# Scientific Appearance in Telegram

**Jeong-Eun Choi**[1, 2], **Karla Schäfer**[1, 2], **York Yannikos**[1, 2]
[1]Fraunhofer Institute for Secure Information Technology
[2]National Research Center for Applied Cybersecurity
jeong-eun.choi, karla.schaefer, york.yannikos@sit.fraunhofer.de

### Abstract

This paper examines the influence of scientific appearance (SA) on post dissemination and analyses a dataset of important actors in Germany, specifically those involved in the dissemination of disinformation on the social media platform Telegram. SA is identified through textual elements such as predefined keywords or digital object identifiers (DOIs). Characteristics and behaviours of actors with and without SA are compared using metadata such as forward counts and original posts. The additional content analysis provides insights into SA's usage and impact. The findings indicate that SA may influence the dissemination of posts and demonstrate how different methods can be applied for studying social media platforms.

## Introduction

The shift of information exchange to the digital sphere poses challenges in observing and studying information dissemination, which is closely related to the behaviours of actors on social media platforms, such as Telegram.

Most social media platforms provide metrics like views and forward counts that reflect the dissemination of information. However, the usability of such metrics depends on the information accessibility of the observed platform. The Telegram API[1], for instance, only discloses the original origin of the post and the user who has forwarded the post to the current channel or group. If a post is forwarded multiple times, it becomes impossible to trace its exact dissemination path. This limitation restricts the examination of both, information flows and interactions among actors on Telegram.

This paper analyses a Telegram dataset of important actors in Germany, specifically those involved in the dissemination of disinformation. It focuses on exploring the impact of posts with scientific information that are identified by their scientific appearance (SA) in text. First we used regex to identify relevant pieces of textual information which can be largely applied to any textual data in other platforms. Then, we observe the impact of SA posts in context of information dissemination.

Our analysis of SA posts using diverse methods reveals how combining these approaches helps to identify various types of actors reflecting trends, behaviours, and interactions within the community, which can be applied for analysing other platforms as well. Moreover, discussing actors in terms of personas enables privacy-preserving analysis, as explained later in this paper. Our findings not only reveal the influence of SA posts on information dissemination but also propose a privacy-preserving analysis of actors in the Telegram dataset, which can be applied to other social media platforms.

## Related Work

**Social Media Analytics and Telegram** Social media analytics (SMA) faces numerous challenges (Sebei, Hadj Taieb, and Ben Aouicha 2018) that are yet to be solved. For instance, the complexity of networks, diversity of platforms, and dynamics of social media platforms are causing difficulties in the application of SMA (Stieglitz et al. 2018).

Prior research has mostly focused on X (Twitter) (Alizadeh et al. 2019; Rajendran et al. 2022) and Facebook (Scrivens and Amarasingam 2020) and only a few addresses Telegram (LaMorgia et al. 2021). Telegram is a messenger application that is growing in popularity due to comparatively fewer regulations[2], especially among users who have been banned from other social media outlets like Facebook, Instagram, X or YouTube (Rogers 2020). In Telegram, there are three **types of actors**, namely individual users, groups, and channels (Schäfer and Choi 2023). Groups and channels can be further categorized to private and public. Telegram supports customizable anonymity features promising data privacy for users. For instance, users can decide not to allow tracking when their posts are forwarded. Such features may result in a lack of transparency and validity of the data, which could complicate the analysis (Zachlod et al. 2022) and limits a range of applicable analysis methods for the researchers. Therefore, this paper presents an opportunity to analyse Telegram despite restricted information caused by data privacy.

**Scientific Appearance (SA)** During the Covid-19 pandemic, the dissemination of scientific health information increased (Islam et al. 2020) and with it the danger caused by the dissemination of scientific false information. Beauvais



[1]The normal Telegram API and not Bot API.

[2]https://telegram.org/blog/ultimate-privacy-topics-2-0

| German [English] | doctor | wissenschaft [science] | forschung [research] | studie [study] | professor | dipl. [diploma] | phd | DOIs |
|---|---|---|---|---|---|---|---|---|
| Total (%) | **185566 (36.60)** | 105425 (20.79) | 51089 (10.08) | 95392 (18.81) | 65607 (12.94) | **845 (0.17)** | 1110 (0.22) | 2009 (0.40) |

Table 1: Overview of the Numbers of Posts with Scientific Appearance and the Terms Used for Classification

(2022) focused on the determinants that lead individuals to believe fake news and found that people tend to trust personalities they see in the media and experts with a scientific background. Tseng and Fogg (1999) identified four types of source credibility, where these types reflect the interplay between perceived credibility (by users or consumers) and generated credibility (by content creators or spreaders). They found that the messages were perceived as credible when the sources of information hold official titles, such as doctor and professor.

Such uses of a "scientific-looking" element is defined in this paper as posts with scientific appearance (SA). While certain visual features like tables or diagrams can contribute to the scientific appearance (Semar 2023), this paper specifically examines textual elements and considers the dissemination of these SA posts as a good indicator of credibility and reputation within Telegram.

## Methods

Our **German Telegram Dataset** consists of 995 public channels or groups (688 channels, 307 groups) crawled via the Telegram API in the period from 25-03-2022 to 30-06-2023. The channels and groups were selected during 15 interviews with journalism and fact-checking experts to identify important actors in the dissemination of disinformation in Germany. The interviews and pre-selection of channels and groups were organized by our partner from Hochschule der Medien as part of the project Dynamo [3]. Thematically, actors are associated with conspiracy theories and radical right-wing ideologies. Therefore, any suggestions or interpretations of our dataset must be approached with caution. Due to privacy policies, we cannot disclose the exact names and IDs of the channels and groups. The dataset is stored on a local server with restricted access only granted to project members. For more details, please refer to our analysis paper on the dataset (Schäfer and Choi 2023).

To address **data privacy concerns** in studying actors on social media platforms, particularly regarding what can be studied and made public, our focus shifts from studying individual users to examining channels and groups that individuals can join or manage. In this paper, channels and groups are treated as singular entities (actors), as channels and groups on Telegram may not always represent a collective, but can also be associated with a specific individual (ex. an official channel of a celebrity). Thus, channels and groups that share similar characteristics are grouped into personas instead of interpreting each of them individually. While the term "persona" can have different meanings depending on the context, in this paper, a persona portrays a singular entity (channel or group) that represents a collective for analysis purposes as proposed by Bounegru, Devries, and Weltevrede (2022). This approach enables us to provide meaningful interpretations about the observed community while maintaining privacy measures, as it does not require us to explicitly mention the names of channels or groups. Furthermore, this approach allows us to focus on relevant and interesting actors, which is particularly beneficial considering the vast amount of data that needs to be interpreted otherwise.

The detailed analysis steps are as follows. First, SA in posts was identified using a predefined **list of terms** (expressions). These terms were collected through an interview with experts involved in the project DESIVE[2] (Dewitz, Stiller, and Peters 2022)[4], which investigates the mechanisms of digital dissemination of supposedly scientific disinformation in the health context. The search for SA also considered relevant abbreviations, such as 'dr.' for 'doctor', or variations, such as 'studien' for the term 'studie'. The list of expressions can be found in Table 1.

Second, three groups of **personas** have been defined: 1) persona having the 0.99 percentile of total SA posts in our total dataset (actors with more than 2920 SA posts), 2) persona having the 0.99 percentile of percentage of SA posts in content posts (actors with more than 31% of SA posts), and 3) persona having 0.01 percentile of both (actors with 0 and thus 0% of SA posts). Telegram has three types of posts: posts, deleted posts and activity posts. Activity posts do not contain any content. Thus, for the percentage of SA posts instead of total posts only posts with content (called content posts) are considered (see Figure 1). The first persona represents the overall contribution in spreading and generating posts within the observed community (our dataset) and the second persona represents those whose content are more focused on SA posts. The last is the opposite of these two personas where no SA posts could be identified.

**Metadata Analysis** For metadata analysis, several popularity metrics were observed that were mostly accessed directly through the API. While there are numerous metrics available through the API, we have selected those that either represent 1) dissemination behaviour of posts, 2) dissemination strength of personas, 3) any other metrics that are related to the posting behaviours of personas, or 4) the textual characteristics of posts that are retrieved from processing textual content, i.e., not directly accessed through the API (see Table 2).

For the dissemination of posts and dissemination strength of personas, the forward count of original posts has been observed. Original forwards measure the forward counts of the posts that were initially posted by the persona and reflects

[3]https://www.dynamo.sit.fraunhofer.de/

[4]https://desive2.org/

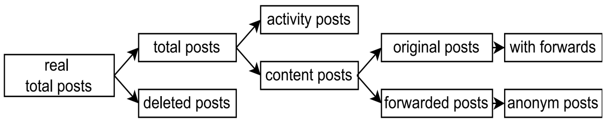


Figure 1: Division of Telegram posts into different post types

the dissemination strength of the observed persona. SA posts that are not originally from the observed persona were being neglected to ensure an accurate reflection of the dissemination strength of the SA posts posted in the observed persona. The percentage of original posts reflects the amount of original content the personas produce, while posts per day indicates the frequency of sharing information. The percentage of self-forwards (reposting original posts), deleted posts, and activity posts (see Figure 1) convey different intentions, such as fostering information dissemination, content monitoring, and actions within Telegram. All metrics that were extracted for each actor (channel or group) have been averaged within the same group (persona), see Table 2.

**Content Analysis** To gain further insights into the three personas, the texts from posts were analysed. First, the texts were preprocessed by removing URLs and emojis. The preprocessed texts (messages) were then used to extract topics for each persona using **BERTopic** (Grootendorst 2022)[5]. The multilingual sentence-transformer model *paraphrase-multilingual-MiniLM-L12-v2* (Reimers and Gurevych 2019, 2020) was used for content embedding, UMAP (McInnes, Healy, and Melville 2018) was applied for dimension reduction, and HDBSCAN (Campello, Moulavi, and Sander 2013) for clustering. Since BERTopic assigns each message to a single topic, only the top ten topics with the highest number of messages were considered as the representative topics of each persona. KeyBERT (Grootendorst 2020) was used to extract topic representations, which were then subjectively compared to assess the topical commonalities within and between groups.

For **sentiment analysis** using the classes positive, negative, and neutral, three different large language models (LLMs) were applied. Before giving the messages to the classifiers, the URLs were removed and the emojis were transformed into its text description using the emoji data descriptions.[6]. The first model used is XLM-T (Barbieri, Espinosa Anke, and Camacho-Collados 2022), which is based on XLM-R (Conneau et al. 2019). XML-R was trained on 2.5 TB of text in 100 languages, while XLM-T is XLM-R additionally fine-tuned on 198 million tweets in 30 languages[7]. The second model is called *multilingual-sentiment-covid19*[8] and is fine-tuned on 9,481,337 samples of multilingual tweets collected between March 2020 and November 2021 using the API of X and Covid-19 related keywords (Lampert and Lampert 2021) with the v2-version of the *stsb-xlm-r-multilingual* model (Reimers and Gurevych 2019) as the base model. The third model is a pre-trained BERT model that was trained on 1,834 million German-language samples from X, Facebook, and film, app and hotel reviews, for the purpose of sentiment classification (Guhr et al. 2020)[9].

The final sentiment label for each message, i.e. post, was assigned based on the sentiment labels with the highest votes among the three classifiers. The results were used to compare the characteristics of SA posts and non-SA posts.

## Results and Discussion

**Metadata Analysis** A total of 507,043 SA expressions were found in our dataset. In terms of posts, a total of 279,140 posts out of 7,992,117 content posts were identified as SA posts, which is only about 3.49%. Thus, it appears as if SA posts are insignificant in the observed community.

Table 2 shows the persona with the highest number of SA posts, referred to as the **contributor** (with a total of 10 actors); the persona with the highest percentage of SA posts, referred to as the **SA-focused** (with total of 10 actors); and the persona without any SA posts, referred to as the **non-SA** (with a total of 80 actors). The unequal distribution of actors among the persona is due to the presence of over 10 actors without SA posts. Since the total number of posts of non-SA was lower than other personas, all 80 actors were observed to provide a better contextual analysis. **All** considers all actors in the dataset and is used here as reference.

The comparison between SA and non-SA posts is referred to as the post-level analysis, while the comparison between persona is called the persona-level analysis. Table 2 displays the results of both analyses, with the top section showing the results from the post-level analysis and the bottom section showing the results from the persona-level analysis.

For measuring the dissemination strength of posts and personas, original forwards and participants were observed. Higher scores for $ori\text{-}forwards_{mean}$, and $participants_{mean}$ indicate higher dissemination strength (see Table 2). Examining the persona-level metrics, it appears that the contributor has a higher overall dissemination strength than the SA-focused (10,555.40 for contributors vs. 551.80 for the SA-focused). However, regarding post-level analysis only, the SA-focused has a much higher dissemination strength (454.00 for the SA-focused vs. 50.42 for the contributor). Thus, the SA-focused may have a significant role in creating and spreading SA posts, although the total number of SA posts and overall dissemination strength is lower than that of the contributors. Therefore, the SA-focused can be considered as the creators of SA posts, while the contributor appear like the spreaders of SA posts.

Regarding other metrics, the contributor behaves differently compared to other personas. They have the lowest mean percentage of original posts (initially posted by the persona), but the highest self-forward (reposting the original posts) and the highest post per day. Again, the con-

[5]https://maartengr.github.io/BERTopic/index.html

[6]https://unicode.org/Public/emoji/14.0/emoji-test.txt

[7]https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment

[8]https://github.com/ISTAustria-CVML/multilingual-sentiment-analysis

[9]https://huggingface.co/oliverguhr/german-sentiment-bert

| Metadata (source) / Persona | Contributor | | SA-focused | | non-SA | All |
|---|---|---|---|---|---|---|
| | SA | non-SA | SA | non-SA | | |
| $ori\text{-}forwards_{mean}$ (metric) | 50.42 | **24.04** | **454.00** | 154.26 | | |
| $emojis_{mean\%}$ (text) | 48.53 | 37.78 | **52.68** | 38.22 | | |
| $original_{mean\%}$ (metric) | 31.73 | 37.98 | **83.32** | 31.28 | | |
| $participants_{mean}$ (metric) | **43258.42** (67448.68) | | 38760.70 (43765.82) | | **4256.60** (24250.03) | 12361.2 (28375.34) |
| $ori\text{-}forwards_{mean}$ (metric) | **10555.40** (15313.39) | | 551.80 (303.32) | | **89.41** (200.38) | 1468.83 (4086.55) |
| $SA_{mean\%}$ (text) | 7.06 (3.31) | | **49.20** (14.37) | | 0 (0) | 4.90 (6.62) |
| $original_{mean\%}$ (metric) | 37.57 (27.97) | | 70.36 (22.78) | | **70.48** (34.38) | 61.83 (29.28) |
| $self\text{-}forward_{mean\%}$ (metric) | **10.08** (18.22) | | 0.95 (1.69) | | 0.07 (0.40) | 1.08 (4.34) |
| $deleted_{mean\%}$ (metric) | 11.82 (21.16) | | 11.23 (8.11) | | **21.56** (29.13) | 16.26 (21.86) |
| $activity_{mean\%}$ (metric) | 0.38 (0.51) | | 0.91 (0.88) | | **10.54** (16.66) | 3.34 (8.64) |
| $emojis_{mean\%}$ (text) | 38.70 (8.8) | | **44.97** (31.88) | | 21.48 (28.04) | 37.63 (23.39) |
| $postsPerDay_{mean}$ (metric) | **666.20** (1210.7) | | 1.43 (0.82) | | 1.19 (6.85) | 23.26 (148.45) |
| $channel_{count}/group_{count}$ (metric) | 5/5 | | 10/0 | | 71/9 | 688/307 |

Table 2: Overview of the Metadata Analysis - $ori\text{-}forwards_{mean}$ is the sum of the average original forwards divided by # of actors. For post-level, $emojis_{mean\%}$ or $original_{mean\%}$ refer to the mean % of the SA/non-SA of content posts. For the persona-level, $participants_{mean}$ ($postsPerDay_{mean}$) is the sum of the average participants (posts per day) over time divided by # of actors. $self\text{-}forward_{mean\%}$ is the mean % of original posts. $deleted_{mean\%}$ and $activity_{mean\%}$ are the mean % in real total posts and total posts. SA, original and emojis are the mean % of the content posts. In brackets are standard variations.

tributor appears to focus more on spreading the posts rather than generating its own content. The non-SA has the highest percentage of original posts. However, it also has the highest number of activity posts (posts without content), which are actions within Telegram such as joining the channel or group. Moreover, it has the highest percentage of deleted posts and the lowest percentage of posts with emojis. As for SA-focused, the percentage of original posts of the SA-focused is comparable to that of the non-SA, but it has a higher percentage of posts with emojis than the non-SA (44.97% for the SA-focused vs. 21.48% for the non-SA). The SA-focused also has the lowest percentage of deleted posts which could be interpreted either as 1) lower self-monitoring behaviours or 2) higher confidence in contextual correctness of posts. Hence, one could deduce that the posts in the SA-focused and the non-SA are likely to differ in their content or their uses. Another interesting insight is that the SA-focused seems to use emojis more frequently than the other personas. This may be related to the tendency for SA posts to generally include emojis.

**Content Analysis** The topic modelling results indicate that the primary topics of the SA-focused are health (vaccine, corona), link to other websites (podcast, streaming, twitter), politics (German, Russia, Ukraine, Palestine and Israel), and references to experts (Dr. X). This supports the assumption that the SA-focused behaves more like a creator of new SA content in Telegram. In contrast, the top topics in the non-SA are discussion (think, need, look), esoteric (oracle), missing person (missing, police), Arabic names (incidents caused by Arabic names), links to other websites, and politics (Trump). Hence, the non-SA tends to focus on discussion, which is also reflected in the higher percentage of original posts and also higher percentage of deleted posts. For the contributor, the primary topics are health (vaccine, corona), politics (Ukraine, Russia, Demonstrations, Q-Anon, Querdenken), mentioning of a well-known person (Karl L., Michael B.), esoteric, advertisements, and links to other websites (YouTube). Therefore, it covers broader topics, whereas advertising is unique to this persona. This could explain the high number of posts per day and self-forwards.

When comparing the sentiments of SA and non-SA posts of the contributor, it appears that 9 out of 10 actors use more neutral tone in SA posts and either a more positive or negative tone in non-SA posts. Within the SA-focused, a similar behaviour can be observed, with 7 out of 10 actors using neutral tones for SA posts, while 8 out of 10 using either a more positive or negative tone for non-SA posts. As a result, SA posts tend to have a more neutral tone.

## Conclusion

This paper identifies and examines SA in Telegram through metadata and content analysis. During the analysis we focused to interpret and understand metadata analysis through content analysis. We argue that a single approach alone does not provide sufficient insights into the complex information exchange on social media platforms. Metadata analysis lacks contextual understanding, while content analysis may lack robustness and interpretability. Our results indicate that SA can influence the information dissemination on Telegram. The use of neutral tones and emojis in SA posts are notable characteristics, that require further investigation. Furthermore, we have identified two interesting personas: the SA-focused, who appears to be the origin of SA posts in the observed community, and the contributor, who focuses on spreading information.

We plan to gather more expressions to identify further SA posts for advanced analysis of SA in Telegram. Additionally, the DOIs and corresponding scientific papers will be investigated, as we have discovered some retracted papers mentioned in the observed community. Furthermore, we intend to identify other connections, such as by identifying copy-pasted texts and the use of external links, or other types of posts, such as advertisements to determine whether the SA-focused really serves as the origin of SA posts.

## Acknowledgments

This research work was supported by the National Research Center for Applied Cybersecurity ATHENE as well as within the German Federal Ministry of Education and Research project DYNAMO.

## References

Alizadeh, M.; Weber, I.; Cioffi-Revilla, C.; Fortunato, S.; and Macy, M. 2019. Psychology and morality of political extremists: evidence from Twitter language analysis of alt-right and Antifa. *EPJ Data Science*, 8(1): 17.

Barbieri, F.; Espinosa Anke, L.; and Camacho-Collados, J. 2022. XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 258–266. Marseille, France: European Language Resources Association.

Beauvais, C. 2022. Fake news: Why do we believe it? *Joint Bone Spine*, 89(4): 105371.

Bounegru, L.; Devries, M.; and Weltevrede, E. 2022. *The Research Persona Method: Figuring and Reconfiguring Personalised Information Flows*, 77–104. Singapore: Springer Nature Singapore. ISBN 978-981-19-2476-7.

Campello, R. J.; Moulavi, D.; and Sander, J. 2013. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, 160–172. Springer.

Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Dewitz, L.; Stiller, J.; and Peters, I. 2022. Fake scientific provenance as a driver for disinformation – Peoples' disinformation behavior in health contexts. *Zenodo*.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Grootendorst, M. 2020. KeyBERT: Minimal keyword extraction with BERT.

Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Guhr, O.; Schumann, A.-K.; Bahrmann, F.; and Böhme, H. J. 2020. Training a Broad-Coverage German Sentiment Classification Model for Dialog Systems. In *Proceedings of The 12th Language Resources and Evaluation Conference*, 1620–1625. Marseille, France: European Language Resources Association.

Islam, M. S.; Sarkar, T.; Khan, S. H.; Kamal, A.-H. M.; Hasan, S. M.; Kabir, A.; Yeasmin, D.; Islam, M. A.; Chowdhury, K. I. A.; Anwar, K. S.; et al. 2020. COVID-19–related infodemic and its impact on public health: A global social media analysis. *The American journal of tropical medicine and hygiene*, 103(4): 1621.

LaMorgia, M.; Mei, A.; Mongardini, A. M.; and Wu, J. 2021. Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. *CoRR*, abs/2111.13530.

Lampert, J.; and Lampert, C. H. 2021. Overcoming Rare-Language Discrimination in Multi-Lingual Sentiment Analysis. In *IEEE International Conference on Big Data (BigData)*. Special Session: Machine Learning on Big Data.

McInnes, L.; Healy, J.; and Melville, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Rajendran, A.; Sahithi, V. S.; Gupta, C.; Yadav, M.; Ahirrao, S.; Kotecha, K. V.; Gaikwad, M.; Abraham, A.; Ahmed, N.; and Alhammad, S. M. 2022. Detecting Extremism on Twitter During U.S. Capitol Riot Using Deep Learning Techniques. *IEEE Access*, 10: 133052–133077.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Reimers, N.; and Gurevych, I. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Rogers, R. 2020. Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, 35(3): 213–229.

Schäfer, K.; and Choi, J.-E. 2023. Transparency in Messengers: A Metadata Analysis Based on the Example of Telegram. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*, HT '23. Association for Computing Machinery. ISBN 9798400702327.

Scrivens, R.; and Amarasingam, A. 2020. *Haters Gonna "Like": Exploring Canadian Far-Right Extremism on Facebook*, 63–89. Cham: Springer International Publishing. ISBN 978-3-030-30138-5.

Sebei, H.; Hadj Taieb, M. A.; and Ben Aouicha, M. 2018. Review of social media analytics process and Big Data pipeline. *Social Network Analysis and Mining*, 8.

Semar, W. 2023. Nachhaltige Information – Information für Nachhaltigkeit.

Stieglitz, S.; Mirbabaie, M.; Ross, B.; and Neuberger, C. 2018. Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39: 156–168.

Tseng, S.; and Fogg, B. 1999. Credibility and computing technology. *Communications of the ACM*, 42(5): 39–44.

Zachlod, C.; Samuel, O.; Ochsner, A.; and Werthmüller, S. 2022. Analytics of social media data – State of characteristics and application. *Journal of Business Research*, 144: 1064–1076.

## Paper Checklist - Ethics Guidelines

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? No, because it does not mention any specific identifiable information
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes
   (f) Did you discuss any potential negative societal impacts of your work? Yes, see methods/approach applied
   (g) Did you discuss any potential misuse of your work? Yes, that is why such methods/approach were applied
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA, we used pre-trained models
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA, we used pre-trained models
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA, we used pre-trained models
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA, we used pre-trained models
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA, we used pre-trained models
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA, we used pre-trained models
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes, also URLs as footnotes.
   (b) Did you mention the license of the assets? Yes, see URLs.
   (c) Did you include any new assets in the supplemental material or as a URL? Yes
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA, we used public channels and groups in Telegram where anyone can join and not the private/secrete ones
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, see methods/approach applied
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA, we are not releasing datasets
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA, we are not releasing datasets
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA