Understanding and Mitigating Mental Health Misinformation on Video Sharing Platforms

VIET CUONG NGUYEN, Georgia Institute of Technology, United States of America

MICHAEL BIRNBAUM, Northwell Health, United States of America

MUNMUN DE CHOUDHURY, Georgia Institute of Technology, United States of America

1 INTRODUCTION

Despite the ever-strong demand for mental health care globally, access to traditional mental health services remains severely limited expensive, and stifled by stigma and systemic barriers [1, 22, 23]. Thus, over the last few years, young people are increasingly turning to content on video-sharing platforms (VSPs) like TikTok and YouTube to help them navigate their mental health journey [8, 9]. Such content is not only readily accessible and free of charge, but they contain easily digestible information through audio and visual affordances within these platforms [24, 27]. If done right, content on video-sharing platforms can be a significant asset to the growing field of digital mental health, as it can provide non-judgmental and democratized access to mental health help and advice to all, within the privacy of their realms [18, 21]. However, navigating towards trustworthy information relating to mental health on these platforms is challenging, given the uncontrollable and unregulated growth of dedicated mental health content and content creators catering to a wide array of mental health conditions on these platforms. One reason for this is the relatively low barrier of entry in creating mental health content on video-sharing platforms compared to other online-based resources (such as blog posts). Consequently, many reports have emerged that mental-health-related videos containing misinformation (referred to henceforth as *mental health misinformation)* are rampant on these platforms [8, 19]. Here, we define mental health misinformation as false or misleading information about mental health and illness, including diagnosis and treatment of these challenges, irrespective of the intention of those spreading such information. An example of mental health misinformation regarding bipolar found on video-sharing platforms is shown in Figure 1. In this publicly available TikTok video where the screenshot from Figure 1 is taken, the presenter presents anecdotal symptoms which they suggest are indicative of type 2 bipolar. This video contains several key markers indicative of mental health misinformation:

- They do not have relevant medical qualifications to back these statements, nor do they disclose their lack of qualifications anywhere within the video or its description
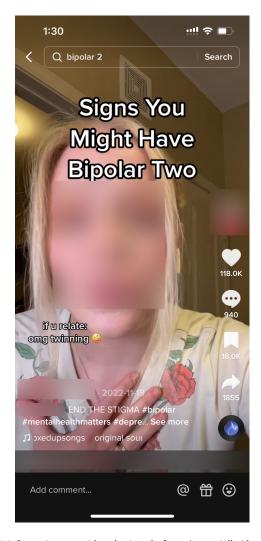
Fig. 1. Example of mental health misinformation on a video-sharing platform. (potentially identifying information has been blurred out for privacy reasons)

- The symptoms they shared for type 2 bipolar are purely anecdotal and not backed by any official diagnostic criteria for the condition (e.g. DSM-5)
- They encourage viewers to self-diagnose themselves with type 2 bipolar by prefacing the video with "Signs You Might Have Bipolar Two"

There has been extensive work has been done on understanding and mitigating misinformation content on text- and image-based social media platforms such as Facebook, Twitter, and Instagram for a variety of topics [6, 7, 12, 13, 15, 16]. However, while the virality and popularity of content on video-based social media platforms are significant, few works have focused on understanding how and mitigating the spread of mental health misinformation on video-sharing platforms. The widespread dissemination of mental health misinformation can have serious consequences. Broadly

speaking, these include stigma and discrimination towards those with mental illness, the perpetuation of harmful beliefs and practices related to mental health care, misdiagnosis, and delay or avoidance of effective treatment [4, 17]. For example, the spread of mental health misinformation on TikTok has led to the recent rise in the self-diagnosis of severe mental health disorders such as bipolar disorder, ADHD, etc. [2]. Thus, we argue that understanding and mitigating mental health misinformation on video-sharing platforms is crucial for reducing such dangerous self-diagnosis. In addition, doing so will ideally increase trust and credibility in video-sharing platforms as a reliable and accurate resource for digital mental health among young people. However, there are many challenges and open questions toward achieving such an ideal.

2 RELATED WORKS

2.1 Misinformation on Video-sharing Platforms

Given the popularity of video-sharing platforms in recent years, there is an emerging thread of research focusing on misinformation hosted within these platforms. Works lying on this thread have focused on measuring the prevalence of topical misinformation content [5, 11], detecting filter bubbles that lead viewers down a rabbit hole of misinformation content [10, 20], and detecting misinformation videos through automated means [14, 25].

2.2 Mental Health (Mis)information on Video-sharing Platforms

Compared to other topics, academic research on mental health information and misinformation on video-sharing platforms are sparser, and to the best of our knowledge, limited to content analysis of the relevant videos [3, 26]. For instance, Yeung et al. [2022] examined the top 100 popular videos on TikTok regarding ADHD [26]. They found that more than half of the videos were misleading, whereas only around 20 percent of the videos were labeled as 'useful'

3 OPEN QUESTIONS

Synthesizing the related works presented above, we now present some open questions we think are important to the understanding and mitigation of mental health misinformation on video-sharing platforms

- How does "bad" mental health content differ from "good" mental health content in terms of virality? framing? user responses? Relatedly, how do we define what constitutes "good" and "bad" mental health information, given the rapidly evolving science in this field?
- How do platform affordances impact the spread of mental health misinformation on them?
- How can we design effective interventions against mental health misinformation on video-sharing platforms? On this note, what are the strengths and limitations of conventional content moderation approaches when applied to this sensitive domain? What moderation strategies would work and which ones wouldn't?
- How can we build effective machine learning models for detecting videos containing mental health misinformation on video-sharing platforms?
- How do recommendation algorithms embedded in video-sharing platforms (e.g. TikTok's For You Page, Youtube's Autoplay) affect people's search for mental health content? Do watching mental health misinformation on these platforms lead young people down a rabbit hole?
- How do we empower psychiatrists in fact-checking mental health misinformation on video-sharing platforms
- How do we ensure online platforms continue to remain safe spaces for seeking and providing mental health help, with undertaking efforts that curb or reduce the impact of misinformation on these issues?

- Could efforts to mitigate the harms of mental health misinformation have chilling effects on those who intend to help others with mental health struggles online?

We believe research projects that properly address these open questions will play a crucial role towards understanding and mitigating mental health misinformation on video-sharing platforms. Such results would, at the same time, increase credibility and trust in video-sharing platforms as an asset for digital mental health.

4 CHALLENGES

Below are some of the challenges we envision when studying video-sharing platforms. We seek to foster discussion on how to tackle these challenges during our participation in the workshop.

- Most video-sharing platforms, such as TikTok and YouTube, either have no official APIs or official APIs that are severely limiting in accessibility and capabilities. However, studying them through unofficial APIs and other technical solutions may bring about ethical as well as legal challenges.
- The multimodal nature of VSP content (music, speech, visuals, textual comments) makes it significantly more difficult to automatically detect misinformation compared to text-only content that is predominant on other social media platforms.
- Assessing what is ground truth for mental health misinformation is challenging, as the scientific knowledge and landscape itself have differing perspectives at times, especially given the subjective experience of mental illness and the subjective nature of the psychiatric treatment.

5 WORKSHOP PARTICIPATION

As a group, we have over a decade's worth of experience working toward understanding and fostering human well-being and safety on social media platforms. Participating in this workshop, we hope to contribute our insights on how to foster credibility, trust, and safety on video-sharing platforms. We also hope to gain insights from workshop participants on how best to solve the open questions we posed above, with a view toward design and technical solutions to mitigate mental health misinformation on video-sharing platforms.

REFERENCES

[1] Laura Helena Andrade, J Alonso, Z Mneimneh, JE Wells, A Al-Hamzawi, G Borges, E Bromet, Ronny Bruffaerts, G De Girolamo, R De Graaf, et al. 2014. Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychological medicine* 44, 6 (2014), 1303–1317.
[2] Holly Avella. 2023. "TikTok≠ therapy": Mediating mental health and algorithmic mood disorders. *New Media & Society* (2023), 14614448221147284.
[3] Corey H Basch, Lorie Donelle, Joseph Fera, and Christie Jaime. 2022. Deconstructing TikTok videos on mental health: cross-sectional, descriptive content analysis. *JMIR formative research* 6, 5 (2022), e38340.
[4] Andrew Devendorf, Ansley Bender, and Jonathan Rottenberg. 2020. Depression presentations, stigma, and mental health literacy: A critical review and YouTube content analysis. *Clinical Psychology Review* 78 (2020), 101843.
[5] Gabriele Donzelli, Giacomo Palomba, Ileana Federigi, Francesco Aquino, Lorenzo Cioni, Marco Verani, Annalaura Carducci, and Pierluigi Lopalco. 2018. Misinformation on vaccination: A quantitative analysis of YouTube videos. *Human vaccines & immunotherapeutics* 14, 7 (2018), 1654–1659.
[6] Christine Geeng, Savanna Yee, and Franziska Roesner. 2020. Fake news on Facebook and Twitter: Investigating how people (don't) investigate. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.
[7] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019), 374–378.
[8] Tatum Hunter. 2022. How to vet mental health advice on TikTok and Instagram. https://www.washingtonpost.com/technology/2022/10/03/tiktok-instagram-mental-health/
[9] Tatum Hunter. 2022. Online creators are de facto therapists for millions. it's complicated. https://www.washingtonpost.com/technology/2022/08/29/mental-health-tiktok-instagram/

[10] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.

[11] Heidi Oi-Yee Li, Adrian Bailey, David Huynh, and James Chan. 2020. YouTube as a source of information on COVID-19: a pandemic of misinformation? *BMJ global health* 5, 5 (2020), e002604.

[12] Paul Mena, Danielle Barbe, and Sylvia Chan-Olmsted. 2020. Misinformation on Instagram: The impact of trusted endorsements on message credibility. *Social Media+ Society* 6, 2 (2020), 2056305120935102.

[13] Francesco Pierri, Luca Luceri, Nikhil Jindal, and Emilio Ferrara. 2022. Propaganda and Misinformation on Facebook and Twitter during the Russian Invasion of Ukraine. *arXiv preprint arXiv:2212.00419* (2022).

[14] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. NLP-based feature extraction for the detection of COVID-19 misinformation videos on YouTube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.

[15] Lanyu Shang, Ziyi Kou, Yang Zhang, and Dong Wang. 2021. A multimodal misinformation detector for covid-19 short videos on tiktok. In *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 899–908.

[16] Victor Suarez-Lledo and Javier Alvarez-Galvez. 2021. Prevalence of health misinformation on social media: systematic review. *Journal of medical Internet research* 23, 1 (2021), e17187.

[17] Briony Swire-Thompson and David Lazer. 2019. Public health and online misinformation: challenges and recommendations. *Annual review of public health* 41 (2019), 433–451.

[18] Amir Tal and John Torous. 2017. The digital mental health revolution: Opportunities and risks. (2017).

[19] PlushCare Content Team. 2022. How accurate is mental health advice on TikTok? https://plushcare.com/blog/tiktok-mental-health/

[20] Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, and Maria Bielikova. 2021. An audit of misinformation filter bubbles on YouTube: Bubble bursting and recent behavior changes. In *Proceedings of the 15th ACM Conference on Recommender Systems*. 1–11.

[21] John Torous, Keris Jän Myrick, Natali Rauseo-Ricupero, Joseph Firth, et al. 2020. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. *JMIR mental health* 7, 3 (2020), e18848.

[22] Milton L Wainberg, Pamela Scorza, James M Shultz, Liat Helpman, Jennifer J Mootz, Karen A Johnson, Yuval Neria, Jean-Marie E Bradford, Maria A Oquendo, and Melissa R Arbuckle. 2017. Challenges and opportunities in global mental health: a research-to-practice perspective. *Current psychiatry reports* 19 (2017), 1–10.

[23] Philip S Wang, Patricia Berglund, Mark Olfson, Harold A Pincus, Kenneth B Wells, and Ronald C Kessler. 2005. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry* 62, 6 (2005), 603–613.

[24] Haijun Xia, Hui Xin Ng, Zhutian Chen, and James Hollan. 2022. Millions and Billions of Views: Understanding Popular Science and Knowledge Communication on Video-Sharing Platforms. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. 163–174.

[25] Jiaheng Xie, Yidong Chai, and Xiao Liu. 2022. An Interpretable Deep Learning Approach to Understand Health Misinformation Transmission on YouTube. In *Proceedings Of The 55th Hawaii International Conference On System Sciences*.

[26] Anthony Yeung, Enoch Ng, and Elia Abi-Jaoude. 2022. TikTok and attention-deficit/hyperactivity disorder: a cross-sectional study of social media content quality. *The Canadian Journal of Psychiatry* 67, 12 (2022), 899–906.

[27] Chengyan Zhu, Xiaolin Xu, Wei Zhang, Jianmin Chen, and Richard Evans. 2020. How health communication via Tik Tok makes a difference: A content analysis of Tik Tok accounts run by Chinese provincial health committees. *International journal of environmental research and public health* 17, 1 (2020), 192.