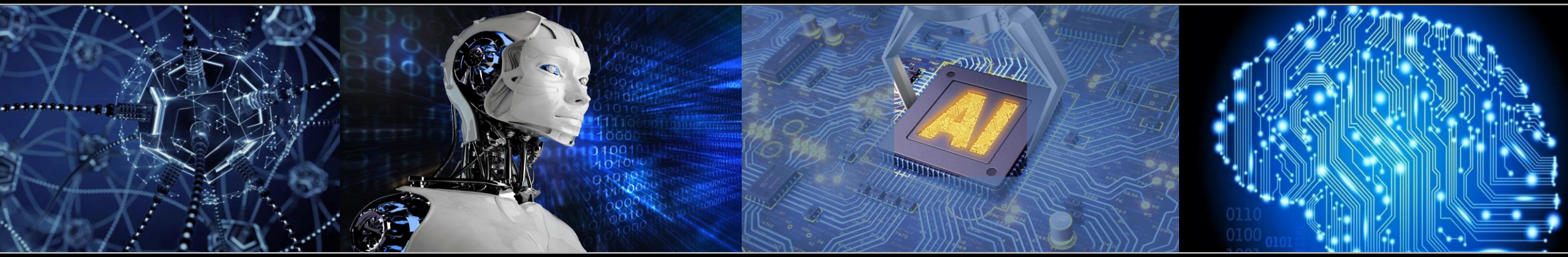


Big social and ethical issues in artificial intelligence

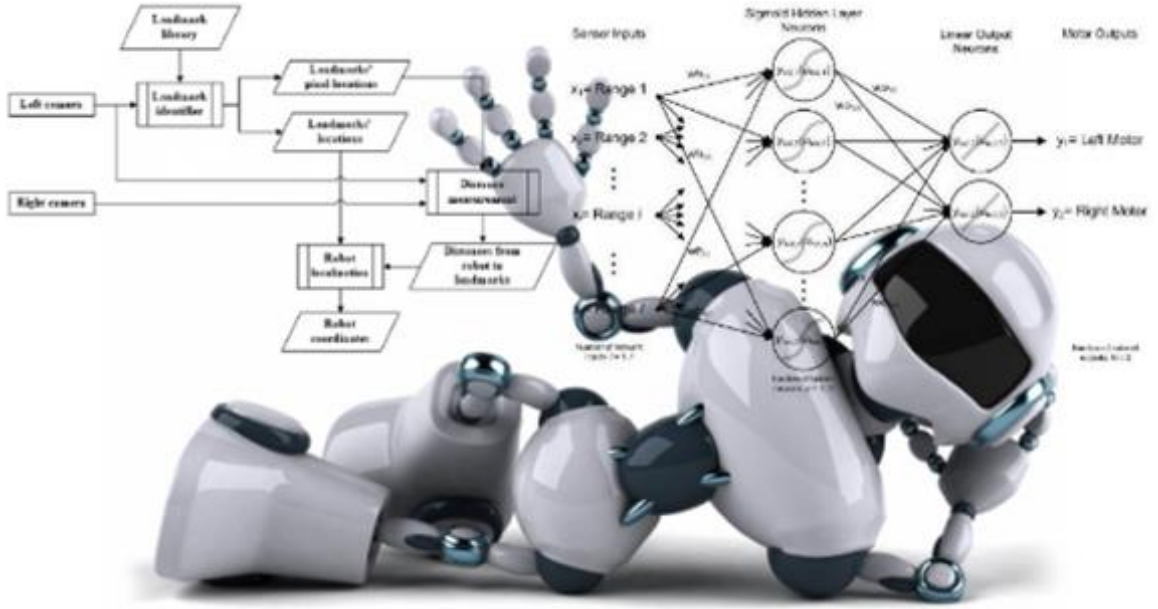
Prof. dr. Philip Brey, University of Twente

Disclaimer: This presentation and its contents reflects the view of the presenter and not the view of SIENNA or the European Union.

sienna.

Presentation outline

1) Defining AI
2) AI systems and techniques
3) General social and ethical issues with AI
4) Ethical issues in AI applications
5) Towards responsible AI



sienna.

Defining AI

Artificial intelligence (AI): the theory and development of computer systems able to perform tasks normally requiring human intelligence …

… such as visual perception, speech recognition, decision-making, and translation between languages.

AI systems and techniques

- Knowledge-based systems

  Systems that can reason and use a knowledge base to solve complex problems

- Natural language processing

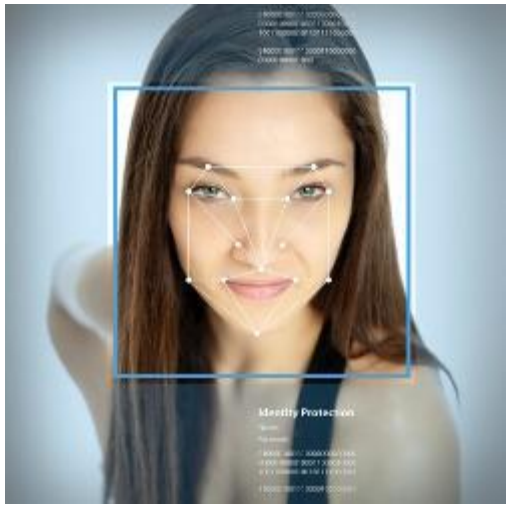
  Systems with capabilities of understanding and producing spoken and written language

- Computer vision

  Systems that are able to r processing, analyze and understand digital images and video

- Intelligent agents

  Software programs that have assigned goals and can perform tasks towards these goals independently, and can make their own decisions doing so.



- Affective Computing

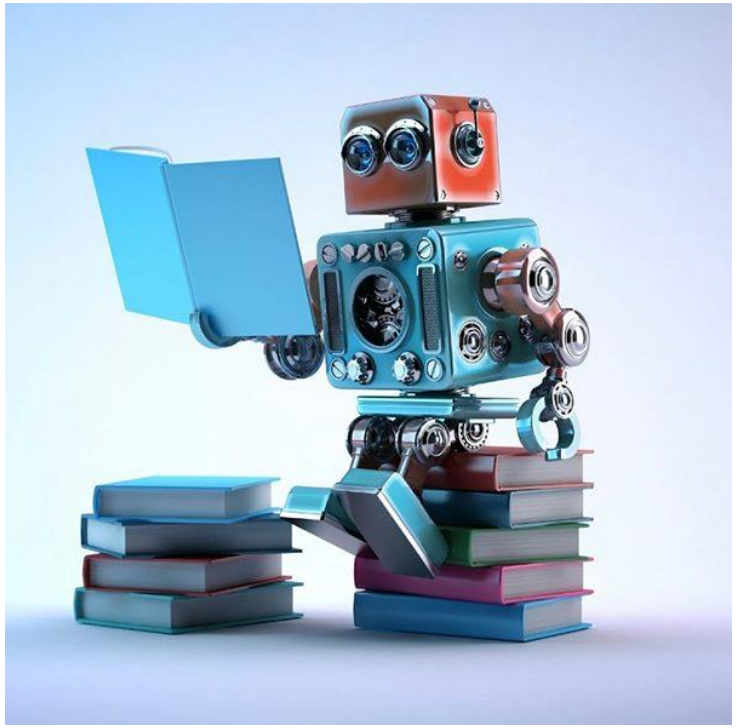
  The development of systems capable of recognizing, understanding and simulating human emotions

- Smart Big Data

  The combination of big data with AI techniques, including intelligent data analytics and machine learning

- Machine learning

  The capability of computer systems to learn by generalizing from a set of examples fed to them, using suitable algorithms and mathematical models



- Embedded AI

  The embedding of AI capabilities in everyday products, making them smart

- Intelligent robotics

  The combination of AI and robotics, resulting in smart electro-mechanical machines that can perform human tasks autonomously

sienna.

General social and ethical issues with AI

(1) Loss of control

Will we still be able to make our own decisions and perform our own actions?

AI systems may decide for us

We are not always the owner

Even if we are, we are not usually the programmer

Even if we are, the system may not act predictably

Will we lose control over our lives?

sienna.

(2) Safety and security

Will intelligent systems (especially those that interact with the real world, like robots and embedded systems) be safe?

Can we prevent the hacking of such systems?

sienna.

(3) Privacy

Will the combination of AI and machine learning with sensors, computer vision, affective computing, natural language processing and big data kill our privacy?

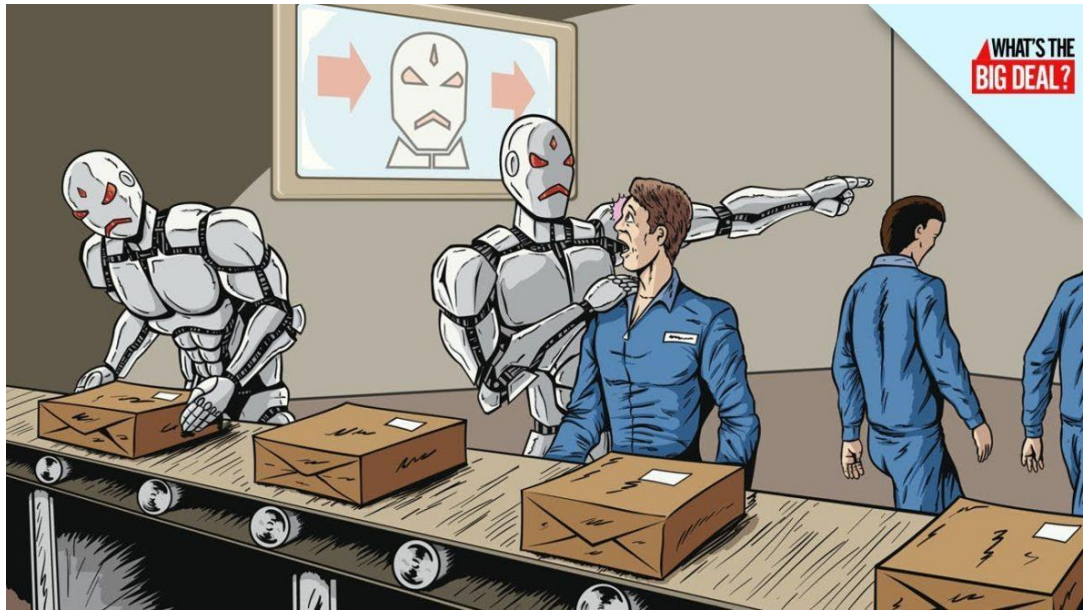
(4) Mass unemployment

Will AI and robotics eliminate a large percentage of existing jobs?

Will there be new jobs to compensate?

Will the needed skills in the new job market be attainable by all?

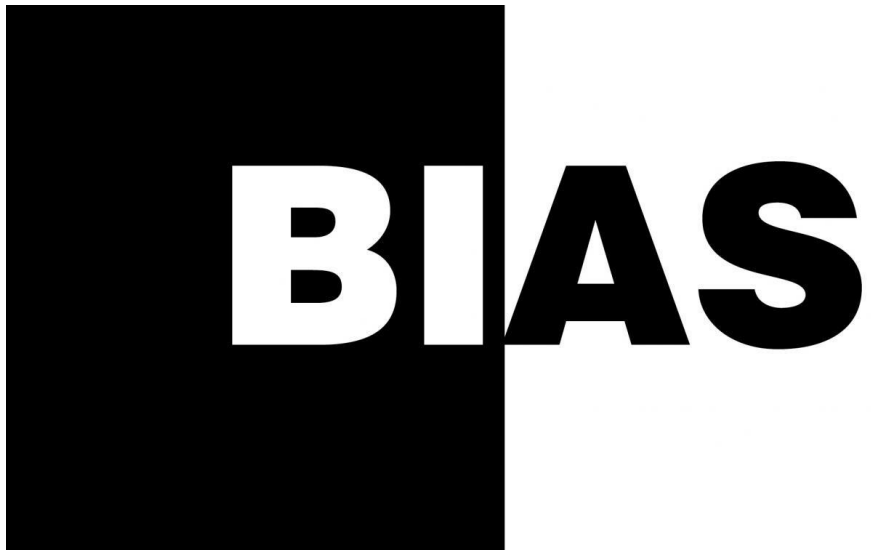
What do we do if many of the disappearing jobs are not replaced?



(5) Bias and inequality

Will AI increase inequality in society or will it decrease it? What can we do to ensure it does not increase inequality?

AI systems (especially big data systems) may contain biases in the way they represent and treat individuals and groups, leading to unfair treatment. How do we prevent this?



(6) Responsibility, accountability and explainability

Who is responsible for the decisions taken by AI system, especially when errors are made and harm is done?

Are there decisions that AI systems should never make?

Should we require algorithmic accountability and transparency?

Should we require that the actions of AI systems are explainable at all times?

Ethical issues in AI applications

Ethical issues that are **specific to particular applications** of AI technology in sectors like education, healthcare and entertainment

Transportation & infrastructure

- Smart scheduling for transportation services
- Energy management, optimization and distribution
- Sensing and predictive analysis for water management
- Predictive modelling of user characteristics for transportation services (future)
- Predictive neighbourhood analysis for urban planning activities (future)

sienna.

Ethical issues in transportation & infrastructure

- Public safety issues
- Issues of privacy and data management
- Issues of responsibility/liability
- Forced-choice decisions (crash ethics)
- Issues in relation to trusting autonomous vehicles
- Issues of responsibility/liability (responsibility gap)

Healthcare

- Clinical decision support
- Patient monitoring and coaching
- Preventative medicine
- Automated image interpretation (future)
- Personalised diagnosis and treatment involving DNA analysis (future)

sienna.

Ethical issues in healthcare

- Issues of privacy and data management (in relation to patient information, data ownership/viewership)
- Issues of quality of care and patient integrity/safety
  (e.g., inaccurate diagnoses made by AI, overconfidence in the use of AI systems by doctors)
- Issues of responsibility/accountability
- Potential inequalities in patient care

Finance & insurance

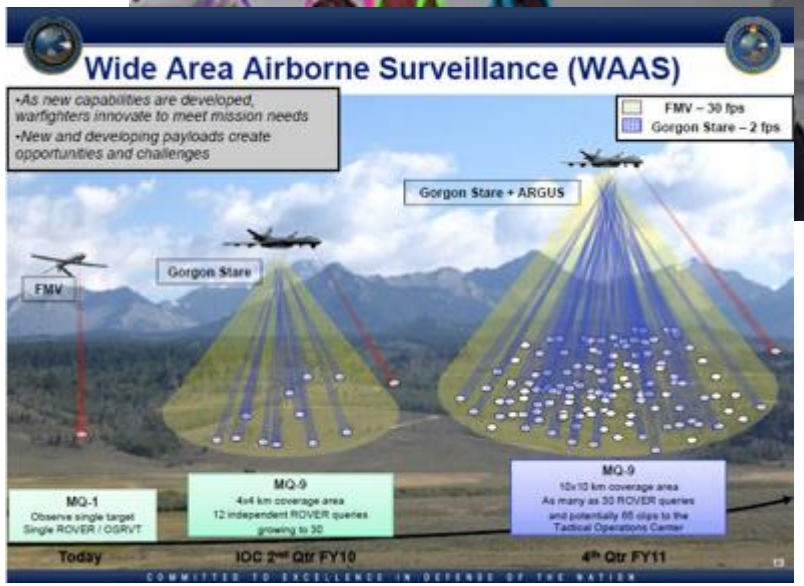
- Algorithmic trading and high-frequency trading
- Automated financial advice and portfolio management
- Underwriting for credit and insurance industries
- Big data for market analysis and automated trading (future)

Ethical issues in finance & insurance

- Risk of catastrophes in algorithmic trading and high-frequency trading due to system errors or design flaws
- Security issues and misuse
  (e.g., hacking, manipulation of markets)
- Issues of responsibility/liability (responsibility gap)
- Objectification of customers in personal finance and insurance
- Issues of privacy and data management
- Algorithmic bias

Law enforcement & defence

- Automated analysis of camera footage for surveillance and predictive policing
- Detection of financial fraud
- Evidence gathering from personal electronic devices
- Advanced cyber defence and weapons systems (future)
- Large-scale surveillance systems using wide-area imagery (future)



sienna.

Ethical issues in law enforcement & defence

- **Law enforcement**
  - Algorithmic bias issues
  - Privacy and data management issues in law enforcement surveillance systems
  - Chilling effect on society due to excessive surveillance
  - Unequal burden of surveillance, discriminatory targeting
  - Risk of function creep in AI systems for public surveillance
  - Potential for abuse and errors (overconfidence in performance of AI surveillance systems)
  - Shifting ethical norms regarding public surveillance
- **Defence**
  - Distancing with regard to military targets
  - Responsibility/accountability for the actions of military AI systems

Towards Responsible AI

AI governed by ethical guidelines and responsible practices

- Responsible automation and job restructuring
- Respecting privacy of citizens, customers and employees
- Avoiding bias and discrimination in AI systems
- Reducing risks of error, safety failures and security risks
- Algorithmic accountability and transparency
- Maintaining meaningful human control