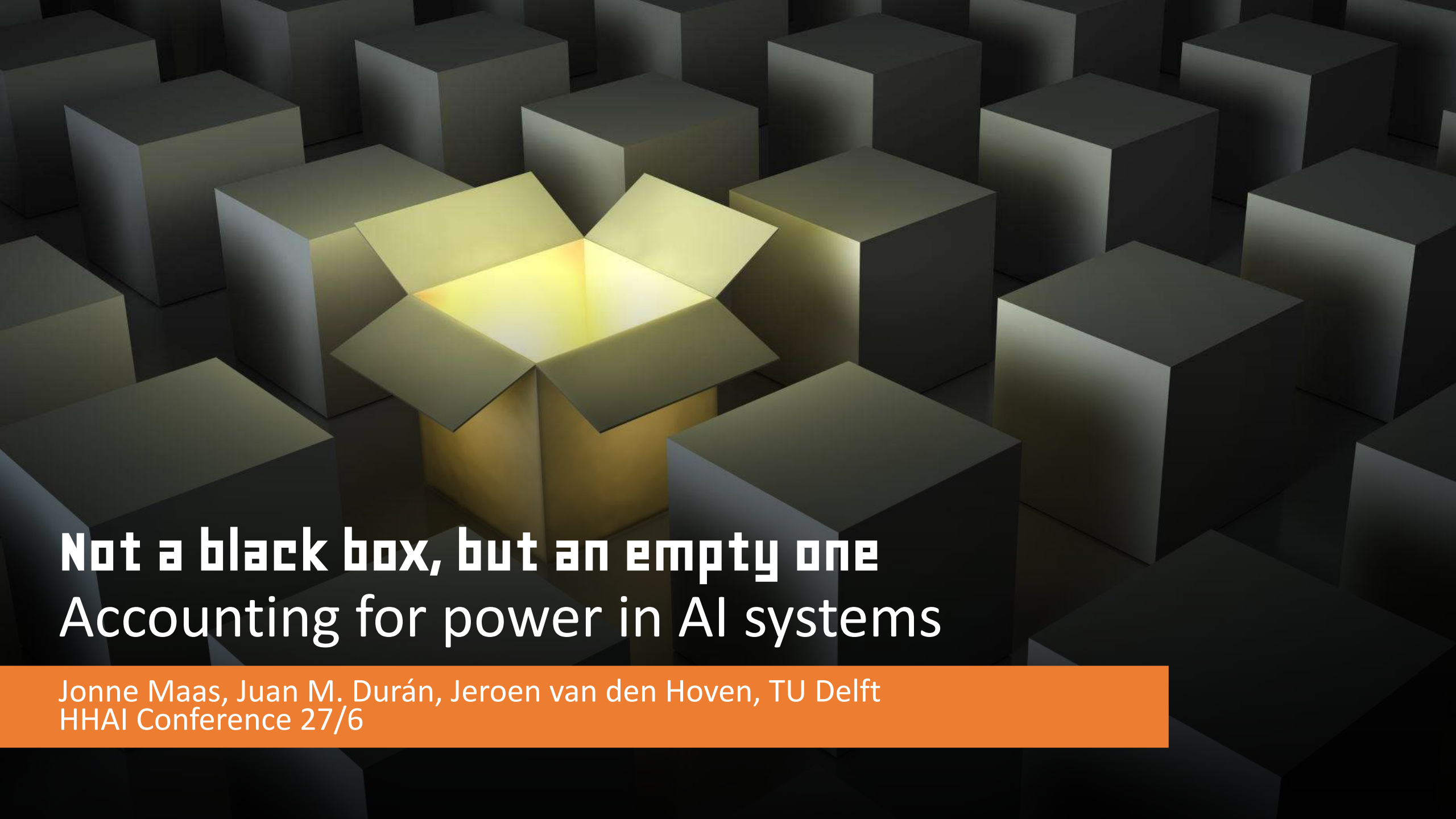


Not a black box, but an empty one
Accounting for power in AI systems

Jonne Maas, Juan M. Durán, Jeroen van den Hoven, TU Delft
HHAI Conference 27/6

Motivation: AI Ethics vs. AI Politics

ETHICS GUIDELINES FOR TRUSTWORTHY AI

High-Level Expert Group on Artificial Intelligence

nature machine intelligence

Explore content · About the journal · Publish with us

nature > nature machine intelligence > perspectives > article

Perspective | Published: 02 September 2019

The global landscape of AI ethics guidelines

RESEARCH-ARTICLE

Gender and Racial Bias in Visual Question Answering Datasets

Authors: Yusuke Hirota, Yuta Nakashima, Noa Garcia Authors Info & Claims

FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency • June 2022 • Pages 1280–1292 • https://doi.org/10.1145/3531146.3533184

XAI—Explainable artificial intelligence

DAVID GUNNING, MARK STEFIK, JAESIK CHOI, TIMOTHY MILLER, SIMONE STUMPF, AND GUANG-ZHONG Y

SCIENCE ROBOTICS · 18 Dec 2019 · Vol 4, Issue 37 · DOI: 10.1126/scirobotics.aay7120

Hard choices in artificial intelligence

Roel Dobbe [a], Thomas Krendl Gilbert [b], Yonatan Mintz [c 1]

Research Article | Open Access | Published: 24 May 2017

Algorithmic Accountability and Public Reason

Stanford University Human-Centered Artificial Intelligence

Reuben Binns

Philosophy & Technology **31**, 543–556 (2018) | Cite this article

About · Centers · Research · Education · Policy · News · Events

Tanner Lecture: AI and Human

Seth Lazar

KATE CRAWFORD

ATLAS OF AI

Plan-de-campagne

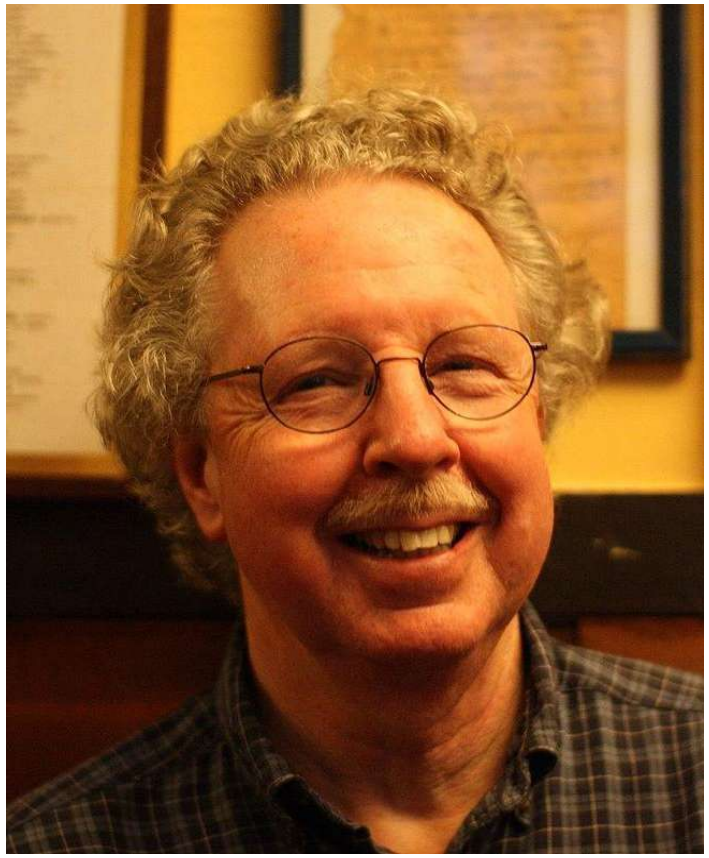
- What is AI Politics?
- Why AI Politics?
- Current limitations
- Ways forward

Not a Black Box, but an Empty One

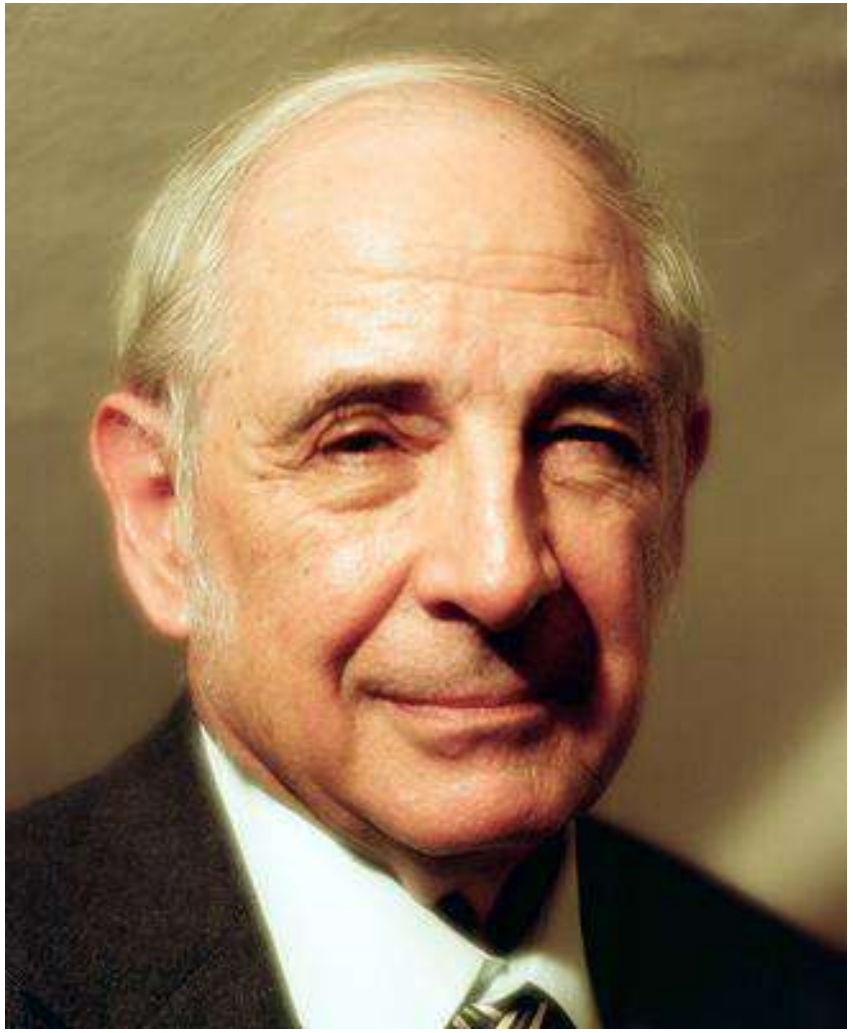
(cf. Winner)

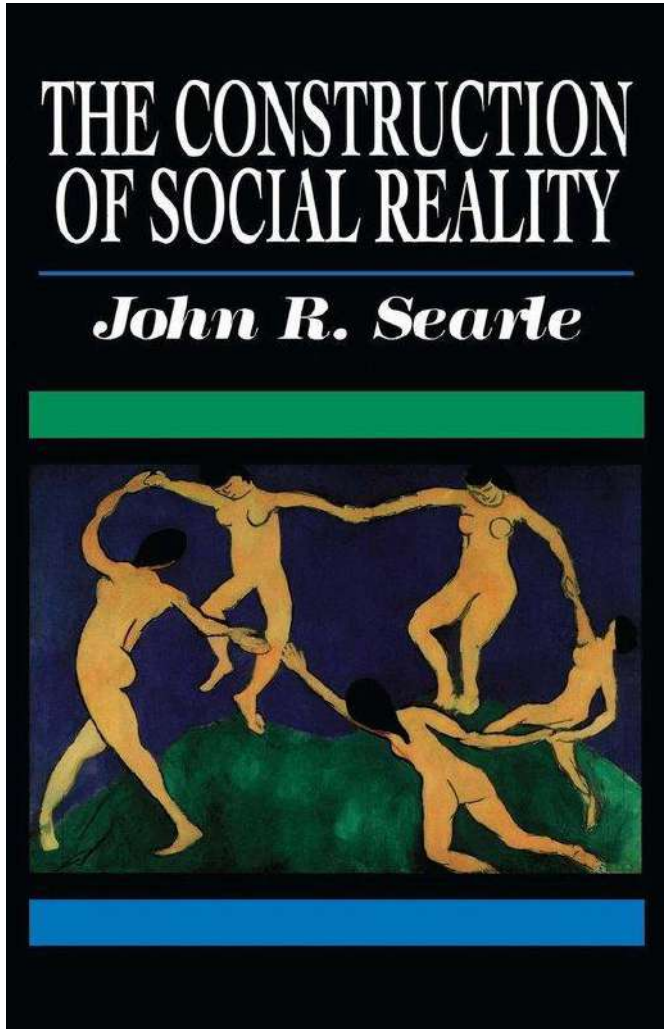
- How the system behaves vs. what's the system's 'right to be around'?

- Why important: system has some kind of epistemic and moral authority/status

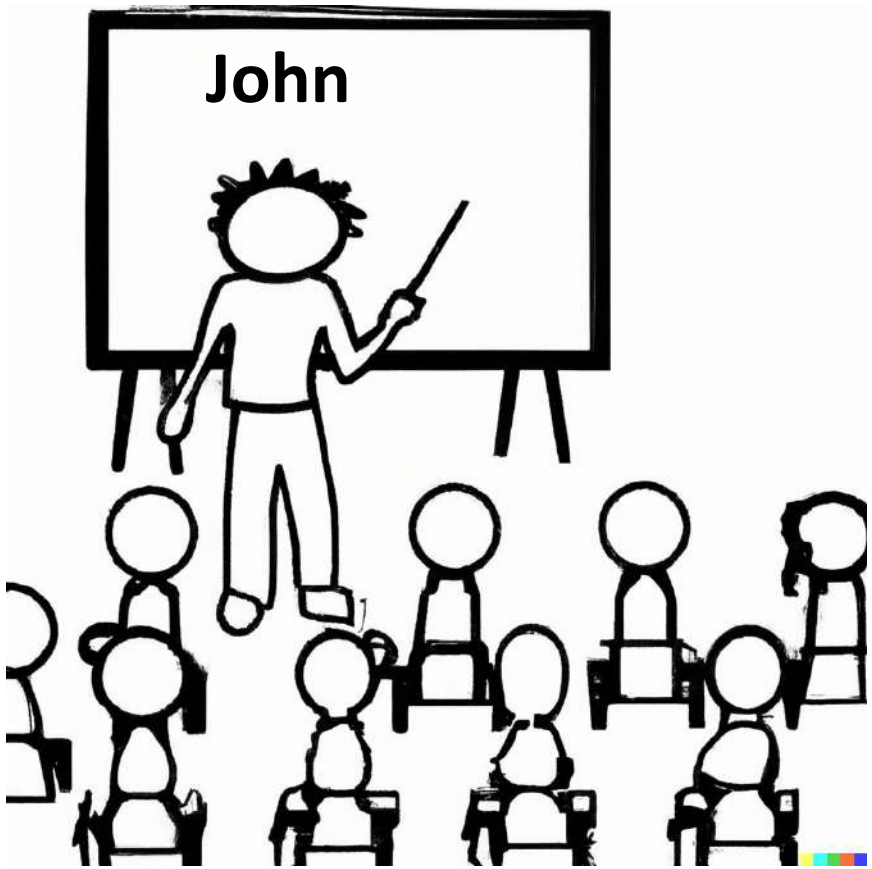


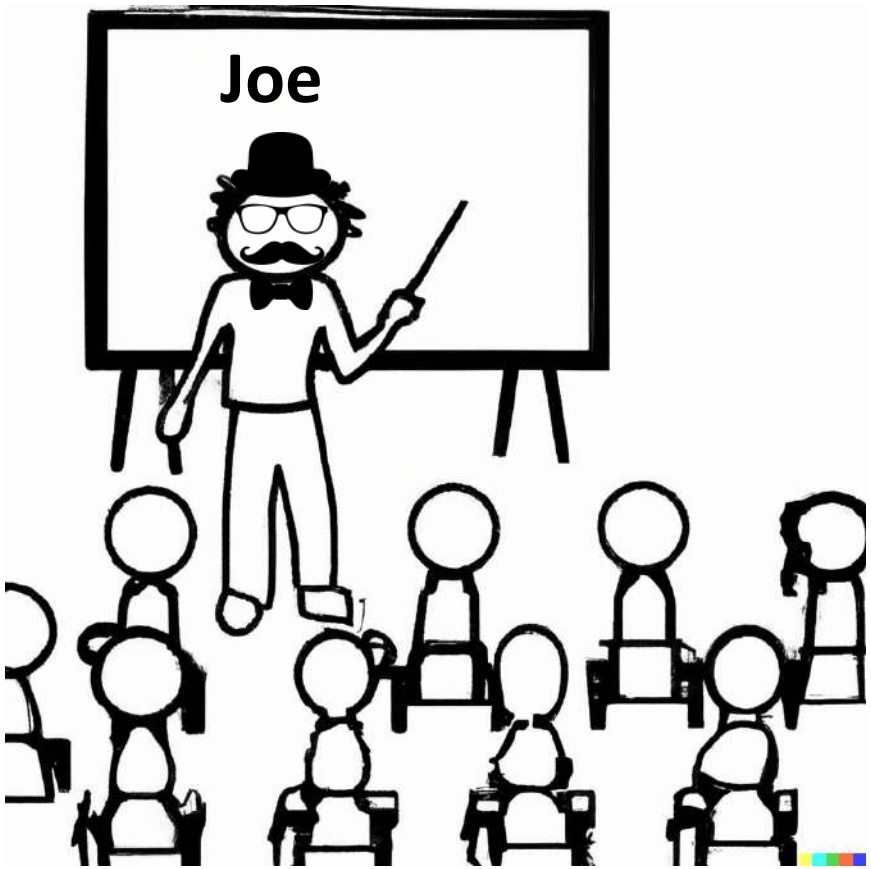
Deontic Power





Teacher John vs. non-teacher Joe





Why the box should be full

- Value-laden design decisions
- ‘fossilization’ of design decisions

Why the box is now empty

- Focus on AI Ethics
- No/unclear regulation (so far)

How the box can be non-empty

- Formal standards for designers?
- Better ethical education for AI developers?
- … (happy to hear your thoughts!)

Thank you!

Happy to hear your thoughts/comments/etc.:

J.j.c.maas@tudelft.nl

AI’s Deontic Power

How does the analogy relate?

How the teacher got its deontic power vs. how AI systems got its deontic power.

In what way do we attribute deontic power to AI?

Epistemic authority: generally perceive these systems are more objective. The fact we include them in our decision-making process necessarily indicates we attribute some kind of epistemic authority to them.

Backwards reasoning: the fact that we have these guidelines assumes we have expectations of how these systems should behave. And so these systems arguably have some obligations and duties, broadly speaking.

In what way is that deontic power now normatively empty?

Developers & deployers just threw these systems in society, and now we’ve reached a point of no return. (point where perhaps the analogy breaks down?

Why is it a problem this deontic power is now normatively empty?

Design choices are value-laden

Not a Black Box, but an Empty One

Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology

Langdon Winner
Rensselaer Polytechnic Institute

What do philosophers need to know about technology? What kind of knowledge do we need to have? And how much? Perhaps it is enough simply to have lived in a society in which a wide variety of technologies are in common use. Drawing upon an everyday understanding of such matters, one can move on to develop general perspectives and theories that may enable us to answer important questions about technology in general. The problem is that one's grasp may be superficial, failing to do justice to the phenomena one wants to explain and interpret. One may seize upon a limited range of vaguely understood examples of technical applications—a dam on a river, a robot in a factory, or some other typification—and try to wring universal implications from a sample that is perhaps too small to carry the weight placed upon it.

An alternative would be to focus one's attention more carefully, becoming expert in the technical knowledge of a specific field, attaining the deeper understanding of, say, a worker, engineer, or technical professional. Even that may prove limiting, however, because the experience available in one field of practice may not be useful in comprehending the origins, character, and consequences of technical practices in other domains. The sheer multiplicity of technologies in modern society poses serious difficulties for anyone who seeks an overarching grasp of human experience in a technological society.

Yet another strategy might be to study particular varieties of technology in a scholarly mode, drawing upon existing histories and contemporary social studies of technological change as one's base of understanding. And one

---

AUTHOR'S NOTE: This article is a shortened version of the presidential address delivered to the Biennial Conference of the Society for Philosophy and Technology, Mayaguez, Puerto Rico, March 1991.

Science, Technology, & Human Values, Vol. 18 No. 3, Summer 1993 362-378