Responsible AI

-

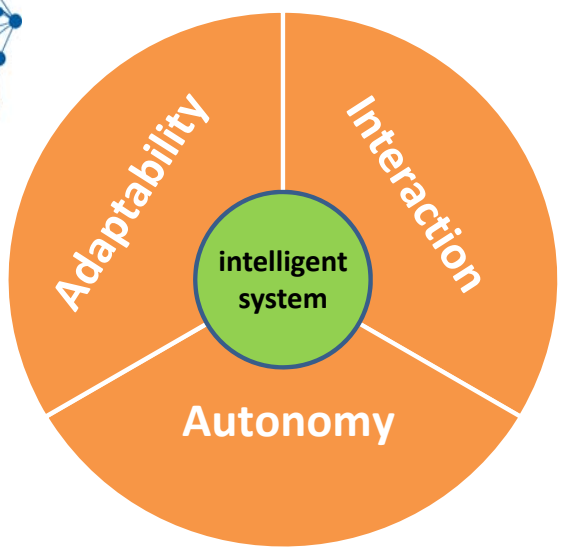
integrating values and ethics into the AI discourse and skills development programmes

Virginia Dignum
Social Artificial Intelligence Lab & Delft Institute Design for Values
Delft University of Technology

Email: m.v.dignum@tudelft.nl
Twitter: @vdignum

http://designforvalues.tudelft.nl/

Artificial Intelligence



Artificial Intelligence

- AI can potentially do a lot. Should it?
- Who should decide?
- Which values should be considered? Whose values?
- What is our role? our well-being?
- Which capabilities do we need?
- …..

Setting boundaries: Technical, ethical, societal

Responsibility - Why care?

- AI systems act autonomously in our world
- Eventually, AI systems will make better decisions than humans
  - Define better!

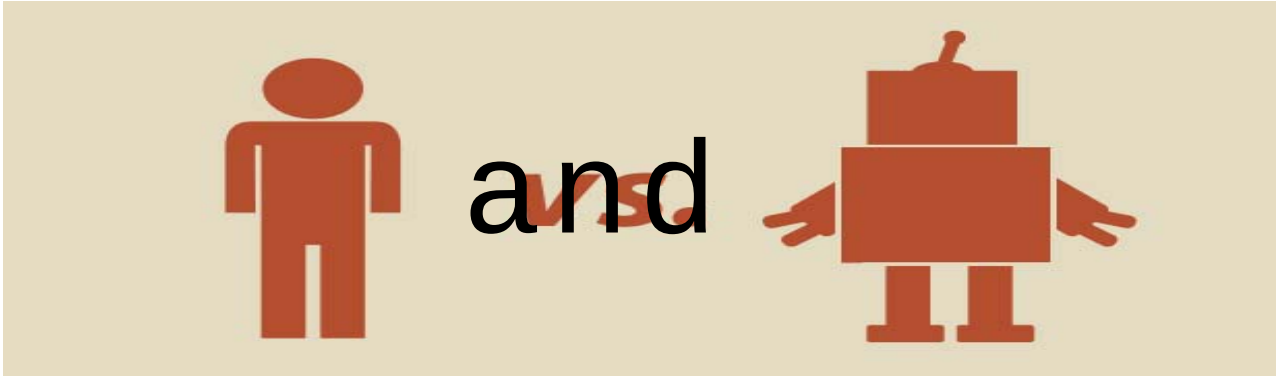
AI is designed, is an artefact

- We need to sure that the purpose put into the machine is the purpose which we really want

Norbert Wiener, 1960 (Stuart Russell)

King Midas, c540 BCE

Machine vs Human



- Creativity
- Flexibility
- Adaptability
- Empathy
- …

- Strength
- Accuracy
- Speed
- Does not tire
- …

Ensuring human well-being

- Not everything that is utility-maximizing is necessarily *better*
- Design for values

Skills for sustainable well-being in the digital age

"Many new jobs will appear for which skilled human workers are needed with a set of skills that combine technical education with humanities, arts and social sciences."

http://news.itu.int/skills-for-sustainable-well-being-in-the-digital-age/

Skills

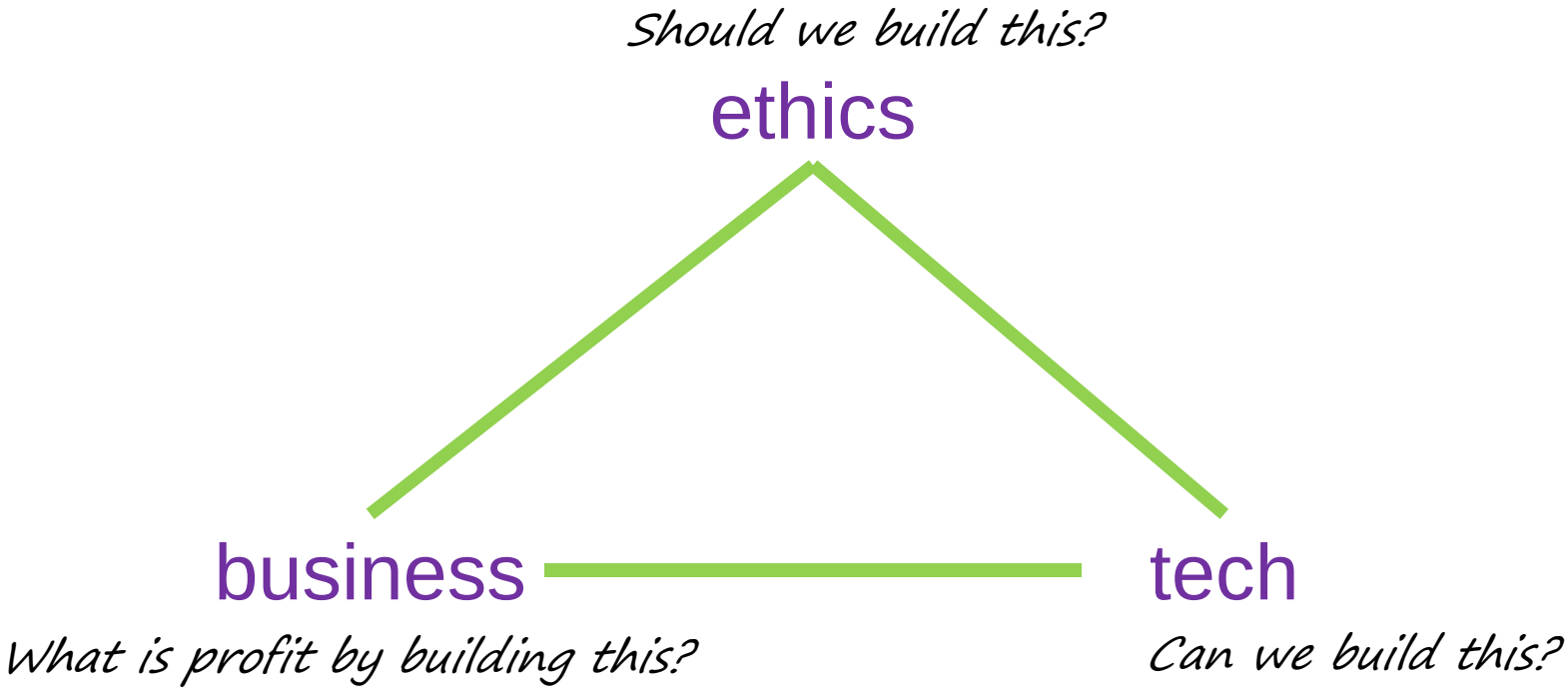
- Decisions made by engineers during the design of products and technologies are not only technical in nature.
- Engineers need the knowledge and tools to include ethical and societal values - such as sustainability, safety, privacy and fairness - into the design of new products and technologies.
- All need the skills to evaluate, decide, use, influence the AI systems one interacts with

Values influence design

Design influences values

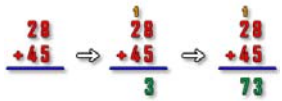
Values determine the future

Skills for AI developers

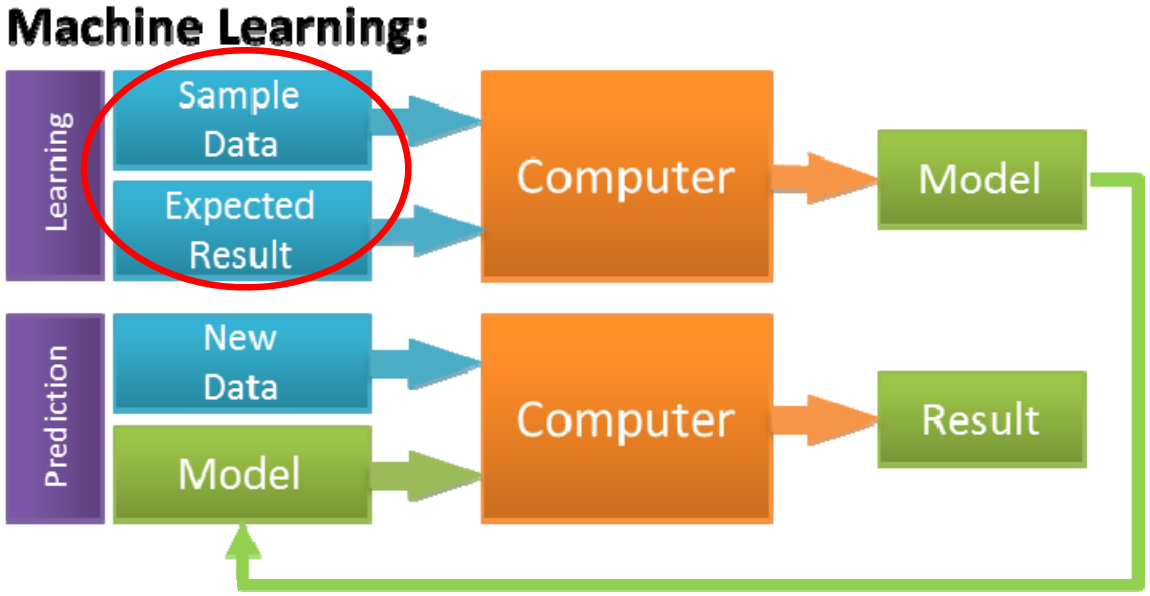


Skills for all – AI is not magic

- An algorithm is a set of rules; an unambiguous specification of how to solve a problem
  - Is not magic, is not the solution for all your problems!
  - Algorithms have been around for thousands of years



- What is new?
  - In AI, an algorithm is a program that can learn from and make predictions on data.

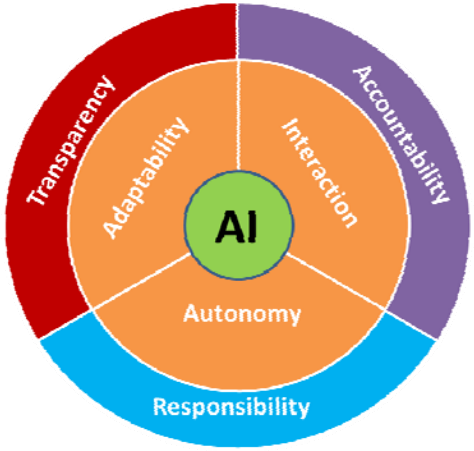


Education – the *right* skills

Taking responsibility

- Ethics **in** Design
  - Ensure that development processes are aligned with ethical principles
  - Consider the ethical implications of artificial intelligence as it integrates and replaces traditional systems and social structures
- Ethics **by** Design
  - Integration of ethical reasoning abilities as part of the behaviour of artificial autonomous systems (such as agents and robots)
- Ethics **for** Design(ers)
  - research integrity of researchers and manufacturers as they design, construct, use and manage artificially intelligent systems,

Ethics **in** Design – the **ART** of AI

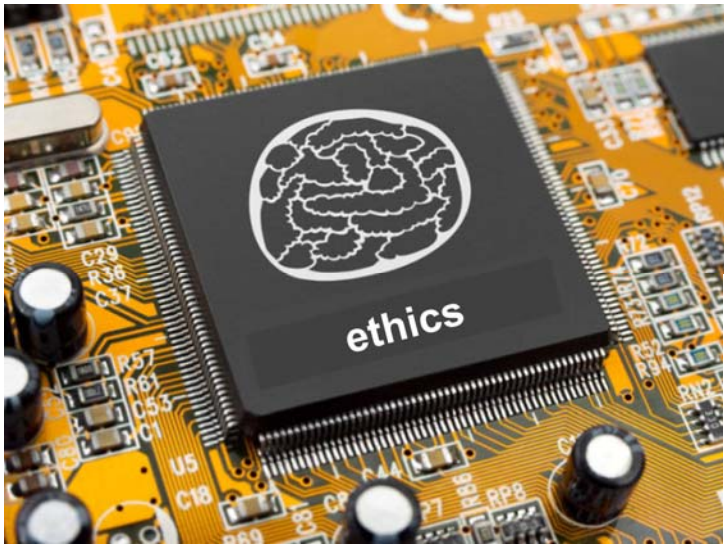


- AI systems (will) take decisions that have ethical grounds and consequences
- Many options, not one 'right' choice
- Need for design methods that ensure

- **A**ccountability
  - Explanation and justification
  - Design for values
- **R**esponsibility
  - Autonomy
  - Chain of responsible actors
  - Human-like AI
- **T**ransparency
  - Data and processes
  - Algorithms

(V. Dignum: "Responsible Autonomy", IJCAI2017)

Ethics by Design



- Can AI artefacts be build to be ethical?
- What does that mean?
- What is needed?

1. **Value alignment**
   - Identify *relevant* human values
   - Are there universal human values?
   - Who gets a say? Why these?

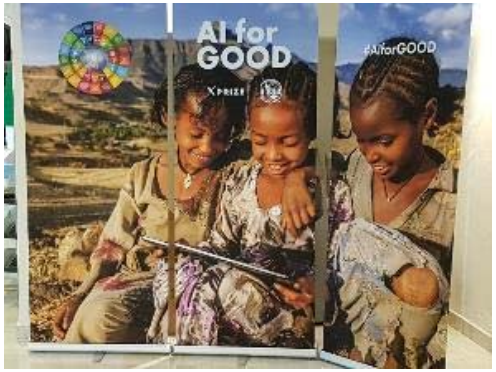
2. **How to behave?**
   - Ethical theories: How to behave according to these values?
   - How to prioritize those values?

3. **How to implement?**
   - Role of user
   - Role of society
   - Role of AI system

TUDelft DESIGN FOR VALUES

Ethics **for** Design(ers)

- Harness the positive potential outcomes of AI in society, the economy: AI for Good
  - Ensure inclusion, diversity, universal benefits
  - Prioritize UN2020 Sustainable Development Goals
- Codes of conduct – coming of age of the professional group
  - Society relies on AI developers
  - As it does with doctors, military, accountants…
- Auditing
  - Certified agencies
  - Algorithmic verification
  - Compliance, risk assessment
- Standardization
  - Processes, data protection, privacy, transparency…



Partnership on AI
to benefit people and society

4 PEOPLE

http://www.ai4people.eu

ETHICS IN ACTION

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

IEEE

https://ethicsinaction.ieee.org/

Well-being in the digital age = Responsible AI

WE ALL ARE RESPONSIBLE