Ethics in Applied AI

Ethics in Applied Artificial Intelligence
POLS200 - Ethics in Social Sciences

YASIN KUTUK, Ph.D.*

*Department of Economics - Altınbaş Univ.

2022-02-03

Outline

Definition of AI

Artificial Intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals (EC, 2011).

How do we define intelligence?

A straightforward definition is that intelligent behaviour is 'doing the right thing at the right time'.

Legg and Hutter (2007) - informal definitions. Intelligence is

- a property that an individual agent has as it interacts with its environment or environments.
- related to the agent's ability to succeed or profit with respect to some goal or objective.
- depends on how able that agent is to adapt to different objectives and environments.

Outline

Narrow-AI?

All present-day AIs and robots are examples of what we refer to as **'narrow' AI**: a term that reflects that fact that current AIs and robots are typically only capable of undertaking one specialised task.

Artificial General Intelligence

A long-term goal of AI and robotics research is so-called **artificial general intelligence (AGI)** which would be comparable to human intelligence.

Outline

Machine Learning (ML)

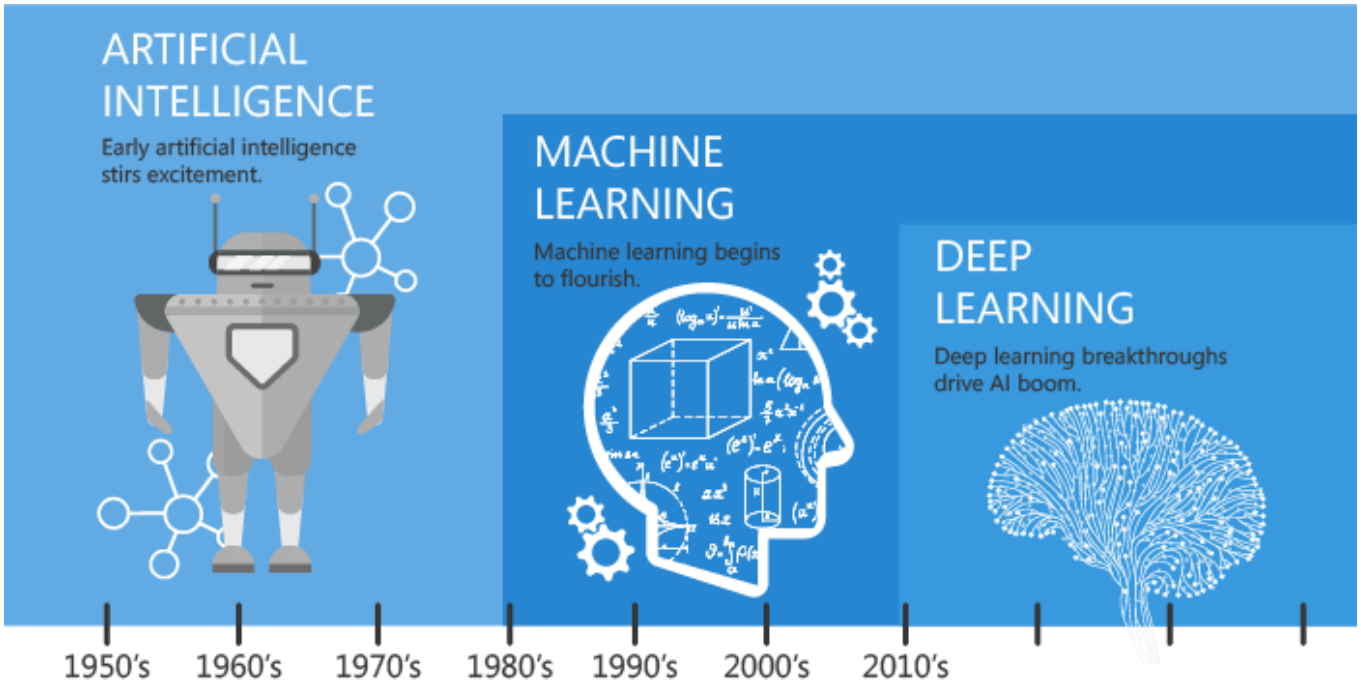
Machine learning is the term used for AIs which are capable of learning or, in the case of robots, adapting to their environment.

Artificial Neural Networks (ANNs)

Supervised learning systems generally make use of **Artificial Neural Networks (ANNs)**, which are trained by presenting the ANN with inputs (for instance, images of animals) each of which is tagged (by humans) with an output (i.e. giraffe, lion, gorilla).

Deep Learning (DL)

The term **deep learning** simply refers to (typically) supervised machine learning systems with large (i.e. many-layered) ANNs and large training data sets.



source: medium.com@alanb_73111

It is important to note the terms AI and ML/DL are not synonymous. Many highly capable AIs and robots do not make use of ML.

Ethics and its Relation to Applied AI

Ethics are moral principles that govern a person's behaviour or the conduct of an activity. As a practical example, one ethical principle is to treat everyone with respect. Philosophers have debated ethics for many centuries, and there are various well-known principles, perhaps one of the most famous being Kant's categorical imperative 'act as you would want all other people to act towards all other people' (Kant, 2008).

Ethical Impacts of Applied AI

Within the last 5 years AI ethics has shifted from an academic concern to a matter for political as well as public debate.

The increasing ubiquity of smart phones and the AI-driven applications that many of us now rely on every day, the fact that AI is increasingly impacting all sectors (including industry, healthcare, policing the judiciary, transport, finance and leisure), as well as the seeming prospect of an AI 'arms race', has prompted an extraordinary number of national and international initiatives, from NGOs, academic and industrial groupings, professional bodies and governments (Bird et al., 2020).

Outline

The labour market

- Automation has often substituted for human labour in the short term, but has led to the creation of jobs in the long term (David, 2015).
- Robotics added an estimated 0.4 percentage (Graetz and Michaels, 2015).
- Vast increases in income inequality, large numbers of unemployable people, and breakdowns in the social order (Smith and Anderson, 2014).

Inequality

- Concentration of power among elites (Nemitz, 2018).

Privacy, human rights and dignity

- Intelligent Personal Assistants
- Big Data
- Facial Recognition, Data Mining
- Brexit Case (Cadwalladr, 2017).

Bias

- AI is created by humans, which means it can be susceptible to bias.
- COMPAS Case (Kirchner et al., 2016).

Democracy

- Fake news and social media (Gorodnichenko et al., 2021).
- At least 28 countries — including both authoritarian states and democracies — employ 'cyber troops' (Bradshaw and Howard, 2017).
- The end of democracies? (Bartlett, 2018).

Outline

Impact on Financial System

- Market manipulation (Spatt, 2014).
- Collusion (Ezrachi and Stucke, 2017).
- Accountability (Wellman and Rajan, 2017).

Outline

Impact on Legal System

The most important near-term legal question associated with AI is who or what should be liable for tortious, criminal, and contractual misconduct involving AI and under what conditions.

Criminal law

- Liability: bankruptcy of the corporation? (Pagallo, 2018)
- Commerce, financial markets and insolvency (Farmer and Skouras, 2013).
- Offences Against the Person, Individual Rights (Citron and Chesney, 2018).

Tort law

- Tort law covers situations where one person's behaviour causes injury, suffering, unfair loss, or harm to another person.
- Two purposes: compensation, deterrioration.

Outline

Use of natural resources

- Exploitation of raw resources.
- Mining and misuse of metals (Khakurel et al., 2018).

Pollution and wast

- Electronic waste.
- Sustainability (Guiltinan, 2009).

Energy concerns

- Supply problems.
- Carboon footprints (Strubell et al., 2019).

Ways AI could help the planet

- Reducing gas emmissions (Igliński and Babiak, 2017).
- Make biodiversification efficient.

Outline

Impact on Trust

- The overwhelming consensus amongst the research community is that trust in AI can only be attained by fairness, transparency, accountability and regulation.
- How much control we want to exert over AI machines, and if, for example we want to always maintain a human-in the loop, or give systems more autonomy.

Fairness

Four requirements: Corbett-Davies et al. (2017)

- Statistical parity,
- Conditional statistical parity,
- Predictive equality,
- Calibration.

Transparency

- Autopiloting.
- Being a **black box**, Kroll (2018).
- Explainable systems.
- Intentional understanding.

Accountability

- Tesla Model S, Uber cases.
- Regulations, (Winfield and Jirotka, 2018).

Control

- Jumping to Superintelligence?
- Human in the loop (HITL) (Rahwan, 2018).
- Hell Yeah: THE BIG RED BUTTON (Orseau and Armstrong, 2016).

Final Thoughts

- We need industrialisation and productivity.
- Make research.
- Re-trainable programs.
- Collaboration (transparent).
- Inclusive Social Development.
- Responsibility.

References I

Bartlett, J. (2018). *The People vs Tech: How the Internet is killing democracy (and how we save it)*. Random House.

Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., and Winfield, A. (2020). The ethics of artificial intelligence: Issues and initiatives. *European Parliamentary Research Service*.

Bradshaw, S. and Howard, P. (2017). Troops, trolls and troublemakers: A global inventory of organized social media manipulation.

Cadwalladr, C. (2017). Revealed: How us billionaire helped to back brexit. *The Guardian*, 26.

Citron, D. K. and Chesney, R. (2018). Deep fakes: A looming crisis for national security, democracy and privacy? *Lawfare*.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., and Huq, A. (2017). Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, pages 797–806.

David, H. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of economic perspectives*, 29(3):3–30.

References II

EC, E. (2011). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions. *A roadmap for moving to a competitive low carbon economy in*, 2050.

Ezrachi, A. and Stucke, M. E. (2017). Two artificial neural networks meet in an online hub and change the future (of competition, market dynamics and society).

Farmer, J. D. and Skouras, S. (2013). An ecological perspective on the future of computer trading. *Quantitative Finance*, 13(3):325–346.

Gorodnichenko, Y., Pham, T., and Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from# brexit and# uselection. *European Economic Review*, 136:103772.

Graetz, G. and Michaels, G. (2015). Robots at work. centre for economic performance. *CEP Discussion Paper*, (1335).

Guiltinan, J. (2009). Creative destruction and destructive creations: environmental ethics and planned obsolescence. *Journal of business ethics*, 89(1):19–28.

References III

Igliński, H. and Babiak, M. (2017). Analysis of the potential of autonomous vehicles in reducing the emissions of greenhouse gases in road transport. *Procedia engineering*, 192:353–358.

Kant, I. (2008). *Groundwork for the Metaphysics of Morals*. Yale University Press.

Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., and Zhang, W. (2018). The rise of artificial intelligence under the lens of sustainability. *Technologies*, 6(4):100.

Kirchner, L., Mattu, S., Larson, J., and Angwin, J. (2016). Machine bias—propublica.

Kroll, J. A. (2018). The fallacy of inscrutability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180084.

Legg, S. and Hutter, M. (2007). A collection of definitions of intelligence.

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180089.

Orseau, L. and Armstrong, M. (2016). Safely interruptible agents.

References IV

Pagallo, U. (2018). Apples, oranges, robots: four misunderstandings in today's debate on the legal status of ai systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180168.

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1):5–14.

Smith, A. and Anderson, J. (2014). Ai, robotics, and the future of jobs. *Pew Research Center*, 6:51.

Spatt, C. (2014). Security market manipulation. *Annu. Rev. Financ. Econ.*, 6(1):405–418.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*.

Wellman, M. P. and Rajan, U. (2017). Ethical issues for autonomous trading agents. *Minds and Machines*, 27(4):609–624.

Winfield, A. F. and Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180085.

Thanks

for your convenience...

yasin.kutuk@altinbas.edu.tr

Google Scholar Profile: Link

Github Profile [1]: Link

[1] includes source code of this presentation and its compiled PDF.