What if blockchain could guarantee ethical AI?

Blockchain has the potential to promote compliance with traditional ethical principles, especially in the fields of healthcare, supply chain management and food safety. As artificial intelligence (AI) companies and other organisations seek ways to comply with ethical principles and requirements, blockchain could be seen as a means to ensure that AI is deployed in an ethically sound manner, under certain specific conditions.

©Elenabsl / AdobeStock

Blockchains are open, decentralised ledgers that record transactions between two parties without the need for third-party authentication. Their ability to ensure that data are secure, well-protected and reliable, and thus can be shared in a secure and auditable manner, mean that blockchain applications are being used in a growing number of domains, such as healthcare management, cross-border payments and supply chain monitoring. Their implementation raises ethical concerns about security vulnerabilities, environmental impact – given the high amount of computing power needed, accountability, privacy and the apparent enabling of cybercrime. Various policy initiatives have been launched to address these challenges in the form of ethical design frameworks, guiding principles and the Blockchain Code of Ethics. At the same time, blockchain has emerged lately as a carrier of ethical values that could resolve societal challenges of high ethical import in several domains. Can the intrinsic features of blockchain technology help AI developers comply with the multiplicity of ethical demands in their field and in effect contribute to the ethical design and deployment of AI applications?

Potential impacts and developments

Blockchain technology has the potential to create ethical value by creating more transparent and traceable food supply chains to tackle major challenges such as unethical labour practices and environmental degradation. Its ethical value also lies in its ability to provide for secured proof of origin and ethical sourcing. It can also facilitate the sharing of medical data via the automation of some aspects of consent and data collection. Beyond the indirect effects of blockchain on the achievement of certain ethical aims within specific policy domains, this emerging technology appears to offer the means to facilitate the compliance of AI, in its various manifestations, with ethical principles and human rights standards.

It is well known that the quality, accuracy and representative nature of the data needed to train algorithms and develop human-centric machine learning models is central to the ethical soundness of AI applications. However, as there is no oversight mechanism and no standard methodologies to review the fairness of these algorithms or the privacy-friendly nature of the data analytics used, multiple calls have been made for the development of ethics standards and frameworks. The opacity of algorithmic operations and the use of self-learning algorithms for predictive policing, social security or diagnostics is currently at the epicentre of the ethical debate at EU level.

This is precisely where blockchain technologies can play an important role in helping AI applications and systems be designed and implemented in an ethically sound manner. One of the main advantages of blockchain lies in its ability to ensure that data are secure, private, reliable and valid, and thus personal data are not compromised. Therefore, blockchain enables cooperative and safe data-sharing, by cryptographically ensuring the trustworthiness of data. It may therefore be seen as a way to enable users to share their data with trusted stakeholders before the data are collected and processed by powerful AI systems in the context of specific decentralised AI platforms. In other words, the introduction of decentralised blockchain solutions in the context of AI may facilitate the removal of false or incorrect data sets, strengthen the privacy-friendly nature of AI data infrastructure and, essentially, contribute to its ethical design and deployment.

Author: Mihalis Kritikos, Scientific Foresight Unit (STOA)

PE 656.334 – December 2020

Given that blockchain can operate as a transparency machine where its users are assured that the data stored, on a datapoint-by-datapoint basis, have not been tampered with through the use of cryptographic hashing, digital signatures or smart contracts, it can increase the trust that is so necessary in the field of AI. Transparency is in fact one of the seven ethical requirements put forward by the High-Level Expert Group on AI and endorsed by the European Commission. This is also a necessary step to promote and protect the principle of explicability: the need for AI processes to be transparent and explainable.

As blockchain technologies offer users a detailed view of how data are being used, the introduction of these properties into the AI context could potentially help developers to design human-centric and responsible algorithms, and citizens to exercise their right to explanation and to effective remedy. In fact, allowing advanced AI models and large datasets to be widely shared, updated and re-trained could boost trust in algorithmic decision-making systems.

Moreover, blockchain's traceability and data integrity features and its capacity to operate in a decentralised manner could be crucial in ensuring that the data used in AI systems are reliable, of high quality and bias-free. Blockchain's use of immutable records of all the data, variables, and processes used by AI for its algorithmic decision-making processes could enable decision-makers in the field of AI to audit the main tenets of the systems/applications used, review and diversify datasets, set aside data that could lead to false negatives, and identify biased algorithms. That could eventually enable AI applications to be viewed as reliable sources of information and knowledge that could not develop any discriminatory or manipulative effects via deep fakes or predictive behaviour algorithms.

As a result, the accessibility and transparency qualities of blockchain programming, making it possible to audit all steps of the process – from data entry to processing outcomes – could serve as a solid basis for demystifying AI, enhancing the ethical nature of algorithmic decision-making systems and solving the AI black box problem through transparency and algorithmic impact assessments. That way, blockchain, by being publicly auditable, would help the public understand machine learning decisions, thus increasing the explainability of AI systems. Blockchain's recording properties can help AI users and decision-makers trace, review and reframe all variables that feed into decisions made on the basis of machine-learning procedures.

Last but not least, not only could blockchain's ability to operate without intermediaries prevent data manipulation, it could also allow small AI companies to obtain trustworthy data directly from their creators through decentralised blockchain data networks. This is particularly important for the ethical development of AI, as blockchain programming can also create an incentive system that could encourage users to contribute and share their data. Such a system could, in effect, enhance the robustness and fairness of data models, strengthen the quality of algorithmic data sets and bring forward a paradigm shift in the ethical governance of AI applications.

Anticipatory policy-making

As the issue of AI ethics has become a key part of discussions on governance and regulatory control of this transformative technology at both organisational and policy-making levels, new and creative ways need to be found to secure the efficient operationalisation of commonly agreed ethical principles. That requires not only the development of practical implementation guidance but also the employment of new tools.

Blockchain, due to its specific design qualities, can become part of ethical problem-solving in the field of AI in various ways. As legislators across the world seek ways to identify the sources and address the effects of data bias in the context of AI, and to introduce a proportionate risk assessment and management framework, blockchain architecture can become an integral element in the ethics-by-design approach. This concept has been proposed repeatedly by the European Parliament and was also reflected in its recent resolution on the framework of ethical aspects of artificial intelligence, robotics and related technologies.

In the light of the Commission's upcoming legislative proposal on the control of AI and the recently proposed data governance act, the regulatory features of blockchain could offer numerous advantages, including anonymisation, enhanced data security, immutability, and consensus-driven tools. Its integration into the AI world could provide developers and users alike with an ecosystem of modalities and features that would enhance the effective implementation of ethical requirements and principles and, in effect, increase public trust in AI systems. Given the potential benefits of the introduction of blockchain properties into AI platforms, the establishment of EU-wide hybrid pilot platforms that could facilitate the convergence of AI and blockchain architectures could unleash the potential of blockchain as an ethical game-changer in the field of AI.

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament. Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy. © European Union, 2020.