



HOW TO BUILD ETHICAL AI?

Dragana Ignjatović Ristić

Faculty of Medical Sciences, University of Kragujevac, Serbia
e-mail: draganaristic4@gmail.com

Abstract:

Artificial intelligence (AI) is a subfield of IT. The goal of AI research is the development of software, which enable computers to behave in a way that could be characterized as intelligent. This may be described as intelligence demonstrated by machines, as opposed to the natural intelligence we attribute to living beings. The term artificial intelligence was coined in 1956, in Dartmouth, Hanover (USA), during a gathering of researchers interested in topics such as intelligence, neural networks and automata theory.

Although it is accepted that AI brings many benefits, if we consider ethical issues, the implicit assumption is that we are talking about AI's morally problematic aspects. Debates about the ethics of AI inextricably follow the progress of the very concept and application of artificial intelligence (Coeckelbergh 2019, Dignum 2019, Müller 2020).

Technology based on AI already made possible significant achievements such as facial recognition, medical diagnosis, and self-driving cars. AI promises enormous benefits for economic growth, social development, as well as the improvement of human well-being and safety. But what are the perceived ethical and moral issues with AI? What are the general and common ethical principles, rules, guidelines, policies, and regulations that can resolve or attenuate those ethical and moral issues? What are some of the necessary features and characteristics of an ethical AI? How to adhere to the ethics in order to build ethical AI?

The ethics of AI is part of the ethics of advanced technology that focuses on robots and other artificially intelligent agents. It can be divided into robot ethics and machine ethics.

Ethical issues that were identified from the case studies and methodological studies such as the Delphi study include: cost of innovation, violation of physical integrity, lack of access to public services, lack of trust, "Awakening" of AI, security problems, lack of quality data, job shortage, power asymmetries, negative impact on health, integrity problems, lack of accuracy of data, privacy problems, lack of transparency, potential for military use, lack of informed consent, bias and discrimination, unfairness, unequal power relations, misuse of personal data, negative impact on justice system, negative impact on democracy, potential for criminal and malicious use, loss of freedom and individual autonomy, contested ownership of data, reduction of human contact, problems of control and usage of data and systems, lack of accuracy of predictive recommendations, lack of accuracy of non-individual recommendations, concentration of economic power, violation of fundamental human rights in the supply chain, violation of fundamental human rights of end users, unintended, unforeseeable adverse impacts, prioritisation of the "wrong" problems, negative impact on vulnerable groups, lack of accountability and liability, negative impact on the environment, loss of human decision-making, lack of access to information and compromised freedom of information.

This paper aims to call attention to the urgent need for various stakeholders to pay attention to the ethics and morality of AI agents. AI ethics should be the central consideration in developing AI agents. The future of humanity may depend on the correct development of AI ethics.

We can build ethical AI only if we behave ethically in the field of AI.

Keywords: artificial intelligence, ethics