On the importance to study fringe social networks and their impending use of GenAI to promote mal-info: Gab as a case study

Florian Barbaro[1,*,†], Andy Skumanich[1,†]

[1]*Innov8ai Inc., California, United States*

Abstract

This article discusses the potential for deleterious effects from the use of AI in the wrong contexts. Because of the impact of social networks on modern society, there is significant potential for "mal-info", such as disinformation, cyberbullying, and hate speech, to affect significant segments of society. This amplification of mal-info can have real-world impacts (such as mass-shooting). Some attempts to introduce moderation on major platforms like Facebook and Twitter have had limited success, with even some backsliding. Even worse, there are "alternative" or fringel social networks such as Gab, Gettr, and Bitchute which have emerged and are now established. The essence of these platforms is to offer a completely unmoderated space which means that any manner of mal-info is allowed. In this article we, provide an introduction of these alternative platforms and present initial results of a qualitative analysis of posts from Gab. We examine several inflammatory terms using text analysis. Importantly, we discusses the use of generative algorithms by Gab (and other fringe platforms) and the impending risks to society. This work can lay the foundation to mitigate these risks.

Keywords

Social Network, Gab, Gettr, Bitchute, Qualitative analysis, Narratives, Generative AI

1. Introduction

Social networks have had a profound impact on the way we communicate, share information, and interact with each other. They have become a central part of modern society, enabling people to connect with each other regardless of their location, share their thoughts and opinions, and participate in online communities [1]. However, the rise of social networks has also led to a number of challenges, including the spread of disinformation [2], cyberbullying [3], and the propagation of hate speech and extremist ideologies [4], along with in general what we call "mal-info". For the latter, some governments (e.g. the EU), with the help of the best-known platforms such as Facebook or Twitter, have tried to introduce some form of moderation. Due to this, some individuals and groups turned to alternative social networks that offer a different set of guidelines and principles, such as Gab, Gettr, and Bitchute. These channels have gained popularity among certain segments of the population because they freely allow any manner of mal-info. While these networks offer a space for so-called *free expression*, they

also raise questions about the role of social networks in shaping our perceptions and attitudes, and the potential consequences of this influence on our democratic institutions and processes. Despite these many issues, very few articles have focused on these new social networks [5], mostly concentrating on Twitter [4]. In addition, there is an impending issue that these actors can implement AI and GenAI for furthering the mal-info elements. This latter is of grave concern and research must begin immediately to develop "guard rails" to mitigate the potential harm.

In this article, we will evaluate three of these new (fringe) social networks, namely Gettr, Bitchute and Gab. Then we conduct a qualitative analysis of posts from Gab containing a typical inflammatory term. Afterwards, we present the new developments driven by these platforms for using generative algorithms. Finally, we discuss the risks for the society of such algorithms. We conclude by alerting the AI community to the need to develop moderating responses to address these risks.

2. New Social Networks

New social networks were created in response to concerns about censorship and moderation on established social media platforms. Some users feel that their "freedom of speech" is being limited on platforms such as Twitter and Facebook and that their content is being unfairly targeted or removed. This is a broader discussion beyond the scope of this paper as it is part of the determination of what constitutes *free speech*. As shown by [6],

more and more Americans are using these platforms for news (6% in 2022). Using tools like Similarweb[1], we can extrapolate that this number must have doubled in 2023.

These new platforms often have little to no moderation and lenient content policies. Although this allows for a wider range of opinions and viewpoints to be shared, some of which can indeed be mal-info. However, they can be legitimately criticized for allowing hate speech, harassment [7], and misinformation to spread unchecked. In this myriad of new social networks, three seem to emerge as frequently used channels, namely Gab, Gettr and Bitchute.

Gab is a social networking platform launched in 2016 and bills itself as an unfettered speech (so-called *free speech*) alternative to mainstream social media sites. It was created in response to the perceived censorship of conservative views on traditional social media sites. Gab allows users to post messages called "gabs," share photos, and interact with other users. It has been observed as being a platform for hate speech and far-right extremism.

Gettr is a newer social media platform that was launched in 2021 by former President Donald Trump's senior adviser, Jason Miller. It is marketed as a "cancel-free" platform that supports unfettered speech and allows users to share their opinions without any type of moderation. Gettr's features are similar to those of Twitter, allowing users to post short messages called "gettrs," share photos and videos, and interact with other users.

Bitchute is a video-sharing platform that was launched in 2017. It was created in response to perceived censorship of fringe or provoking views on traditional video-sharing sites like YouTube. Bitchute allows users to upload, share, and view videos on various topics, including news, politics, and entertainment. It has been observed as being a platform for conspiracy theories and hate speech.

The main difference between these platforms is their focus and features. Gab and Gettr are primarily social media platforms that allow users to share short messages and interact with other users. Bitchute, on the other hand, is a video-sharing platform that allows users to upload and view longer-form content. Additionally, Gab has been associated with far-right extremism, while Gettr is marketed as a nominally more mainstream platform. All three platforms have been observed to tolerate hate speech and conspiracy theories, but Gab has faced the most scrutiny.

3. Analysis of Gab

In this section, we will analyse posts from Gab.

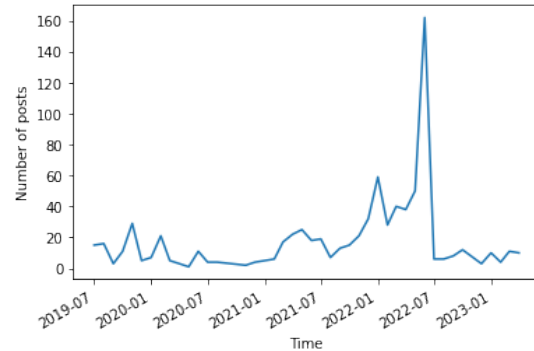To showcase that this fringe social network does contribute to the spread of hate speech, we selected a

Figure 1: Timeline of posts containing #*Cuckservative*

term from the Glossary of Extremism from The Anti-Defamation League (ADL)[2]. ADL is an international non-governmental organization based in the United States specializing in civil rights law (historically focusing on anti-semitism). In their glossary, they provide an overview of many of the terms most frequently used by a variety of extremist groups and movements.

Based on this lexicon, we selected the term *Cuckservative*. In 2015, alt-right white supremacists began disparaging members of the conservative movement with the derogatory term *cuckservative*, a combination of *conservative* and *cuckold*, to describe a white conservative who putatively promotes the interests of Jews and non-whites over those of whites. The Groypers, a group that attracts white supremacists and other far-right activists, also employ the term.

For the purposes of this paper, as no API is available for Gab, a tailor-made scraper based on Python's Selenium library [8] was used to automate and scale up this process. Using the #*Cuckservative*, we retrieved 788 messages from July 4, 2019, to April 22, 2023. Figure 1 shows the number of posts over time. We can observe a peak around May-June 2022 and we could link this peak to the Mass Shooting in Uvalde, Texas and the reaction it created in the political sphere and for white supremacists and other far-right activists. At the political level, more than 600 conservatives[3], mostly in Texas, called for gun reform. Instead, the segment of white supremacists and other far-right activists used the term *Cuckservative* to describe these politics of attempting to develop a gun control solution and to defend their rights to carry a weapon with unfettered access.

Figure 2 shows the fifteen most used hashtags. It is interesting to note that the hashtags used for the term *Cuckservative* are in agreement with the definition by

---

Figure 2: Top # in post

Figure 3: Example of Post published the 12 June 2022
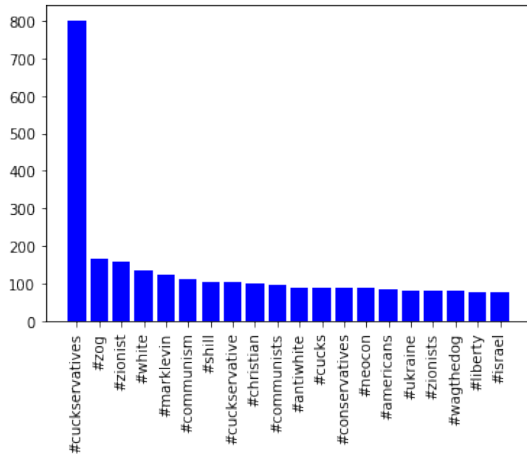
ADL. Users post using mostly hashtags concerning the *Jewish population* and their feeling about *white racism.*

In order to analyse posts, we applied a classical text pre-processing [9]. We removed all non-alphabetical characters (numbers, punctuations, ...) and stopwords. Then we applied lemmatization. Afterwards, we selected texts from March to August 2022 to understand the different posts around the peak of May-June 2022.

Figure 4 shows the fourth most used words by month. As expected, the posts make extensive use of the terms *cuckservative* and *white* over time in support of their ideology. It is interesting to note that the posts take up current issues. In March 2022, one of the main topics is the war in Ukraine. For the months of May-June, we can observe that the posts are directed against the American right, notably with the word *rat*, because according to them they betray their ideology. Finally, it is interesting to observe a certain consistency in their anti-Semitic language, with the words *jew* and *israel*. Figure 3 shows a post published on June 12, 2022, highlighting all these topics.

In order to do a deep dive into the different narratives spread by these posts, we applied the method of *keyness* analysis [10]. This approach from the fields of corpus linguistics and corpus-based discourse analysis is directed at identifying *key* items (e.g. words) in a target corpus in relation to a reference corpus based on the frequencies of items in both corpora. As such, a *keyness* analysis can support an exploratory approach to texts that gives an indication of their *aboutness.* The keyness metric chosen for this paper is that of Log Ratio, which is defined as the *binary log of the ratio of relative frequencies* [11]. This gives a measure of the actual observed difference between two corpora for a key item (rather than a measure of statistical significance). The advantage of this is that
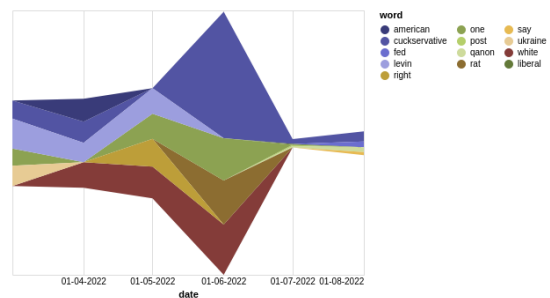

Figure 4: Streamgraph of most used words in posts by month

it allows for the sorting of items by the size of the actual frequency difference between the corpora, enabling us to find the top *N* most key items.

Table 1 shows the key items by month. It is interesting to note the different changes in narratives. For example, in March, users are mostly talking about *Covid* by using words such as *wuhan* and *thecurrentthing.* Also, they are referring to the war in Ukraine with the word *quagmire.* Then, in May-June, users seem to react to the Mass shooting and the political reaction in favour of gun control using terms such as *withheld* or *unum* and criticizing people betraying their ideology (*snitch*). Finally, in July-August, they return to more usual narratives (in their ideology) by criticizing The Establishment (*shitlibs*), anti-Semitic organizations such as *adl* or *aipac*[4] and other far-wing ideology like *Qanon.*

---

[4]https://www.aipac.org/

Table 1
Key items by month from March 2022 to August 2022

Month	Key items
March	wuhan, thecurrentthing, quagmire
April	regimechange, extortion, falseflags
May	withheld, outlawed, unum
June	snitch, plunder, accountability
July	shitlibs, thinktank, fedbois
August	qanoncuck, adl, aipac

4. Concerning Generative AI development

In the previous section, we saw how an unmoderated social network can propagate messaging hatred against a targeted population or conspiracy ideas.

In addition to the dangers of unmoderated content, the use of AI systems in social networks also presents a significant risk to democratic institutions and processes. The political and demographic biases embedded in widely used AI systems can degrade the quality of democratic discourse and decision-making. As humans increasingly rely on AI-generated content to make decisions, AI systems will have an enormous amount of influence to shape human perceptions and manipulate human behavior.

In this sense, Gab's CEO, Andrew Torba, wishes to propagate his ideologies through his platform. Indeed, he published an article on 27 January 2023 entitled: "Christians Must Enter the AI Arms Race"[5]. In his article, Andrew Torba, is discussing the potential for building a new AI system that is not "skewed" with a *liberal/globalist/talmudic/satanic* worldview like many current AI systems. He argues that if the enemy is going to use AI for evil, then they should build an AI system for good. He suggests that if people with the same ideology don't build their own AI system, then their enemies will dominate this space and use it as a weapon against the minds of the people. The author believes that they need to develop their AI system for the glory of God that can communicate the Truth of the Gospel to millions of people. This is an almost Orwellian flip of the notions of good and evil, however it shows the very nature of the potential of Adverse Impacts (AI) for AI.

This development of a soon-to-be-available Text Generation AI called Based AI[6] is part of a comprehensive plan to develop tools for like-minded people. Indeed, Gab already launched a service for Image Generation called Gabby[7] and a service for Movie Generation called Mel[8].

---

[5] https://news.gab.com/2023/01/christians-must-enter-the-ai-arms-race/
[6] https://gab.com/basedai
[7] https://gab.com/AI
[8] https://gab.com/movie

5. Discussion

To conclude, we emphasise the importance of studying the new social networks that were fringe but are becoming increasingly important as having an inpact of the degradation of social discourse. We have shown using an example of analysis of Gab's posts, that this unmoderated network propagates hate and mal-info. Moreover, it is typical that these networks can serve as an echo chamber for the most radical ideologies.

At last, we should be cautious about the potential impact of AI systems that have fringe political and demographic biases [12]. As humans increasingly rely on AI-generated content to make decisions, these systems will have an enormous influence to shape our perceptions and manipulate our behavior if left unchecked. Public-facing AI systems that exhibit fringe political bias will contribute to societal polarization, as users seeking confirmation bias may gravitate towards politically aligned systems, while those with different viewpoints may avoid them.

Commercial and political interests may be tempted to fine-tune, as ideologies spread on Gab, and deploy AIs to manipulate individuals and societies, underscoring the need for caution in how these systems are integrated into our technological landscape. To avoid contributing to societal polarization, AI systems should remain largely neutral on normative questions where there is no conclusive scientific evidence or a variety of legitimate and lawful human opinions.

Instead of being used by fringe elements to advance an agenda, AI systems can be directed to provide factual information on empirically verifiable issues. If these are based on legitimate elements, the content can offer diverse viewpoints and sources on contested topics that are often under-determined. By doing so, these systems can help users gain insight, overcome in-group biases, and broaden their perspectives, potentially playing a useful role in defusing societal polarization. It is essential that language models claiming political neutrality and accuracy, like GPT-4 based models [13], remain transparent about any biases they exhibit on normative questions, or be used by ideologues to drive a fringe narrative. It is the responsibility of the broader AI community to carefully monitor these types of mal-info driving elements, and develop the necessary guard rails to ensure quality social engagements for the betterment of all individuals.

References

[1] K. K. Osman, F. Barbaro, A. Skumanich, The potential for social media analysis to assess and optimize water management, in: 9 th International Conference on Computational Social Sci-

ence IC2S2, Copenhagen, Denmark, 2023. URL: https://hal.science/hal-04084263.

[2] F. Barbaro, A. Skumanich, Addressing socially destructive disinformation on the web with advanced ai tools: Russia as a case study, in: Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion, Association for Computing Machinery, New York, NY, USA, 2023, p. 204–207. URL: https://doi.org/10.1145/3543873.3587348. doi:10.1145/3543873.3587348.

[3] G. W. Giumetti, R. M. Kowalski, Cyberbullying via social media and well-being, Current Opinion in Psychology 45 (2022) 101314. URL: https://www.sciencedirect.com/science/article/pii/S2352250X22000161. doi:https://doi.org/10.1016/j.copsyc.2022.101314.

[4] J. Govers, P. Feldman, A. Dant, P. Patros, Down the rabbit hole: Detecting online extremism, radicalisation, and politicised hate speech, ACM Comput. Surv. (2023). URL: https://doi.org/10.1145/3583067. doi:10.1145/3583067, just Accepted.

[5] M. Peucker, T. J. Fisher, Mainstream media use for far-right mobilisation on the alt-tech online platform gab, Media, Culture & Society 45 (2023) 354–372. URL: https://doi.org/10.1177/01634437221111943. doi:10.1177/01634437221111943. arXiv:https://doi.org/10.1177/01634437221111943.

[6] G. Stocking, A. Mitchell, K. E. Matsa, R. Widjaya, M. Jurkowitz, S. Ghosh, A. Smith, S. Naseer, C. S. Aubin, The role of alternative social media in the news and information environment, Pew Research Center. URL https://www. pewresearch. org/journalism/2022/10/06/therole-of-alternative-social-media-in-the-news-andinformation-environment (2022).

[7] S. Abarna, J. Sheeba, S. Jayasrilakshmi, S. P. Devaneyan, Identification of cyber harassment and intention of target users on social media platforms, Engineering applications of artificial intelligence 115 (2022) 105283.

[8] S. S. Salunke, Selenium Webdriver in Python: Learn with Examples, 1st ed., CreateSpace Independent Publishing Platform, North Charleston, SC, USA, 2014.

[9] F. Barbaro, Analyse exploratoire et classification de textes, Theses, Paris 1 - Panthéon-Sorbonne, 2022. URL: https://hal.archives-ouvertes.fr/tel-03708173.

[10] C. Gabrielatos, Keyness Analysis: nature, metrics and techniques, Routledge, United Kingdom, 2018, pp. 225–258.

[11] A. Hardie, Log ratio: An informal introduction, ESRC Centre for Corpus Approaches to Social Science (CASS) (2014) 1–2. URL: http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/.

[12] F. Suguri Motoki, V. Pinho Neto, V. Rodrigues, More human than human: Measuring chatgpt political bias (2023).

[13] OpenAI, Gpt-4 technical report, 2023. arXiv:2303.08774.