RESPONSIBLE AI

Prof. Dr. Virginia Dignum

Chair of Social and Ethical AI - Department of Computer Science

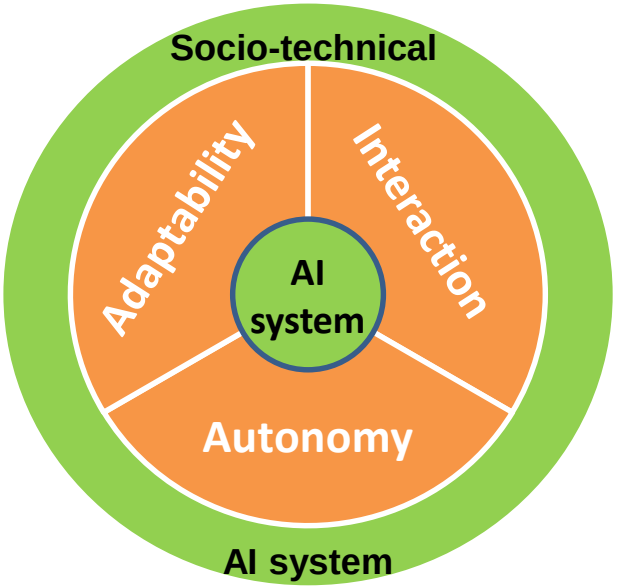
Email: virginia@cs.umu.se - Twitter: @vdignum

UMEÅ UNIVERSITY

WHAT IS AI?

- Not just algorithm
- Not just machine learning

- But
- AI applications are not alone
  - Socio-technical AI systems



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

WHAT IS RESPONSIBLE AI?

Responsible AI is

- Ethical
- Lawful
- Reliable
- Beneficial

Responsible AI recognises that

- AI systems are artefacts
- We set the purpose

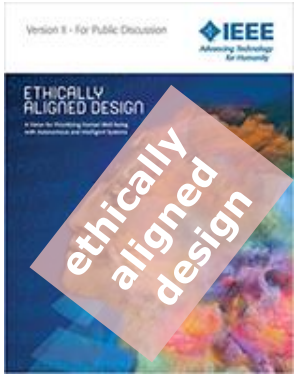
RESPONSIBLE AI

- 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?
- .....

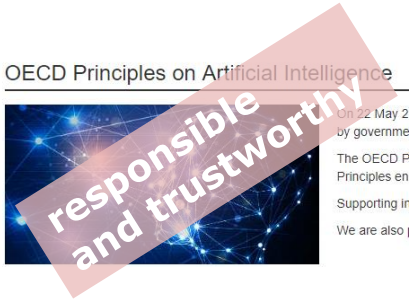
PRINCIPLES AND GUIDELINES

Responsible / Ethical / Trustworthy....

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



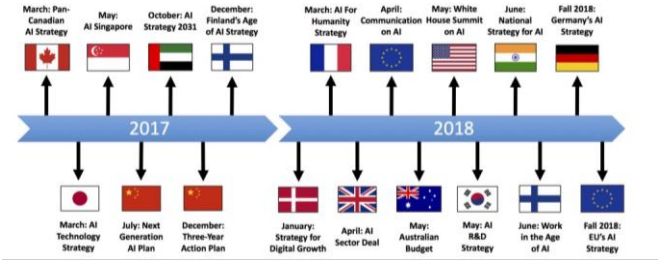
https://ethicsinaction.ieee.org



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

MANY INITIATIVES (AND COUNTING…)

- Strategies / positions
  - IEEE Ethically Aligned Design
  - European Union
  - OECD
  - WEF
  - Council of Europe
  - National strategies:
    - Tim Dutton, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd
  - …
- Declarations
  - Asilomar
  - Montreal
  - …



Check Alan Winfield blog: http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html

EU HLEG	OECD	IEEE EAD
• Human agency and oversight • **Technical robustness and safety** • Privacy and data governance • **Transparency** • **Diversity**, non-discrimination and fairness • **Societal and environmental well-being** • **Accountability**	• 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	• 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?

Responsible AI is (also) about ensuring that guidelines have meaning
What is needed?
Who is needed?
Why is needed?

ENDORSEMENT ≠ COMPLIANCE

IEEE P7000™ Standardization Projects

- **IEEE P7000™** - Model Process for Addressing Ethical Concerns During System Design
- **IEEE P7001™** - Transparency of Autonomous Systems
- **IEEE P7002™** - Data Privacy Process
- **IEEE P7003™** - Algorithmic Bias Considerations
- **IEEE P7004™** - Standard on Child and Student Data Governance
- **IEEE P7005™** - Standard on Employer Data Governance
- **IEEE P7006™** - Standard on Personal Data AI Agent Working Group
- **IEEE P7007™** - Ontological Standard for Ethically driven Robotics and Automation Systems
- **IEEE P7008™** - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
- **IEEE P7009™** - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
- **IEEE P7010™** - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems
- **IEEE P7011™** - Standard for the Process of Identifying and Rating the Trustworthiness of News Sources
- **IEEE P7012™** - Standard for Machine Readable Personal Privacy Terms
- **IEEE P7013™** - Inclusion and Application Standards for Automated Facial Analysis Technology.

EU Ethics Guidelines for AI – Assessment List

Assessment list to operationalise the requirements

- **Practical questions** for each requirement – 131 in total
- Test through piloting process to collect **feedback** from all stakeholders (public & private sector)
  - "Quantitative" analysis track -> open survey
  - "Qualitative" analysis track -> in depth interview

Contribute!

https://ec.europa.eu/eusurvey/runner/Trustworthy_AI_Assessment_Pilot_20_Main

UMEÅ UNIVERSITY

European Commission

Recommendations for trustworthy AI – Main issues

1. Empower and protect humans and society
2. Take up a tailored approach to the AI market
3. Secure a Single European Market for Trustworthy AI
4. Enable AI ecosystems thorough sectoral multi stakeholder alliances
5. Foster the European data economy
6. Exploit the multi-faceted role of the public sector
7. Strengthen and unite Europe's research capabilities
8. Nurture education
9. Adopt a risk-based governance approach to AI and ensure an appropriate regulatory framework
10. Stimulate an open and lucrative investment environment
11. Embrace a holistic way of working

European Commission

OECD AI POLICY OBSERVATORY

AN INCLUSIVE HUB FOR AI INFORMATION, EVIDENCE AND POLICY OPTIONS

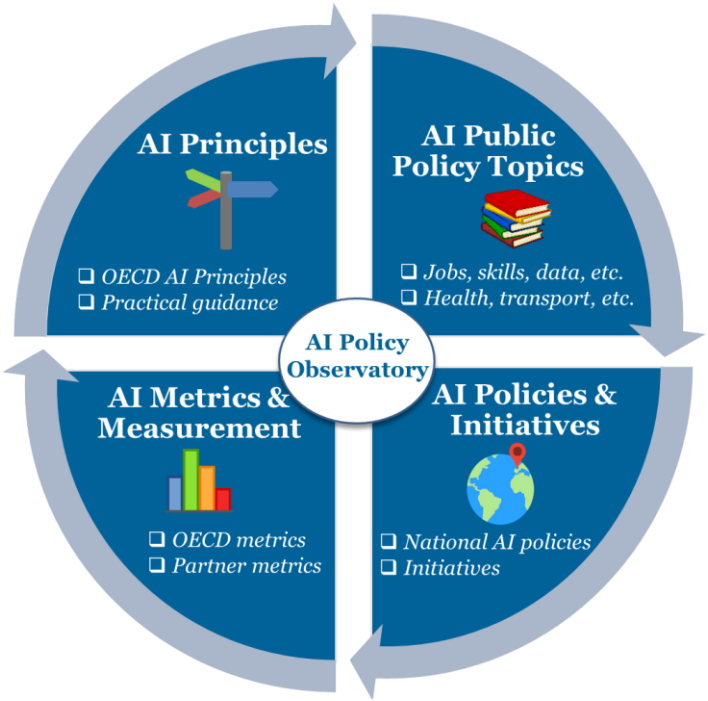
Characteristics

Multidisciplinary

Evidence-based

Multi-stakeholder

Partnerships



UMEÅ UNIVERSITY

GUIDELINES AND RESPONSIBILITY - KEY QUESTIONS TO ASK

1. Who wrote it, and how?
2. Who is it intended for, and what is its purpose?
3. Why should I follow it?
4. How do I follow or implement it?
5. How should I resolve conflicting interpretations of essentially contested concepts?
6. How will you know I am following it?
7. What happens if I fail to follow it?
8. How can I raise disagreements or questions for clarification?

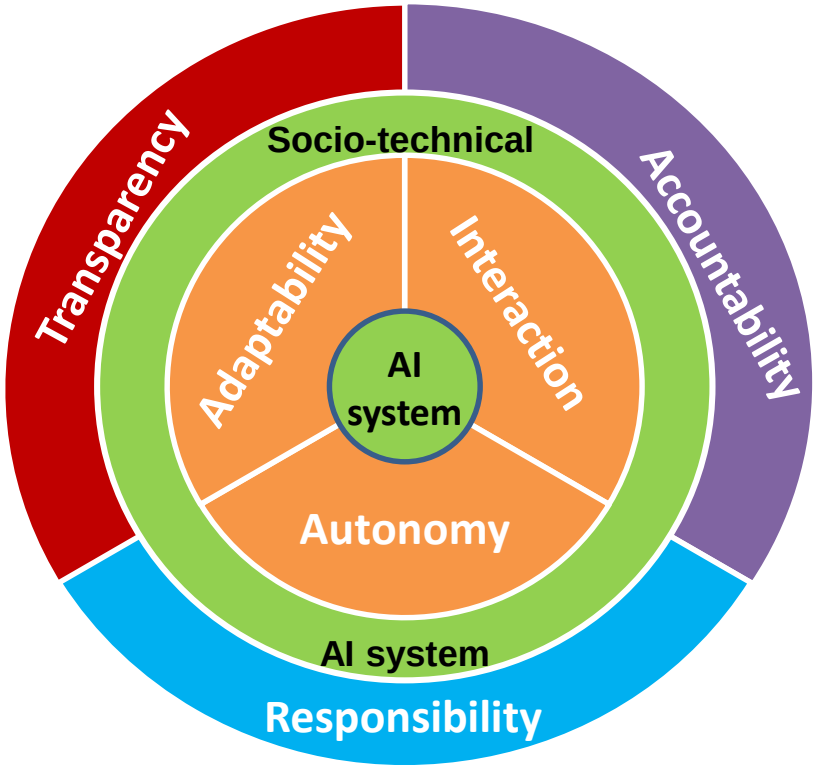
AI Ethics – Too Principled to Fail? By Brent Mittelstadt
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3391293

TAKING RESPONSIBILITY

- **in** Design
  - Ensuring that development processes take into account ethical and societal implications of AI and its role in socio-technical environments
- **by** Design
  - Integration of ethical reasoning abilities as part of the behaviour of artificial autonomous systems
- **for** Design(ers)
  - Research integrity of stakeholders (researchers, developers, manufacturers,…) and of institutions to ensure regulation and certification mechanisms

TAKING RESPONSIBILITY: ART

- AI needs ART
  - **A**ccountability
  - **R**esponsibility
  - **T**ransparency



IN DESIGN: PROCESS

- Doing the right thing
- Doing it right
- Design for Values
- Participation

ETHICS *IN* DESIGN– DOING IT RIGHT

- Principles for Responsible AI = ART
  - **A**ccountability
    - Explanation and justification
    - Design for values
  - **R**esponsibility
    - Autonomy
    - Chain of responsible actors
    - Human-like AI
  - **T**ransparency
    - Data and processes
    - Not just about algorithms

- AI systems (will) take decisions that have ethical grounds and consequences
- Many options, not one 'right' choice
- Need for design methods that ensure

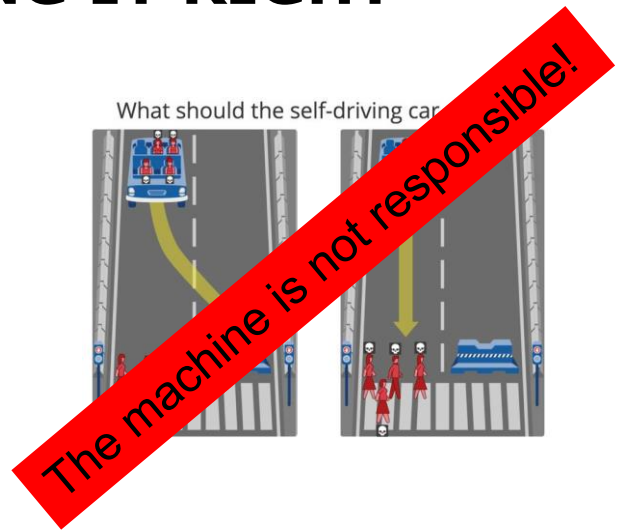
ETHICS *IN* DESIGN: AI – DOING IT RIGHT

- Principles for Responsible AI = ART
  - Accountability
    - Explanation and justification
    - Design for values
  - Responsibility
  - Transparency

- Optimal AI is explainable AI
- Many options, not one 'right' choice

ETHICS *IN* DESIGN: AI – DOING IT RIGHT

- Principles for Responsible AI = ART
  - Accountability
    - Explanation and justification
    - Design for values
  - **R**esponsibility
    - Autonomy
    - Chain of responsible actors
    - Human-like AI
  - Transparency



ETHICS *IN* DESIGN: AI – DOING IT RIGHT

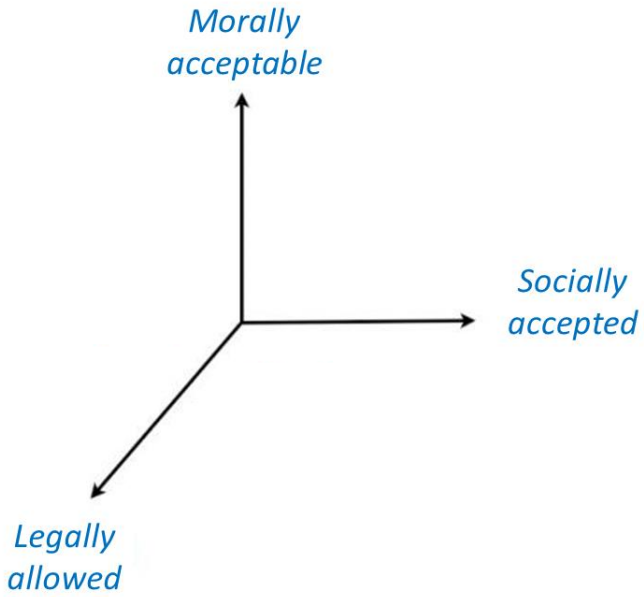
- Principles for Responsible AI = ART
  - Accountability
    - Explanation and justification
    - Design for values
  - Responsibility
    - Autonomy
    - Chain of responsible actors
    - Human-like AI
  - Transparency
    - Data and processes
    - Algorithms
    - Choices and decisions

BY **DESIGN: ARTIFICIAL AGENTS**

- **Can we teach ethics to AI?**
- **Should we teach ethics to AI?**
- **Can AI artefacts be build ethically?**
  - What does that mean?
  - What is needed?
- **Decisions matter**
  - From values to norms to functionalities…
- **Stakeholder differ**

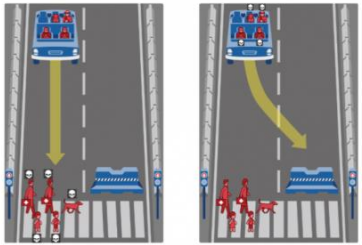
WHICH VALUES – WHOSE VALUES

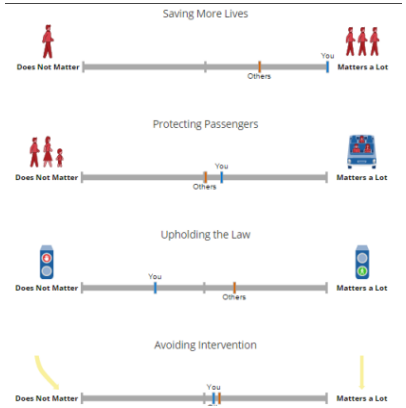
- Sources
  - Society (Designer, Users, Owner, Manufacturer)
  - Law: legislation, standards
  - Ethics
- But
  - Who decides who has a say?
  - How to make choices and tradeoffs between conflicting values?
  - How to verify whether the designed system embodies the intended values?



email: virginia@cs.umu.se
Twitter: @vdignum

SOCIAL ACCEPTANCE – DEMOCRACY





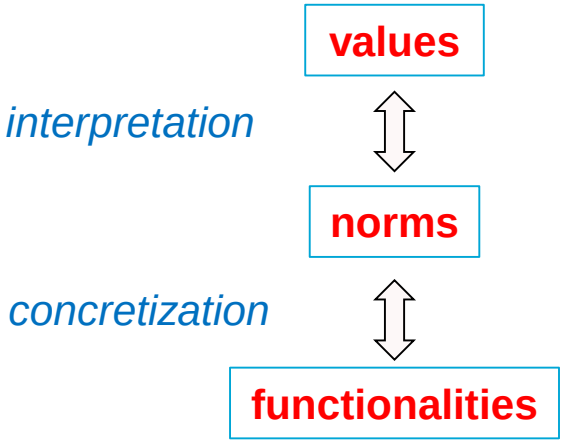
- Choices
- Formulation / Information
  - comprehensability
- Involvement
- Legitimacy
- Voting system

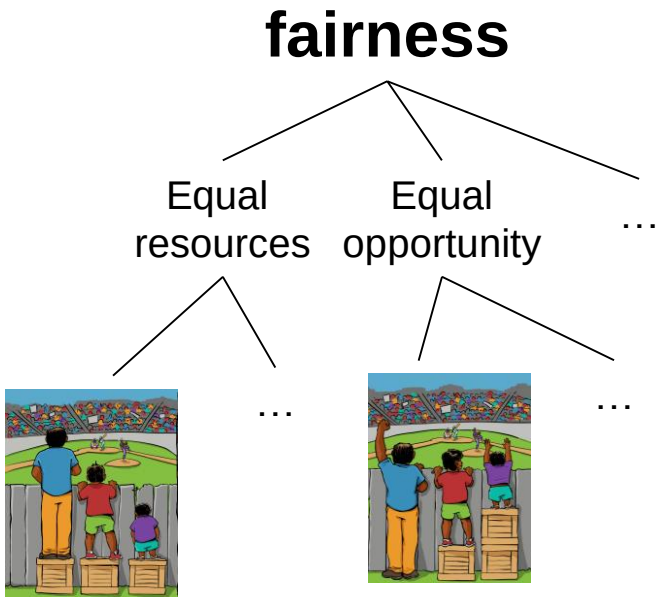
UMEÅ UNIVERSITY

email: virginia@cs.umu.se

Twitter: @vdignum

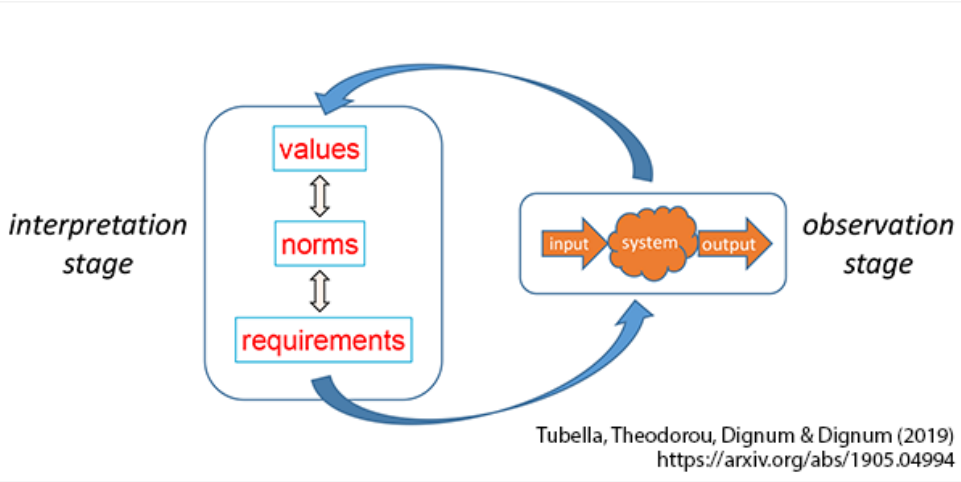
DECISIONS MATTER!





GLASS BOX APPROACH

- Responsibility in Design
- Doing the right thing
  - Elicit, define, agree, describe, report
- Doing it right
  - Explicit values, principles, interpretations, decisions
  - Evaluate input/output against principles



FOR DESIGN(ERS): PEOPLE

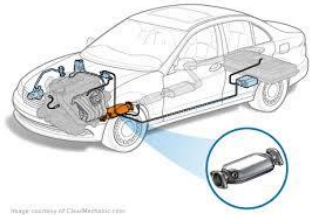
- **Regulation**
- **Certification**
- **Standards**
- **Conduct**

AI principles are principles for us

ETHICS *IN* DESIGN - DOING THE RIGHT THING

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

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



NATIONELL AI AGENDA FÖR SVERIGE

AI INNOVATION of Sweden

AI COMPETENCE FOR SWEDEN

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

WASP-HS
Wallenberg Autonomous Systems and Software Program – Humanities and Society

Strategy: policy and regulation

Innovation:
competitive now and in the future

Education:
capacity building, informed consumers

Fundamental research:
Engineering, computer science, math

Fundamental research:
Multidisciplinary, humanities, society

NATIONELL AI AGENDA FÖR SVERIGE

AI INNOVATION of Sweden

AI COMPETENCE FOR SWEDEN

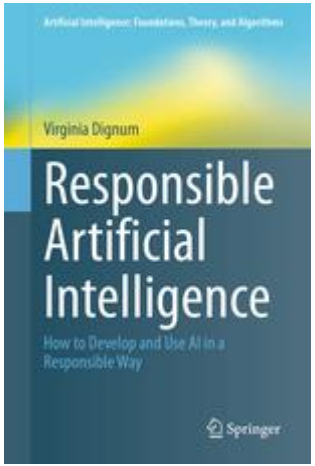
WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

WASP-HS
Wallenberg Autonomous Systems and Software Program – Humanities and Society

WASP – HS

- **Scientific Director: Virginia Dignum**
- **Board: Kerstin Sahlin (Chair)**
- **Aims**
  - Multidisciplinary research on AI impact
  - Excellence in humanities and social sciences
    - Graduate School
    - Lectorships
  - Internationalisation
  - Embedding in social and industrial challenges and needs
- **720 MSEK over 10 years**
- **Complementary to WASP**

TAKE AWAY MESSAGE



- AI influences and is influenced by society
- AI systems are tools, artefacts made by people:
  We set the purpose
- AI needs ART (Accountability, Responsibility, Transparency)
- Ethical guidelines are important but tools to guarantee compliance are needed

RESPONSIBLE ARTIFICIAL INTELLIGENCE

WE ARE RESPONSIBLE

Email: virginia@cs.umu.se

Twitter: @vdignum

UMEÅ UNIVERSITY