

Human-Subject Research Strategies on Social Media Misinformation

Filipo Sharevski^{1,2}

¹DePaul University, Chicago, IL

²Adversarial Cybersecurity Automation Lab (ACAL)

LASER workshop 2023



Misinformation: Science/System Perspective

What → Online misinformation is *not* the information we want *ergo* has to go (mistaken information, fake, factoids, out-of-context facts?) [13, 14]

Why → Misinformation is “sticky” (*a thousand times repeated a lie...*) and impersonates the truth (is this an absolute concept?) [3, 4]

Where → Mainstream social media, alt-platforms and fringe communities, boards, chatrooms, discords, *chans... [7, 12]

When → All the time, only for big topics, never (what if it's Onion news?)

Who → Nation-states, political parties, biased media, corporations, people,... (language models, ChatGPT?)

How → Prebunking, debunking, moderating, removing (censoring?) [10]



Misinformation: User Perspective

What → How do I know *what* is misinformation in a first place? [9]

Why → Is this a "plandemic" or someone is "stealing the vote?" [2, 1]

Where → Used to be Facebook, Twitter, Instagram, YouTube... is it now on TikTok, Discord, and Eve Online? [6]

When → Every time there are elections, a pandemic, a Supreme Court decision, a strange baloon in the sky...? [8]

Who → Russians for sure, but... also people from the *other side* [9]

How → Troll farms, think tanks, political figureheads, Alexa/Siri/Cortana, now chatbots too [11, 5]



Misinformation: User Perspective



source



Dimension: Research Settings

Platform Dimension

Platform Perception Matters → Users are attuned to platform's misinformation image

Favoritism → Objectivity in perception of misinformation handling

Misinformation Dimension

Concepts → Multiple competing *folk models* of misinformation exist

Self-Image → Misinformation is part of users' *identity* on social media

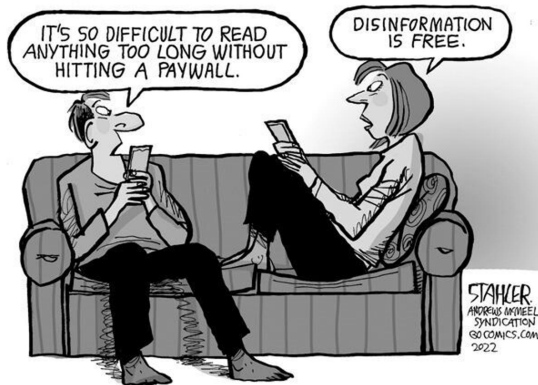
Participation Dimension

Engagement → Users subjectively *engage/disengage* with misinformation per *topic* per platform of *choice*

Assessment → Misinformation is both *analytically* and *heuristically* assessed



Economics Dimension: Research Settings



source



Misinformation: Assumptions, Limitations, Delimitations

Assumptions

Literacy → Information literacy, social media literacy

Exposure → Subjective designation of content as “fake news”

Interaction → Lived experiences interacting with “fake news”

Limitations

Expiry Date → falsehoods/myths/rumors proven true and factual

Production Date → facts proven falsehoods

Platform Interference → Data availability, moderation decisions, rollbacks

Delimitations → low hanging fruit (e.g. COVID-19, elections, QAnon...)



Misinformation: Instruments and Data Collection

Surveys

Self-presentation → Misinformation is internalized and people fear of being portrait in pro/con lights

Open-Ended Questions → People are *opinionated* on misinformation

Debriefing → Misinformation constitutes “deception” in the view of IRB (err on the side of caution)

Interviews

Agnostic posture → People are *not keen* on researchers

Anonymity → Allow for dropping our and removing answers at any time of the study even after you send them your draft paper

Trust and Credibility → Critically important for longitudinal studies

Focus Groups → topic, party, platform, clique



Misinformation: Data Analysis and Interpretation

Analysis

Themes → People use analogies, vernacular, offensive language

Tools → Tools help (e.g. nvivo, RQDA, MAXRQDA), but inter-coding and manual inference helps a lot

Interpretation

Context → Do we do this for psychology, cybersecurity (misinformation as attack on *integrity*?), policy, law, economics...?

Contradictions → Many misinformation effects are brief, strange, and contradictory evidence abound all the time

Implications → How the human input from the study changes the future shape of *misinformation* or platform's (lack of) handling it?



Misinformation: Experimental Lessons Learned

Animosity → Non-negligible number of people harbor quite an animosity surrounding misinformation

Threatening → It's not entirely foreign for researchers to get real threats to their well beings and families

Volatility → Misinformation evolves and rapidly reshapes even during the short period of study data collation and analysis

Ephemeral Solutions → Results and solutions for one form of misinformation might not work for another (text, mime, trolling, deepfakes, ChatGPT outputs)



Discussion



Discussion Talking Points

Did you use experimentation artifacts borrowed from the community?

Did you attempt to replicate or reproduce results of earlier research as part of your work?

What can be learned from your methodology and your experience using it?

What did you try that did not succeed before getting to the results you presented?

Did you produce any intermediate results including possible unsuccessful tests or experiments?

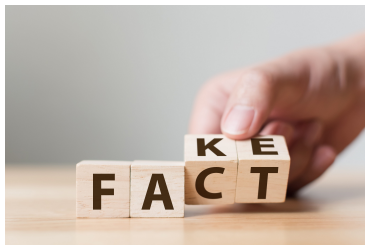
Did you share experimentation artifacts with the community?



Discussion Next Steps

Next steps

Plans for post-workshop paper



Thank you!

Questions, Comments, Concerns



Twitter: @ACALaboratory

References I

- [1] Peter Jachim, Filip Sharevski, and Emma Pieroni. “TrollHunter2020: Real-time Detection of Trolling Narratives on Twitter During the 2020 US Elections”. In: *International Workshop on Security and Privacy Analytics 2021*. IWSPA '21. <https://doi.org/10.1145/3445970.3451158>. Online, USA: Association for Computing Machinery, 2021, pp. 1–11. DOI: 10.1145/3445970.3451158.
- [2] Peter Jachim, Filip Sharevski, and Paige Treebridge. “TrollHunter [Evader]: Automated Detection [Evasion] of Twitter Trolls During the COVID-19 Pandemic”. In: *New Security Paradigms Workshop 2020*. NSPW '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 59–75.
- [3] Stephan Lewandowsky and Sander van der Linden. “Countering Misinformation and Fake News Through Inoculation and Prebunking”. In: *European Review of Social Psychology* 32.2 (2021), pp. 348–384.
- [4] Rakoena Maertens et al. “Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments.”. In: *Journal of Experimental Psychology: Applied* 27.1 (2021), p. 1.
- [5] Filip Sharevski and Donald Gover. “Two Truths and a Lie: Exploring Soft Moderation of COVID-19 Misinformation with Amazon Alexa”. In: *Proceedings of the 16th International Conference on Availability, Reliability and Security*. ARES 21. Vienna, Austria: Association for Computing Machinery, 2021. ISBN: 9781450390514. DOI: 10.1145/3465481.3470017. URL: <https://doi-org.ezproxy.depaul.edu/10.1145/3465481.3470017>.



References II

- [6] Filipo Sharevski and Benjamin Kessell. *Fight Fire with Fire: Hacktivists' Take on Social Media Misinformation*. 2023. DOI: 10.48550/ARXIV.2302.07788. URL: <https://arxiv.org/abs/2302.07788>.
- [7] Filipo Sharevski et al. “Gettr-ing” User Insights from the Social Network Gettr. https://truthandtrustonline.com/wp-content/uploads/2022/10/TTO_2022_proceedings.pdf. Boston, MA, 2022.
- [8] Filipo Sharevski et al. *Abortion Misinformation on TikTok: Rampant Content, Lax Moderation, and Vivid User Experiences*. 2023. DOI: 10.48550/ARXIV.2301.05128. URL: <https://arxiv.org/abs/2301.05128>.
- [9] Filipo Sharevski et al. *Folk Models of Misinformation on Social Media*. 2022. DOI: 10.48550/ARXIV.2207.12589. URL: <https://arxiv.org/abs/2207.12589>.
- [10] Filipo Sharevski et al. *Meaningful Context, a Red Flag, or Both? Users' Preferences for Enhanced Misinformation Warnings on Twitter*. 2022. DOI: 10.48550/ARXIV.2205.01243. URL: <https://arxiv.org/abs/2205.01243>.
- [11] Filipo Sharevski et al. *Talking (Mis)Information with ChatGPT on TikTok*. 2023. DOI: 10.48550/ARXIV.2301.05128. URL: <https://arxiv.org/abs/2301.05128>.



- [12] Filipo Sharevski et al. “VoxPop: An Experimental Social Media Platform for Calibrated (Mis)Information Discourse”. In: *New Security Paradigms Workshop*. NSPW '21. Virtual Event, USA: Association for Computing Machinery, 2021, pp. 88–107. ISBN: 9781450385732. DOI: 10.1145/3498891.3498893. URL: <https://doi.org/10.1145/3498891.3498893>.
- [13] Liang Wu et al. “Misinformation in Social Media: Definition, Manipulation, and Detection”. In: *SIGKDD Explor. Newsl.* 21.2 (Nov. 2019), pp. 80–90. ISSN: 1931-0145. DOI: 10.1145/3373464.3373475. URL: <https://doi.org/10.1145/3373464.3373475>.
- [14] Savvas Zannettou et al. “What is Gab: A Bastion of Free Speech or an Alt-Right Echo Chamber”. In: *Companion Proceedings of the The Web Conference 2018*. WWW '18. Lyon, France: International World Wide Web Conferences Steering Committee, 2018, pp. 1007–1014. ISBN: 9781450356404. DOI: 10.1145/3184558.3191531. URL: <https://doi.org/10.1145/3184558.3191531>.

