Ethics in Computing in the Era of Generative AI

A Reflective Piece on Correa et al.'s 2023 article

1. Introduction

Artificial intelligence has never been a more relevant topic than it is today. And with the boom in AI usage, is the question of what is ethical to do when using AI? Deckard (2023) states that those working with AI need to be able to balance their own interests with the principles and values that ensure that AI technologies are "fair, transparent, and accountable".

Correa et al. (2023) have written an article that covers 200 different articles to understand the guidelines and recommendations for AI governance around the world. When reading this article, the following figure was alarming. Out of the 200 documents Correa et al. (2023) analyzed, 91.6% opt for "soft" forms of regulation. When considering the damage that artificial intelligence can cause, which will be detailed further within the article, there must be harsher regulations on the usage of AI. Fjeld et al. (2020) state that while there is an established set of "AI principles", scholars have not focused on whether these principles are actually being followed or adapted by those who actually guide the development and usage of AI technologies. More needs to be done by governing bodies and international organizations to protect individuals from the harm that AI can cause.

2. The Risks of Generative AI

One of the most frightening developments in generative AI is its ability to create highly realistic but false media, otherwise known as deepfakes. Masood et al. (2022) mention that Generative Adversarial Networks (GANs) are easily available, and it is possible to generate deepfakes with the intent to spread misinformation, create pornography, and disrupt government functioning.

A current case that is occurring is the usage of deepfakes in South Korea to create pornographic images of women without their consent. Barr (2024) states that it is "…an epidemic of digital sex crimes, hundreds of women and girls targeted…". These individuals include those under the age of consent in South Korea. Barr also argues that government inaction is fueling the abuse, and strongly urges the South Korean to hold perpetrators accountable. One victim went to the police and was told that her case could not be pursued as due to the images being fake, the case was not truly criminal (Mackenzie & Choi, 2024). The social impact this has had on South Korea is astounding. South Korea already has a gender divide, with Kim & Lee (2022) stating that the gender divide could "fuel democratic deconsolidation", and that there is an alarming rise in anti-feminist sentiments in young men. Women in South Korea do not feel safe around men, and generative AI has made the issue even worse. And South Korea is not alone in this issue. Countries around the world face problems with the misuse of generative AI.

3. The Case for AI Moratorium

While it may be controversial, there is a case for pausing giant AI experiments. Pausing or even stopping all generative AI is "similar to discussing halting fundamental computer

science." (Belle, 2023), however the larger and more powerful models could be halted, or at the very least their use strictly regulated for research purposes only. AI is developing faster than the legislation is, with governments deliberating and discussing while AI has gone from creating hands with dozens of extra fingers to generating false media of presidential candidates that is so realistic it is hard for the average person to know what they are looking at is fake. There is an open letter from Future of Life Institute that calls for "all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4." (Future of Life Institute, 2023). Yudkowsky (2023) mentions how it took humans 60 years to go from the notion of Artificial Intelligence to today's capabilities, and it will take us more than a 6 months long pause to solve the issue of safety when it comes to generative AI.

While the author of this article may not fully understand the risks that artificial intelligence has for humanity when it comes to the existential crises that has driven the creation of the open letter and Yudkowsky's call to shut down all AI development until we have figured out how to do it safely, the current reality displays enough reason for at least a pause, likely longer than just 6 months. South Korean women would benefit, the world would benefit, and governments would finally get a chance to catch up to the rapid pace of generative AI development.

4. Conclusion

After reading the original article by Correa et al. (2023) in order to reflect upon it, originally the author thought that simply harsher regulations were needed to ensure that artificial intelligence could be used ethically. However, the research done has proven

that more drastic steps may need to be taken. Can individuals be trusted with easy access to powerful generative AI models such as GANs, when it can result in pornographic images of non-consenting women and minors? Or when it can cause individuals to believe that they're watching media of influential political figures, when it is all fake? And will governments ever be able to catch up, if not given the chance to do so with a pause in the development of AI technologies? This reflection has given the author much to think about in the future, and hopefully you, the reader, will also think more about the ethical risks generative AI has.

References:

Barr, H. (2024) South Korea's Digital Sex Crime Deepfake Crisis. Available from: https://www.hrw.org/news/2024/08/29/south-koreas-digital-sex-crime-deepfake-crisis [Accessed September 4 2024]

Belle, V. (2023) AI Moratorium – a multifaceted issue. Available from: https://medium.com/@vaishakbelle/ai-moratorium-a-multifaceted-issue-0aa512034f58 [Accessed September 1 2024]

Correa, K. N. et al. (2023) Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns* 4(10): 100854. Available from: https://doi.org/10.1016/j.patter.2023.100857

Deckard, R. (2023) What are Ethics in AI? Available from: https://www.bcs.org/articles-opinion-and-research/what-are-ethics-in-ai/ [Accessed August 23 2024]

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2020) Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *Berkman Klein Center Research Publication* 2020-1 Available from: http://dx.doi.org/10.2139/ssrn.3518482

Future of Life Institute. (2023) Pause Giant AI Experiments: An Open Letter. Available from: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ [Accessed September 4 2024]

Kim, J. H. & Lee, C. (2022) The 2022 South Korean Presidential Election and the Gender Divide among the Youth. *Pacific Affairs* 95(2): 285-308 Available from: https://doi.org/10.5509/2022952285

Mackenzie, J. & Choi, L. (2024) Inside the deepfake porn crisis engulfing Korean schools. Available from: https://www.bbc.com/news/articles/cpdlpj9zn9go [Accessed September 4 2024]

Masood, M., Nawaz, M., Malik, M. K., Javed, A., Irtaza, A. & Malik, H. (2022) Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence.* 53(): 3974-4026. Available from: https://doi.org/10.1007/s10489-022-03766-z

Yudkowsky, E. (2023) Pausing AI Developments Isn't Enough. We Need to Shut it All Down. Available from: https://empoweruohio.nyc3.cdn.digitaloceanspaces.com/Pausing%20AI%20Development.pdf [Accessed September 5 2024]