ARTIFICIAL INTELLIGENCE AND HUMANITY'S FUTURE: RISKS, RESPONSIBILITIES, AND GLOBAL IMPLICATIONS

Artem Nazarchuk

Faculty of Electronics,

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Artificial intelligence (AI) has become significant research focus due to its profound impact on humanity's future. In Superintelligence: Paths, Dangers, Strategies, Nick Bostrom examines scenarios where superintelligent AI – capable of surpassing human intellect – could present both challenges and threats. He warns that AI's goals might not align with human values, risking loss of control. Thus, Bostrom stresses the need for safety mechanisms to manage AI, as the potential consequences could be global.

Other researchers emphasize the importance of international collaboration in AI development. Bostrom introduces the "singleton" concept – a centralized organization overseeing superintelligence. This approach could help avoid an AI arms race and reduce conflicts among nations or corporations that may develop powerful systems without ensuring their safety.

In Life 3.0, Max Tegmark views AI development as a transformative step in human evolution, reshaping society fundamentally. He highlights AI's promising applications in science and medicine but stresses the need for responsible development and ethics. Tegmark describes three stages of life – biological, cultural, and technological – where AI could significantly enhance life, though it also demands safeguards to prevent misuse.

In addition to these risks and challenges, AI has unprecedented potential to benefit humanity if developed with socially valuable goals. For instance, in healthcare, AI assists in diagnosing diseases, predicting health risks, and creating new treatments. Economically, AI optimizes processes, reduces costs, and supports sustainable development. However, as researchers note, maximizing these benefits requires not only technological progress but also efforts in ethics, law, and

international cooperation to prevent global inequalities and potential misuse.

In summary, the issue of AI control and ethical alignment is increasingly critical. Developing transparent and safe algorithms, alongside broad societal involvement in decision-making, is essential to harness AI's benefits without endangering humanity's future.

References:

1.   Miller, Z. (2015, January 18). *Superintelligence – Paths, Dangers, Strategies.* Yale Scientific Magazine. Retrieved from https://www.yalescientific.org/2015/01/book-review-superintelligence-paths-dangers-strategies/

2.   The AI Track Team. (2024, July 5). *Superintelligence by Nick Bostrom: Detailed Summary.* The AI Track. Retrieved from https://theaitrack.com/superintelligence-paths-dangers-strategies-summary/

3.   Nonfiction Summary. (n.d.). *Superintelligence by Nick Bostrom Book Summary.* Retrieved from https://www.nonfictionsummary.com/book-summary/superintelligence-by-nick-bostrom/

4.   Bryson, J. J. (2019, December 18). *The future of AI's impact on society.* MIT Technology Review. Retrieved from https://www.technologyreview.com/2019/12/18/102365/the-future-of-ais-impact-on-society/