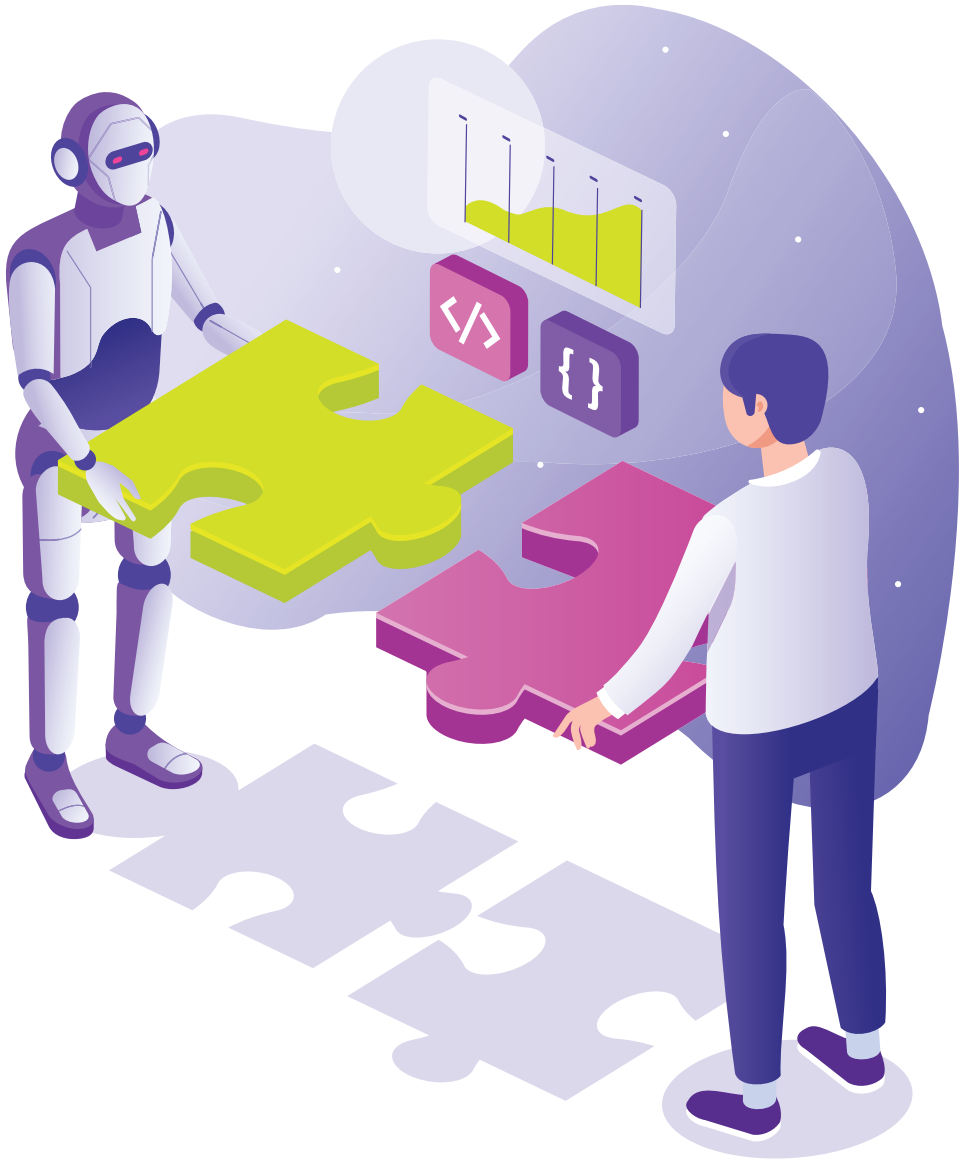
Ethics and AI: Teaching Our Machines to Tell Right from Wrong

DR HAN YU
Nanyang Assistant Professor, School of Computer Science and Engineering, Nanyang Technological University

From recommending products on Taobao.com to assessing an individual's creditworthiness based on their behaviour on social media app WeChat, Artificial Intelligence (AI) is becoming an integral part of daily lives for millions of people. As a result of its rapid development and growing significance, ethical issues in AI have become a point of public debate and discussion.



In particular, a number of recent accidents related to autonomous vehicles has brought the topic to the forefront of public attention. In addition, the recent large-scale study conducted by the MIT Moral Machine project[1] further reveals the ethical dilemmas facing autonomous vehicle designers, passengers, and other road users and the complexity involved in getting a society to agree on the ethics governing AI applications. These developments suggest that there is a need for a social contract between AI and society.

[1] MIT Moral Machine project http://moralmachine.mit.edu/

STEP 01

Figuring out What is Right

As of now, the AI research community has agreed on some desirable qualities of an ethical AI system. These include guidelines such as: AI applications should respect and protect user privacy; decisions made by AI should be fair, unbiased and explainable to human beings; and for the purpose of accountability, responsibility attribution should be possible if something goes wrong. Various groups are also researching ethical dilemmas, individual and collective ethical AI decision-making, and ethical human-AI interactions[2]. However, while pockets of advances have been made, few of them have reached the stage where research outcomes can be deployed in real-world AI applications.

Notably, acknowledgement of the importance of incorporating ethics into AI has also led to the engagement of the AI research and engineering community in a number of global initiatives. One such initiative is the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, established by the Institute of Electrical and Electronics Engineers (IEEE). Since its inception, the Initiative has produced the Ethically Aligned Design (EAD) report which outlines principles, guidelines and best practices for developing and governing future AI empowered systems.

STEP 02

Getting Buy-in to Best Practices

The Association for the Advancement of Artificial Intelligence (AAAI) has also teamed up with the Association for Computing Machinery (ACM) in 2018 to organise the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)[3]. The Conference provides a platform for AI researchers and social scientists to come together and work out interdisciplinary solutions to ethical challenges in AI applications.

STEP 03

Taking Positive (Not Limiting) Actions

However, without waiting for ethical AI technologies to be ready, the legislative landscape has already evolved. In 2016, the European Union (EU) established one of the most stringent privacy protection laws targeting AI applications with the General Data Protection Regulation (GDPR). The GDPR specifies many terms aimed at protecting user privacy and prohibiting organisations from exchanging data without explicit user approval. Similar laws have also since emerged in China and the US. These harsh legal environments threaten to impede AI development by making it infeasible for different companies who own diverse types of user data to collaborate and build new business.

Fortunately, the AI research community has an appropriate response to the imposed legal challenges. Through introducing a new paradigm of machine learning – federated learning, different data owners can continue to collaborate and collectively train a model by storing data locally and observing secure protocols such as homogeneous encryption, differential privacy, and secret sharing to prevent user privacy breach[4].

STEP 04

Empowering Continuous Improvement

One of the proponents for federated learning is the Federated AI Ecosystem (FedAI)[5] led by Professor Qiang Yang, Chief AI Officer (CAIO) of WeBank. Besides providing a global open platform for research, development, and deployment of federated learning technologies in areas with strong data privacy concerns such as banking and healthcare, FedAI is also dedicated to fostering an inclusive environment for open source development by making available source code for the building blocks of federated learning – the Federated AI Technology Enabler (FATE) – as well as tutorial materials which enable AI researchers and engineers to create more complex and capable privacy preserving machine learning technologies compliant with stricter laws governing AI.

Indeed, this approach has the potential to enable AI to continue its strong development trajectory forward even as the legal environment becomes tougher. The future of AI looks bright and it is an exciting time to work in this domain. But much of the road ahead tests our resolve to not only make AI a fair tool, but also a sustainable systemic technology that brings about a more just society for all.

---

2 H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser & Q. Yang, "Building Ethics into Artificial Intelligence," in Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18), pp. 5527–5533, 2018.

3 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) http://www.aies-conference.com/

4 Q. Yang, Y. Liu, T. Chen & Y. Tong, "Federated Learning: Concepts and Applications," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, no. 2, pp. 12:1–12:19, 2019.

5 Federated AI Ecosystem (FedAI) https://www.fedai.org/