

Association for Information Systems

AIS Electronic Library (AISeL)

ECIS 2024 TREOS

AIS TREO Papers

6-14-2024

Generativity For Bad: Towards Generative Chains of Disinformation On Social Media

Christine Abdalla Mikhaeil

IESEG School of Management, c.abdallamikhaeil@ieseg.fr

Roxana Ologeanu-Taddei

TBS Education, r.ologeanu-taddei@tbs-education.fr

Follow this and additional works at: https://aisel.aisnet.org/treos_ecis2024

Recommended Citation

Abdalla Mikhaeil, Christine and Ologeanu-Taddei, Roxana, "Generativity For Bad: Towards Generative Chains of Disinformation On Social Media" (2024). *ECIS 2024 TREOS*. 41.

https://aisel.aisnet.org/treos_ecis2024/41

This material is brought to you by the AIS TREO Papers at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2024 TREOS by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

GENERATIVITY FOR BAD: TOWARDS GENERATIVE CHAINS OF DISINFORMATION ON SOCIAL MEDIA

TREO Paper

Christine Abdalla Mikhaeil, IESEG School of Management, Lille, France,
c.abdallamikhaeil@ieseg.fr

Roxana Ologeanu-Taddei, TBS Business School, Toulouse, France,
r.ologeanu-taddei@tbs-education.fr

Abstract

Algorithmic content management, social media affordances and generative AI used all together can unlock innovative usages but also vulnerabilities. While the innovation potential is held as a positive consequence of generativity and as its negative side effect, the features offered by social media such as TikTok together with generative AI can foster innovation as vulnerability. Our aim is to investigate how the generativity of social media leads to a disinformation chain in which information contents (videos) are altered to spread disinformation.

Keywords: Social Media, Generativity, Disinformation, Videos.

1 Introduction

More than 12,000 videos have been posted between 2018 and 2023 across 96 YouTube channels that have promoted climate denial and disinformation (Ramirez, 2024). Climate change is one of the crucial topics which, prompted by populist movements, have created burgeoning opportunities for individuals to frame their media consumption around their own beliefs, opinions, and prejudices (Davies, 2016). These claims are built through increasingly visual narratives shared on social media platforms.

Moreover, recent technological advances and in particular in the domain of generative AI has raised concerns that people will not be able anymore to distinguish what is true from what it is not (Groh et al., 2022; Metz, 2023). Generative AI (GenAI) can be defined as “a type of artificial intelligence that can learn from and mimic large amounts of data to create content such as text, images, music, videos, code, and more, based on inputs or prompts.”¹ Recent advances include GenAI solutions that apply machine learning to large datasets to not only generate texts (e.g. OpenAI’s ChatGPT) but also visuals and audio (e.g. respectively DALL-E and VALL-E). As sophisticated as the results can be when giving the right prompts, these AI-generated output are increasingly easy and quick to produce and misleading as they are harder to set apart from human-generated contents (Groh et al., 2022; Langguth et al., 2021). The increasing accessibility, ease of mastery and availability of the features to change or produce content as well as of Generative AI tools raises common concerns over the spread of an increasing volume of as well as increasingly realistic and misleading disinformation.

2 Disinformation on social media

A first concern is related to the volume of false contents available. GenAI makes disinformation easier, and cheaper to produce at an unprecedented scale (Hsu & Thompson, 2023). Yet, disinformation already abounds online as it was shown during the 2016 U.S. Presidential Election (Allcott & Gentzkow, 2017)

¹ Generative Artificial Intelligence (AI) | Harvard University Information Technology

or during the COVID-19 pandemic (Zarocostas, 2020), only to cite a couple of cases. Hence, the facilitation of production and dissemination of disinformation by GenAI questions if the volume and accessibility of disinformation raises the risks of more consumption of misleading and false contents. Some research suggests that accessibility does not play a key role in the consumption of fake news, but rather predispositions or traits leading individuals to intentionally seek out disinformation (Broniatowski et al., 2023; Osmundsen et al., 2021).

A second concern relates to the increasing level of sophistication of the produced content, both more convincing and persuasive. GenAI can improve content quality in the visual domain beyond what Photoshop capabilities could do for malicious actors (Kapoor & Narayanan, 2023). Yet, the availability of generative technologies that can create realistic yet fake contents to malicious actors can also be used to discredit the fake content and such a strategy may payoff (Christoher, 2023).

A third concern relates to the personalization capabilities to target specific individuals (Hsu & Thompson, 2023; Zarouali et al., 2022). Training GenAI on large datasets to understand people beliefs, values and preferences could make it easier to persuade specific targeted individuals of the authenticity of a message. Political campaigning has thrived on the monetization of social media platforms to make use of political microtargeting (Hackenburg & Margetts, 2023; Tappin et al., 2023; Zarouali et al., 2022). Yet, in their study of persuasion effects of personality-congruent political microtargeting, Zarouali et al. (2022) showed that people can be more strongly persuaded by political ads that match their personality by looking into extraversion. Personalization might be the next frontier for GenAI as current evidence of persuasive effect of political microtargeting remains limited and context-dependent (Hackenburg & Margetts, 2023; Tappin et al., 2023; Zarouali et al., 2022).

3 Generativity: innovation as vulnerability

These features to change or produce content increase the generativity attributes. Generativity of technology is “*a technology’s overall capacity to produce unprompted change driven by large, varied and uncoordinated audiences*” (Zittrain, 2006, p.1980). Generativity fosters innovation but also vulnerability through security breaches and malware threats (Zittrain, 2006). While the innovation potential is held as a positive consequence of generativity and as its negative side effect, the features offered by social media such as TikTok together with GenAI can foster *innovation as vulnerability*. More specifically, those features are used to generate user content which can be altered as disinformation through an information chain by large and uncoordinated audiences.

In parallel, several studies have shown how algorithmic content management on social media (mostly Facebook) leads to filter bubbles and echo chambers (Flaxman et al., 2016), leading to opinion polarization (Bessi et al., 2016; Wade et al., 2020). But the generativity and new affordances of social media (for instance the exposure to content on TikTok according to different algorithmic paths than on Facebook) together with GenAI can give rise to different phenomena that have not yet been investigated.

Our aim is to investigate how the generativity of social media leads to a disinformation chain in which information contents (videos) are altered to spread disinformation. This disinformation chain can manifest by connected bubbles, meaning that the disinformation shared on a specific channel or group can switch (through users’ actions) to other channels and groups, in the same or different countries.

References

- Allcott, H., and Gentzkow, M. (2017). “Social media and fake news in the 2016 election,” *Journal of Economic Perspectives* 31 (2), 211-236.
- Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., and Quattrociocchi, W. (2016). “Users polarization on Facebook and Youtube,” *PLOS ONE* 11 (8), 1-24.
- Broniatowski, D. A., Simons, J. R., Gu, J., Jamison, A. M., and Abrams, L. C. (2023). “The efficacy of Facebook’s vaccine misinformation policies and architecture during the COVID-19 pandemic,” *Science Advances* 9 (37), 1-17.

- Christoher, N. (2023). *An Indian politician says scandalous audio clips are AI deepfakes. We had them tested*. URL: <https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/> (visited on December 15, 2023).
- Davies, W. (2016). *The age of post-truth politics*. URL: <https://www.nytimes.com/2016/08/24/opinion/campaign-stops/the-age-of-post-truth-politics.html/> (visited on November 30, 2021).
- Flaxman, S., Goel, S., and Rao, J. M. (2016). "Filter bubbles, echo chambers, and online news consumption," *Public Opinion Quarterly* 80 (S1), 298-320.
- Groh, M., Sankaranarayanan, A., Lippman, A., and Picard, R. (2022). "Human detection of political deepfakes across transcripts, audio, and video," *arXiv preprint arXiv:2202.12883*.
- Hackenburg, K., and Margetts, H. (2023). "Evaluating the persuasive influence of political microtargeting with large language models," OSF Preprints.
- Hsu, T., and Thompson, S. A. (2023). *Disinformation researchers raise alarms about A.I. chatbots*. URL: <https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html/> (visited on December 15, 2023).
- Kapoor, S., and Narayanan, A. (2023). *How to prepare for the deluge of generative AI on social media*, Knight First Amendment Institute at Columbia University.
- Langguth, J., Pogorelov, K., Brenner, S., Filkuková, P., and Schroeder, D. T. (2021). "Don't trust your eyes: Image manipulation in the age of deepfakes," *Frontiers in Communication* 6, 632317.
- Metz, C. (2023). *'The Godfather of A.I.' leaves Google and warns of danger ahead*. URL: <https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton/> (visited on December 15, 2023).
- Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). "Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter," *American Political Science Review* 15 (3), 999-1015.
- Ramirez, R. (2024). *What is 'new denial?' An alarming wave of climate misinformation is spreading on YouTube, watchdog says*. URL: <https://edition.cnn.com/2024/01/16/climate/climate-denial-misinformation-youtube/index.htm/> (visited on February 29, 2024).
- Tappin, B. M., Wittenberg, C., Hewitt, L. B., Berinsky, A. J., and Rand, D. G. (2023). "Quantifying the potential persuasive returns to political microtargeting," *Proceedings of the National Academy of Sciences* 120 (25), e2216261120.
- Wade, J. T., Roth, P. L., Thatcher, J. B., and Dinger, M. (2020). "Social media and selection: Political issue similarity, liking, and the moderating effect of social media platform," *MIS Quarterly* 44 (3), 1301-1357.
- Zarocostas, J. (2020). "How to fight an infodemic," *The Lancet* 395 (10225), 676.
- Zarouali, B., Dobber, T., De Pauw, G., and de Vreese, C. (2022). "Using a personality-profiling algorithm to investigate political microtargeting: Assessing the persuasion effects of personality-tailored ads on social media," *Communication Research* 49 (8), 1066-1091.
- Zittrain, J. L. (2006). "The generative Internet," *Harvard Law Review* 119 (7), 1974-2040.