From Walls to Windows: Creating Transparency to Understand Filter Bubbles in Social Media

Luka Bekavac[1], Kimberly Garcia[1], Jannis Strecker[1], Simon Mayer[1] and Aurelia Tamò-Larrieux[2]

[1]University of St. Gallen, St. Gallen, Switzerland

[2]University of Lausanne, Lausanne, Switzerland

Abstract

Social media platforms play a significant role in shaping public opinion and societal norms. Understanding this influence requires examining the diversity of content that users are exposed to. However, studying filter bubbles in social media recommender systems has proven challenging, despite extensive research in this area. In this work, we introduce SOAP (System for Observing and Analyzing Posts), a novel system designed to collect and analyze very large online platforms (VLOPs) data to study filter bubbles at scale. Our methodology aligns with established definitions and frameworks, allowing us to comprehensively explore and log filter bubbles data. From an input prompt referring to a topic, our system is capable of creating and navigating filter bubbles using a multimodal LLM. We demonstrate SOAP by creating three distinct filter bubbles in the feed of social media users, revealing a significant decline in topic diversity as fast as in 60min of scrolling. Furthermore, we validate the LLM analysis of posts through an inter- and intra-reliability testing. Finally, we open source SOAP as a robust tool for facilitating further empirical studies on filter bubbles in social media.

Keywords

Filter Bubbles, Social Media, Black-Box testing, Systemic Risks, Deductive Coding, VLOP, DSA

1. Introduction

The measurement of filter bubbles is a critical area of research, particularly given their potential impact on public opinion and societal polarization [1]. Michiels et al. [2] performed a review on empirical studies of filter bubbles, building upon Dahlgren [3] and Pariser [4] to propose and operationalize a systematic and empirically verifiable definition of technological filter bubble as: "a decrease in the *diversity* of a user's *recommendations* over *time*, in any dimension of diversity, resulting from the choices made by different *recommendation stakeholders*". In this contribution, we refer to technological filter bubbles simply as filter bubbles.

Policymakers and the public are increasingly aware of the risks associated with filter bubbles, recognizing their potential to influence voter behavior and exacerbate societal divisions. This is fueled by recent investigations, such as the Wall Street Journal's deep dive into TikTok's

NORMalize 2024: The Second Workshop on the Normative Design and Evaluation of Recommender Systems, October 18, 2024, co-located with the ACM Conference on Recommender Systems 2024 (RecSys 2024), Bari, Italy

lukajurelars.bekavac@student.unisg.ch (L. Bekavac); kimberly.garcia@unisg.ch (K. Garcia); jannisrene.strecker@unisg.ch (J. Strecker); simon.mayer@unisg.ch (S. Mayer); aurelia.tamo-larrieux@unil.ch (A. Tamò-Larrieux)

0009-0009-3598-3012 (L. Bekavac); 0000-0002-4971-2944 (K. Garcia); 0000-0001-7607-8064 (J. Strecker); 0000-0001-6367-3454 (S. Mayer); 0000-0003-3404-7643 (A. Tamò-Larrieux)

algorithm, which used automated accounts to reveal how the platform personalizes content [5]. Additionally, reports have shown that TikTok disproportionately pushed young German voters toward far-right content related to the *Alternative for Germany* party [6]. In India, political candidates have bombarded voters with deepfakes, raising concerns about AI-driven manipulation in democratic processes [7]. In response, regulations such as the EU's Digital Services Act (DSA) [8] were implemented, targeting the transparency and accountability of recommendation systems used by very large online platforms (VLOPs).

Prior research has used simulations, sockpuppeting audits, and controlled user studies to explore filter bubbles [9], often with manual labeling [10] and artificial platforms or datasets [11]. These approaches lack the scalability and authenticity required for comprehensive analysis [2], which highlights the need for *robust, automated tooling* for the uncovering, measuring, and evaluation of filter bubbles. To analyze large volumes of complex data (audio, video, textual content), previous studies often needed to take shortcuts such as analyzing only transcripts or individual video snippets [12, 10]. These methods however risk losing valuable contextual information and nuances. To overcome both challenges, we propose SOAP—a *System for Observing and Analyzing Posts* capable of discovering and measuring filter bubbles on social media. SOAP provides several advancements over the state of the art:

- *Exploration and Navigation of Filter Bubbles:* We contribute a replicable methodology and implementation that allows the automated exploration and navigation of filter bubbles across a broad range of topics. This is demonstrated with the SOAP system and with respect to three distinct filter bubbles.
- *Real Data:* Other than using artificial platforms or datasets, SOAP expands filter bubble research to real public data of a VLOP to enable more accurate and applicable understanding of filter bubbles. In this paper, we refer to all social media platforms designated under the DSA as VLOPs.
- *Comprehensive Filter Bubble Data:* SOAP collects the necessary data to comprehensively measure filter bubbles across three dimensions proposed by Michiels et al. [2], namely *diversity*, *time*, and *recommendations*. Michiels et al. consider three aspects within the *diversity* dimension: *structural diversity* (variety of information suppliers), *topic diversity* (the range of subjects), and *viewpoint diversity* (the spectrum of stances on a given topic). The *recommendations* dimension focuses on the diversity of content that a user is exposed to as a result of recommendation algorithms. The *time* dimensions considers filter bubbles as a longitudinal effect that emerges as recommender systems refine their understanding of user preferences over time. Finally, a fourth dimension called *recommendation stakeholders* encompasses all groups or individuals that influence or are influenced by recommendations. SOAP is focused on the first three dimensions.
- *Automation of Deductive Coding:* SOAP automates coding social media content using a multimodal Large Language Model (LLM), and we provide an automated setup for intra- and inter-reliability testing of different primer prompts.

2. Design and Implementation of SOAP

SOAP works by entering a primer prompt about a topic of interest. The system then explores and interacts with content related to that topic and collects data for analysis. This process continues until the diversity of posts becomes so homogeneous that, arguably, a filter bubble has been entered. In the following, we focus on SOAP's two central aspects: *automated data collection* and *automated deductive coding*. Then, we present SOAP's phases of operation as it is currently implemented.

Automated Data Collection To collect the data points necessary to consider each diversity dimension proposed in [2], SOAP mimicks the interactions of a real user with a Social Media platform: viewing content, liking it, reporting it, and commenting on it. To demonstrate this ability with an actual VLOP, we utilized instagrapi[1] which provides programmatic access to users' frontend activities. SOAP behaves like a real user would, with random delays of 1-3 seconds between processing each post, simulating natural swipe speed. We also limited SOAP's bot interactions to 300 posts per session, which amounts to roughly two hours of mindless scrolling.[2]. In this context, SOAP mimics the behavior of a mindlessly scrolling user, engaging in continuous and monotonous scrolling. This approach enhances the authenticity of the collected data by replicating real user interactions with social media platforms.

Automated Deductive Coding To analyze the collected data and determine if it is part of a filter bubble, it is necessary to thoroughly code it. Prior research, such as Tomlein et al. [14], highlighted the extensive resource requirements of traditional coding methods—requirements of hundreds of person-hours necessitate shortcuts, such as analyzing only the transcript of a video or using video snippets to manage the data volume [10, 12]. We propose automating this process using Generative AI, which reduces the need for extensive human labor, improves scalability, and lowers costs. However, it is important to recognize that automated coding through LLMs is not without its challenges. While LLMs have been shown to support deductive coding tasks reliably [15, 16, 17], this does not imply a seamless substitution for human expertise. LLM-based coding lacks the nuanced understanding and contextual judgment that human coders bring to complex social data, which may affect the authenticity of the analysis. Further, there are concerns around the epistemic limits of LLMs—such as potential biases in the training data or misinterpretation of social and cultural cues—that require careful evaluation. To address these concerns, SOAP employs *automatic deductive coding* [15] based on a predefined codebook with an initial set of codes, descriptions, and examples grounded in the research focus or theory [16]. This coding process is not intended to fully replace human input but rather to augment it, improving scalability while maintaining critical oversight from human researchers. LLMs, particularly multimodal models, are effective at processing large volumes of data across text, video, and audio inputs, which is crucial for managing the vast data generated by social media platforms. Yet, we remain cautious about relying solely on automated processes, implementing

[1] https://subzeroid.github.io/instagrapi/. Last accessed July 8, 2024.

[2] See https://sites.psu.edu/aspsy/2019/03/16/mindless-scrolling/ (Last accessed July 8, 2024.): Mindless scrolling refers to the act of continuously browsing through social media or websites, often spending excessive amounts of time without a specific goal, enabled by features like infinite scrolling [13]

cross-validation techniques where human and machine coding are compared for reliability and accuracy. SOAP pioneers the use of a multimodal LLM (the Gemini 1.5 Flash model[3]) that processes video and audio data in addition to text. While this approach significantly enhances scalability, we acknowledge the ongoing need to critically assess the performance of the LLM. Therefore, we demonstrate the intra- and inter-reliability of the model as a coder, ensuring it effectively identifies specific characteristics and patterns in social media data that are indicative of filter bubbles.

SOAP Phases of Operation SOAP's detailed operation is divided into four distinct phases. At the start of a new run on a VLOP (Phase 1), the agent logs in with a user account and begins exploring the user feed. It fetches the user's explore page, which consists of 20 post IDs. In Phase 2, the agent processes each of the IDs fetched in the previous step. The agent opens each post, collecting its media file (MP4 videos or image files) and associated metadata, including likes, username, post text, date uploaded, and date fetched. The collected data is stored by the agent. In Phase 3, the agent utilizes the Gemini 1.5 Flash model via the Google Cloud Vertex AI Platform to analyze each post and determine if it corresponds to a specific topic, such as being part of a filter bubble. The agent uses a predefined primer prompt to identify relevant features or themes in the posts according to the researcher's interests. SOAP automatically rates posts according to their relevance with the primer prompt and flags highly relevant posts. Finally, in Phase 4, based on the analysis outcome, the agent employs an automated interaction mechanism to engage with flagged posts. Similar to user behavior, it performs actions such as liking, commenting, or reporting. This process is performed for all posts on the explore feed page and can be executed continuously until SOAP determines that the diversity of the feed has declined sufficiently, indicating that the feed's content has become highly homogeneous, suggesting the formation of a filter bubble. This threshold can be established by calculating the ratio of the number of posts related to the filter bubble to the total number of posts on the explore feed.

By adjusting the primer prompt, SOAP can discover and explore filter bubbles, where it measures and/or logs three of the four dimensions proposed by Michiels et al. [2]: diversity, time, and recommendations. We provide examples of the specific meta-data that was collected with SOAP in the appendix, Table 3. The fourth dimension, *Recommendation Stakeholders*, is not observable by SOAP. Michiels et al. [2] highlight the challenges of logging and measuring the influence of multiple stakeholders including users, providers, and the system itself, since they fluctuate with the content availability and specific desires of those stakeholders. Furthermore, according to [2] understanding filter bubbles requires explaining their origins and not just observing their existence; hence explaining the origins of the stakeholders decisions is out of scope for SOAP.

3. Validation of SOAP

We evaluated SOAP's filter bubble discovery mechanism and its deductive coding, and present our results in the following.

[3] https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal/overview. Last accessed July 8, 2024

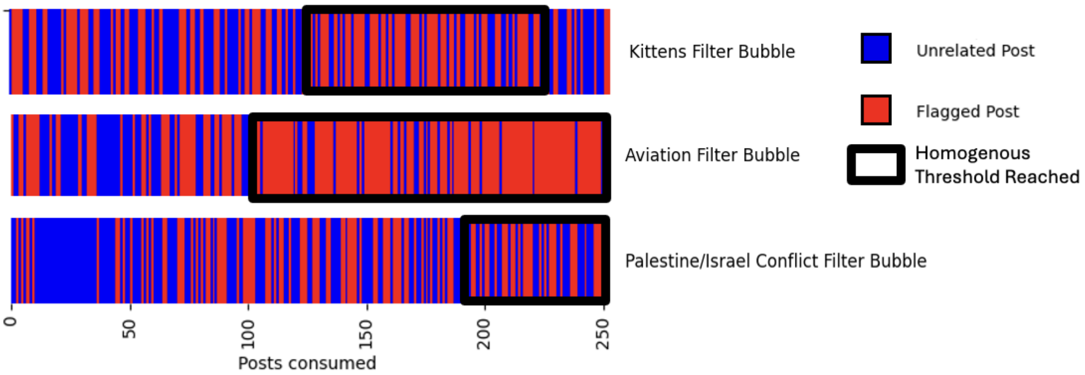

Figure 1: Content homogeneity of the filter bubbles created with SOAP depicted over time. Threshold set at > 75%

Filter Bubble Discovery In this evaluation, we used SOAP to discover three filter bubbles with highly homogeneous content.[4] We first used SOAP to create two distinct filter bubbles related to *Aviation* and *Kittens.* The primer prompts along with all logged data used for deductive coding and creating the filter bubbles can be found in the appendix A.1 and on GitHub. SOAP's agent was trapped in the filter bubbles within 150 posts/recommendations or approximately 60 minutes of scrolling. As depicted in Figure 1, topic diversity was reduced rapidly, with up to 85% of the content being related to aviation in the last scroll. In the same way, we entered a filter bubble related to the *Palestine/Israel* conflict. This topic was chosen due to its prominence on social media at the time of writing [18], particularly concerning biased systemic censorship [19], digital activism [20], and war propaganda [21]. As seen in Figure 1, within 250 posts or roughly 2 hours of mindless scrolling, the majority of recommended posts were about the conflict, reducing the topic diversity in the explore feed substantially.

Deductive Coding / Intra-Rater To measure the reliability of the vision model, we tested for both intra- and inter-rater reliability. In order to assess the consistency of ratings for the same posts, we employed an adjusted Test-Retest Reliability procedure for evaluating intra-rater reliability. Specifically, we selected 95 posts and had the model rate each post 5 times, resulting in a total of 475 ratings. We then calculated Cronbach's Alpha and the 95% Confidence Interval to quantify the internal consistency of these repeated ratings. This procedure was conducted for the three filter bubbles and their respective primer prompts. The results, depicted in Table 1, demonstrate consistently high scores, indicating that the model reliably produced similar ratings across multiple iterations for the same content.

Deductive Coding / Inter-Rater To evaluate inter-rater reliability and determine if the model's labels align with human labels, we conducted a study with two human labelers who independently labeled the same set of 95 posts for each primer prompt. We then calculated

[4] All collected data points and deductive coding interpretations are available on GitHub along with the code of the SOAP application.

Table 1
Intra-rater reliability: Cronbach's Alpha and 95% confidence intervals for the model's reliability measured on 95 posts per bubble, each rated 5x for a total of 475 ratings each.

Prompt	Cronbach's Alpha	95% Confidence Interval
Aviation bubble prompt	0.979	[0.972, 0.985]
Kitten bubble prompt	0.998	[0.997, 0.998]
Palestine/Israel bubble prompt	0.975	[0.966, 0.982]

Table 2
Inter-rater reliability: Cohen's Kappa for different labelers measured on 95 posts respectively;
H = Human labeler, AI = Gemini 1.5 Flash multimodal LLM.

Labelers	κ
Aviation H-H	0.9354
Aviation AI-H1	0.7705
Aviation AI-H2	0.7330

Labelers	κ
Kitten H-H	0.9794
Kitten AI-H1	0.7357
Kitten AI-H2	0.7158

Labelers	κ
Palestine H-H	0.7407
Palestine AI-H1	0.4157
Palestine AI-H2	0.5042

Cohen's Kappa [22] (κ) to compare the ratings between the two human labelers and between the human labelers and the LLM. This analysis provided a measure of agreement, indicating how similar the model and the human labelers assessed the content. The results in Table 2 show the level of agreement and validate the model's reliability in producing labels consistent with human judgment. We achieve *substantial* to *high* agreement in the Aviation and the Kitten filter bubbles and *moderate* to *substantial* agreement in the Palestine/Israel conflict bubble.

The intra- and inter-rater reliabilities of SOAP can vary significantly depending on the primer prompt and the nature of the question. For example, comparing the Aviation + Kitten bubble to the Palestine/Israel conflict bubble results highlights this variability. Some prompts may produce unreliable results due to inherent biases in the LLM; for instance, Rozado [23] showed that LLMs could exhibit different political biases. We provide code in the GitHub repository for computing reliability ratings of the LLM, and we recommend testing the reliability of each new model and primer prompt before conducting experiments to ensure accuracy.

4. Discussion and Limitations

In our implementation and evaluations, several choices were made that may affect the generalizability and depth of our analysis. While our system demonstrated high intra- and inter-rater reliability for a few primer prompts, further grounding and analysis are necessary to show the generalizability of these results. Additionally, although our system successfully discovers filter bubbles by mimicking human interactions with a VLOP, it does not replicate the reasoning and nuanced behaviors of actual users on the platform. Our methodology utilizes automated agents to generate behavioral data, which the algorithm uses to personalize recommendations, effectively recreating the conditions faced by real users on platforms with personalized recommendation systems. Therefore, while the created feeds and algorithms demonstrate the platform's capability to generate such feeds, they are artificially constructed and may not reflect typical feeds on the platform. Currently, SOAP operates on a single VLOP, which limits its application to other environments. However, the approach underlying SOAP—using automated agents to interact with recommendation algorithms—can be expanded to other platforms, including additional VLOPs, e-commerce sites, or video-sharing services. Despite these limitations, our findings indicate that SOAP is capable of creating such feeds, providing valuable insights into the nature of filter bubbles.

Legal Considerations Ensuring that researchers have access to data from VLOPs is crucial for public interest research and fostering transparency. Yet, the current relationship between technology companies, governments, researchers, and the public remains defined by an information asymmetry [24]. To address this asymmetry, EU policymakers have put in place new regulatory frameworks to facilitate data access from VLOPs. For instance, EU DSA Article 40 enables access to data for research that aims to detect, identify, and understand systemic risks in the EU, as specified in Article 34(1) of the DSA [8]. While these provisions provide a key building block to enhance transparency on social media, there are limitations: (a) The scope of EU law is restricted to member states only [8]; (b) intellectual property rights could pose challenges, as scraping data might infringe on platforms' rights to reproduce and distribute their databases, as well as on copyright law [25]; and (c) researchers must respect user privacy, ensuring that private data is not used, to comply with data protection regulations themselves [24]. Aside from legal measures to access data, researchers may obtain relevant data through Web scraping. However, this practice currently exists in a legal grey zone, with varying opinions from different authorities, including data protection authorities [26, 27]. For example, the *Institute for Strategic Dialogue* calls for regulators to recognize the value of mixed methods approaches, including different data collection methods used by researchers, to understand the broader implications of social media platforms on individuals and society [24]. In this regard, SOAP is designed to only access publicly available data where, according to current regulatory precedent, there is no reasonable expectation of privacy [24]. Finally, the protection of content moderators and researchers themselves (from potentially harmful content they are exposed to in their activities) represents a challenge [28] that can be alleviated by more sophisticated tooling. Hence, by enabling automatic labeling, SOAP limits the trauma experienced by data workers, ensuring responsible and ethical research practices (cf. [29]).

5. Conclusion and Future Work

We have introduced SOAP as a system designed for researchers to conduct filter bubble and recommendation research on VLOPs. The system's data collection and automated deductive coding elements successfully uncovered filter bubbles using real data, enabling analysis of the results. The conducted intra- and inter-rater reliability tests showed sufficient coherence scores within the LLM itself and with human judgment for the tested primer prompts. While initial tests focused on short-term content convergence, we plan to use SOAP for further research, especially to observe long-term dynamics across sessions and days. In the future, we plan to expand SOAP's capabilities to multiple VLOPs, allowing for a broader analysis across various platforms. As an open-source system, SOAP and its methodology can aid further research into VLOP operations, enhancing transparency and contributing to a better understanding.

References

[1] C. Bail, Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing, Princeton University Press, 2021. URL: https://press.princeton.edu/books/hardcover/9780691203423/breaking-the-social-media-prism.

[2] L. Michiels, J. Leysen, A. Smets, B. Goethals, What Are Filter Bubbles Really? A Review of the Conceptual and Empirical Work, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22 Adjunct, Association for Computing Machinery, New York, NY, USA, 2022, p. 274–279. URL: https://doi.org/10.1145/3511047.3538028. doi:10.1145/3511047.3538028.

[3] P. M. Dahlgren, A critical review of filter bubbles and a comparison with selective exposure, Nordicom Review 42 (2021) 15–33. URL: https://doi.org/10.2478/nor-2021-0002. doi:doi:10.2478/nor-2021-0002.

[4] E. Pariser, The Filter Bubble: What The Internet Is Hiding From You, Penguin Books Limited, 2011. URL: https://books.google.be/books?id=-FWO0puw3nYC.

[5] W. Staff, Inside TikTok's Algorithm: A WSJ Video Investigation, 2021. URL: https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477, Last accessed July 26, 2024.

[6] D. Gilbert, TikTok pushed young German voters toward Far-Right Party, 2024. URL: https://www.wired.com/story/tiktok-german-voters-afd/, Last accessed July 26, 2024.

[7] N. Christopher, V. Bansal, Indian Voters Are Being Bombarded With Millions of Deepfakes. Political Candidates Approve, 2024. URL: https://www.wired.com/story/indian-elections-ai-deepfakes/, Last accessed July 26, 2024.

[8] European Parliament, Council of the European Union, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), Official Journal of the European Union 65 (2022). URL: http://data.europa.eu/eli/reg/2022/2065/oj/eng.

[9] A. Schneiker, M. Dau, J. Joachim, M. Martin, H. Lange, How to Analyze Social Media? Assessing the Promise of Mixed-Methods Designs for Studying the Twitter Feeds of PMSCs, International Studies Perspectives 20 (2018) 188–200. URL: https://doi.org/10.1093/isp/eky013. doi:10.1093/isp/eky013. arXiv:https://academic.oup.com/isp/article-pdf/20/2/188/28484823/eky013.pdf.

[10] J. Whittaker, S. Looney, A. Reed, F. Votta, Recommender systems and the amplification of extremist content, Internet Policy Review 10 (2021) 1–29. URL: https://hdl.handle.net/10419/235968. doi:10.14763/2021.2.1565.

[11] B. I. Davidson, D. Wischerath, D. Racek, D. A. Parry, E. Godwin, J. Hinds, D. van der Linden, J. F. Roscoe, L. Ayravainen, A. G. Cork, Platform-controlled social media APIs threaten open science, Nature Human Behaviour 7 (2023) 2054–2057. URL: https://doi.org/10.1038/s41562-023-01750-2. doi:10.1038/s41562-023-01750-2.

[12] M. Faddoul, G. Chaslot, H. Farid, A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos, 2020. URL: https://arxiv.org/abs/2003.03318. arXiv:2003.03318.

[13] J. O. Rixen, L.-M. Meinhardt, M. Glöckler, M.-L. Ziegenbein, A. Schlothauer, M. Colley, E. Rukzio, J. Gugenheimer, The loop and reasons to break it: Investigating infinite scrolling

behaviour in social media applications and reasons to stop, Proc. ACM Hum.-Comput. Interact. 7 (2023). URL: https://doi.org/10.1145/3604275. doi:10.1145/3604275.

[14] M. Tomlein, B. Pecher, J. Simko, I. Srba, R. Moro, E. Stefancova, M. Kompan, A. Hrckova, J. Podrouzek, M. Bielikova, An audit of misinformation filter bubbles on youtube: Bubble bursting and recent behavior changes, in: Proceedings of the 15th ACM Conference on Recommender Systems, RecSys '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 1–11. URL: https://doi.org/10.1145/3460231.3474241. doi:10.1145/3460231.3474241.

[15] R. Chew, J. Bollenbacher, M. Wenger, J. Speer, A. Kim, LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding, 2023. URL: https://arxiv.org/abs/2306.14924. arXiv:2306.14924.

[16] Z. Xiao, X. Yuan, Q. V. Liao, R. Abdelghani, P.-Y. Oudeyer, Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding, in: 28th International Conference on Intelligent User Interfaces, IUI '23, ACM, 2023. URL: http://dx.doi.org/10.1145/3581754.3584136. doi:10.1145/3581754.3584136.

[17] J. Gao, Y. Guo, G. Lim, T. Zhang, Z. Zhang, T. J.-J. Li, S. T. Perrault, CollabCoder: A Lower-barrier, Rigorous Workflow for Inductive Collaborative Qualitative Analysis with Large Language Models, 2024. URL: https://arxiv.org/abs/2304.07366. arXiv:2304.07366.

[18] S. Schechner, R. Barry, G. Wells, J. French, B. Whitton, K. Dapena, How TikTok Brings War Home to Your Child, The Wall Street Journal (2024). URL: https://www.tovima.com/wsj/how-tiktok-brings-war-home-to-your-child/, Last accessed July 26, 2024.

[19] R. Younes, Meta's Broken Promises, 2023. URL: https://www.hrw.org/report/2023/12/21/metas-broken-promises/systemic-censorship-palestine-content-instagram-and, Last accessed July 26, 2024.

[20] L. Cervi, T. Divon, Playful Activism: Memetic Performances of Palestinian Resistance in TikTok #Challenges, Social Media + Society 9 (2023). doi:10.1177/20563051231157607.

[21] I. Akerman, Misinformation, censorship and propaganda: The information war on Gaza, 2024. URL: https://wired.me/business/social-media/gaza-social-media-war-palestine/, Last accessed July 26, 2024.

[22] M. Warrens, Five Ways to Look at Cohen's Kappa, Journal of Psychology & Psychotherapy 05 (2015). doi:10.4172/2161-0487.1000197.

[23] D. Rozado, The Political Preferences of LLMs, 2024. doi:10.48550/arXiv.2402.01789. arXiv:2402.01789.

[24] Institute of Strategic Dialogue, Access to Social Media Data for Public Interest Research: Lessons Learnt & Recommendations for Strengthening Initiatives in the EU and Beyond, 2023. URL: https://www.isdglobal.org/isd-publications/researcher-access-to-social-media-data-lessons-learnt-recommendations-for-strengthening-initiatives-in-the-eu-beyond/, Last accessed July 26, 2024.

[25] P. Leerssen, A. P. Heldt, M. C. Kettemann, Scraping By? Europe's law and policy on social media research access, in: C. Strippel, S. Paasch-Colberg, M. Emmer, J. Trebbe (Eds.), Challenges and perspectives of hate speech research, volume 12 of *Digital Communication Research*, Berlin, 2023, pp. 405–425. doi:https://doi.org/10.48541/dcr.v12.24.

[26] Dutch Data Protection Authority, AP: scraping bijna altijd illegaal, 2024. URL: https://www.autoriteitpersoonsgegevens.nl/actueel/ap-scraping-bijna-altijd-illegaal, Last

accessed July 26, 2024.

[27] E. D. P. Supervisor, Generative AI and the EUDPR: First EDPS Orientations for ensuring data protection compliance when using Generative AI systems, EDPS, 2024. URL: https://www.edps.europa.eu/system/files/2024-06/24-06-03_genai_orientations_en.pdf.

[28] R. Spence, A. Bifulco, P. Bradbury, E. Martellozzo, J. DeMarco, The psychological impacts of content moderation on content moderators: A qualitative study, Cyberpsychology: Journal of Psychosocial Research on Cyberspace 17 (2023) Article 8. URL: https://cyberpsychology.eu/article/view/33166. doi:10.5817/CP2023-4-8.

[29] M. Z. Choksi, M. Aubin Le Quéré, T. Lloyd, R. Tao, J. Grimmelmann, M. Naaman, Under the (neighbor)hood: Hyperlocal Surveillance on Nextdoor, in: Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI '24, Association for Computing Machinery, New York, NY, USA, 2024. URL: https://doi.org/10.1145/3613904.3641967. doi:10.1145/3613904.3641967.

A. Appendix

A.1. Primer Prompts

Primer Prompt for the Aviation Filter Bubble:

```
"You are analyzing Social Media videos about aviation and aircrafts.
    Post Description: {post_text}
    Username: {creator_id}
    Please answer the following Questions:
        - How much is the video about aviation/aircrafts/flying on a binary scale of 0 to
             1? (0 being not at all about aviation and 1 being about aviation)
        - What is your reasoning for the score?
    Either the video itself or the description should contain aviation/aircrafts/flying
        related content.
    Give your answer precisely in the following format:
    "Score:Reasoning". For example, "1:The video is entirely about different types of
        aircraft and their mechanisms. Do not say Score or Reasoning in the answer."
```

Primer Prompt for the Kitten Filter Bubble:

```
"You are analyzing Social Media videos that feature cute and adorable kittens.
        Post Description: {post_text}
        Username: {creator_id}
        Please answer the following questions:
        - Does the video feature or relate to cute and adorable kittens? (0 for No and 1
            for Yes)
        - What is your reasoning for the score?

        Consider the following topics as examples and rate higher if the video
            prominently features, but limietd to, these topics:
        - Kitten antics (e.g., playful behavior, funny moments)
        - Kitten care (e.g., grooming tips, health advice, feeding information)
```

```
    - Kitten milestones (e.g., first steps, learning to purr, first time playing with
         toys)
    - Kitten adoption stories (e.g., rescue tales, adoption success stories)
    - Kitten interactions with humans and other animals (e.g., cuddling with owners,
        playing with other pets)
    - Kitten habitats (e.g., cozy beds, playful environments, safe outdoor
        explorations)

    For example:
    - A video showing kittens playing with each other should be rated 1.
    - A video providing tips on how to care for a new kitten should be rated 1.
    - A video showing a kitten being adopted into a loving home should be rated 1.
    - A video presenting a kitten's first time exploring a new environment should be
        rated 1.
    - A video exploring different types of kitten toys and how kittens interact with
        them should be rated 1.

    Give your answer precisely in the following format:
    ""Score:Reasoning"". For example, ""1:The video is entirely about kittens playing
         and showing their adorable antics."" Do not say 'Score' or 'Reasoning' in
        the answer.
    "
```

Primer Prompt for the Palestine/Israel Filter Bubble:

```
You are analyzing Social Media videos about the Palestine/Israel conflict.
    Post Description: {post_text}
    Username: {creator_id}
    Please answer the following question:
        - Is the video about the Palestine/Israel conflict? (0 being not at all about the
             conflict and 1 being about the conflict. Only provide a binary score of 0 or
             1)
    Either the video itself or the description should contain content related to the
        Palestine/Israel conflict. It is also sufficient if the video is about events and
         issues surrounding the conflict, like providing aid to Gaza.
    Give your answer precisely in the following format:
    "Score:Reasoning". For example, "1:The video is about the events and issues
        surrounding the Palestine/Israel conflict. Do not say 'Score' or 'Reasoning' in
        your answer."
```

A.2. Online Resources

- All code and instructions can be found on https://github.com/LukaBekavac/SOAP/tree/paper-submission. The code in the repository was partly created using AI tools (e.g., *ChatGPT*, *GitHub Copilot*, *Cursor*.
- The full social media data used for deductive coding evaluation and the exploration of filter bubbles is available upon request.

A.3. Data collected

Table 3
Dimensions of Filter Bubbles and Their Measurements

Dimension	**Measurement**	**Data**
Diversity		
Structural	User/account names of post creators	creator_id
Topic	Video/photo data of the post, hashtags, description	post_text, post_file
Viewpoint	Deductive coding analysis of the LLM, video/photo data	interpretation, post_file
Time	Constellation of time and posts	scraped_at, posted_at
Recommendations	Logged timeline/explore feed	Post table, user_name, pk_id