



Misinformation Mitigation Praxis: Lessons Learned and Future Directions from Co-Insights

Scott A. Hale
scott@meedan.com
Meedan & University of Oxford
Oxford, UK

Kiran Garimella
kiran.garimella@rutgers.edu
School of Communication and
Information
Rutgers University
New Brunswick, New Jersey, USA

Shiri Dori-Hacohen*
shiridh@uconn.edu
Reducing Information Ecosystem
Threats Lab
University of Connecticut
Storrs, Connecticut, USA

ABSTRACT

Misinformation is a global challenge, but successful mitigations must come from the communities affected and not be imposed by external entities. Co-Insights is a multi-year NSF-funded initiative building capacity to respond to misinformation and harmful narratives in Asian American and Pacific Islander (AAPI) communities, based on a deep involvement with grassroots organizations and a co-construction of tools grounded on community needs. Co-Insights' unique cross-sectoral, cross-disciplinary collaboration is a convergence of information retrieval, computational social science, and ethnographic inquiry with a unique platform that enables community organizations, fact-checkers, and academics to work together to respond effectively to harmful content targeting communities. In this SIRIP talk designed for a technical audience, we will share lessons learned from the first 2.5 years of Co-Insights, and how we are bridging the academic-praxis divide by integrating state-of-the-art research developments into large-scale deployed systems. Topics covered will include: the challenges and joys of collaborating across disciplinary and sectoral boundaries; community-driven approaches that utilize information retrieval techniques such as claim matching in concert with emerging best practices in misinformation mitigation; open problems we have encountered in the space; and future directions we find promising. Co-Insights is led by Meedan, a global non-profit providing award-winning software solutions to mitigate misinformation; and in partnership with AAPI community organizations and several academic institutions.

CCS CONCEPTS

• **Information systems** → **Information retrieval**; **World Wide Web**; **Web mining**.

KEYWORDS

information ecosystem threats; misinformation; community-led; minoritized communities

*Senior corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.

SIGIR '24, July 14–18, 2024, Washington, DC, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0431-4/24/07.
<https://doi.org/10.1145/3626772.3661352>

ACM Reference Format:

Scott A. Hale, Kiran Garimella, and Shiri Dori-Hacohen. 2024. Misinformation Mitigation Praxis: Lessons Learned and Future Directions from Co-Insights. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*, July 14–18, 2024, Washington, DC, USA. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3626772.3661352>

1 INTRODUCTION

Misinformation is a global challenge; yet, minoritized communities are often neglected by mainstream misinformation mitigation approaches for a variety of reasons. In order to avoid paternalism [1, 4] or bias [6, 9], and to construct successful mitigations, solutions must come from the communities affected, rather than imposed by external entities [1]. Computational research on the topic should be informed both by the social sciences and by real-world needs and constraints, and—ideally—integrated into misinformation mitigation approaches utilized in the wild [2]. In practice, working across sectoral and disciplinary boundaries is far easier said than done; academics trained in different fields are effectively speaking different “languages” [3], and it is reasonable to expect that these barriers are similarly reproduced in the academic-practitioner divide, as well as between academia and civil society.

Our response to these challenges is **Co-Insights: Fostering Community Collaboration to Combat Misinformation**.¹ Co-Insights is a multi-year NSF-funded initiative constructing technology to counter misinformation and harmful narratives in Asian American and Pacific Islander (AAPI) communities, based on a deep involvement with grassroots organizations and a co-construction of tools based on community needs. Co-Insights' unique cross-sectoral, cross-disciplinary collaboration is a convergence of information retrieval, computational social science, and ethnographic inquiry with a unique platform that enables community organizations, fact-checkers, and academics to work together to respond effectively to harmful content targeting communities. In this SIRIP talk designed for a technical audience, we will share lessons learned from the first 2.5 years of Co-Insights, and how we are bridging the academic-praxis divide by integrating state-of-the-art research developments into large-scale deployed systems.

The remainder of this paper will be organized as follows. First, we will introduce the Co-Insights initiative. Next, we will describe the talk contents in detail. Finally, we will share descriptions of the organizations presenting this talk as well as speaker bios.

¹<https://meedan.com/project/co-insights>

2 THE CO-INSIGHTS INITIATIVE

Co-Insights is a multi-year NSF-funded initiative constructing technology to counter for misinformation and harmful narratives in Asian American and Pacific Islander (AAPI) communities, based on a deep involvement with grassroots organizations and a co-construction of tools based on community needs. Co-Insights' unique cross-sectoral, cross-disciplinary collaboration is a convergence of information retrieval, computational social science and ethnographic inquiry with a unique platform that enables community organizations, fact-checkers, and academics to work together to respond effectively to harmful content targeting communities. The Co-Insights initiative is led by Meedan,² a technology non-profit that builds software for fact-checkers and journalists worldwide. Participant organizations include the **Annenberg Public Policy Center** at the University of Pennsylvania, which operates **FactCheck.org**,³ and several **AAPI grassroots nonprofits** operating in the fact-checking space:

- **Piyaoba**,⁴ Chinese for Affirmative Action (CAA), is a centralized fact-checking portal to combat misinformation and disinformation directed at Chinese-Americans. “Piyaoba” means “let’s fact-check it” in Mandarin Chinese. Piyaoba offers in-language resources and tools to help Chinese Americans to combat misinformation and disinformation effectively, such as real-time alerts, reports, and in-depth fact-checking articles. CAA has three full-time and 15 part-time staff, including social scientists, data analysts, software engineers, fact-checkers, and Chinese-language writers. CAA is also a founding partner of Stop AAPI Hate.
- **Tayo**,⁵ a project of the Filipino Young Leaders Program (FYLPRO), is an innovative data hub that empowers Filipino-American communities by collecting data, fostering partnerships, publishing culturally relevant insights, fact-checking, and developing leaders to create an equitable and sustainable future. Tayo has been running a tipline service for over three years.
- **Viet Fact Check**⁶ has a non-partisan network of volunteers across the United States who seek to empower Vietnamese Americans with fact-checked, source-verified analysis and rebuttals in English and Vietnamese to combat the onslaught of misinformation that is circulating widely in our Vietnamese American communities.
- **DesiFacts**,⁷ a project of the non-profit Indian American Impact, regularly publishes fact-checks and other resources in Hindi, Bengali, Tamil, and English based on the latest claims going viral on WhatsApp. They have also developed social media kits and graphics users can use to proactively spread accurate content in their own networks as well as media literacy resources. Their “TrueCrew” network of volunteers help promote trusted content and identify potentially problematic content at the national and local levels.

²<https://meedan.com/>

³<https://factcheck.org/>

⁴<https://www.piyaoba.org/>

⁵<https://www.tayohelp.com/>

⁶<https://vietfactcheck.org/>

⁷<https://www.desifacts.org/>

Co-Insights also includes several **academic institutions**: University of Massachusetts Amherst, University of Connecticut, Rutgers University, University of Colorado Boulder and the aforementioned University of Pennsylvania.

3 TALK DESCRIPTION

In this SIRIP talk, geared towards a technical, IR-expert audience, we will share lessons learned from the first 2.5 years of Co-Insights, and emerging best practices in the misinformation mitigation space. We will also discuss open problems we have encountered, and future directions we find promising. We will share our approach and details on how we are integrating state-of-the-art IR and AI/ML research developments into large-scale deployed systems, bridging the academic-praxis divide.

During our talk, we plan to cover the following topics:

(1) Lessons Learned:

- The considerable challenges, and joys, of collaborating across disciplinary and sectoral boundaries;
- Strategies to improve team alignment and collaboration;
- Ethnographic and customer discovery approaches that can be used by anyone to uncover unmet needs in your constituents and partners;
- Best practices in misinformation mitigation, such as utilizing pre-bunking, inoculation, and preemptive explainers for common misinformation themes;
- Utilizing social media APIs and LLMs to rapidly generate targeted topic-specific, synthetically labeled datasets for model fine-tuning purposes [12]; and
- Innovative data collection strategies for studying information ecosystem threats on platforms used by minoritized communities (such as WhatsApp [5]).

(2) Open Problems and Future Directions:

- Detecting and mitigating cross-platform [11], cross-lingual [7, 8], and diasporic community [10] misinformation spread;
- The nuanced and intertwined nature of misinformation, bias, polarization, and harmful narratives [2];
- How state-of-the-art claim matching approaches can be modified to serve up high-quality explainers; and
- Constructing successful interventions on a variety of platforms used in practice by diasporic communities, especially end-to-end encryption messaging platforms such as WhatsApp.

Finally, we will share our sustainability plans to keep Co-Insights operational long after the end of our NSF contract, before opening the floor up for questions and discussion.

4 SPEAKERS

Our speakers bring a wealth of industry, non-profit and academic expertise to the topic at hand.

4.1 Institutions

The organizations directly contributing to this talk bridge misinformation praxis and academic research.

Meedan is a global technology not-for-profit that builds software and programmatic initiatives to strengthen journalism, digital literacy and accessibility of information online and off. Meedan’s

fact-checking and annotation platform, Check, received three Online Journalism Awards including for supporting Electionland, a collaborative verification project during the 2016 US elections, and Verificado, a collaborative verification project during the 2018 Mexican national elections. Meedan's tipline system is used by over 50 organizations, serves over 230K users annually, and automatically distributes human-written fact-checks via messaging apps such as Whatsapp. The **Rutgers University** School of Communication and Information is an interdisciplinary school which combines traditional methods from communication research with advances in computer science. The **Reducing Information Ecosystem Threats (RIET) Lab**, part of the **School of Computing** at the **University of Connecticut** focuses on threats to the information ecosystem online and to healthy public discourse from an information retrieval lens, informed by insights from the social sciences. The **Oxford Internet Institute** is a multidisciplinary research and teaching department of the **University of Oxford**, dedicated to the social science of the Internet.

4.2 Speaker Bios

Prof. Scott A. Hale is the Director of Research at Meedan, an Associate Professor at the Oxford Internet Institute of the University of Oxford, and a Fellow at the Alan Turing Institute. In these roles, he works to bridge academia and industry divides and increase equitable access to quality information online with a focus on misinformation and hate speech. Prof. Hale is currently serving as PI on Co-Insights, which as is a \$5.75M NSF Convergence Accelerator contract (now in Phase II).

Prof. Kiran Garimella is an Assistant Professor at Rutgers University in the School of Communication and Information. His Computational Social science research contributes to the understanding of large-scale human behavior on social networks. His latest work studies information threats on encrypted social networks such as WhatsApp using innovative data collection and analysis methods. He is a Co-PI of the Co-Insights project.

Prof. Shiri Dori-Hacohen is an Assistant Professor at the School of Computing at the University of Connecticut, where she leads the Reducing Information Ecosystem Threats (RIET) Lab, and is an inaugural member of a first-of-its-kind entrepreneurial cluster, bridging the startup and academic worlds. She was Founder & Executive Chairwoman at AuCoDe for 2 years, after serving as Founder/CEO for over 5 years. Until it closed its doors in 2023, AuCoDe was also a member organization of the Co-Insights project. She has 20 years of experience in industry, startups, and academia, including at Google and Facebook; and was named to the 2023 D-30 Disability Impact List. Prof. Dori-Hacohen has served as PI or Co-PI on over \$7.7M worth of federal funds from the US National Science Foundation and is a Co-PI of the Co-Insights project.

ACKNOWLEDGEMENTS

The authors thank our many collaborators in the Co-Insights project for sharing their rich expertise, which greatly informed this paper. We also thank our anonymous reviewers for providing thoughtful feedback. This material is based in part upon work supported by the National Science Foundation Convergence Accelerator (contract 49100421C0035), and by the National Science Foundation Program on Fairness in Artificial Intelligence in Collaboration with Amazon (#IIS-2147305). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] J.L.A. Donohue. 2023. Click-Gap, paternalism, and tech giants' relationships with their users. *AI and Ethics* (2023), 1–12.
- [2] Shiri Dori-Hacohen, Keen Sung, Jengyu Chou, and Julian Lustig-Gonzalez. 2021. Restoring healthy online discourse by detecting and reducing controversy, misinformation, and toxicity online. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2627–2628.
- [3] Patricia A Duff. 2010. Language socialization into academic discourse communities. *Annual review of applied linguistics* 30 (2010), 169–192.
- [4] Isabelle Freiling, Nicole M Krause, and Dietram A Scheufele. 2023. Science and ethics of “curing” misinformation. *AMA Journal of Ethics* 25, 3 (2023), 228–237.
- [5] Kiran Garimella. 2022. Community-Driven Fact-Checking on WhatsApp: Who Fact-Checks Whom, Why, and With What Effect? (2022).
- [6] Nolan Higdon and Allison Butler. 2023. (Fake) News is Racist: Mapping Culturally Relevant Approaches to Critical News Literacy Pedagogy. *Critical Education* 14, 3 (2023), 78–97.
- [7] Ashkan Kazemi, Kiran Garimella, Devin Gaffney, and Scott Hale. 2021. Claim Matching Beyond English to Scale Global Fact-Checking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 4504–4517. <https://doi.org/10.18653/v1/2021.acl-long.347>
- [8] Ashkan Kazemi, Zehua Li, Verónica Pérez-Rosas, Scott A Hale, and Rada Mihalcea. 2022. Matching tweets with applicable fact-checks across languages. In *DeFactify: Workshop on Multimodal Fact Checking and Hate Speech Detection, co-located with AAAI 2022*. CEUR Workshop Proceedings. <https://ceur-ws.org/Vol-3199/paper1.pdf>
- [9] Janet Kemei, Dominic A Alaazi, Mia Tulli, Megan Kennedy, Modupe Tunde-Byass, Paul Bailey, Ato Sekyi-Otu, Sharon Murdoch, Habiba Mohamud, Jeanne Lehman, et al. 2022. A scoping review of COVID-19 online mis/disinformation in Black communities. *Journal of Global Health* 12 (2022).
- [10] Sarah Nguyễn, Rachel Kuo, Madhavi Reddi, Lan Li, and Rachel E Moran. 2022. Studying mis- and disinformation in Asian diasporic communities: The need for critical transnational research beyond Anglocentrism. *Harvard Kennedy School Misinformation Review* (2022).
- [11] Irene V Pasquetto, Briony Swire-Thompson, Michelle A Amazeen, Fabrício Benvenuto, Nadia M Brashier, Robert M Bond, Lia C Bozarth, Ceren Budak, Ulrich KH Ecker, Lisa K Fazio, et al. 2020. Tackling misinformation: What researchers could do with social media data. *The Harvard Kennedy School Misinformation Review* (2020).
- [12] Michael Shliselberg, Ashkan Kazemi, Scott A Hale, and Shiri Dori-Hacohen. 2024. SynDy: Synthetic Dynamic Dataset Generation Framework for Misinformation Tasks. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24)*. Association for Computing Machinery, New York, NY, USA.