Tracking and Identifying International Propaganda and Influence Networks Online

Hans W. A. Hanley

Stanford University
hhanley@cs.stanford.edu

Abstract

Misinformation and propaganda undermine trust in institutions, spread falsehoods, and sometimes incite violence. However, recent advancements in transformer-based AI models can help combat the proliferation of disinformation globally and in real time. In this work, I propose and develop a system using these models to scalably identify, track, and analyze the spread of narratives from over 40,000 international news websites. First, by employing novel multilingual Matryoshka embeddings and hierarchical level-wise clustering, my proposed system identifies news stories, topics, and themes across these thousands of news websites. Second, by utilizing multilingual stance detection, my system assesses the biases and factual inconsistencies in news articles, enabling the identification of websites that spread propaganda or misinformation. Finally, through network inference methods, my system uncovers connections among websites disseminating slanted or false content. My approach illustrates how AI can be utilized to mitigate the global spread of harmful misinformation and propaganda.

Introduction

Throughout the past decade, unscrupulous individuals and websites alike have spread disinformation, propaganda, and garbage science in an unprecedented fashion on social media, eroding faith in institutions and causing a supposed "Golden Age of Conspiracy Theories" (Hanley, Kumar, and Durumeric 2023). Misinformation and propaganda have promoted dangerous supposed health cures, obscured facts during wars (Hanley, Okabe, and Durumeric 2025), and incited violence (Hanley, Kumar, and Durumeric 2023). However, despite this, most investigations into misinformation remain limited in scope and reactive, primarily conducted through time-consuming qualitative approaches (Hanley, Okabe, and Durumeric 2025). To combat the threat posed by misinformation and propaganda, I argue that researchers can leverage AI systems to track the spread of false narratives globally and in real time.

To this end, for my thesis, I am developing a large language model-based automated system to scalably identify, track, and understand the spread of news stories, propaganda, and misinformation (Hanley, Kumar, and Durumeric 2024; Hanley, Okabe, and Durumeric 2025). To do this, I have built a large language model-based and statistical approach to analyze, understand the biases of, and monitor the stories published by 40,000 international news websites.

Tracking Propaganda Narratives

To track the spread of international news narratives, I created and continue to maintain an evolving catalog of popular news-related websites to scrape. To do so, every three months, I use the CommonCrawl website index and the Cloud Domain Intelligence API to compile and refine a list of websites dedicated to "news." Upon generating this list of news websites, I crawl them daily, collecting each site's homepage, RSS feed, and the corresponding linked news articles, extracting each page's article contents.

To identify the different narratives and stories being spread amongst these news websites, I propose utilizing customized Matryoshka embeddings and hierarchical clustering. Besides decoder-based large language models like OpenAI's GPT-4 and Anthropic's Claude, which remain prohibitively expensive for processing millions of documents for topic analysis, current encoder-based models used for topic analysis often simply identify the "similarity" overlap of two news articles via the cosine-similarity of their embeddings. To address this challenge, and better define the similarity between two news articles for later clustering, I recently proposed a novel multilingual adaptation of Matryoshka embeddings for news story differentiation as well as a novel clustering algorithm to take advantage of the hierarchical nature of Matryoshka embeddings. These Matryoshka embeddings' upper dimensions progressively learn higher levels of detail, allowing for the identification and differentiation of multilingual news articles at varying levels of story and event similarity. In my approach, the upper dimensions of our embeddings are used to determine if two news articles are about the same event, the middle dimensions to determine if they address the same topic, and the lower dimensions if they address the same theme. I utilized the SemEval-22-t8 dataset (Chen et al. 2022), which grades news articles on their similarity at various levels, to train and evaluate my approach, achieving state-of-the-art results (Pearson correlation of $\rho = 0.833$ with human labels of article similarity on SemEval-22-t8 test set). Compared to traditional embedding approaches, my approach makes sim-

ilarity calculations at various granularity levels more interpretable and reduces the cost of concurrent similarity calculations at these different levels of granularity. For clustering these embeddings, I developed a novel bottom-up agglomerative approach based on an online version of the Reciprocal Nearest Neighbor algorithm (Monath, Zaheer, and McCallum 2023) that leverages the naturally hierarchical structure of Matryoshka embeddings. To ensure the quality of the resulting news story clusters, I performed a large-scale manual evaluation to ensure that the resulting clusters were coherent (Hanley, Kumar, and Durumeric 2024). I further benchmarked our approach against Miranda et al. (2018)'s news article clustering dataset, achieving state-of-the-art results.

Upon identifying individual news stories, topics, and themes, to make these clusters human-interpretable, I propose utilizing an open-source decoder LLM fine-tuned for multilingual summarization. Based on each cluster's texts, after summarizing each cluster, I further propose extracting descriptive keywords using pointwise mutual information. Using this approach, for instance, I found that during 2022-2023, the most prominent stories on English-language outlets concerned inflation and the Russo-Ukrainian War (Hanley, Okabe, and Durumeric 2025).

Understanding Propaganda, Biased Information, and Misinformation Influence Networks. Upon identifying the different narratives being spread amongst different websites, I propose applying a mixture of zero-shot multilingual stance detection and network-based inference (Leskovec, Backstrom, and Kleinberg 2009) to identify the influence networks promoting particular narratives. This is such that my approach first identifies the stance/attitude of each clustered news article to its respective extracted topics. In this way, this approach not only determines the overall valence of each website to its topic (*e.g.*, the overall valence of a given website to vaccines) but also the changing attitudes of websites to a given topic over time. This enables my approach to then identify and track the spread of particularly slanted news stories towards particular entities (*e.g.*, Ukraine, political figures, vaccines, artificial intelligence). For instance, using this technique, I observed that attitudes in online newspapers toward artificial intelligence began to polarize in December 2022. The news website hollywoodreporter.com was the most negative largely due to AI's potential impact on writers and actors. Similarly, after identifying particular topics and subsequently determining different factual claims of interest (*e.g.*, The 2020 US Presidential Election was stolen) using the stance detection aspect of my work, based on these claims, my system can further identify articles that seek to undermine these facts and spread misinformation. By then framing these clusters as time cascades and subsequently applying the open-source NETINF (Leskovec, Backstrom, and Kleinberg 2009) network inference algorithm, my system finally will uncover underlying relationships between my set of websites based on the patterns by which they post news articles. By applying this approach to the set of articles with a given stance on a given topic, my system subsequently uncovers networks of websites that spread particular slanted content and narratives. For instance, by determining the eigenvector centralities in the resulting network, I found in recent work (Hanley, Okabe, and Durumeric 2025) that the most influential anti-vaccine websites included pandemic.news and the childrenshealthdefense.org.

Future Work. For future work, I plan to apply my proposed methodology to understand how particular valenced information spreads between and within individual countries (*e.g.*, how articles and attitudes published in Russia correlate with coverage in the United States and elsewhere). While I currently have developed an English stance detection model, which I have utilized in prior work (Hanley, Okabe, and Durumeric 2025), I am working towards the development of a multilingual model that works with high accuracy in the 54 languages upon which my Matryoshka embeddings were trained. Similarly, given that keyword extraction with pointwise mutual information only works with monolingual texts, I am currently exploring different avenues for the extraction of keywords from multilingual texts for subsequent use in a multilingual stance detection model (*e.g.*, using a decoder LLM to predict keywords based on the Matryoshka embedding of a given text). Finally, I will publicly deploy my system so that journalists and researchers can reliably reliably understand the spread of misinformation.

References

Chen, X.; Zeynali, A.; Camargo, C.; Flöck, F.; Gaffney, D.; Grabowicz, P.; Hale, S.; Jurgens, D.; and Samory, M. 2022. SemEval-2022 Task 8: Multilingual news article similarity. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, 1094–1106.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2023. A Golden Age: Conspiracy Theories' Relationship with Misinformation Outlets, News Media, and the Wider Internet. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2): 1–33.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2024. Specious Sites: Tracking the Spread and Sway of Spurious News Stories at Scale. In *2024 IEEE Symposium on Security and Privacy (SP)*, 180–180. IEEE Computer Society.

Hanley, H. W.; Okabe, E.; and Durumeric, Z. 2025. Tracking the Takes and Trajectories of News Stories from Trustworthy and Worrisome Websites. In *34th USENIX Security Symposium*.

Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Memetracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 497–506.

Miranda, S.; Znotiņš, A.; Cohen, S. B.; and Barzdins, G. 2018. Multilingual Clustering of Streaming News. In Riloff, E.; Chiang, D.; Hockenmaier, J.; and Tsujii, J., eds., *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4535–4544. Brussels, Belgium: Association for Computational Linguistics.

Monath, N.; Zaheer, M.; and McCallum, A. 2023. Online Level-wise Hierarchical Clustering. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1733–1745.