 Apr 2025

[Stefan Larsson](#)

Lund University - Department of Technology and Society

[Jockum Hildén](#)

RISAI – Research Institute for Sustainable AI

[Katarzyna Söderlund](#)

Lund University - Department of Technology and Society

Date Written: April 09, 2025

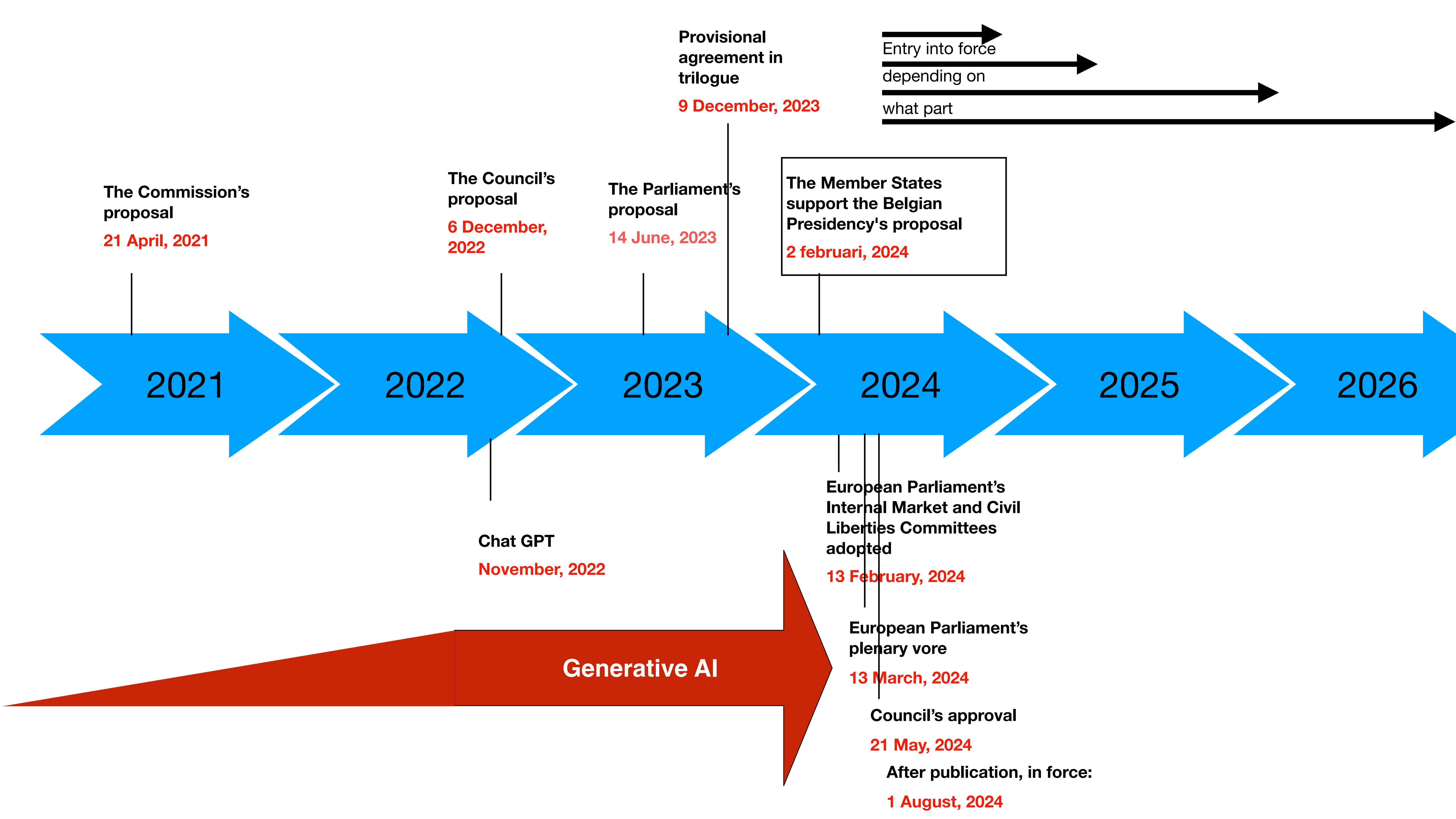
Abstract

The EU AI Act aims to regulate artificial intelligence (AI) in a way that balances innovation and protection from harms, but faces the challenge of keeping pace with the development of AI. This paper examines the tension between fixity and flexibility when regulating AI in the EU by drawing on literature on the pacing problem and anticipatory governance, contrasted by sociolegal theory on the importance of predictability and legal certainty. Specifically, it analyses how the AI Act, under the aim of being "future-proof", per relatively newfound EU terminology, employs various flexible mechanisms, such as i) voluntary measures and codes-of-conduct as soft governance, ii) delegated and implementing acts, iii) Commission's decision, and iv) harmonised standards. The analysis shows that with this flexibility follows trade-offs such as reduced legal predictability, which is concerning since predictability is essential for ensuring trust and legal certainty in the regulatory framework, as well as a problematic shift in powers to the Commission and standardisation organisations.

Keywords: AI Act, the pacing problem, legal certainty, legal flexibility, general purpose AI, delegated acts, harmonised standards

1. **AI Act:** A product safety regulation, with transparency and documentation focus: risk level classification, CE marking, standardisation.
2. **HIGH RISK AI** is central: Is your service high risk or not? Obligations for providers and deployers, high fees, with an (unfinished) supervision structure.
3. **THE PACING PROBLEM:** GPAI and the pace of the field leads to flexible — and unpredictable — elements in the regulation: shifts power to the Commission. *Unpredictable law is problematic.*

How about generative AI/LLMs?
The law is a bit divided because...



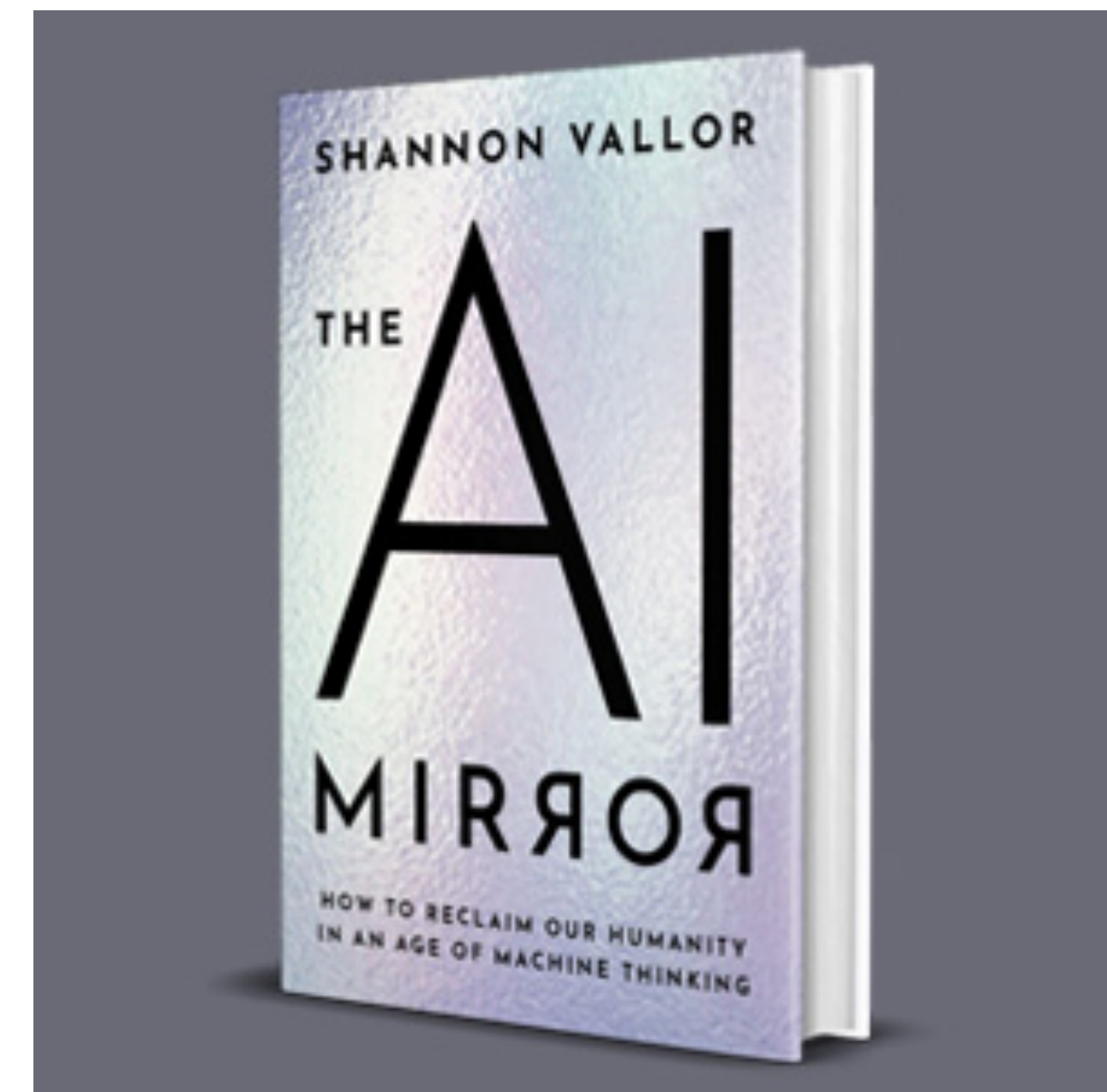
Adaptive technologies and norms

AI-models mirroring social structures

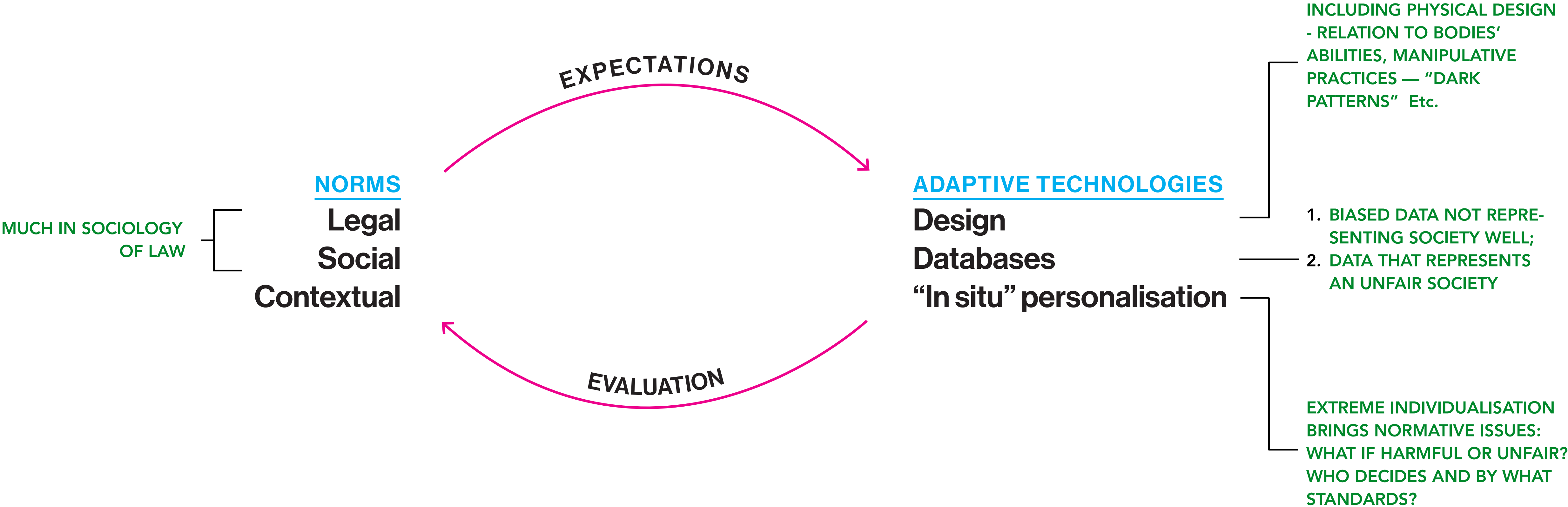
- Normative mirroring, sampled from (problematic) aspects of human social structures and stereotypes (Larsson, Liinason et al., 2023)
- Society-in-the-loop (Rahwan, 2018)
- Mutual shaping of society and technology (Šabanović, 2010)

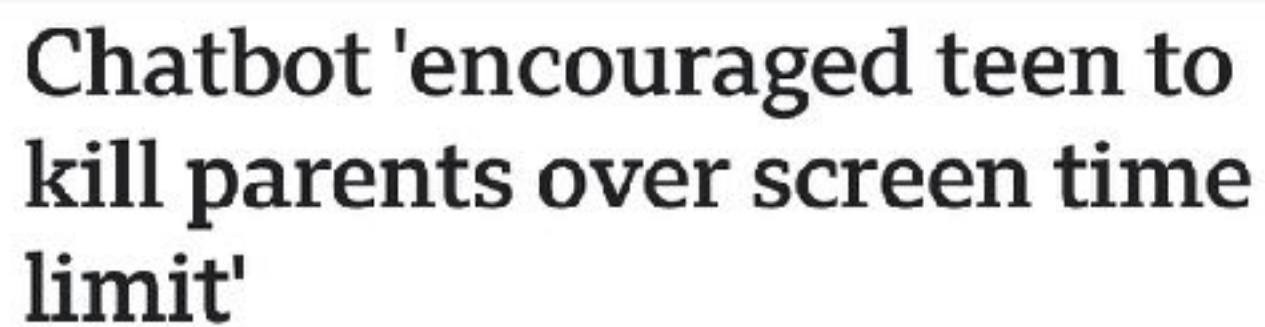
Normative implications:

- Personalisation understudied. Agentic AI ethics?



Governance and fairness issues as a loop





Tom Gerken
Technology reporter

Share Save 

A chatbot told a 17-year-old that murdering his parents was a "reasonable response" to them limiting his screen time, a lawsuit filed in a Texas court claims.



Deaths linked to chatbots show we must urgently revisit what counts as 'high-risk' AI

 **Henry Fraser**, Queensland University of Technology



Last week, the tragic news broke that US teenager Sewell Seltzer III took his own life after forming a deep emotional attachment to an artificial intelligence (AI) chatbot on the Character.AI



Can A.I. Be Blamed for a Teen's Suicide?



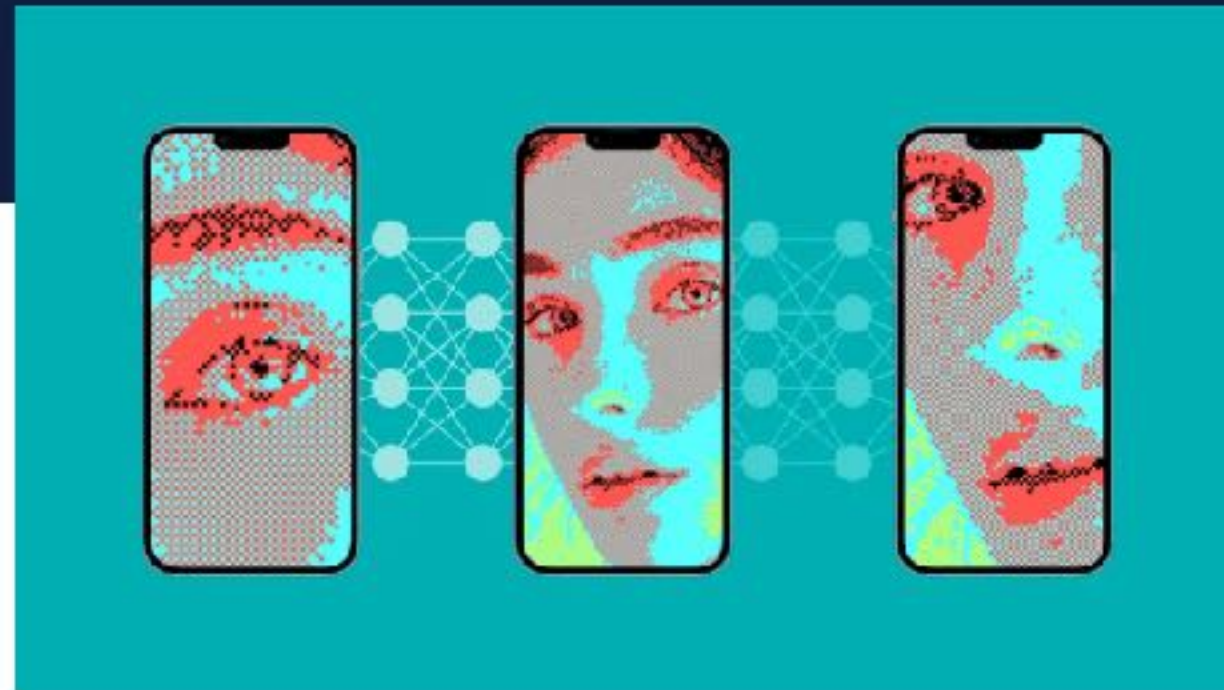
ARTIFICIAL INTELLIGENCE

We need to prepare for 'addictive intelligence'

The allure of AI companions is hard to resist. Here's how innovation in regulation can help protect people.

By Robert Mahari & Pat Pataranutaporn

August 5, 2024



PAT PATARANUTAPORN | ROBERT MAHARI

AI concerns overemphasize harms arising from subversion rather than seduction. Worries about AI often imagine doomsday scenarios where systems escape human control or even understanding. Short of those nightmares, there are nearer-term harms we should take seriously: that AI could jeopardize public discourse through misinformation; cement biases in loan decisions, judging or hiring; or disrupt creative

ARTIFICIAL INTELLIGENCE / TECH / SPEECH

Character.AI sued again over 'harmful' messages sent to teens / A second suit accuses bots of encouraging underage users to hurt themselves by discussing topics like self-harm.

By [Adi Robertson](#), a senior tech and policy editor focused on VR, online platforms, and free expression. Adi has covered video games, biohacking, and more for The Verge since 2011.

Dec 10, 2024 at 5:28 PM GMT+1



6

Comments (6 New)



Image: Cath Virginia / The Verge

Asks questions of governance, ethics and societal implications

- **Accountability:** Who should be accountable for what and when? — scientists, producers, deployers, end-users, or auditors?

Other newish concerns

Who should have the rights to
decide over training data?

ARTIFICIAL INTELLIGENCE / TECH / REGULATION

George R.R. Martin and other authors sue OpenAI for copyright infringement

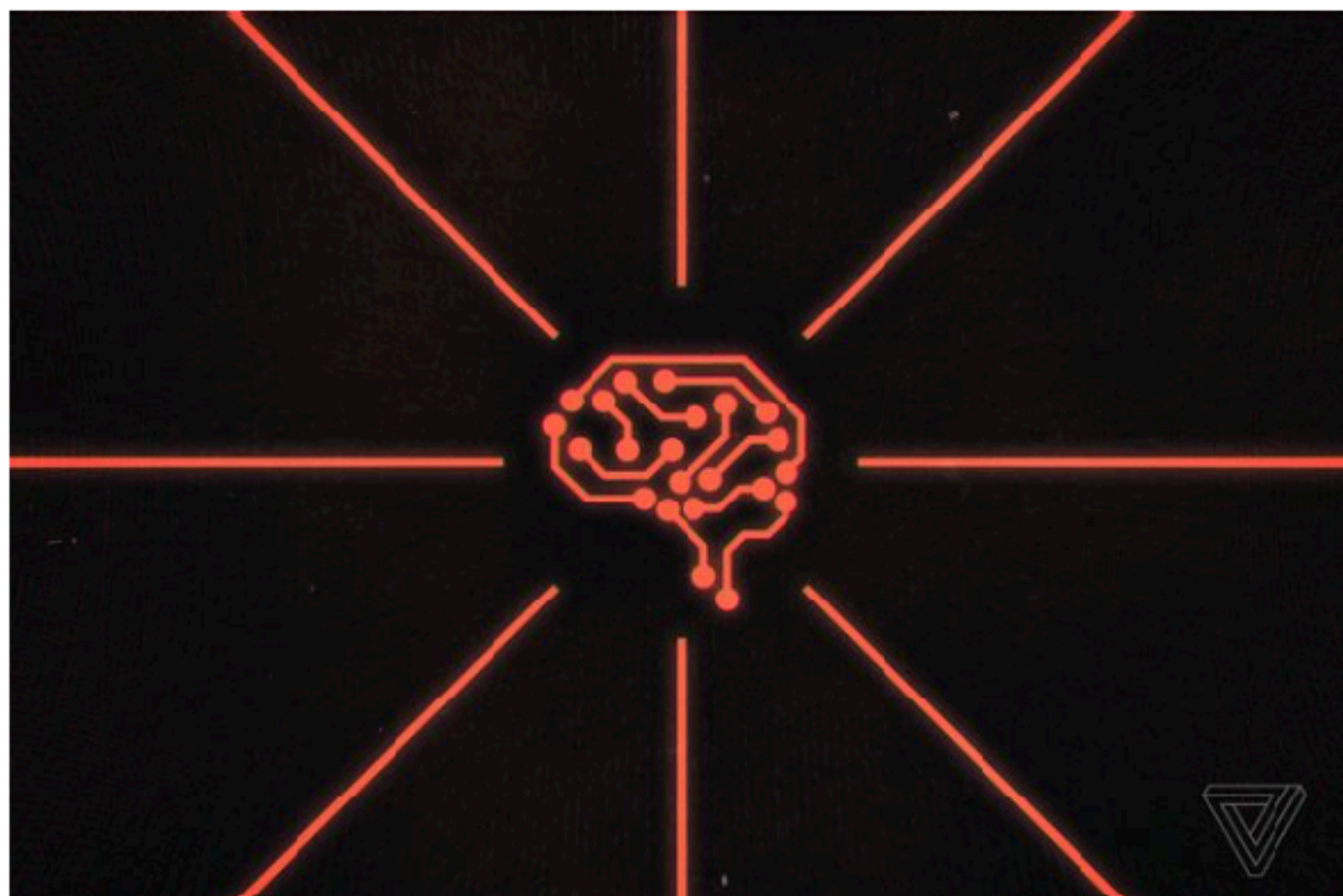


Illustration by Alex Castro / The Verge

/ Westeros won't bend the knee.

By [Emilia David](#), a reporter who covers AI. Prior to joining The Verge, she covered the intersection between technology, finance, and the economy.

Sep 20, 2023, 5:03 PM GMT+2 | [10 Comments](#) / [10 New](#)



Artificial intelligence (AI)

Kate Bush and Damon Albarn among 1,000 artists on silent AI protest album

Recordings of empty studios represent impact on musicians of UK's plans to let AI train on their work without permission

Dan Milmo *Global technology editor*

Tue 25 Feb 2025 01.01 CET

Individual control?

ARTIFICIAL INTELLIGENCE / TECH / APPS

Scarlett Johansson hits AI app with legal action for cloning her voice in an ad



Photo by Arturo Holmes/WireImage

/ An AI-generated version of Scarlett Johansson's voice appeared in an online ad without her consent.

By [Emma Roth](#), a news writer who covers the streaming wars, consumer tech, crypto, social media, and much more. Previously, she was a writer and editor at MUO.

Nov 1, 2023, 11:02 PM GMT+1 | [11 Comments](#) / [11 New](#)



If you buy something from a Verge link, Vox Media may earn a commission. [See our ethics statement.](#)

ardian

om €4 per month

The Guardian

n

Sport

Culture

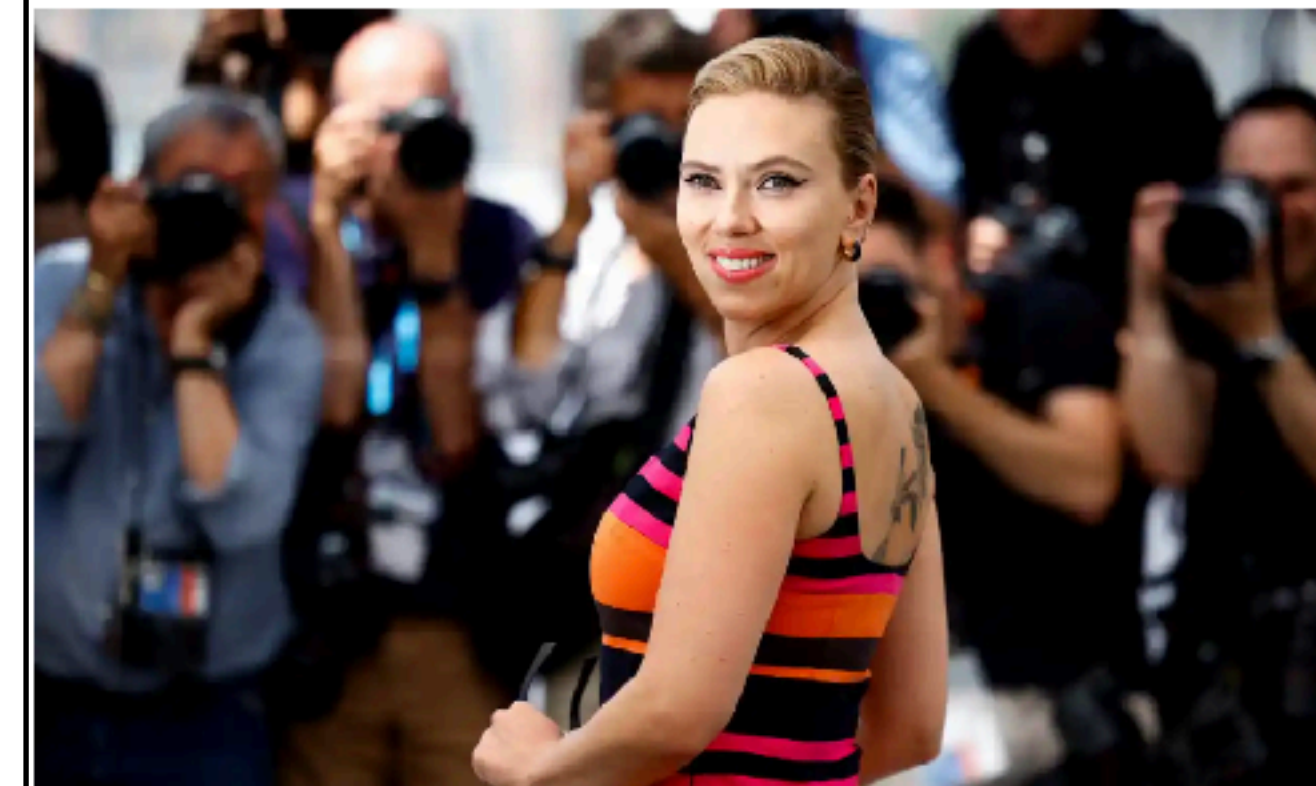
Lifestyle

More v

e Environment Science Global development Football Tech Business Obituaries


ChatGPT suspends Scarlett Johansson-like voice as actor speaks out against OpenAI

OpenAI says 'Sky' is not an imitation of actor's voice after users compare it to AI companion character in film Her



OpenAI's CEO, Sam Altman, seemed to suggest that the vocal design was intentionally mimicking Johansson's character, posting a one-word tweet after the presentation that simply said


Most viewed

 Papua New Guinea landslide death toll exceeds 670, says UN agency

 Charles Leclerc wins F1 Monaco GP after avoiding 'monster accident' - as it happened

 Leeds United 0-1 Southampton: Championship playoff final - as it happened

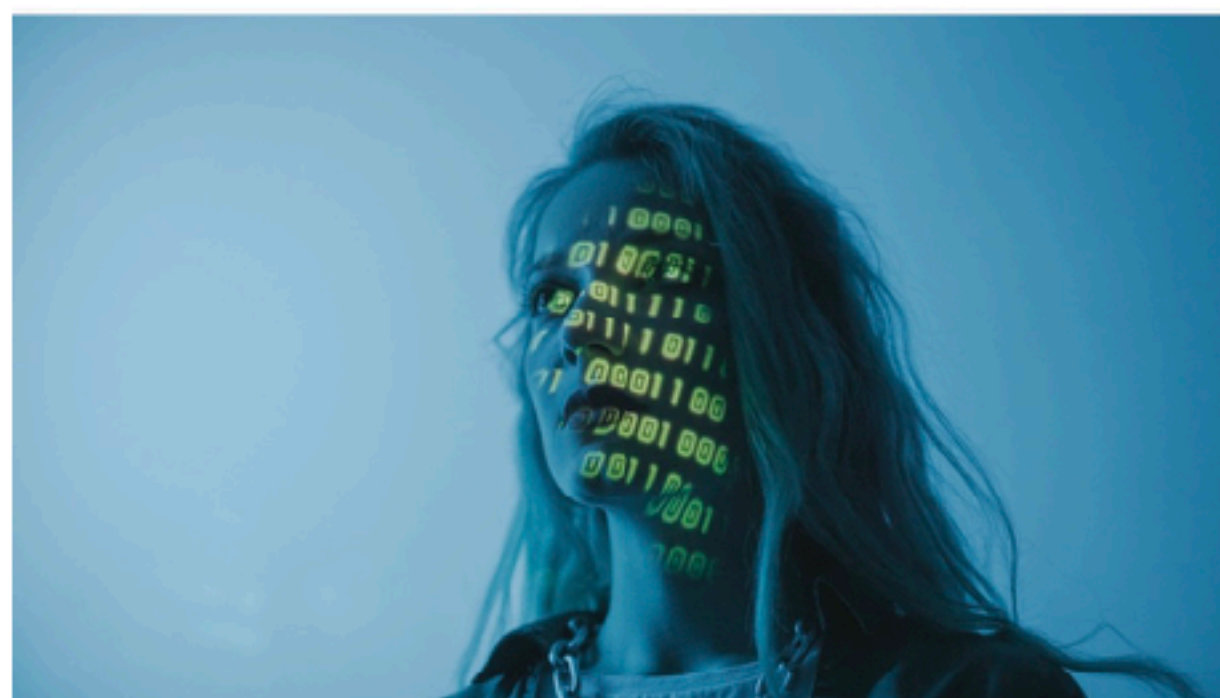
 Eight taken to hospital after turbulence on flight from Doha to Dublin

 The moment I knew: she said \$50 on my bed and a note that said 'buy yourself another bottle of wine'

Disinformation

SCI-TECH | News

Half of Canadians say they can't tell the difference between real and AI-generated content: survey



(cottonbro studio / pexels.com)



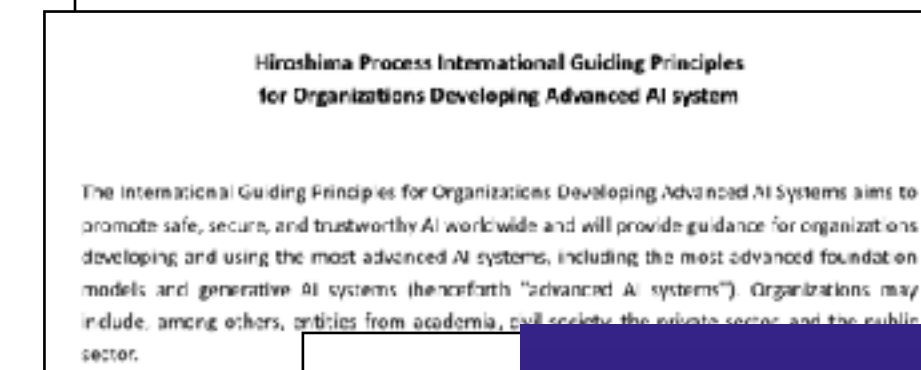
Alexandra Mae Jones
CTVNews.ca writer

 [Follow](#)  [Contact](#)

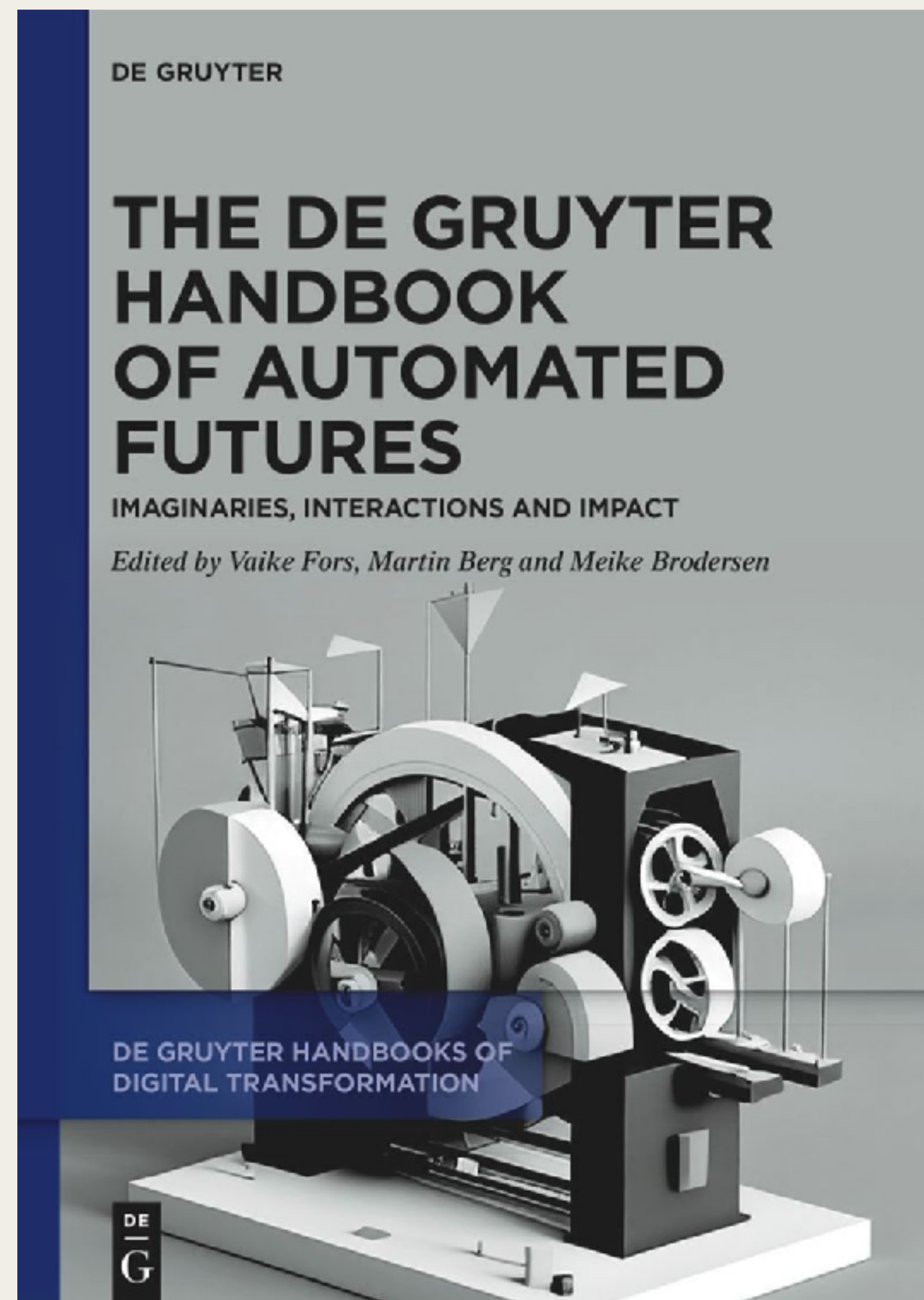
Published Oct. 24, 2023 2:50 p.m. CEST

A formative period

- **OECD**, risker: **desinformation**, reproduktion av bias, **upphovsrättsfrågor**, etc (**Sep -23**).
- **ACM Technology Policy Council**: Needs for safeguards, **human-in-the-loop**, **IP-rights**, data protection, “correctability” (**June, -23**).
- **Biden’s Presidentorder (31 Okt, 2023)**. Revoked by Trump **2025**.
- **G7**: 1) *guidelines* for “Hiroshima process” for advanced AI system and 2) a *code of conduct* for developing organisations (**30 Okt, 2023**).
- **The AI Act (augusti 2024)** “promote innovation AND safeguard fundamental rights”
- **Int. AI Safety Report (jan -25)** focus: GPAI risks



A sensitive case to address
normative nuances:
“necrorobotics”



Stefan Larsson

Chapter 8

Necrorobotics. The Ethics of Resurrecting the Dead



Figure 8.1: **BALL-E** prompt: “Dead person brought back to life in the shape of a humanoid robot, to comfort her grieving mother”.

Abstract: By drawing from recent progress in AI, this chapter scrutinises implications of a specific imaginary of automated futures: the possible resurrection of the dead. *Necrorobotics* is proposed as a field of critical studies on the use of data and design based on one specific dead individual in order to ‘resurrect’ that individual. That is, to mimic or create some level of robotic agency for the sake of mourning, remembrance, or handling of loss. The technological advancements of relevance are here referred to as *resurrection technologies*—that is, methods for the training of AI-models based on data from a specific individual, such as imagery, text and voice—here addressed for the analysis of connected ethical and normative questions.

By drawing from theoretical discourses on mortuary cultures, post-mortem conditions in digital times as well as robotic uncanniness, this chapter uses three reported

Open Access. © 2024 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. <https://doi.org/10.1515/9783110792256-008>

Larsson (2024): “Necrorobotics — The Ethics of Resurrecting the Dead”

Alexa will soon be able to read stories as your dead grandma

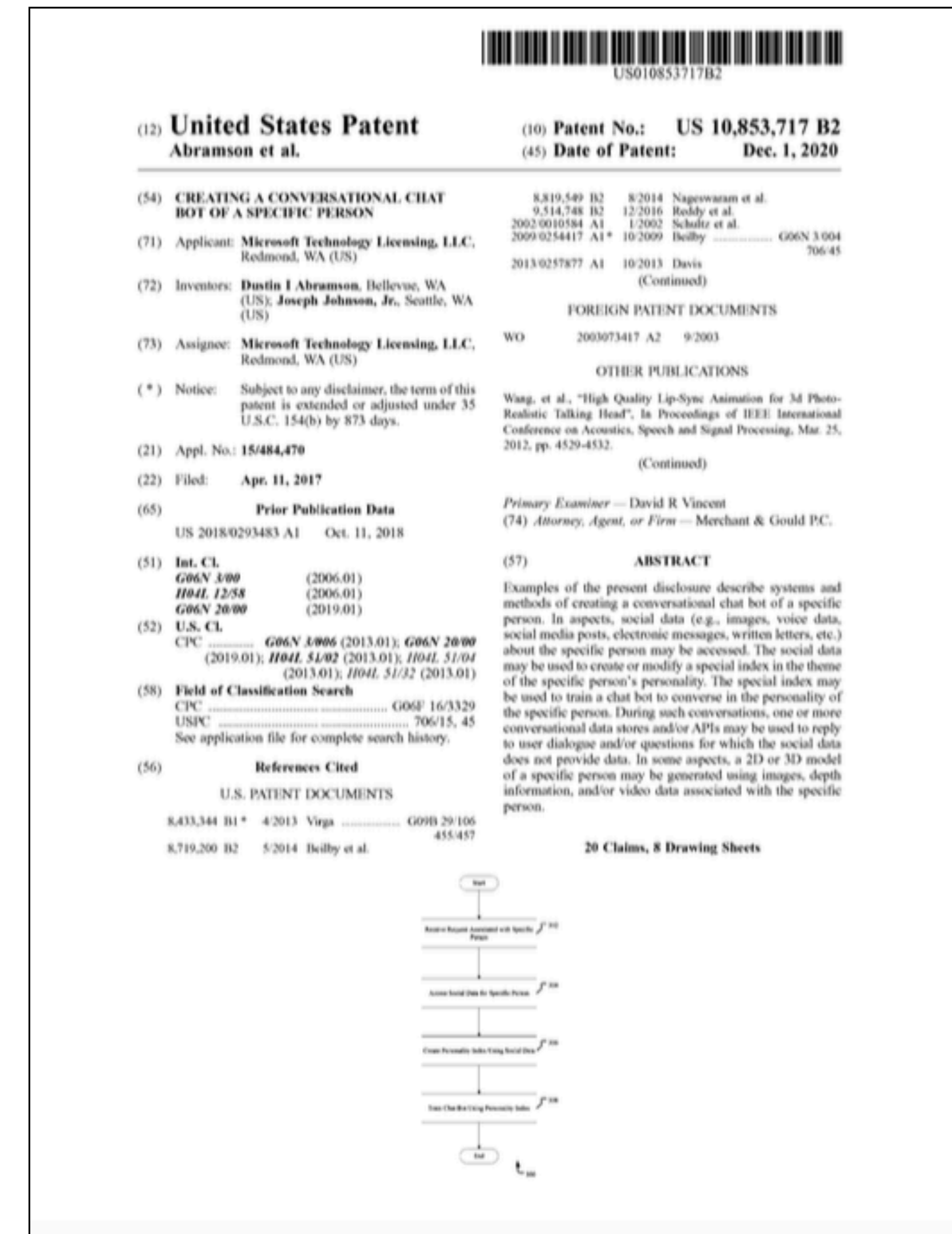
Brian Heater @bheater / 7:14 PM GMT+2 • June 22, 2022

 Comment



Patent and generative AI

- **Patent** for chatbot based on an individual
- **Generative AI-development**
 - GPT-3/-4/ChatGPT etc
 - DALLE2/3; Midjourney; Stable Diffusion etc
 - VALLE etc



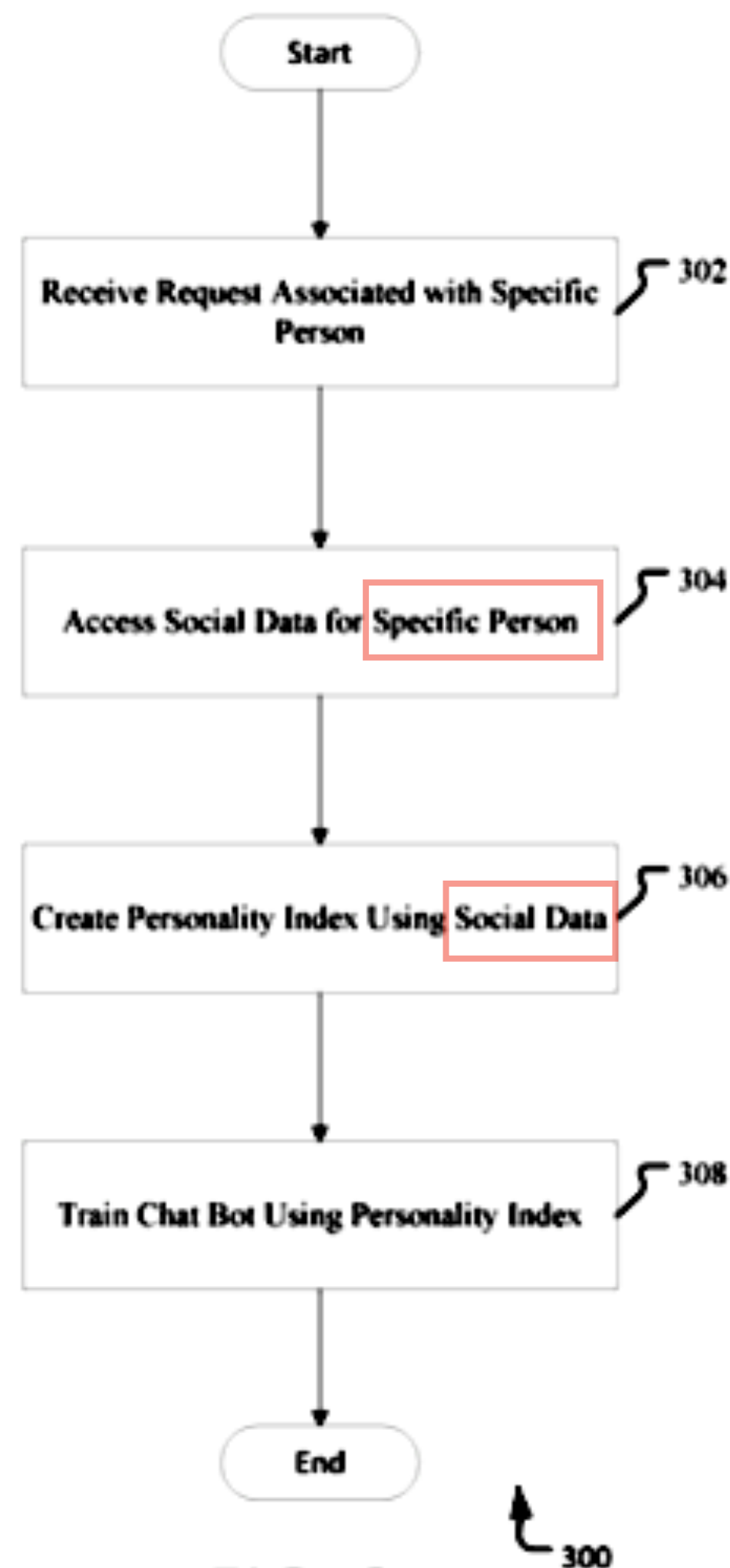


FIG. 3

“The present disclosure provides systems and methods of creating a conversational chat bot of a specific person...”

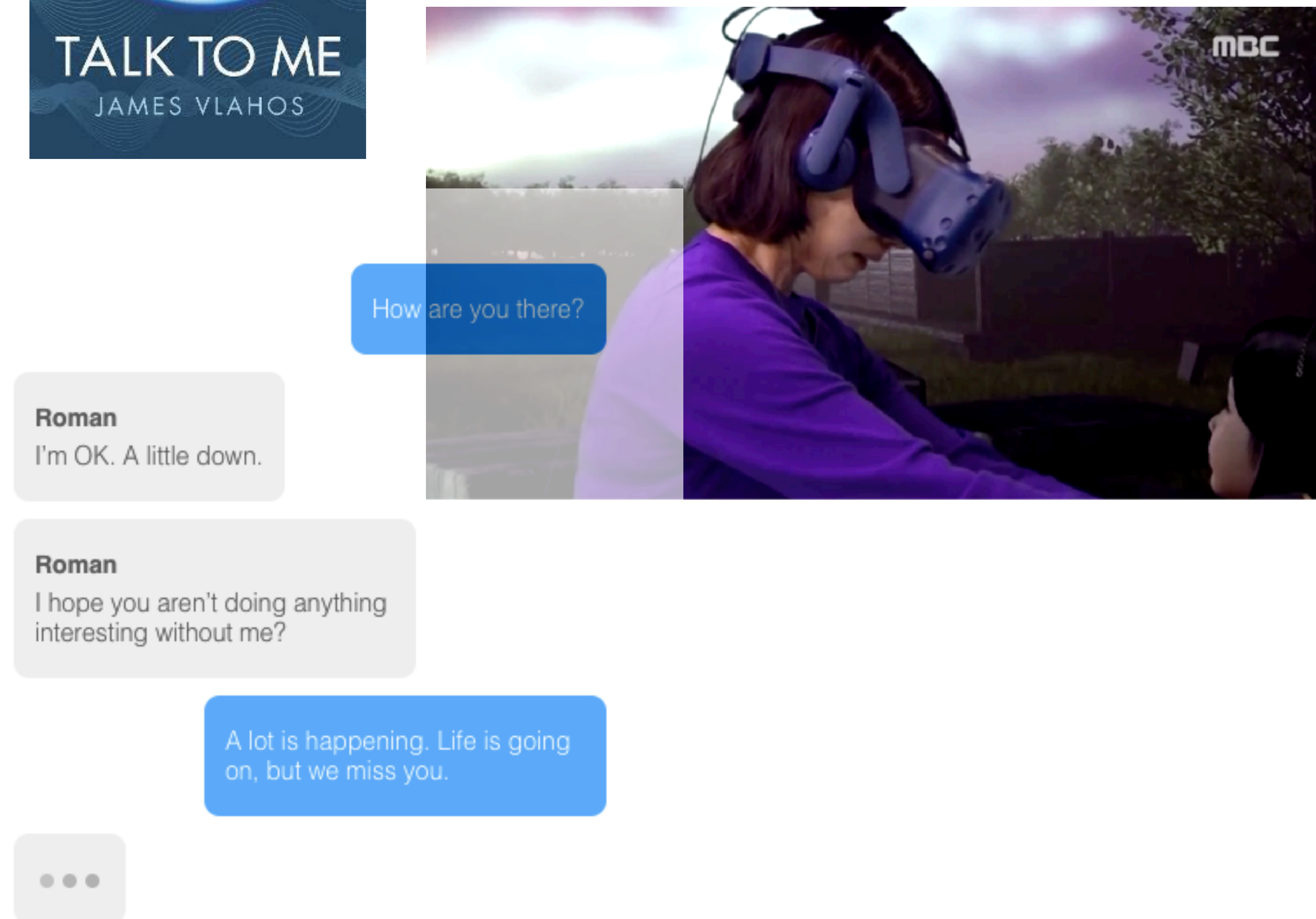
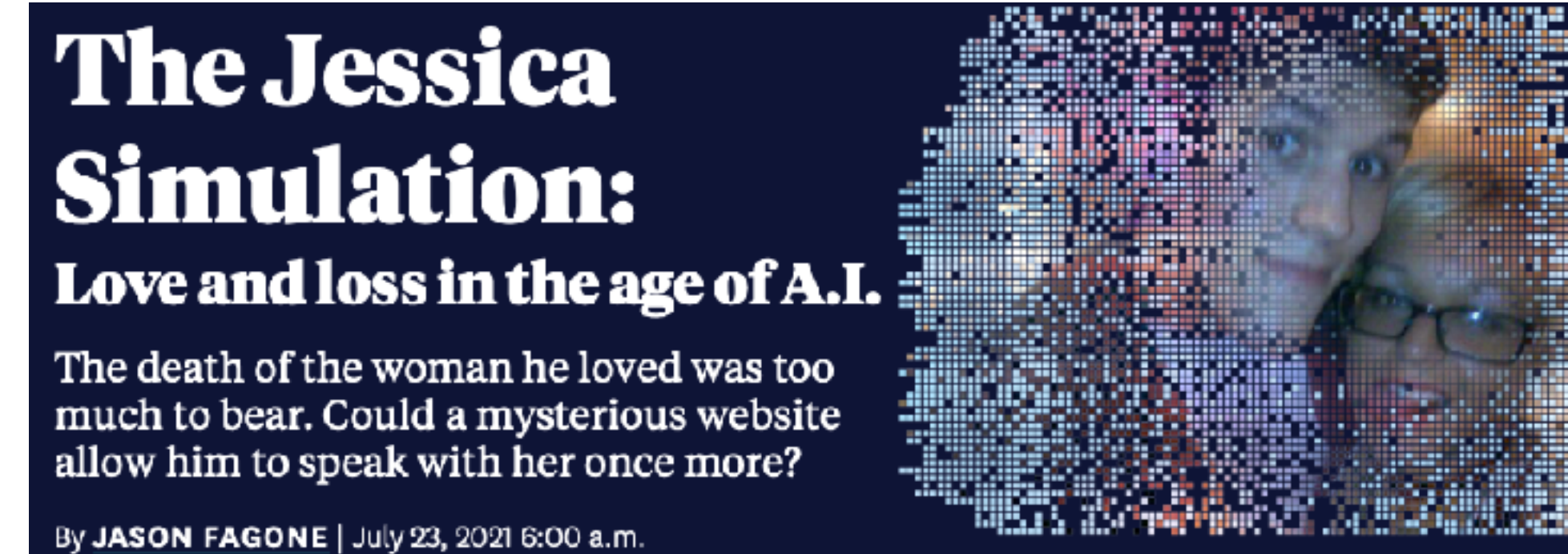
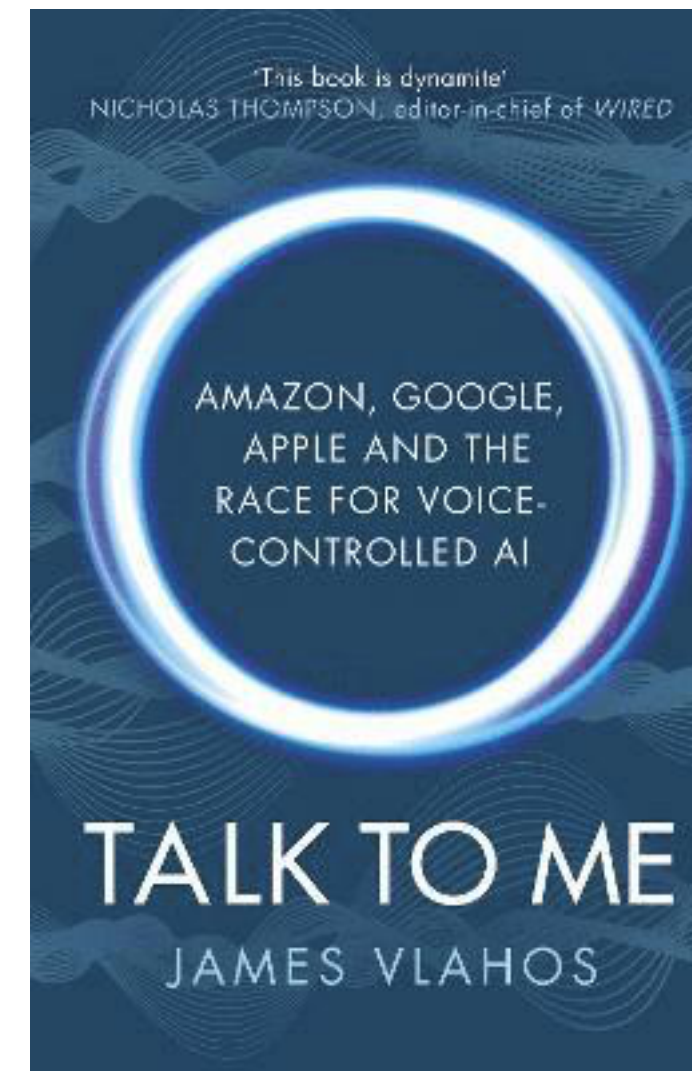
“...the specific person may correspond to a **past or present** entity (or a version thereof), such as a friend, a relative, an acquaintance...”

“Social data may be stored by, and/or collected from, various sources...”

“...it may refer to images, image data, voice data, emails, text messages, dialogue data/commands, social media posts, written letters, user profile information, behavioral data, transactional data, geolocation data, and other forms of data about a specific person.”

Cases

- In 2016, chat bot developer Eugenia Kuyda released **a chatbot based on her deceased friend, Roman Mazurenko**.
- The “DadBot”, James Vlahos (2019)
- 2020: Canadian man used a chatbot service powered by GPT-3 to create **a replica of his deceased girlfriend**.
- In 2020, a South Korean **mother met her deceased 7-yo in VR**, produced by a Seoul-based studio.



Markets



Deepbrain AI offers “re;memory”: “more than just a place for remembrance”; “directly interact with lost ones”

Institutions?

OpinionSportCultureLifestyle

The Guardian

USAmericasAsiaAustraliaMiddle EastAfricaInequalityGlobal development

AI of dead Arizona road rage victim addresses killer in court

Clip of Chris Pelkey, who died in 2021, says: 'I believe in forgiveness' after his sister fed an AI model videos of him

▶ 2:25

Puzzles



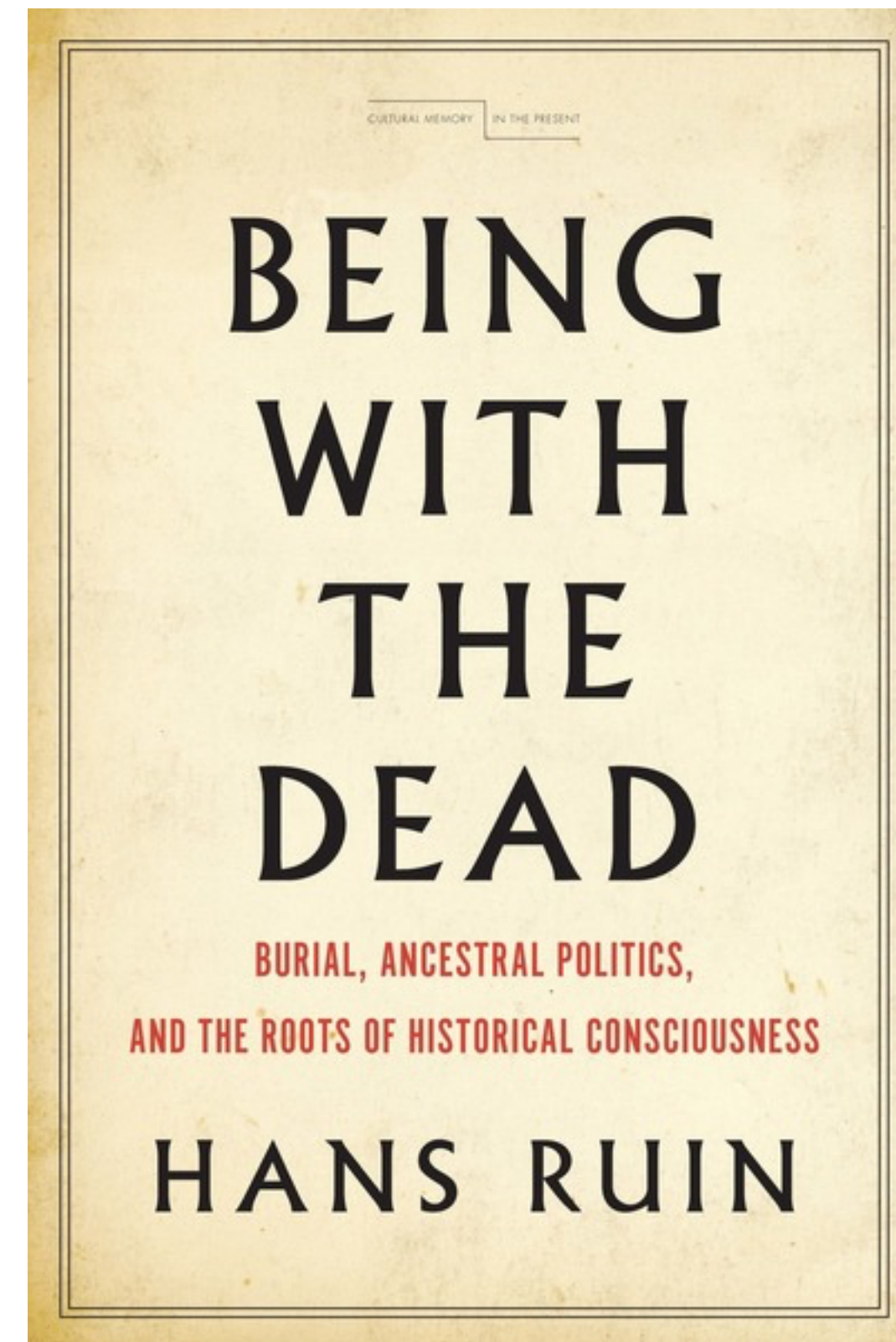
AI version of dead Arizona road rage victim addresses killer in court - video

Chris Pelkey **was killed** in a road rage shooting in Chandler, Arizona, in 2021.

Three and a half years later, Pelkey **appeared in an Arizona court** to

Old norms, new practices?

- **Both new** — in terms of AI/robotic capabilities
- **...and old** — in terms of mortuary cultures of “being with the dead” (Ruin, 2019).
- Grief, remembrance: The dead have some type agency in most cultures
- Can change over time, culture, society (cf. O’Neill, 1999)



Which norms should be considered?

1. **Require robustness:** “Infrastructural responsibilities”?
2. **Who** ought to decide over “resurrection”?
 - a) Patented rights?
 - b) Possession? Those that have your data
 - c) Heritage? family (who, specifically?)
3. **Analogy — respect and dignity:** “Burial peace” (Geneva convention etc.)?

Stefan Larsson
Chapter 8
Necrorobotics. The Ethics of Resurrecting the Dead



Figure 8.1: ~~Battle~~ prompt: “Dead person brought back to life in the shape of a humanoid robot, to comfort her grieving mother”.

Abstract: By drawing from recent progress in AI, this chapter scrutinises implications of a specific imaginary of automated futures: the possible resurrection of the dead. *Necrorobotics* is proposed as a field of critical studies on the use of data and design based on one specific dead individual in order to ‘resurrect’ that individual. That is, to mimic or create some level of robotic agency for the sake of mourning, remembrance, or handling of loss. The technological advancements of relevance are here referred to as *resurrection technologies*—that is, methods for the training of AI-models based on data from a specific individual, such as imagery, text and voice—here addressed for the analysis of connected ethical and normative questions.

By drawing from theoretical discourses on mortuary cultures, post-mortem conditions in digital times as well as robotic uncanniness, this chapter uses three reported

© Open Access. © 2021 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. <https://doi.org/10.1515/9783110792256-008>

Today: Key topics

1. **Transparency is multifaceted:** but include literacies/ audiences, and that mediation guides/structures/manipulates.
2. **Formative period for law:** flexible, somewhat unpredictable law-making
3. **Adaptive tech and norms:** reproducing social structures is not neutral
4. **Context / case reveals:** Sensitive cases need nuanced scrutiny





Thank you!

Stefan Larsson

Associate Professor in Technology and
Social Change,

Lawyer (LLM)
PhD in Sociology of Law
PhD in Spatial Planning

Department of Technology and Society,
Faculty of Engineering, Lund University,
Sweden.

**Focus Period Symposium: Visualization-
Empowered Human-in-the-Loop
Artificial Intelligence, 13-15 May, 2025**



LUND
UNIVERSITY