Building a case for decentralised AI ethics

James Brusseau

Published on Oct 18, 2021

> A user-centric bottom-up approach to ethics in AI could be the answer to regulating AI innovation that is usually too fast and dynamic for a central regulator to keep up with

The idea behind decentralised Artificial Intelligence (AI) ethics is simple: Instead of a single regulating authority, masses of individuals actively direct technologies through their own informed decisions and uses. Countries can build robust frameworks to shape the trajectory of new AI technologies by empowering all users to participate. Decentralised finance provides an analogy: For cryptocurrency exchanges and smart-contract enforcement, it is users themselves who verify the transactions and legitimise or ostracise the practices and participants. The difference is that where decentralised finance promotes economic growth, decentralised technological ethics facilitates human-centered AI innovation.

The bureaucratic processes of centralised regulation are too slow for the pace of contemporary innovation. In AI medicine, for example, the University of Copenhagen is rendering health diagnoses from language patterns detected in emergency phone calls. However, regulatory categories fitting this new kind of healthcare have not yet been conceived, much less passed into law. By nature, when centralised regulation does extend to cover today's breakthroughs, it will stifle future innovation. This is because rules and regulations are made to fit technologies that already exist, meaning they misfit—and consequently hinder—the most inventive projects.

> **By nature, when centralised regulation does extend to cover today's breakthroughs, it will stifle future innovation. This is because rules and regulations are made to fit technologies that already exist, meaning they misfit—and consequently hinder—the most inventive projects.**

Both these problems—speed and innovation suffocation—are solved by decentralised AI ethics. Centralised AI ethics works from the top down, from experts and their determinations down to users and their actions. An illustrative example emerged from the Frankfurt Big Data Lab in 2020. A PhD-level team of philosophers, computer scientists, doctors, and lawyers united to approach AI-intensive startup companies in the field of medicine, and to collaboratively explore the ethical aspects of technological development. The group's work began with lengthy deliberations guided by AI ethics principles, and eventually concluded with case-study reports. These reports have been published in academic journals, where they may be read by institutional lawyers and administrators, and eventually contribute to the promulgation of AI regulation by conventional means. So, the process starts with high-level experts and works down to shape the user experience along a timeline of months and years.

Decentralised AI ethics starts from the ground up. Instead of experts and high-level discussions, the process begins with common and public information. Companies routinely publish quarterly statements, which may include details about installed privacy protections, or efforts to ensure that their products work fairly across diverse populations. There are also news reports and investigative journalism, which may reveal a platform's sloppy privacy safeguards or oppressive censorship

practices. Then there is the endless flow of social media, where users relate their own experiences with AI in medicine, banking, insurance, and entertainment. Regardless of the domain, the process starts with accessible information about real technologies functioning for tangible human beings.

The second element of decentralised AI ethics is universal ethical evaluation. Instead of a boutique service offered at certain times to a select group of companies and products, evaluation of technological innovation occurs always and everywhere. To achieve the speed and breadth, expert deliberators are replaced by natural language processing. Textual machine learning filters common and public information for indicators that reveal how specific technologies are affecting human lives. Ethics is automatic: AI applies AI ethics to AI-intensive companies.

Training AI for this task is a challenge. Nevertheless, initial efforts are set to begin in September 2021 at the University of Trento, Italy. The project builds on six ethical principles widely recognised as well-tailored to AI's interface with human experience. They are: Individual autonomy, individual privacy, social wellbeing, social fairness, technological performance, and technological accountability. The idea is that machine learning can continuously and broadly scour public data and detect whether an AI is serving or subtracting from these principles. So, instead of human experts occasionally analysing a specific technology, we have ethical information about every mainstream digital technology, flowing all the time.

> **The AI Human Impact platform translates the findings of machine learning analyses into results that are open to all, and into a format that is meaningful in financial terms. The result is that individuals—including those managing their own money through fintech—can directly and intelligently participate in monetarily rewarding human-centred AI companies**

The third aspect of decentralised AI ethics is implementation. Diverse users must be empowered to shape new technology through their own informed actions. Ideally, everyone should have access to the ethical insights generated by natural language processing and then be able to apply them on their own. However, in reality, current opportunities are limited. Nevertheless, one example of empowered individuals is AI Human Impact. This project leverages AI ethics to help investors make financial decisions. The premise is that sustainable economic success will accrue to those companies and technologies that serve human purposes as opposed to nudging, manipulating, or exploiting users. Humanist AI is technology that supports user autonomy, ensures data privacy, operates fairly, contributes to social wellbeing, and performs well and accountably. These ethical qualities both cause and predict economic profit. The prediction becomes increasingly confident as accurate information about the ethical performance of technologies becomes more accessible.

The AI Human Impact platform translates the findings of machine learning analyses into results that are open to all, and into a format that is meaningful in financial terms. The result is that individuals—including those managing their own money through fintech—can directly and intelligently participate in monetarily rewarding human-centred AI companies. So, as opposed to a top-down approach, where a narrow range of experts and regulators shape technology by deciding what everyone else is allowed to do, now what diverse people can freely do is the shaping of innovation.

Conclusion

The question this piece posed was, "How can countries build robust frameworks to shape the trajectory of new technologies?" Within the domain of AI, the answer is decentralised AI ethics, which means, first, that the source of the ethical evaluation is not human experts so much as common, public data. Second, the evaluation does not occur through arduous human discussion, but in real

time and as constantly updated by natural language processing and machine learning. Third, the implementation of ethical standards is not executed through governmental or regulatory authorities, but by independent users making informed, personal decisions that collectively shape the trajectory of artificial intelligence.