

AI: RESPONSIBILITY IN A CHANGING WORLD

Prof. Dr. Virginia Dignum

Chair Responsible AI – Director AI Policy Lab, Umeå University

Guest professor SIAS, University of Amsterdam

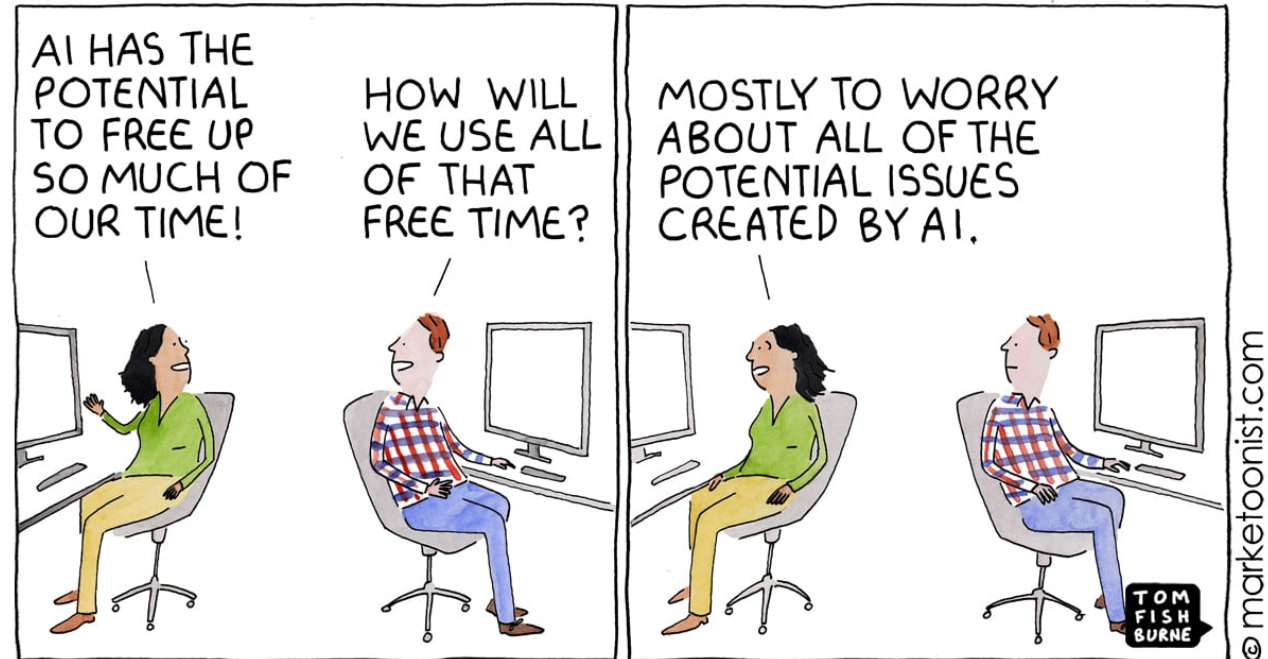
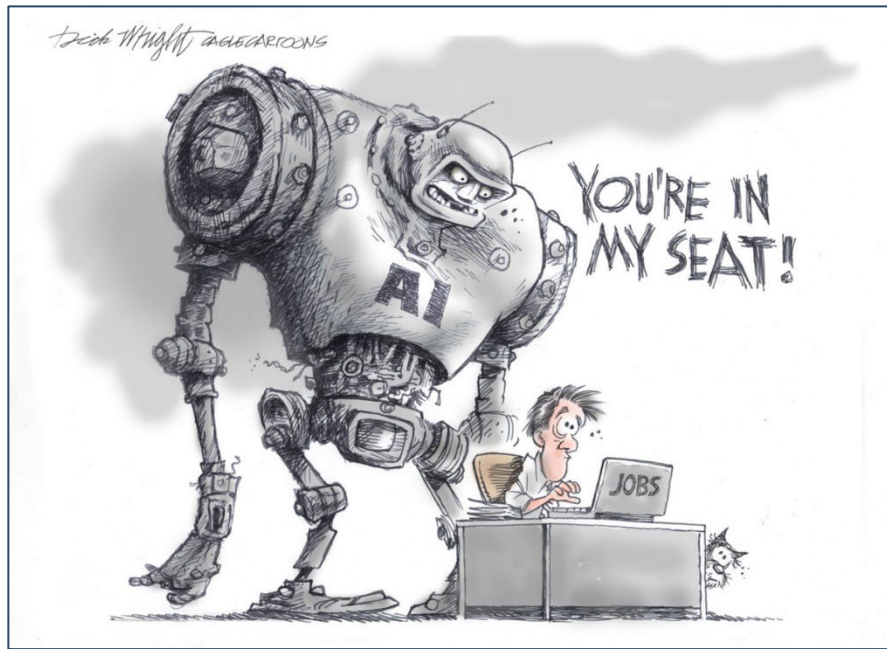
Email: virginia@cs.umu.se



UMEÅ UNIVERSITY

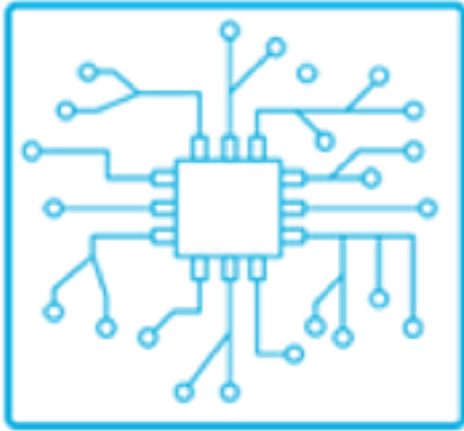
AI Policy  **Lab**
at Umeå University

ARE WE SHAPING AI, OR IS AI SHAPING US?



- AI raises as many concerns as it solves
- Need human insight to address AI's ethical, social, and unintended consequences

AI – MORE IS BETTER?



Computing Power

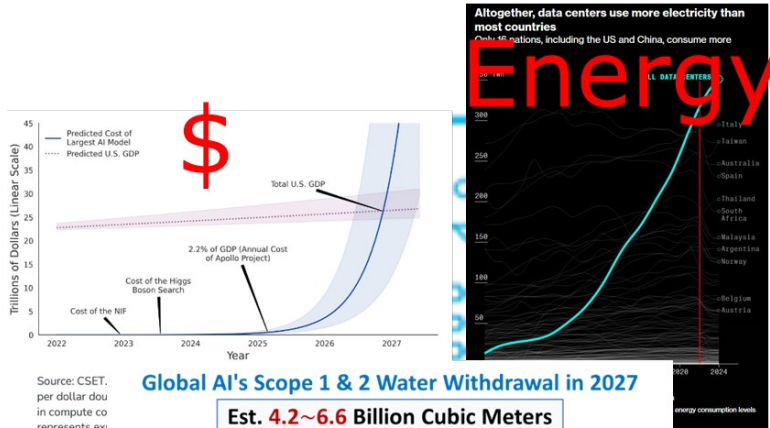


Algorithm Power



Data Availability

AI – MORE IS BETTER?



Source: CSET. per dollar dou in compute ex represents exj with error sha

Global AI's Scope 1 & 2 Water Withdrawal in 2027

Est. 4.2~6.6 Billion Cubic Meters

AI ≈ [6 Danish flags]

4~6x Annual Water Withdrawal of Denmark

Computing Power

Original Result

0 200 400 600 800 1000

0 250 500 750 1000

Google exploited homeless black people to develop the Pixel 4's facial recognition AI

Russia Tests New Disinformation Tactics in Africa to Expand

Amazon's facial recognition matched 28 members of Congress to criminal mugshots

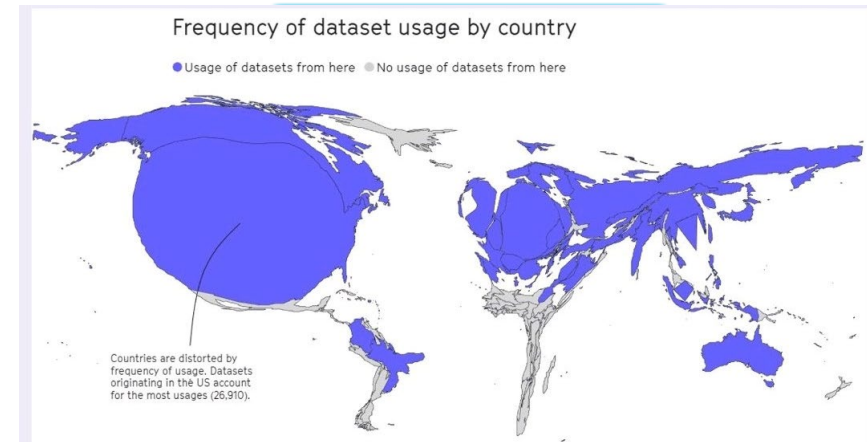
Flawed Algorithms Are Millions of Students' Enemies

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

Indigenous elder slams 'hollow and tokenistic' consultation by Side

WRITERS GUILD ON STRIKE! NO AI!!

Algorithm Power



Data Availability

AI IS NOT INTELLIGENT

DIFFERENT

- Correlation rather than causal mechanisms
- A language ‘Frankenstein monster’
 - Incapable of moral thinking, to determine what ought and ought not to be
 - Incapable of distinguishing the possible from the impossible

- Built by people for a given purpose
- Dependent on the labor of many
- Using natural resources

AI IS NOT ARTIFICIAL

JUST

Manipulation of language is not a proxy for intelligence!

<https://fof.se/artikel/overtygande-sprak-ar-inget-belagg-for-intelligens/>

CAN WE BE RESPONSIBLE?

Responsibility

AI IS NOT NEUTRAL

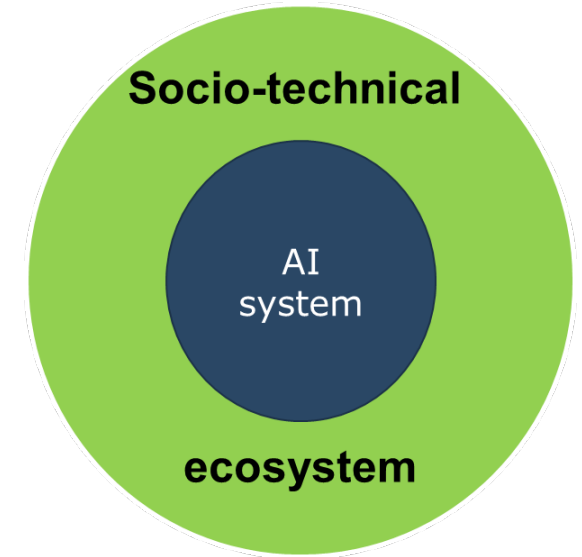
- AI can advance human dignity, or undermine it
- embodies human choices
- Governing AI = Governing ourselves



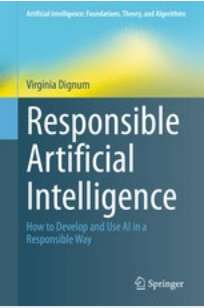
RESPONSIBLE ABOUT AI

AI does not happen to us!

- AI is designed. We make the choices
 - Who should decide?
 - Which values should be considered?
 - Whose values? How to prioritize?
- Ask **Question Zero!**
 - Is AI the best option here?
 - Ask why before asking how
- AI does not exist in a vacuum
 - There is no technology fix for ill effects!
 - Ethics, regulation, governance concern the ecosystem



Responsible AI solutions need to be social rather than technical!

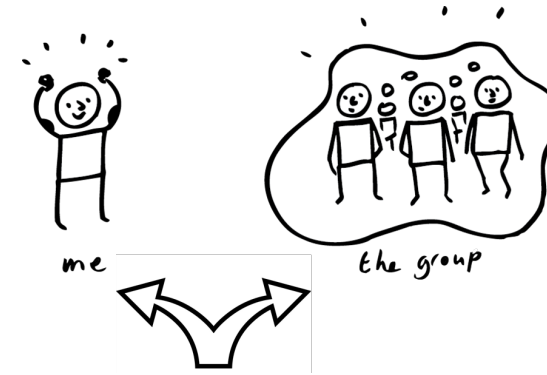


WHAT DO WE WANT AI TO BE?

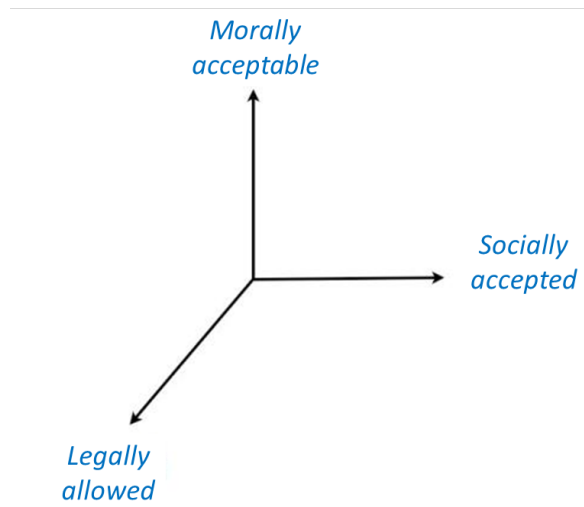
WHAT IS AI?

- Human-like?
 - Why?
 - What does this mean?
- Tool?
 - For what? For who?
- Simulation or operation?
 - Understand intelligence by building intelligence, or
 - Active intervention in real world
- Normative or descriptive?
 - Do as we say or do as we do?

WHICH DECISIONS SHOULD AI MAKE?



HOW SHOULD AI TAKE DECISIONS?



RAI IS MULTIDISCIPLINARY

understanding and critiquing the intended and unforeseen, positive and negative, socio-political consequences of AI for society in terms of equality, democracy and human rights

- **governance**, not only in terms of competences and responsibilities, but also in terms of **power, trust and accountability**;
- **societal, legal and economic** functioning of socio-technical systems;
- **value-based design** approaches and of ethical frameworks;
- **inclusion and diversity** in design, and how such strategies may inform processes and results;
- **distributed and increasingly ubiquitous nature of AI** applications and developing new scholarly perspectives on human-machine communication.

GOVERNANCE – WHY? WHAT FOR?

- Regulation as **incentive for responsible innovation, sustainability, and fundamental human rights**
 - powerful stepping stone for innovation with societal benefits
 - signaling expected ambitions enhancing innovation, competitive power

Cars drive faster with brakes

-

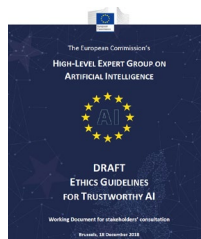
In a game without rules, no one wins

- *ensuring transparency, security, standards, and governance approaches to AI systems, products, and results*
 - Need for better understanding and integration of existing frameworks alongside introducing more regulation
- Avoidance of an "arms race" narrative in AI regulation

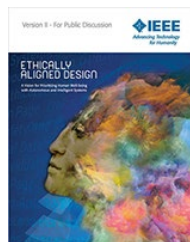
PRINCIPLES AND GUIDELINES

EU HLEG	OECD	IEEE EAD
<ul style="list-style-type: none"> Human agency and oversight Technical robustness and safety Privacy and data governance Transparency Diversity, non-discrimination and fairness Societal and environmental well-being Accountability 	<ul style="list-style-type: none"> benefit people and the planet respects the rule of law, human rights, democratic values and diversity, include appropriate safeguards (e.g. human intervention) to ensure a fair and just society. transparency and responsible disclosure robust, secure and safe Hold organisations and individuals accountable for proper functioning of AI 	<ul style="list-style-type: none"> How can we ensure that A/IS do not infringe human rights? effect of A/IS technologies on human well-being. How can we assure that designers, manufacturers, owners and operators of A/IS are responsible and accountable? How can we ensure that A/IS are transparent? How can we extend the benefits and minimize the risks of AI/AS technology being misused?

Level	Framework & Reach
Global	UNESCO (194 countries), OECD (>70 jurisdictions), GPAI (25+ members), CoE treaty (50+ countries), G7 principles
Regional	EU AI Act (27 EU states), Santiago Declaration (Latin America/Caribbean)
National	930+ initiatives in 71 countries
Others	200+ guidelines across NGOs, academic private bodies



<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>



<https://ethicsinaction.ieee.org>

OECD Principles on Artificial Intelligence



On 22 May 2019
by governance
The OECD Principles on Artificial Intelligence
Supporting innovation
We are also |

<https://www.oecd.org/digital/ai/principles/>



<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>



<https://www.un.org/ai-advisory-body>

Well over 1000 published worldwide!

GOVERNANCE: REGULATION AND MORE

- Legislation
 - Compulsory, formal enforcement
 - Many options: ex ante, ex post, risk based, principle based...
- Standards
 - soft governance; non mandatory to follow
 - demonstrate due diligence and limit liability
 - user-friendly integration between products
- Organisational processes and functions
 - Set and monitor ethical guidelines
 - able to veto any projects or deliverables that do not adhere to guidelines
- Monitoring and assessment
 - responsible AI is more than ticking boxes
 - Means to assess maturity are needed
- Education and training
 - Awareness
 - Participation / civic duty

ISSUES IN AI GOVERNANCE

- **AI Safety (preventing harm from AI itself)**
 - **Technical robustness & reliability:** preventing unintended failures, adversarial attacks
 - **Scalability & control:** ensuring AI behaves safely as it becomes more capable
 - **Alignment:** making AI systems follow human intent
 - **Security risks:** preventing AI from being misused or becoming uncontrollable
 - **Existential risks:** managing risks from superintelligence, loss of control
- **AI Ethics & Responsible AI (focus on societal impact of AI)**
 - **Bias & fairness:** ensuring AI doesn't reinforce social inequalities (bias, discrimination, exclusion)
 - **Transparency & explainability:** making AI understandable
 - **Accountability & governance:** who is responsible for AI's actions
 - **Misinformation & manipulation:** Preventing AI from spreading false information or being used for harmful persuasion.
 - **Economic & labour impacts:** Addressing issues of job loss, surveillance risks,
 - **Power Concentration:** Preventing AI from amplifying inequalities by benefiting only a few powerful entities.
 - **Autonomous Weapons & Misuse:** Preventing AI from being used in warfare or malicious activities.

RESPONSIBLE AI APPROACH

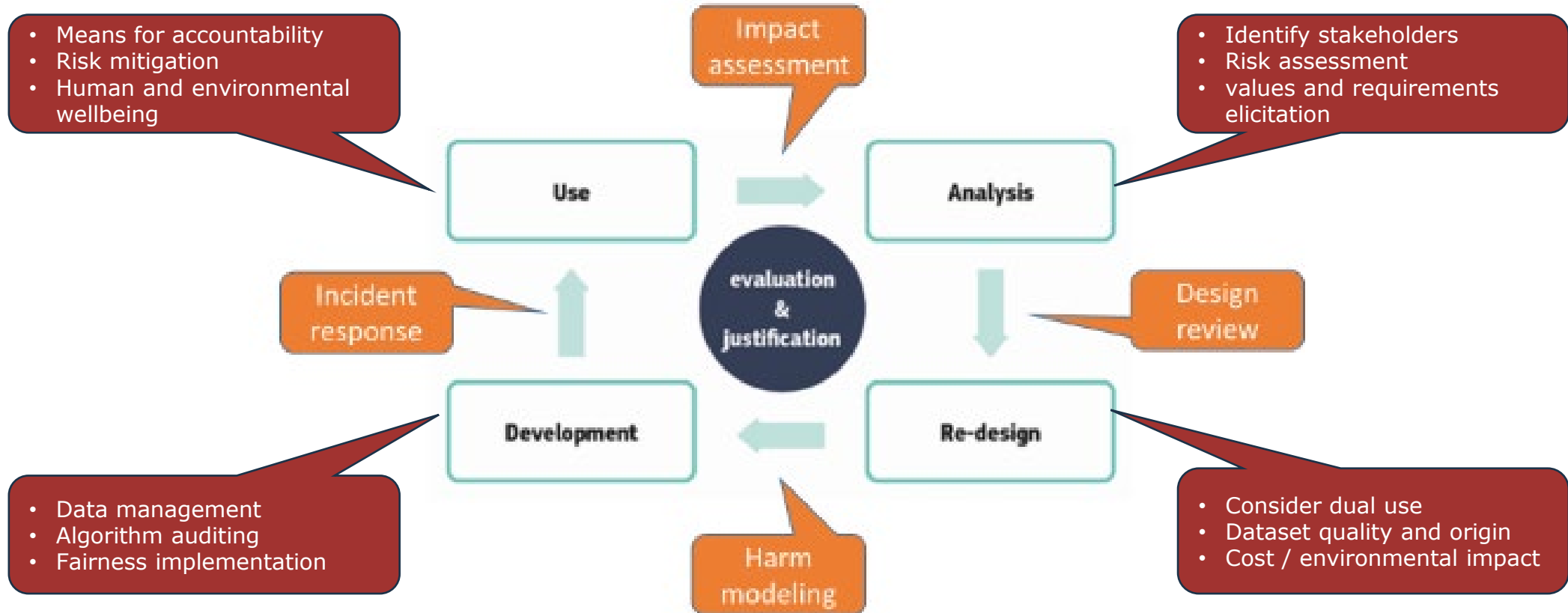
Committing to responsible and trustworthy AI principles for innovative, ethical AI that respects human rights

Recognizing potential tensions between ethical principles and the need to prioritize them

Developing a culture of responsibility within organizations

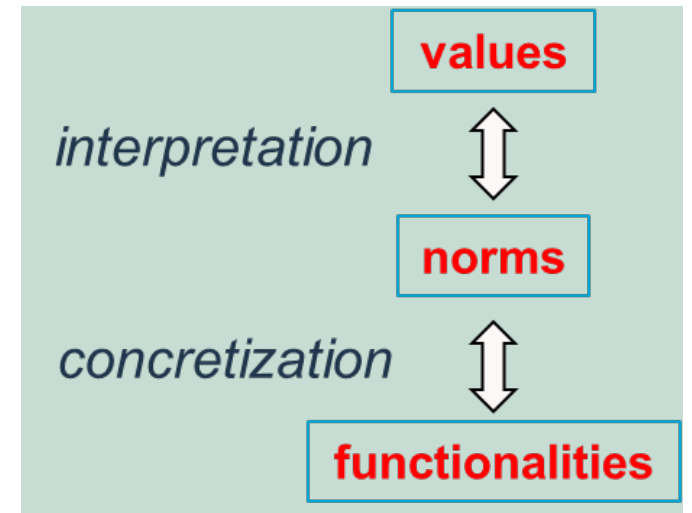
Prioritizing stakeholder interaction and communication

RESPONSIBLE AI LIFECYCLE



RAI IS ABOUT BEING EXPLICIT

- Design for Values
 - Legal and ethical aspects are not an add-on!
- Governance
 - External monitoring and control
 - Agreements, contracts, norms
- Design
 - Question your options and choices
 - Motivate your choices
 - Document your choices and options



<https://medium.com/@virginiadignum/on-bias-black-boxes-and-the-quest-for-transparency-in-artificial-intelligence-bcde64f59f5b>

AI ETHICS

- Is AI all that different from other technologies?
 - Overlap with tech ethics, business ethics, medical ethics...
- Some specific challenges to AI brought by:
 - Autonomy.
If executive control is delegated, how to ensure responsibility for actions?
 - Complexity.
If models are black boxes how can training lead to 'good' results? How to avoid reproduction and amplification of biases?
 - Deployment in open environments.
If we cannot predict the context which the system will encounter, how do we ensure ethical behaviour?

ETHICS AND RESPONSIBILITY

- If there is no “one ethics”, whose ethics do we use?
- Responsible development: transparently exposing which factors have been considered, how they have been implemented.
- Adherence to general principles: Lawfulness, Transparency, Accountability, Privacy, Diversity, Explainability...
- Soft governance e.g. EU guidelines for trustworthy AI, IEEE standards...
- Trade-offs: who decides?

RESPONSIBLE AI – MORE THAN ETHICS

- **Not philosophising about ethics**
 - Ethics is not about the answer but about recognizing the issue
 - Ethics is a (social) process not a solution
- **Not technification of ethics**
 - Your implementation does not 'solve' ethics
 - It is a (context specific) solution for a possible interpretation
- **Fundamentally, is about choices, priorities, tradeoffs**
 - Accuracy / Explanation
 - Accuracy / Computational resources
 - Security / privacy
 - Equity / equality
 - Long term benefit / Short term
 - ...



Better metrics are needed!

RESPONSIBLE AI IS NOT A CHOICE!

Not *innovation vs governance* but
governance as stepping-stone for innovation

- Innovation is not a race!
 - There is no finish line
 - Exploration rather than one fixed direction
- Adopting responsible AI
 - Builds trust
 - Drives transformation
 - Support business differentiation



THANK YOU!