Artificially disinformed and radicalised: How AI produced disinformation could encourage radicalisation

Thomas James Vaughan Williams, Maria Ioannou & Calli Tzani

Key digested message

The rapid advancements in artificial intelligence technologies have enhanced the ability for individuals who want to generate a substantial amount of seemingly genuine discussions, images and/or videos that are tailored to promote specific narratives. Unfortunately, this advancement has also provided a valuable tool for actors who seek to promote potentially harmful ideologies and share disinformation to large online audiences. By leveraging AI, these individuals can significantly enhance their recruitment efforts and bolster their perceived credibility, by producing seemingly legitimate but artificially fabricated evidence that supports their proposed narrative. This pressing issue is discussed in terms of its potentially negative consequences on the encouragement of radicalisation in users exposed to this artificially produced disinformation. Not only does it pose a risk to the integrity of people's perception of truth, but it also has the potential to exacerbate the likelihood of radicalisation occurring.

Keywords: Radicalisation, Disinformation, Artificial Intelligence, Online, Ideology, Misinformation.

https://doi.org/10.53841/bpsadm.2024.16.1.29

Introduction and background

Disinformation

DISINFORMATION being promoted and shared is a prominent issue on the internet (Anderson & Rainie, 2017), with research highlighting that false narratives discussing world issues and encouraging conspiracy theories are some of the most prominent forms of disinformation on social media (Constella Intelligence, 2022). When defining disinformation, it is important to clarify the difference between disinformation and its commonly mistaken counterpart, 'misinformation'. Misinformation refers to instances when false information is created without the intent to misguide or harm, whereas disinformation is when false information is purposely created to mislead, manipulate and/or harm others (Roberts-Ingleson & McCann, 2023). Research has highlighted that the spread of this harmful content has a detrimental impact on the improvement efforts of all areas and disciplines, from public health to environmental policy and even maintaining stability within a democratic nation (APA, n.d: Matthes et al, 2022). The internet has increased the prevalence of disinformation and misinformation exposure drastically, especially as cyberspace platforms such as social media become more essential and a common daily interaction for a majority of people (Auxier & Anderson, 2021).

Disinformation and AI

Recently the issues and discussions surrounding disinformation have been amplified by the growing emergence and utilisation of artificial intelligence (AI), which provides promoters of disinformation with a valuable tool to further their efforts (Gonzalez, 2023). The main issue that arises is the potential capability of AI to create realistic imagery and produce convincingly authentic videos of any inputted command, such as having prominent figures or celebrities believably delivering dialogue chosen by the creator (Helmus, 2022). A popular example of this video manipulation (coined 'deepfake' – a combination of the terms 'deep learning' and 'fake') was a viral collection of deepfake videos of the actor Tom Cruise, which many believed to be the real actor talking but was instead an AI generated deepfake (Metz, 2021). Although this particular case was for humoristic purposes, the problems arise from this utilisation of AI is when believable deep fakes are created for the purpose of promoting disinformation. If AI provides an individual or a group with the ability to potentially enhance their argument by artificially manufacturing visually credible evidence of their claims, this has the potential to allow promoters of harmful rhetoric and narratives to develop their promotion tactics.

Disinformation and radicalisation

One of the growing concerns surrounding this potential enhancement of harmful narrative promotions through the use of AI tools, is the notion that it may result in individuals who are exposed to this manipulated content becoming radicalised. Radicalisation is the process of an individual adopting ideologies and eventually justifying the use of violence to obtain the societal changes this ideology aims for (Doosje et al, 2016). The radicalisation process has been claimed to be encouraged and facilitated by both misinformation and disinformation (Roberts-Ingleson & McCann, 2023). This suggests that exposure to this perceived credible but falsified information can potentially encourage not only a change in an individual's belief system but also potentially strengthen

their already established ideology. Looking at how disinformation may be utilised, research that has explored the language utilised by promoters of extreme ideologies (harmful belief systems about particular groups) has highlighted that one of the key tactics utilised by these online promoters is the notion of building credibility about their ideological belief system (Williams & Tzani, 2022). This is often achieved through creating and disseminating content that promotes narratives that supports the ideology's views and use of violence, often depicted through the use of claimed 'evidence' that the group the ideology discriminates against is responsible for the societal issues the ideology is focused on (Williams & Tzani, 2022). Therefore, if AI has the ability to enhance the efforts and perceived legitimacy of disinformation, then this could have the potential to facilitate and even enhance the process of radicalisation (Roberts-Ingleson & McCann, 2023). Three areas where AI has the potential to increase the efficiency of harmful ideological promoters are: The increased quality and quantity of the content, utilisation of deepfakes and the fabrication of forums and discussions.

Increased quality and quantity

Before the rapid advancement in AI technologies, if someone wanted to edit or create a falsified photograph and/or image to promote disinformation, they would have to use traditional manual editing tools (e.g. Photoshop) to create or enhance an image (Butrym, 2023). However, AI capabilities have surpassed these traditional methods by providing the ability to create and/or edit images on a much shorter timescale while simultaneously improving the quality of the product through the automation capabilities the AI software provides (Butrym, 2023). This results in not only higher quality content to be produced more consistently and frequently but allows anyone to create and/or enhance images without the need to be technically skilled with the software, unlike previous methods where the editing would have to be conducted manually by the author. However, this ease of use allows bad actors with no photo editing skillset to also utilise this enhancement technology, permitting promoters of harmful ideologies to create consistent and credible looking images to utilise for disinformation purposes. Furthermore, it is not just falsified imagery that can be utilised for disinformation purposes.

Deepfakes

O'Sullivan (2019) highlighted in their report of deep fakes and disinformation the potential dangers that the ability to manipulate prominent political figures can have on the belief systems of individuals, especially if these manipulations are utilised to have prominent figures support the narratives of a particular extreme ideology. If prominent influential figures can be digitally manipulated to drive certain narratives that support or encourage the arguments of harmful ideologies, then this has the potential to influence and facilitate radicalisation in individuals who are exposed to this content (Roberts-Ingleson & McCann, 2023).

An example of this was highlighted in a deepfake that went viral on the social media platform Tik Tok in November, where it depicted US President Joe Biden stating in a speech that he was invoking the selective services act, drafting women into the military and going to war, stating 'You are not sending your sons and daughters to war. You are sending them to freedom', with this speech allegedly being in response to the conflict in Israel (McCarthy, 2023). This falsified video, that many believed to be real, resulted in

far-right commentators and Joe Biden opposers sharing this content, utilising it to support their relevant political narratives and ideologies (Ibrahim, 2023). Unfortunately, this is becoming a more common occurrence, especially in political spaces.

Forums and discussions

Another aspect AI can enhance that doesn't focus on the manufacturing of imagery or video to supply evidence of a harmful ideology credibility, is the creation of falsified discussions and forums. One element of prominent AI tools (such as Chat GPT) is that it allows users to generate authentic looking written content, which will be tailored to the prompts the user enters into the AI system (Lacroix, 2023). Therefore, this AI function could potentially be utilised to create falsified discussions and forum threads that focus and promote a certain narrative that the author entered as a prompt into the machine (Lacroix, 2023). If then these authentic looking dialogues and comments were presented on prominent social media sites, it would provide the appearance that there are numerous users who are engaging with this ideological conversation and simulate a community aspect to the ideology (Lacroix, 2023). This could have the potential to aid in recruitment for that particular ideology, as research has highlighted that individuals who had been radicalised and introduced to harmful ideologies via the internet were first introduced to the content of the ideology through being exposed to active discussions and forums (Bauget & Neumann, 2020: Williams & Tzani, 2022). AI has provided ideological promoters the means to create large volumes of content that can be shaped into an active discussion creating the illusion of a thriving online community to entice interested online users to learn more or even engage in the fabricated discussion.

Conclusion

By providing the means for anyone to quickly create a large volume of authentic looking discussions, images and videos all focused and promoting any ideological belief the author desires to promote, AI has supplied bad actors promoting harmful ideologies with a valuable tool to enhance their recruitment efficiency and strengthen their perceived credibility. The necessity for practical implementations to combat this prominent issue and its potential negative outcomes is of high priority not just because of the risk of harm it provides in regard to diluting peoples perception of the truth but because of the potential life-threatening implications this possible enhancement of radicalisation likelihood could cause.

The authors

Thomas James Vaughan Williams is a PhD student at the University of Huddersfield, exploring online radicalisation language and the promotion of extremist ideologies on social media.

Dr Maria Ioannou is a Professor of Investigative and Forensic Psychology at the University of Huddersfield, Course Director of the MSc Investigative Psychology and Course Director of the MSc Security Science. She is a Chartered Forensic Psychologist, Chartered Manager, Fellow of the Higher Education Academy, HCPC Registered Practitioner and Associate Fellow of the British Psychological Society.

Dr Calli Tzani is a senior lecturer of Investigative and Forensic Psychology at the University of Huddersfield, Deputy Director of the Applied Criminology and Policing Centre, and the Forensic Consultant Editor of ADM. She is a Fellow of the Higher Education Academy, and Associate Fellow of the British Psychological Society.

References

Anderson, J. & Rainie, L. (2017). *The Future of Truth and Misinformation Online*. Pew Research Center. https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online

APA. (n.d). *Misinformation and disinformation*. American Psychological Association. https://www.apa.org/topics/journalism-facts/misinformation-disinformation

Auxier, B. & Anderson, M. (2021). Social Media Use in 2021. Pew Research Center. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Baugut, P. & Neumann, K. (2020). Online propaganda use during Islamist radicalization. *Information, Communication & Society, 23*(11), 1570–1592. https://doi.org/10.1080/1369118X.2019.1594333

Butrym, N. (2023). Why Are AI Photo Enhancers Better Than Photoshop Or Other Manual Image Editors? DeepImage. https://deep-image.ai/blog/why-ai-photo-enhancers-are-better-than-photoshop-or-other-manual-image-editors/

Constella Intelligence. (2022). The 5 Dangers of Misinformation & How Disinformation Campaigns are Related. Constella Intelligence. https://constellaintelligence.com/how-words-can-disrupt-your-business-the-5-dangers-of-misinformation/

Doosje, B., Moghaddam, F.M., Kruglanski, A.W., de Wolf, A., Mann, L. & Feddes, A. R. (2016).Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology, 11*(1), 79–84. https://doi.org/10.1016/j.copsyc.2016.06.008

Gonzalez, O. (2023). AI Misinformation: How It Works and Ways to Spot It. CNET. https://www.cnet.com/news/misinformation/ai-misinformation-how-it-works-and-ways-to-spot-it/

Helmus, T.C. (2022). *Artificial Intelligence, Deepfakes, and Disinformation.* RAND Corporation, , 1–24. https://www.rand.org/pubs/perspectives/PEA1043-1.html

Ibrahim, N. (2023). Did Biden Call for a Military National Draft?. Snopes. https://www.snopes.com/fact-check/biden-military-national-draft/

Lacroix, J. (2023). *AI Will Make Extremists More Effective, Too.* Online Secruity. https://inkstickmedia.com/ai-will-make-extremists-more-effective-too/

Matthes, J., Corbu, N., Jin, S., Theocharis, Y., Schmer, C. et al., (2022). Perceived prevalence of misinformation fuels worries about Covid-19: A cross-country, multi-method investigation. *Information, Communication & Society,* 1–24. https://doi.org/10.1080/1369118X.2022.2146983

Mccarthy, B. (2023). Biden deepfake announcing US draft resurfaces amid Israel-Hamas war. AFP Fact Check. https://factcheck.afp.com/doc.afp.com.33YP34M

Metz, R. (2021). How a deepfake Tom Cruise on TikTok turned into a very real AI company. CNN. https://edition.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company/index.html

O'Sullivan, D. (2019). When seeing is no longer believing: Inside the Pentagon's race against deepfake videos. CNN. https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

Roberts-Ingleson, E.M. & McCann, W.S. (2023). The Link between Misinformation and Radicalisation: Current Knowledge and Areas for Future Inquiry. *Perspectives on Terrorism, 17*(1), 36–49. https://www.jstor.org/stable/27209215

Williams, T.J.V. & Tzani, C. (2022). How does language influence the radicalisation process? A systematic review of research exploring online extremist communication and discussion. *Behavioral Sciences of Terrorism and Political Aggression, 1*(1), 1–21. https://doi.org/10.1080/19434472.2022.2104910