IN5620 Group assignment fall 2024 - Module 1

Group name: AI-sian

Ingeborg Sili Østgaard - ingebsos@uio.no, An Diep Vo - anvo@uio.no, Shannen Leigh Finnøy - shannenf@uio.no, Fredrik Kin Shing Ma - fkma@uio.no, Thanawut Pathawanich - thanawup@uio.no

We are interested in learning how we get affected by fake information that is generated by AI (Hajli et al., 2022). With the Internet and social media, we get affected all the time by what we see. However, not all information out there is true and one reason for that is people's use of AI to generate content. The spread of disinformation by this type of content is a challenge we are facing today.

We aim to focus on the user group of students who actively use social media. Specifically, we want to explore their behaviors, motivations, and experiences with social media platforms. This includes understanding how they interact with AI-generated content, such as filters or deepfakes as they are very popular in applications that use selfie images (Hedman et al., 2022). We also wish to focus on how social media influences their self-image, decision-making and their perceptions of authenticity.

Questions to investigate

How do AI-generated images influence people's ability to separate real content from fake content online?

We want to explore how the spread of AI-generated images impacts individuals' ability to differentiate between real and manipulated visual content. With the increasing of AI tools, it becomes crucial to understand how people interact with and interpret content. Are there patterns in how people misjudge AI-generated images? What cognitive or contextual factors lead to confusion, and how do platforms or user experiences affect this?

How can misinformation or deepfake content on social media negatively affect someone's beliefs or decisions?

This question seeks to examine the psychological and societal impact of deepfakes and misinformation. We are curious about how this content affects perceptions of truth and reality, and

how it may influence decision-making processes in contexts like politics, personal relationships, or social behavior.

What are the positive and negative ways AI-generated content can be utilized, and how can these opposing uses shape society?

We aim to understand the dual nature of AI-generated content: its potential for creativity, efficiency, and problem-solving, as well as its risks in spreading harmful or false information. How can use of AI enhance communication, education, or entertainment while minimizing the risks of manipulation or break of trust?

Potential methods

We wish to conduct a literature review to examine existing academic research on AI-generated content, deepfakes, misinformation, and their psychological and societal impacts. This will help identify key trends, challenges and knowledge gaps that can guide our own investigation.

Additionally, semi-structured interviews or focus groups with individuals who frequently encounter AI-generated content online (e.g., social media users, content creators) will provide valuable insights into their experiences, awareness, and ability to distinguish between real and fake content (Kornouskus, S. 2020). These discussions will offer different perspectives on identifying manipulated content.

Another method is reviewing existing AI tools and platforms, such as deepfake generation tools or content moderation services, we get an understanding on how the tools work, and their potential to spread false or misleading content. AI tools we want to use could be beauty filters from different social media, such as Tiktok, Instagram and Snapchat etc (Javornik et al., 2022). There are also different applications for beauty that we may want to discover and have a look at. Other possibilities could be AI image generation tools such as DALL-E and MidJourney, or in general deepfake tools.

The **Double Diamond** is a structured framework used to navigate the complexity of design problems by moving through phases of divergent and convergent thinking. It is divided into four stages: **discover**, **define**, **develop**, and **deliver**, guiding the investigation from problem exploration to solution implementation (Gwangwava, 2021). This framework will help structure our approach to investigating the influence of AI-generated content and misinformation.

Literature

- Javornik, A., Marder, B., Barhorst, J., McLean, G., Rogers, Y., Marshall, P., & Warlop, L. (2022). What lies behind the filter? Uncovering the motivations for using augmented reality (AR) face filters on social media and their effect on well-being. *Computers in Human Behavior, 128*, 107126. https://www.sciencedirect.com/science/article/pii/S0747563221004490
- Gwangwava, N. (2021). Learning design thinking through a hands-on learning model. International Journal of Innovative Teaching and Learning in Higher Education, 2(1), 1-19. https://doi.org/10.4018/ijitlhe.20210101.oa4
- Hedman, P., Skepetzis, V., Hernandez-Diaz, K., Bigun, J., & Alonso-Fernandez, F. (2022). On the effect of selfie beautification filters on face detection and recognition. *Pattern Recognition Letters, 163*, 104-111. https://libkey.io/libraries/269/articles/535937992/content-location
- Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society, 1*(3), 138-147. https://ieeexplore.ieee.org/document/9123958
- Hajli, N., Saeed, U., Tajvidi, M., & Shirazi, F. (2022). Social bots and the spread of disinformation in social media: The challenges of artificial intelligence. *British Journal of Management, 33*(4), 1238-1253. https://onlinelibrary.wiley.com/doi/epdf/10.1111/1467-8551.12554