29. ETHICAL ISSUES OF AI – DEEPFAKE THREATS, BIAS IN MODELS, COPYRIGHT

Dorosh M.D., Yevtukh T.V.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Ryabykh V.A. – Lecturer, Master of Arts

This article discusses three key ethical challenges of AI: deepfake threats, bias in AI models, and copyright issues. It examines approaches to solve these problems in order to make AI safer and more reliable for common users.

Artificial Intelligence (AI) is transforming the world, offering incredible opportunities across industries. However, its rapid development also brings significant ethical challenges. Three of the most pressing issues are the rise of deepfake threats, bias in AI models, and copyright concerns. These problems highlight the need for careful regulation and ethical guidelines to ensure AI is used responsibly and fairly.

Deepfakes are AI-generated videos, images, or audio that appear incredibly realistic. Although they offer benefits in areas like entertainment and education, they are increasingly being misused for harmful purposes. Cybercriminals use deepfakes to deceive people, spread misinformation, and commit fraud. For example, a deepfake video might impersonate a CEO, instructing employees to transfer funds or share sensitive information. Deepfakes are particularly dangerous because they rely on social engineering—manipulating people's emotions to make them act quickly without thinking. To combat this threat, organizations need a combination of technology and human awareness. AI tools can help detect deepfakes by identifying inconsistencies, such as unnatural facial movements or mismatched audio. At the same time, employees must be trained to recognize and respond to potential deepfake attacks [1].

AI systems learn from data, but if the data is biased, the AI produces biased results. This can lead to unfair or discriminatory outcomes. For instance, the facial recognition systems have been shown to perform poorly on people with darker skin tones, reflecting a lack of diversity in training data. Similarly, AI-driven hiring tools may favor certain demographics, perpetuating workplace inequality. Bias in AI often stems from incomplete or unrepresentative datasets, as well as the unconscious biases of developers. When left unchecked, these biases can reinforce systemic inequalities and erode trust in AI systems. For example, biased algorithms used in predictive policing or loan approvals can disproportionately harm marginalized communities [2].

Developers need to emphasize inclusivity and equity in their data to resolve this issue. Regular audits of AI systems can help identify and correct biases. Additionally, ethical guidelines and regulations should be established to ensure AI is used in a way that promotes equality and justice.

Generative AI, such as DALL·E and ChatGPT, can create new content, including images, music, and text. However, this raises complex questions about copyright and intellectual property (IP). Who owns the

content produced by AI? Is it the user, the AI developer, or the original creators whose works were used to train the AI? The current copyright laws were not designed to handle AI-generated content, leading to legal uncertainties. For example, if an AI creates a piece of art that resembles a copyrighted work, does it infringe on the original creator's rights? Several lawsuits have already been filed by the artists and companies claiming that their work was used unauthorised use to train AI models.

To resolve these issues, clear rules and guidelines are needed. AI developers should obtain proper licenses for the data used to train their models and ensure that original creators are compensated. Policymakers must also update copyright laws to address the unique challenges posed by AI-generated content. AI has the potential to revolutionize society, but it also presents significant ethical challenges. Deepfakes can be used to deceive and manipulate, biased AI models can perpetuate inequality, and copyright disputes can harm creators. To harness the benefits of AI while minimizing its risks, we need strong ethical frameworks, transparency, and collaboration among stakeholders.

Governments, companies, and individuals must work together to address these issues. Key steps include funding technology that identifies deepfakes, fostering diversity and fairness in AI design, and reforming copyright policies to uphold creators' protections. By taking these steps, we can ensure that AI is used in a way that benefits everyone and upholds ethical standards.

.

References:

1. Deepfakes: an emerging cyber threat that combines AI, realism and social engineering [Electronic resource]. - Mode of access: https://axaxl.com/fast-fast-forward/articles/deepfakes_an-emerging-cyber-threat-that-combines-ai-realism-and-social-engineering – Date of access: 17.03.2025.

2. Generative AI Has an Intellectual Property Problem [Electronic resource] - Mode of access: https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem – Date of access: 14.03.2025.