

From Bad to Worse: Using Private Data to Propagate Disinformation on Online Platforms with a Greater Efficiency

Protik Bose Pranto
ppranto@asu.edu
Arizona State University
Tempe, Arizona, USA

Waqar Hassan Khan
wkhan17@asu.edu
Arizona State University
Tempe, Arizona, USA

Sahar Abdelnabi
sahar.abdelnabi@cispa.de
CISPA Helmholtz Center for
Information Security
Saarbrücken, Saarland

Rebecca Weil
weil@cispa.de
CISPA Helmholtz Center for
Information Security
Saarbrücken, Saarland

Mario Fritz
fritz@cispa.de
CISPA Helmholtz Center for
Information Security
Saarbrücken, Saarland

Rakibul Hasan
rhasan3@asu.edu
Arizona State University
Tempe, Arizona, USA

ABSTRACT

We outline a planned experiment to investigate if personal data (e.g., demographics and behavioral patterns) can be used to selectively expose individuals to disinformation such that an adversary can spread disinformation more efficiently compared to broadcasting the same information to everyone. This mechanism, if effective, will have devastating consequences as modern technologies collect and infer a plethora of private data that can be abused to target with disinformation. We believe this research will inform designing policy and regulation for online platforms.

CCS CONCEPTS

• **Security and privacy** → **Social aspects of security and privacy**; **Usability in security and privacy**.

1 PROBLEM SPACE

Technologies collect huge amount of personal and potentially sensitive data, including demographic, bio-metric, behavioral patterns, and (dis)interests. These data may be collected, shared, or used with or without people’s consent [1, 5, 11] and digitally stored, making them permanent, replicable, and re-shareable [23], heightening the concern regarding numerous privacy risks, such as data breaches, hacking, identity theft, and other forms of unauthorized access (e.g., [7, 15, 25, 30, 31, 37]). We plan to investigate whether personal information can be abused to propagate disinformation, by exposing individuals to specific disinformation based on their demographic attributes, personality traits, behavioral patterns, physiological and emotional states, or other properties. Thus, our research is at the intersection of privacy and disinformation, where we aim to investigate if private data can be used to “target” individuals with disinformation to elicit desired outcomes, such as increasing the likelihood of believing the information or further propagating it to others, or both.

We hypothesize that an individual, when targeted with disinformation based on some property p (e.g., gender, socio-economic status, interests, personality traits, etc.), will *react differently* compared to other people who do not possess property p . For example, health-conscious individuals may react differently to health-related disinformation than individuals with less interest in such news.

The reaction can be either positively or negatively correlated with the level of interest (i.e., a higher or lower level of trust in or engagement with the provided information). We aim to investigate the magnitude and direction of change in reaction from individuals targeted with disinformation compared to showing the same information to a random set of people.

Prior research has reported that certain properties, e.g. age, gender, emotional intelligence, etc., are associated with a higher level of trust in fake news. For instance, individuals with low emotional intelligence are more susceptible to false information than individuals with high emotional intelligence [29]. Additionally, older people were found to be more likely to believe false information than younger people [4, 14, 27]. Previous research also found a positive correlation between education level and false news acceptance [8, 27]. Thus, it is likely that by serving different disinformation to different groups of the population that possess these properties, an adversary will be able to propagate disinformation more efficiently than by broadcasting the same message to the whole population.

Moreover, the advancement of Extended Reality, such as Augmented Reality (AR) [6], Virtual Reality (VR) [43] technologies are closing the gap between the physical and virtual worlds [44]. Even in the Metaverse [28], interpersonal engagement is more intense and scalable than in traditional social media environments [38]. Such interactions may allow adversaries to exploit an unprecedented amount of personal information and use them to target people with disinformation.

Conversely, in some cases, targeting may result in a lower level of trust in disinformation. E.g., if someone is targeted based on their interests or hobbies, they might be less likely to believe in that information because of their prior knowledge on that topic of interest [33]. Yet, a vast amount of research suggests that the exact opposite might be true. Due to a processing advantage of familiar information, familiarity with the topic may lead to the impression that the information is true [12, 13]. However, identifying characteristics of people who are resilient to targeted disinformation will benefit future research in crowdsourced fact-checking (e.g., by directing potential disinformation to fact-checkers with these characteristics).

For the cases where targeting leads to a higher level of trust or engagement, the consequences can be disastrous. Since individuals' preexisting ideas and interests are positively correlated with their acceptance of false information [34], targeting may even create opportunities to mislead sub-populations who were previously believed to be resilient to disinformation, e.g., highly educated individuals [8] and young adult [4], if they find the served information aligned with their personal characteristics and interests. For example, young adults who are concerned about health, climate, etc. can be misled by deceptive articles [10, 36], especially if they are emotionally invested in that topic [22]. Furthermore, platform users build networks with others who have similar interests [21] and share information they find interesting with their connections [8, 32], massively scaling up the number of affected people who can further propagate the information at no cost to the adversary.

Additionally, posts generated by Artificial Intelligence (AI) can be highly emotional [9] and can be used to target individuals with low emotional intelligence [29] that can be inferred from breached data [18, 24], sensor data [19], or online personality tests [5]. Moreover, AI is currently used for personalized recommendations on digital platforms, healthcare, and marketing, which depend heavily on user characteristics, behaviors, demographics, and preferences data [16, 20, 35]. Since AI technologies such as DeepFake [39], botnet [2], ChatGPT [40] etc., are cheap, fast, scalable, and able to create personalized content [17], adversaries may soon be capable of generating targeted disinformation automatically and faster than a human being.

2 EXPERIMENTAL DESIGN

As a first step towards understanding the effect of targeted disinformation, we designed a study where we target based on people's demographic attributes and topical interests (e.g., healthy diet, cooking, celebrity news, gardening, movie, books, etc.). Each participant will read 20 recently published news articles manually collected from high [26] and low-credible [42] news sources selected depending on NewsGuard's reliability score [41]. Also, the low-credible news sources selected for this study tend to publish false news [3]. However, which news item stems from which source will not be revealed to participants. They will be asked whether or not they believe the news to be true. Suppose we find that people believe false information that matches their self-reported interests more than false information that doesn't match their self-reported interests. In that case, we can infer that targeted disinformation influences people more than any non-targeted misinformation. In particular, if targeted disinformation turns out to be effective in propagating disinformation, this finding will strengthen the case for better protection of consumer data with regulations as well as platform design choices. As such, this study will inform the design of online platforms and policies regarding the collection and use of personal data.

REFERENCES

- [1] Elizabeth Aguirre, Dominik Mahr, Dhruv Grewal, Ko De Ruyter, and Martin Wetzels. 2015. Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of retailing* 91, 1 (2015), 34–49.
- [2] K Alieyan, A Almomani, R Abdullah, B Almutairi, and M Alauthman. 2021. Botnet and Internet of Things (IoT): A definition, taxonomy, challenges, and future directions. In *Research Anthology on Combating Denial-of-Service Attacks*. IGI Global, 138–150.
- [3] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & Politics* 6, 2 (2019), 2053168019848554.
- [4] Nadia M Brashier and Daniel L Schacter. 2020. Aging in an era of fake news. *Current directions in psychological science* 29, 3 (2020), 316–323.
- [5] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The guardian* 17, 1 (2018), 22.
- [6] J Carmigniani and B Furht. 2011. Augmented reality: an overview. *Handbook of augmented reality*. (2011), 3–46.
- [7] Rajarshi Chakraborty, Jaeung Lee, Sharmistha Bagchi-Sen, Shambhu Upadhyaya, and H Raghav Rao. 2016. Online shopping intention in the context of data breach in online retail stores: An examination of older and younger adults. *Decision Support Systems* 83 (2016), 47–56.
- [8] Xinran Chen, Sei-Ching Joanna Sin, Yin-Leng Theng, and Chei Sian Lee. 2015. Why students share misinformation on social media: Motivation, gender, and study-level differences. *The journal of academic librarianship* 41, 5 (2015), 583–592.
- [9] De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of*.
- [10] Tara Culp-Ressler. [n.d.]. How Food Companies Trick You Into Thinking You're Buying Something Healthy — archive.thinkprogress.org. <https://archive.thinkprogress.org/how-food-companies-trick-you-into-thinking-youre-buying-something-healthy-4b5c3cc960b/>. [Accessed 24-Feb-2023].
- [11] Asunción Esteve. 2017. The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *Int. Data Priv. Law* 7, 1 (Feb. 2017), 36–47.
- [12] Lisa K Fazio, Sarah J Barber, Suparna Rajaram, Peter A Ornstein, and Elizabeth J Marsh. 2013. Creating illusions of knowledge: learning errors that contradict prior knowledge. *Journal of Experimental Psychology: General* 142, 1 (2013), 1.
- [13] Rainer Greifeneder, Mariela Jaffe, Eryn Newman, and Norbert Schwarz. 2021. *The psychology of fake news: Accepting, sharing, and correcting misinformation*.
- [14] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances* 5, 1 (2019), eaau4586.
- [15] Ashiq JA. [n.d.]. Hackers Selling Healthcare Data in the Black Market | Infosec Resources — resources.infosecinstitute.com. <https://resources.infosecinstitute.com/topic/hackers-selling-healthcare-data-in-the-black-market/>. [Accessed 23-Feb-2023].
- [16] Kevin B Johnson, Wei-Qi Wei, Dilhan Weeraratne, Mark E Frisse, Karl Misulis, Kyu Rhee, Juan Zhao, and Jane L Snowdon. 2021. Precision medicine, AI, and the future of personalized health care. *Clinical and translational science* 14, 1 (2021), 86–93.
- [17] Daniel Kang, Xuechen Li, Ion Stoica, Carlos Guestrin, Matei Zaharia, and Tatsunori Hashimoto. 2023. Exploiting Programmatic Behavior of LLMs: Dual-Use Through Standard Security Attacks. *arXiv preprint arXiv:2302.05733* (2023).
- [18] Mark Keierleber. [n.d.]. Trove of L.A. Students' Mental Health Records Posted to Dark Web After Cyber Hack — the74million.org. <https://www.the74million.org/article/trove-of-l-a-students-mental-health-records-posted-to-dark-web-after-cyber-hack/>. [Accessed 23-Feb-2023].
- [19] David Kotz, Carl A Gunter, Santosh Kumar, and Jonathan P Weiner. 2016. Privacy and security in mobile health: a research agenda. *Computer* 49, 6 (2016), 22–30.
- [20] Vipin Kumar, Bharath Rajan, Rajkumar Venkatesan, and Jim Lecinski. 2019. Understanding the role of artificial intelligence in personalized engagement marketing. *California Management Review* 61, 4 (2019), 135–155.
- [21] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.
- [22] Cameron Martel, Gordon Pennycook, and David G. Rand. 2020. Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications* 5, 1 (Dec. 2020), 47. <https://doi.org/10.1186/s41235-020-00252-3>
- [23] Alice E Marwick and Danah Boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New media & society* 16, 7 (2014), 1051–1067.
- [24] Marianne Kolbasuk McGee. [n.d.]. More Breaches Expose Mental Health, Substance Abuse Data — careersinfosecurity.com. <https://www.careersinfosecurity.com/more-breaches-expose-mental-health-substance-abuse-data-a-9390/>. [Accessed 23-Feb-2023].
- [25] Bernard Meyer. 2021. COMB: largest breach of all time leaked online with 3.2 billion records. *Cybernews, February* (July 2021).
- [26] Nicholas Micallef, Mihai Avram, Filippo Menczer, and Sameer Patil. 2021. Fakey: A game intervention to improve news literacy on social media. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–27.

- [27] Sophie Morosoli, Peter Van Aelst, Edda Humprecht, Anna Staender, and Frank Esser. 2022. Identifying the drivers behind the dissemination of online misinformation: a study on political attitudes and individual characteristics in the context of engaging with misinformation on social media. *American Behavioral Scientist* (2022), 00027642221118300.
- [28] Stylianos Mystakidis. 2022. Metaverse. *Encyclopedia* 2, 1 (2022), 486–497.
- [29] Stephanie Preston, Anthony Anderson, David J Robertson, Mark P Shephard, and Narisong Huhe. 2021. Detecting fake news on Facebook: The role of emotional intelligence. *Plos one* 16, 3 (2021), e0246757.
- [30] Brian Quarumby. 2022. 400m Twitter users' data is reportedly on sale in the black market. <https://cointelegraph.com/news/400m-twitter-users-data-is-reportedly-on-sale-in-the-black-market>
- [31] Jim Reed. 2019. EE Data Breach 'led to stalking'. <https://www.bbc.com/news/technology-46896329>
- [32] Samuel C Rhodes. 2022. Filter bubbles, echo chambers, and fake news: How social media conditions individuals to be less critical of political misinformation. *Polit. Commun.* 39, 1 (Jan. 2022), 1–22.
- [33] Tobias Richter, Sascha Schroeder, and Britta Wöhrmann. 2009. You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of personality and social psychology* 96, 3 (2009), 538.
- [34] Laura D Scherer, Jon McPhetres, Gordon Pennycook, Allison Kempe, Larry A Allen, Christopher E Knoepke, Channing E Tate, and Daniel D Matlock. 2021. Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychol.* 40, 4 (April 2021), 274–284.
- [35] Donghee Shin. 2020. User perceptions of algorithmic decisions in the personalized AI system: perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media* 64, 4 (2020), 541–565.
- [36] Nicole D. Sintov, Victoria Abou-Ghalioum, and Lee V. White. 2020. The partisan politics of low-carbon transport: Why democrats are more likely to adopt electric vehicles than Republicans in the United States. *Energy Research & Social Science* 68 (Oct. 2020), 101576. <https://doi.org/10.1016/j.erss.2020.101576>
- [37] Daniel J Solove and Danielle Keats Citron. 2017. Risk and anxiety: A theory of data-breach harms. *Tex. L. Rev.* 96 (2017), 737.
- [38] Rand Waltzman. 2022. Facebook misinformation is bad enough. The metaverse will be worse. <https://www.rand.org/blog/2022/08/facebook-misinformation-is-bad-enough-the-metaverse.html>. Accessed: 2023-2-22.
- [39] Mika Westerlund. 2019. The emergence of deepfake technology: A review. *Technol. Innov. Manag. Rev.* 9, 11 (Jan. 2019), 39–52.
- [40] Wikipedia. 2023. ChatGPT — Wikipedia, The Free Encyclopedia. <http://en.wikipedia.org/w/index.php?title=ChatGPT&oldid=1141030814>. [Online; accessed 23-February-2023].
- [41] Wikipedia. 2023. NewsGuard — Wikipedia, The Free Encyclopedia. <http://en.wikipedia.org/w/index.php?title=NewsGuard&oldid=1140342433>. [Online; accessed 24-February-2023].
- [42] Jiding Zhang, Ken Moon, and Senthil K Veeraraghavan. 2022. Does Fake News Create Echo Chambers? Available at SSRN 4144897 (2022).
- [43] JM Zheng, KW Chan, and Ian Gibson. 1998. Virtual reality. *Ieee Potentials* 17, 2 (1998), 20–23.
- [44] F Zünd, M Ryffel, S Magnenat, A Marra, M Nitti, M Kapadia, G Noris, K Mitchell, M Gross, and R W Sumner. 2015. Augmented creativity: bridging the real and virtual worlds to enhance creative play. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*. 1–7.