AI4EO

Special Topics in Applications (AIL861)
Artificial Intelligence for Earth Observation
Lecture 28

Instructor: Sudipan Saha

Opportunity or Risk

“We Teach A.I. Systems Everything, Including Our Biases“
– The New York Times (Nov 2019)

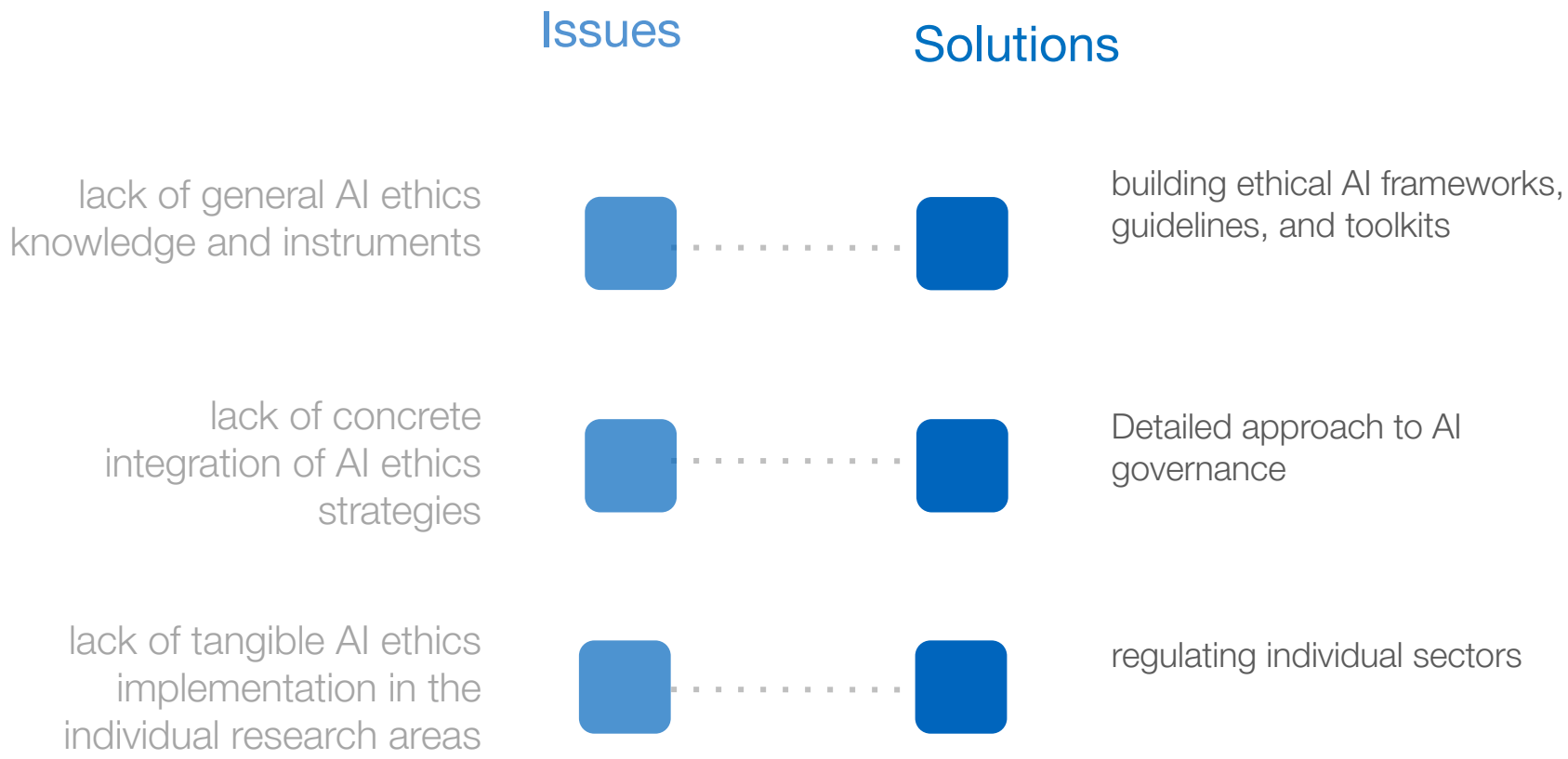
“This is the Stanford vaccine algorithm that left out frontline doctors“
– MIT Technology Review (Dec 2020)

“When Self-Driving Cars Can’t Help Themselves, Who Takes the Wheel?”
– The New York Times (Mar 2018)

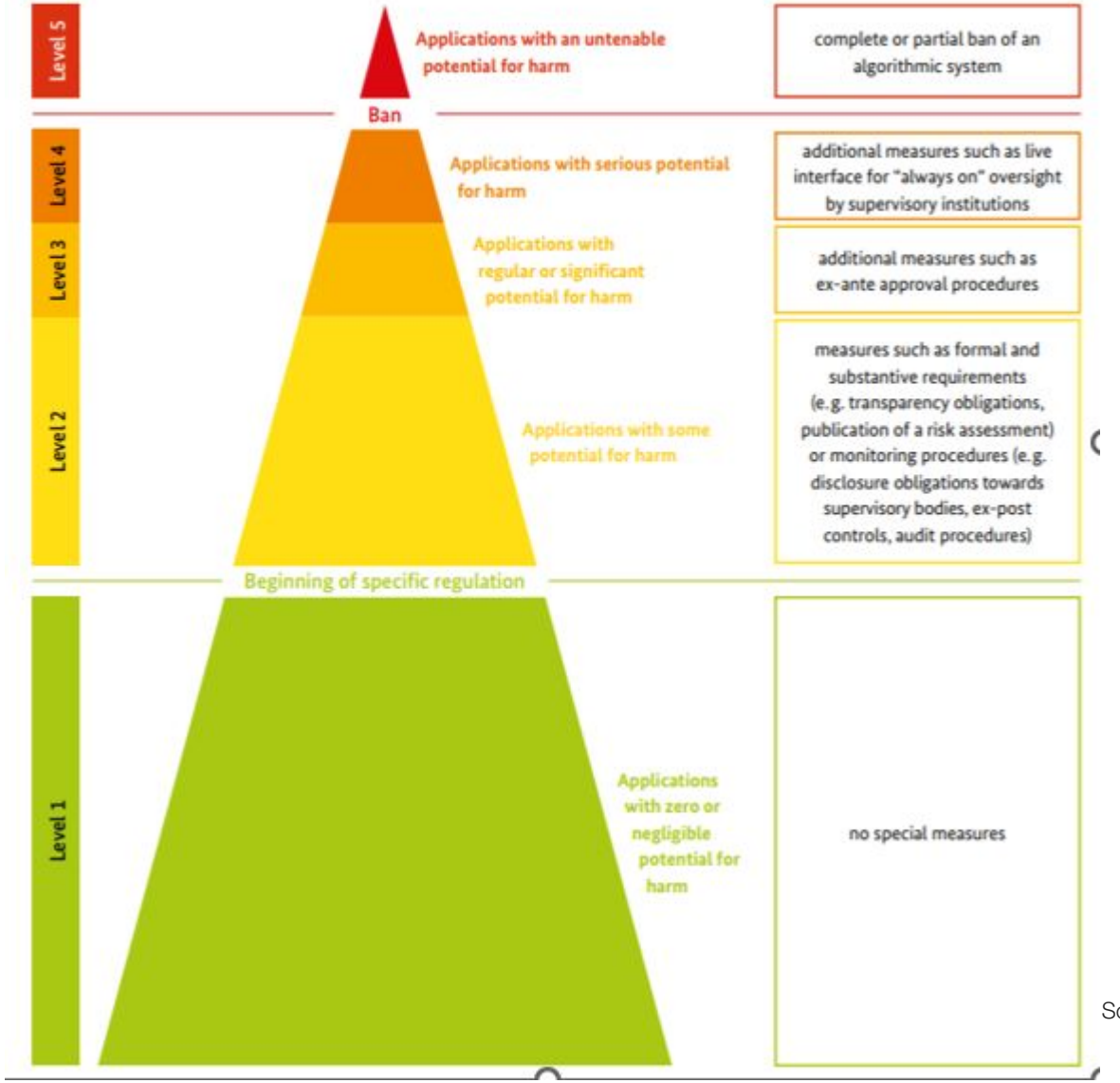
Some Technical Challenges

- Dependence on the accuracy of technical systems, danger of technical errors
    - Telemedicine
    - Autonomous driving
- Increased vulnerability against cyber attacks
- Privacy and danger of data misuse
    - Development of General Data Protection Regulation (GDPR)
- Digital literacy
    - Education in digital literacy often lacking

Issues and Solutions



Criticality pyramid and risk-adapted regulatory system
for the use of algorithmic systems



Source: Datenethikkommission der Bundesregierung (2019)

AI4People

- ✔ AI4People is the first European forum bringing together key stakeholders as academia, industry, civil society organizations and the European Parliament to lay the foundations for a “Good AI Society” shaping the impact of Artificial Intelligence

- ✔ Three-year roadmap for founding principles, policies and practices for a “good AI society” and “good AI governance”

AI4People: Human Centric Approach

Beneficence

Promoting well-being, preserving dignity and sustaining the planet

Non-maleficence

Ensuring privacy, security and “capability caution” (upper limit of future AI capabilities)

Autonomy

Striking a balance between the decision-making power we retain for ourselves and which we delegate to AI

Justice

Creating benefits that are (or could be) shared, preserving solidarity

Explicability

Enabling the other principles through intelligibility and accountability

Beneficience

Promoting well-being, preserving dignity and sustaining the planet

Non-maleficence

Ensuring privacy, security and “capability caution” (upper limit of future AI capabilities)

- ✔ Are space agencies’ or private companies’ image storage safe?
- ✔ Data sharing between different space agencies?
- ✔ How reliable are individual datasets?

Non-maleficence

Ensuring privacy, security and “capability caution” (upper limit of future AI capabilities)

- ✔ Risk with excessive dependence on publicly available models: following text from “BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain”

Outsourced training introduces new security risks: an adversary can create a maliciously trained network (a backdoored neural network, or a Badnet) that has state-of-the-art performance on the user's training and validation samples, but behaves badly on specific attacker-chosen inputs.

Autonomy

Striking a balance between the decision-making power we retain for ourselves and which we delegate to AI

- ✔ Can we rather reformulate this as a combination of AI-driven model and traditional physical models?

Justice

Creating benefits that are (or could be) shared, preserving solidarity

- ✔ Law or livelihood: Recall the case of small-scale illegal mining

Explicability

Enabling the other principles through intelligibility and accountability

- ✔ AI Explainability and (to some extent) Uncertainty Quantification is contributing to it.

AI Bill of Rights (US)

Source: https://www.whitehouse.gov/ostp/ai-bill-of-rights/

- ✔ Safe and Effective Systems
- ✔ Algorithmic Discrimination Protections
- ✔ Data Privacy
- ✔ Notice and Explanation
- ✔ Human Alternatives, Consideration, and Fallback