

RESPONSIBLE ARTIFICIAL INTELLIGENCE: FROM PRINCIPLES TO ACTION

Prof. Dr. Virginia Dignum

Chair Responsible AI - Department of Computing Science

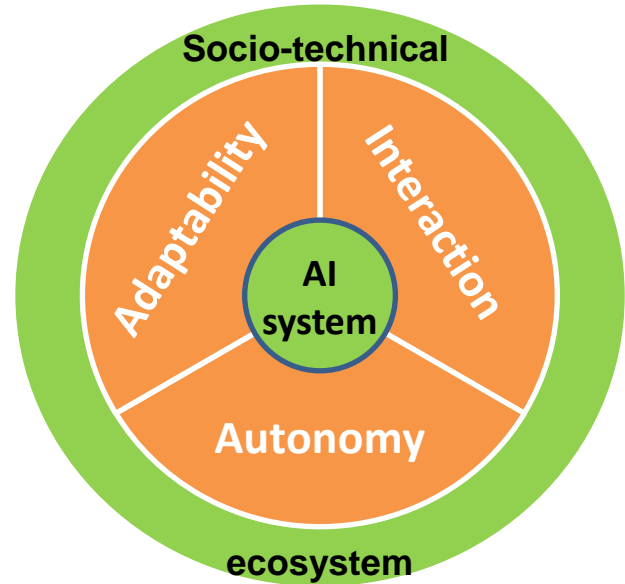
Email: virginia@cs.umu.se - Twitter: [@vdignum](https://twitter.com/vdignum)



UMEÅ UNIVERSITY

WHAT IS AI?

- Business as usual ?
- Magic ?
- Not just algorithm
- Not just machine learning
- Not just statistics or data analytics
- **AI applications are not alone!**



AI IS NOT INTELLIGENCE!

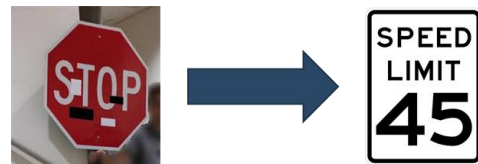
- What AI systems cannot do (yet)
 - Common sense reasoning
 - Understand context
 - Understand meaning
 - Learning from few examples
 - Learning general concepts
 - Combine learning and reasoning
- What AI systems can do (well)
 - Identify patterns in data
 - Images
 - Text
 - Video
 - Extrapolate those patterns to new data
 - Take actions based on those patterns



school bus 1.0 garbage truck 0.99 punching bag 1.0 snowplow 0.92



UMEÅ UNIVERSITY



RESPONSIBLE AI

- AI can potentially do a lot. **Should it?**
- Who should decide?
- Which values should be considered? Whose values?
- How do we deal with dilemmas?
- How should values be prioritized?

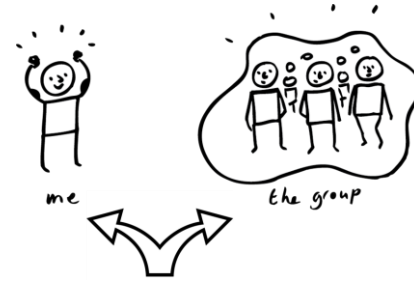


RESPONSIBILITY - SOME CASES

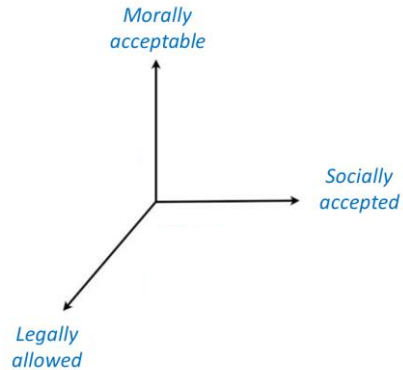
- Chatbots
 - Mistaken identity (is it a person or a bot?)
 - Manipulation of emotions / nudging / behaviour change support
- Automated decision making
 - Accountability: never “*computer says no*”
 - Repeating the past (bias, exclusion...)
- Meaning:
 - Ex: What is *fair*? Equal resources or equal opportunities or ... or...
 - Can we agree? Who decides?
- Dilemmas
 - Privacy or security?
 - Accuracy (more computer power) or climate?



WHICH DECISIONS SHOULD AI MAKE?



HOW SHOULD AI TAKE DECISIONS?

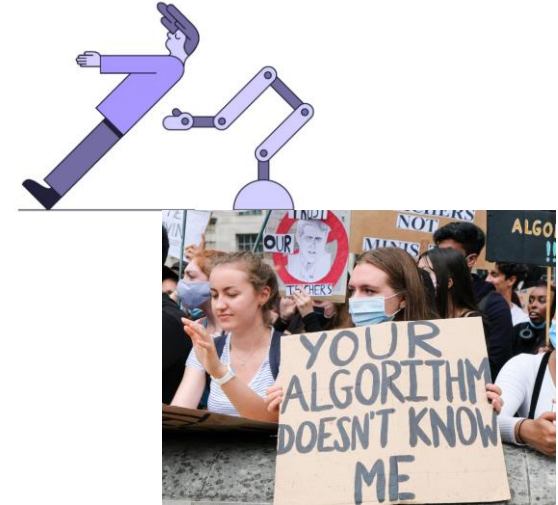
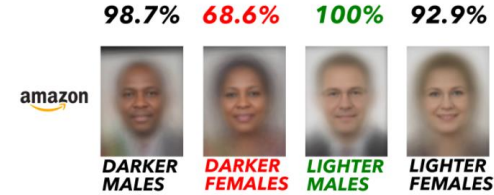


TRUST

AI and Democracy - Social acceptance



Choices
Formulation
Information
Involvement
Legitimacy
Aggregation

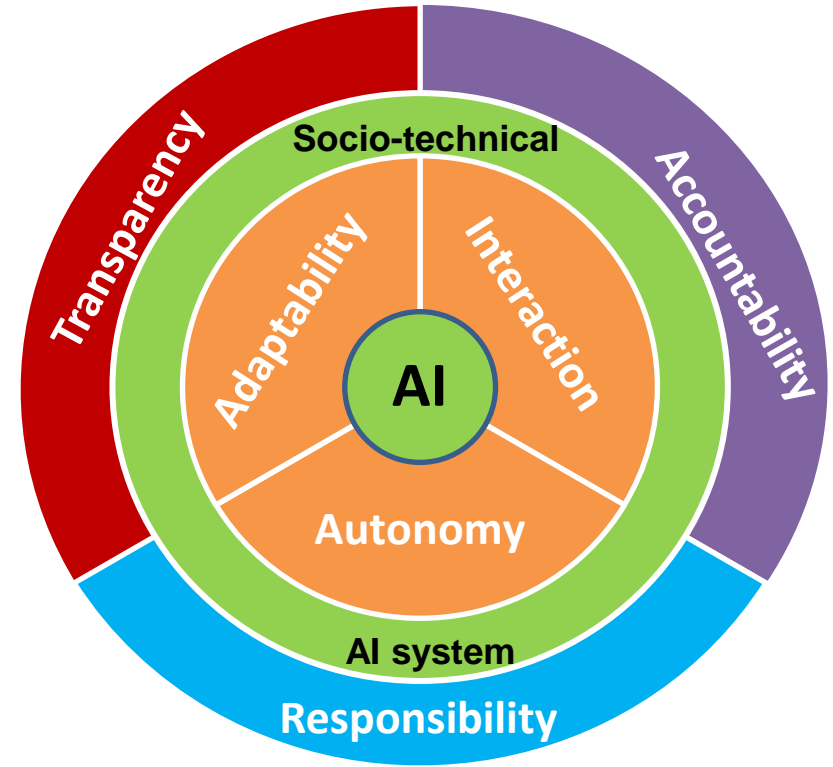


DESIGN IMPACTS DECISIONS IMPACTS SOCIETY

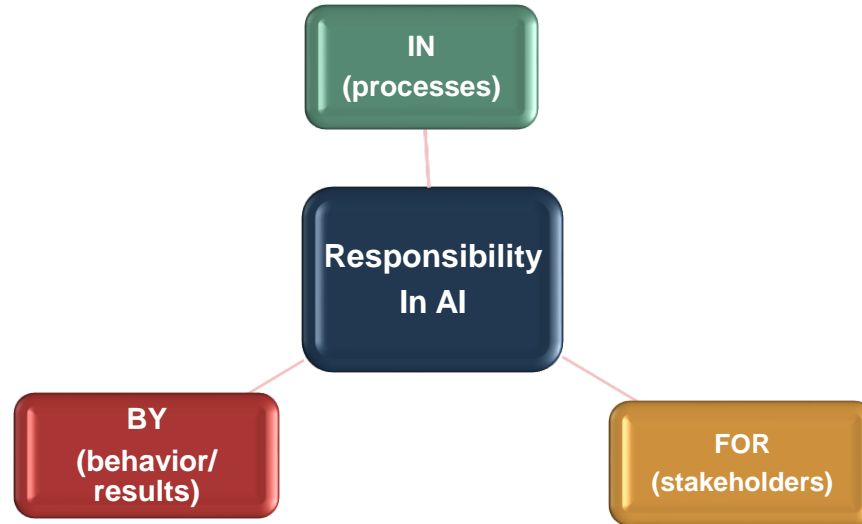
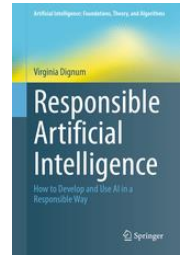
RESPONSIBLE AI = ART!

Responsible AI is

- Lawful
- Ethical
- Reliable
- Beneficial
- Verifiable



TAKING RESPONSIBILITY



PRINCIPLES AND GUIDELINES

- UNESCO
- European Union
- OECD
- WEF
- Council of Europe
- IEEE Ethically Aligned Design
- National strategies
- ...

AI principles
are principles for us!



<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 2020
by governance
The OECD P
Principles en
Supporting In
We are also |

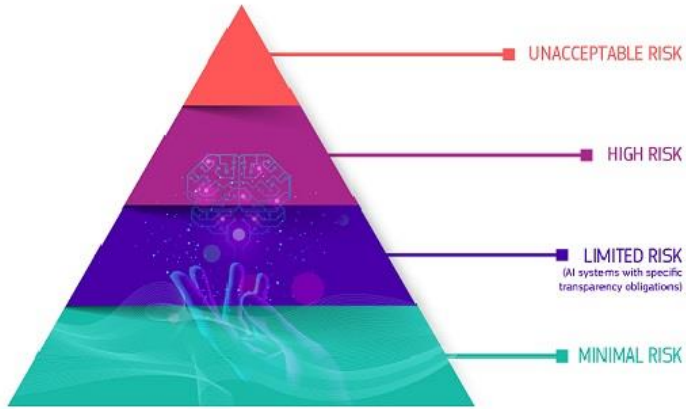
<https://www.oecd.org/going-digital/ai/principles/>

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?



UMEÅ UNIVERSITY

EUROPE'S REGULATION PROPOSAL



- Human-centered: Fundamental rights and principles are central
- Blurry red lines, blurry definition -> loopholes
- AI definition: what is it aiming to regulate?



ALTERNATIVES / COMPLEMENTS TO REGULATION

- Standards (IEEE, ISO)
 - soft governance; non mandatory to follow
 - demonstrate due diligence and limit liability
 - user-friendly integration between products
- Advisory panels and ethics officers (Industry)
 - Set and monitor ethical guidelines
 - able to veto any projects or deliverables that do not adhere to guidelines
- Assessment for trustworthy, ethical, AI (EU)
 - responsible AI is more than ticking boxes
 - Means to assess maturity are needed
- Awareness and Participation
 - Education and training
 - Appeal to civic duty / voluntary implementation



TRUSTWORTHY AI

Not *innovation vs regulation / ethics* but
Regulation/ ethics as stepping stone for innovation

- Taking an ethical perspective
 - Business differentiation (“Ethics is the new green”)
 - Certification to ensure public acceptance

- Principles and regulation are drive for transformation
 - Better solutions
 - Return on Investment



THANK YOU!

virginia@cs.umu.se

Twitter: @vdignum



UMEÅ UNIVERSITY