

Trustworthiness of AI: From ethical concerns to democratic interests

Dr. Eugenia Stamboliev

(Universität Wien & Prague University of Economics and Business)

Women in AI Austria board member

Eugenia.stamboliev@univie.ac.at



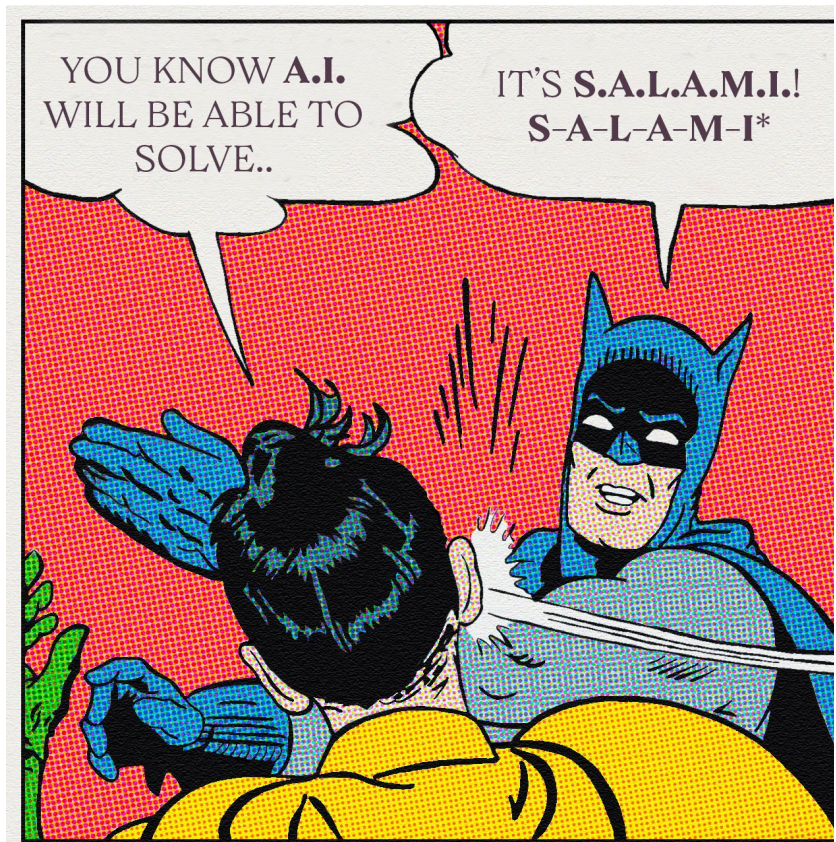
From narrative to ethical norms



What is the difference between Artificial Intelligence and S.A.L.A.M.I.?

... the expectations and storyline behind the label?!

<https://blog.quintarelli.it/2019/11/lets-forget-the-term-ai-lets-call-them-systematic-approaches-to-learning-algorithms-and-machine-inferences-salami/>



***S.A.L.A.M.I.**

**SYSTEMATIC APPROACHES TO LEARNING
ALGORITHMS AND MACHINE INFERENCE**

by *Stefano Quintarelli*

Cate Crawford & Vladan Joler
(2018)



Estampa is a collective of programmers, filmmakers and researchers working in the fields of audiovisual media and digital environments.

AI Ethics?

Responsibility

Agency

Norms

Autonomy

Authorship

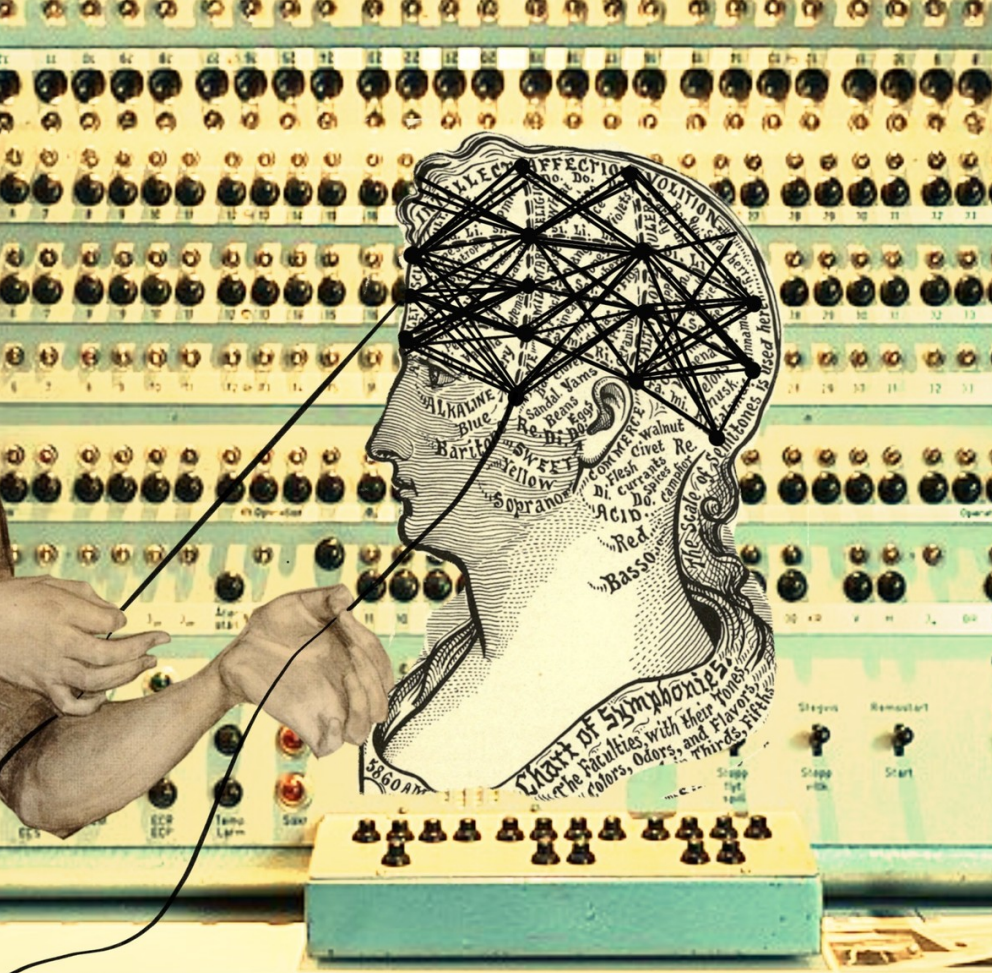
Consciousness

Morality

Discrimination (Bias, inequality)

Labour

Environment (!)





Autonomy

Autonomy to do what? Do? Think? Work? Vote?

Men versus machine

Active versus passive (Illusion? Or human right?)

Free versus automatic (Convenience)

Democracy versus autocracy

Doing (Techno & humans) **and/or** Thinking (only humans)

AI “agents” systems

Sovereignty (technical, individual, societal, intellectual)

Responsibility

How do we know what for?

Responsible: who versus what?

How do we know?

Accountability $\neq$ responsibility

“Many hands” & “many things” theory

Transparency versus opacity?





Environment & Climate

ENERGY. ENERGY. ENERGY.

Resources and water (affecting communities and climate)

Sustainable AI **versus** sustainability of AI (A. v. Wynsberghe) What is the price for running more AI

AI shapes relationship between humans and/as animals and the technical world

*What is the position of AI in “Umwelt”? (Technology versus Human or **versus** Human **as** Environment)*



Discrimination (Gender)

Bias $\neq$ discrimination.

Discrimination of someone based on their gender is a democratic issue.

Problem is structural (not about “ONE IT guy not liking women”)

Difficult: Who is responsible and accountable?
What or who is affected? Mostly, non-white male.

Where?

- Medical data, Job descriptions, Image generators (GenAI)

NEW: Gendered discrimination leading to online erasure of women – **Synthetic discrimination**” (Stamboliev, 2025, Mehr Bilder, weniger Präsenz”, UNESCO report)

[illegible]

Autonomy to think independently versus learning to use a tech/company product

Autonomy / sovereignty / responsibility
of teachers? Schools?

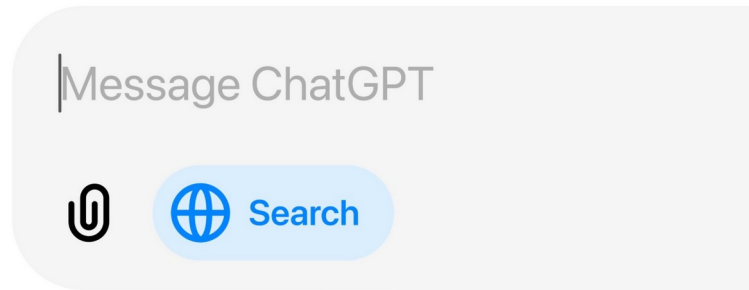
Time saving? (quicker Admin?)



Trust in AI? Yes, no, maybe.

Are hallucinations the most pressing trust issue with AI?

For many, yes. Truth? Facts?
Accurate?



Trustworthy AI? Somewhat of a research overview

Trustworthiness of AI is a question of...

- Is AI robust, ethical and legal? (EGTAI guidelines, 2019)
- Is AI transparent? (OECD)
- Is AI explainable?
- Is AI reliable in terms of output accuracy?
- Is AI performing legally dependable?



TAI research goals

- Increasingly tech-oriented
- Debated as design or expectation matter
- Focusing on the output or results of AI
- Trustworthiness gets increasingly technified (not wrong, just limited)

Trend: Trust in “factuality”

Trust of AI is often grounded in the “factual accuracy of statements” that AI produces (Buchanan & Hickmann, 2024)

- Factual accuracy is viewed semantically or algorithmically?

Issue: Depoliticised and unclear metrics (not like a relativist view)

- Trustworthiness = epistemic credibility of information
- Metric or benchmark for accuracy?

Danger: Algorithmic and predictive “accuracy” is favoured over contextual and human interpretation

Trust is political and ideological, long before AI

Trust in media institutions is always contested: Not new, not linked to AI!

Fenton, N. “Democratic Delusions: How the Media Hollows Out Democracy and What We Can Do About It” (2024)

- representation, corporate interests and power are important factors
- informational value is political, not static
- Distrust is a reaction to disempowerment, not informational disagreement
 - perceived/actual disenfranchisement

Democratic trustworthiness of AI?

From knowledge of the user to protection of the citizen

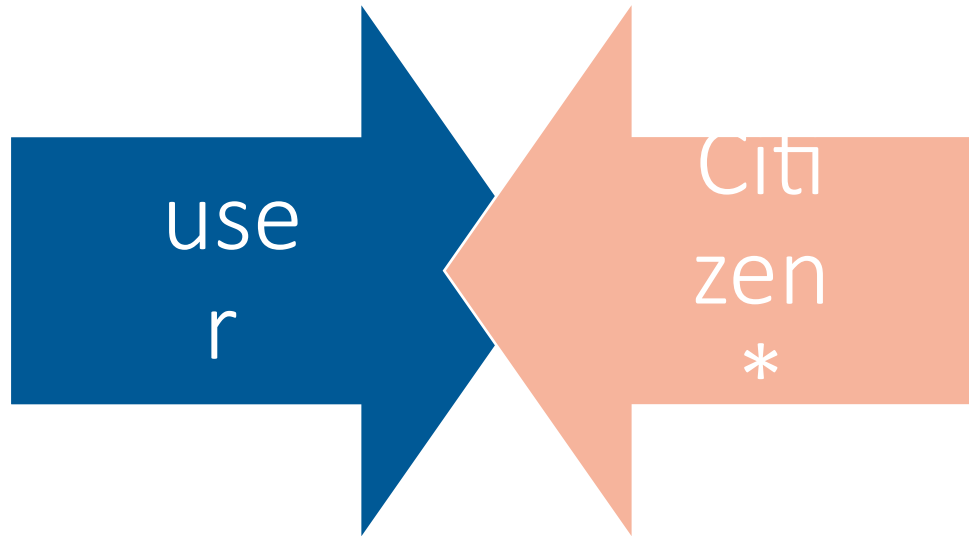
Democratic TAI?

Looking at AI development, implementation and use through the angle of

Citizen interests

Not user-centric perspectives of trust in AI, but from view of rights and societal value that AI has holistically.

Democratically speaking, these two figures have different interests to protect)



Trustworthiness as “encapsulating interests”

Interests? (Rights)

- “I trust you because your interest **encapsulates mine**. You have assessable incentives to protect my interests as if they are yours. This enables our future relationship.” (Russell Hardin (2012, p. 7)

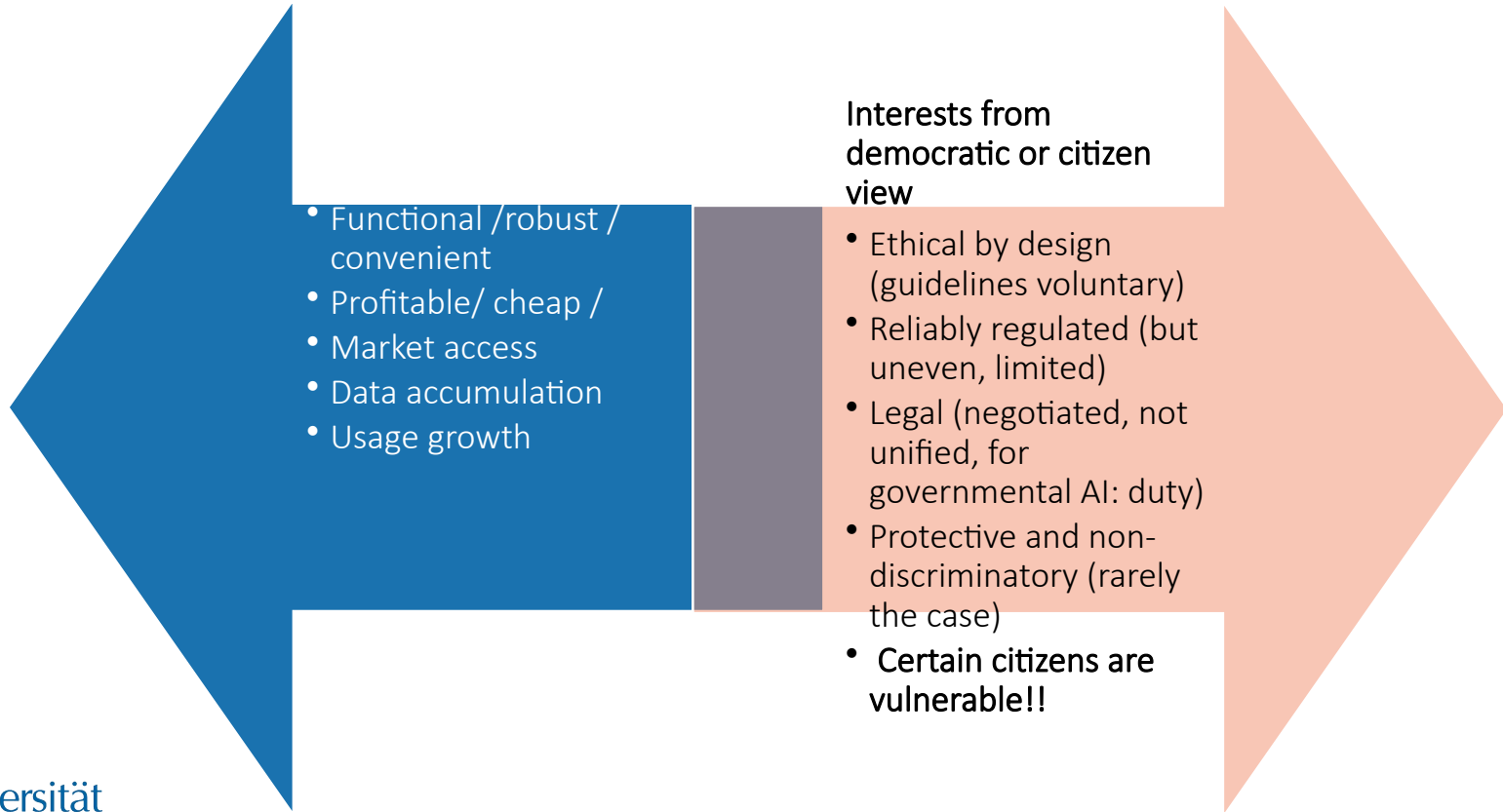
New Questions:

- Is AI made and used in a way that protects my interests as citizen?
- Do the developers/companies/regulators encapsulate my interests in the making of AI?

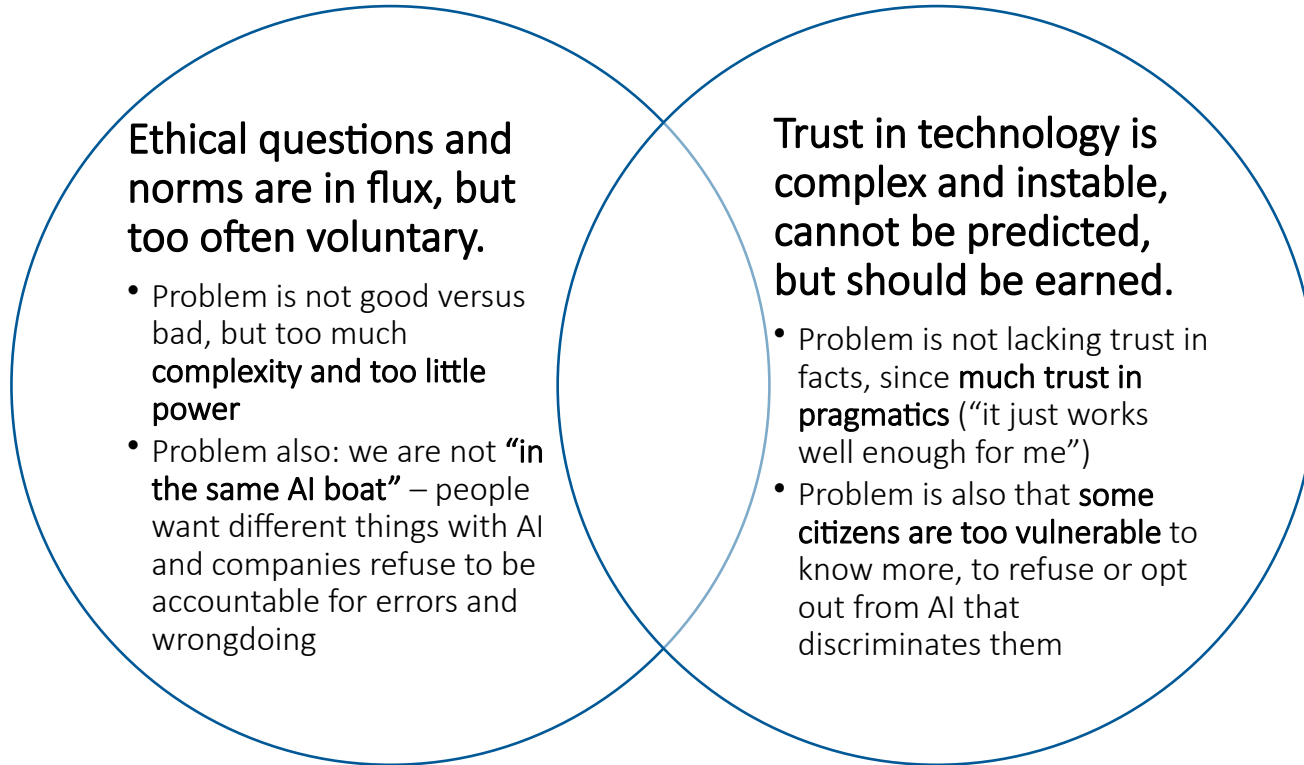
Specific rights / interests:

- Non-discriminatory / non-exploitative AI
- AI that enables free access to free press, impartial juridical system etc.

How could user and citizen differ within AI?



To conclude...





universität
wien

Thank you.
Questions?

Eugenia Stamboliev

Eugenia.stamboliev@univie.ac.at