Week 1: What is the ethics of AI?

What is Artificial Intelligence (AI)?

- Coined by John McCarthy in 1956 at the Dartmouth conference.
- Defined as "the science and engineering of making intelligent machines".

What can AI systems do?

- Identify objects and people in images and videos
- Identify cancer in tissues
- Beat humans at Jeopardy, chess, and Go
- Control robots, vehicles, and weapons
- Translate between languages
- Generate images and compose music
- Draft an academic essay
- Etc.

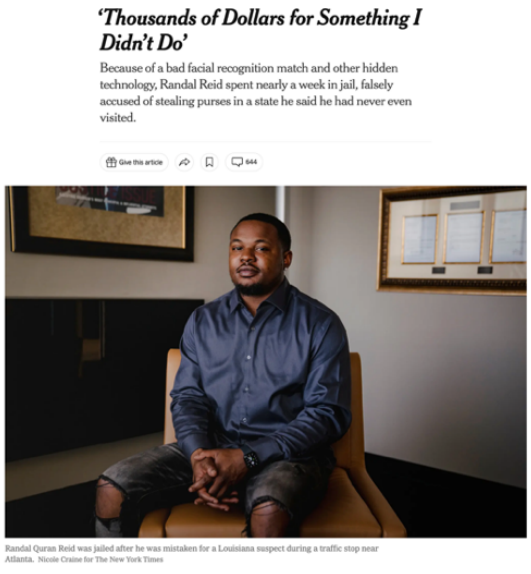
Different types of AI algorithms

- Symbolic AI, GOFAI, or rule-based algorithms:
  - Dominated AI research from the 1950s to the 1980s.
  - Implement explicitly programmed if-then rules (i.e., a decision tree) that transform inputs to outputs.
- Machine Learning (ML) algorithms:
  - Learns a function that transforms inputs to outputs by training on data sets.
  - Perform cognitive tasks without modelling symbolic reasoning or logic.
  - Driver of recent breakthroughs in AI research.

Why is AI a topic of ethics and philosophy?* *Reasons listed below are by no means exhaustive.

Reason 1: AI systems are already causing ethically troublesome consequences.

Hill, K. & Mac, R. (2023, March 31) 'Thousands of Dollars for Something I Didn't Do' *The New York Times*. https://www.nytimes.com/2023/03/31/technology/facial-recognition-false-arrests.html



'Thousands of Dollars for Something I Didn't Do'

Because of a bad facial recognition match and other hidden technology, Randal Reid spent nearly a week in jail, falsely accused of stealing purses in a state he said he had never even visited.

Randal Quran Reid was jailed after he was mistaken for a Louisiana suspect during a traffic stop near Atlanta. Nicole Craine for The New York Times

Ethical challenge: How should we develop and implement AI systems to avoid them from bringing about ethically problematic consequences?

Ethical guidelines for AI development and implementation

- AI research cannot just aim to develop more powerful and scalable systems.
- AI systems that take on social cognitive tasks must also meet social requirements:
  - Transparency
  - Predictability
  - Robustness against manipulation
  - Responsibility
  - Etc.

But these are not just technological problems, but also social political problems.

→ *Module 3 Fairness, biases, and discriminations in AI systems.*

Reason 2: AI systems are starting to raise questions about are relationship with them.

Moral status of artificial agents

Grades of moral status

- Rocks
- Human person
- Embryos – moral permissibility of abortion
- Non-human animals – moral permissibility of animal exploitation

Ethical challenge:

- Should we grant moral status to AI systems?
- Should we treat them like rocks, persons, non-human animals, or differently from them any of them?

Two criteria for moral status:

- Sentience: the capacity for phenomenal experience or qualia
- Sapience: a set of capacities associated with higher intelligence

Following this view, should the moral status of AI systems depend on their intrinsic capacities?

➔ *Module 2: Living with artificial agents*

Reason 3: Future AI systems might pose unprecedented existential risks to humanity.

Artificial General Intelligence (AGI)

- Locally pre-programmed machines (e.g. toaster)
- Domain-specific AI (e.g. chess algorithm Deep Blue)
- Artificial General Intelligence – Applicable beyond prespecified tasks and domains

Superintelligence

- Agents equipped with intelligence that exceed human capacity.
- Intelligence explosion (Good 1965): AI intelligent enough to redesign its design will eventually turn into a system much smarter than humans.
- An example of "existential risks" – "an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential" (Bostrom and Yudkowsky 2014, p. 329).
- E.g. Paperclip maximizing superintelligent agents (Bostrom 2014)

Ethical challenge: How can we avoid such negative consequences?

- Machine ethics (or machine/artificial morality)
- How can we make machines ethical?
- How can we engineer ethical cognition in machines?

➔ *Module 1: Long-term issues in AI ethics*

Discussion question

How might the rapid development of large language models like ChatGPT affect education? What risks and benefits can we anticipate them to bring about? Would they affect students, educators, and other stakeholders differently? If so, how?