https://doi.org/10.2352/EI.2024.36.1.VDA-359

Creating Visual Persona Profiles in Telegram using NLP

Jeong-Eun Choi, Karla Schäfer, Martin Steinebach; Fraunhofer Institute for Secure Information Technology SIT — ATHENE - National Research Center for Applied Cybersecurity; Darmstadt, Germany

Abstract

Numerous studies of social media analytics (SMA) shed light upon interesting insights into the information flow in social media. As social media becomes a crucial part of human society, bridging and merging these studies could shape ideas and designs for real-world applications that allow more transparency and understanding of social media. Among several challenges of SMA, this paper focuses on two issues of 1) invasive and greedy analysis methods concerning user privacy, and 2) lack of comprehensive representations of analysis results. We use our analysis on Telegram data to propose that pursuing persona profiling using generalizing contextual analysis via Natural Language Processing (NLP) technologies could address the first problem. For the second problem, we propose to visualize the analysis results, i.e. persona profiles, to increase both comprehensibility and interpretability.

Introduction

The rise of the internet and the shift of information sources from traditional media outlets to various social media platforms facilitated the dissemination of information to an unprecedented scale that also led to negative consequences. For instance, multiple attempts have been discovered that influenced national elections or endangered public health in the Covid-19 pandemic through spreading (dis-)information in social media platforms [1, 2, 3, 4]. Different scholars point out that through the introduction of the internet, the trust in the information provided by authorities or elites has been degraded [5, 6], which is not only caused by the internet alone but by the inability of authorities and elites to provide the necessary verification, integrity, and authentication measures [6], and by the numerous negative consequences that are caused by the lack of information transparency [7, 8, 9, 10, 11, 12]. This paper focuses on one of these practical goals, by supporting stronger transparency for users.

Providing transparency for users raises concerns about data privacy. Therefore, rather than profiling individuals in social media, we propose to pursue persona profiling. The term “persona” can have different meanings depending on the context. In general, a persona is a singular entity that represents a collective [13]. In social media, personas can represent specific groups sharing similar characteristics that are considered as a single entity during the analysis. Furthermore, rather than solely depending on analysis methods based on metadata analysis, we propose to apply generalizing contextual analysis using NLP technologies that provide a comprehensive summary of each persona.

Moreover, to provide transparency for users, it is important to enhance the usability, interpretability, and comprehensibility of the analysis results. We believe that a suitable visualization will provide necessary transparency, allowing users of social media to raise awareness for more responsible decisions and actions.

Related Work

Social media analytics (SMA) faces numerous challenges that are yet to be solved, which are often caused by the volume, variety, veracity, and velocity of data [14]. For any type of SMA research, it is important to clearly describe each step of discovery, tracking, preparation, and analysis [15] which are not always provided due to the data privacy protection measures or lack of transparency and validity of the data and methods used [16]. SMA also suffers from the lack of applicability which is mainly due to the complexity (network), diversity (platforms), and dynamics of social media platforms, but also because new methods and findings are not evaluated in terms of their actual usability [15].

The examination of network analysis methods, specifically used for SMA is called social network analysis (SNA). Traditionally, SNA has not included online social networks but has focused more on the analysis of physical social networks. The aim of SNA is to “detect and to understand the structure identified by a set of relations defined on a set of individuals” [17] and consists of 1) understanding the link between the environment of a person and the properties and behaviours of that person (relation theory), and 2) understanding the network structure (graph theory) [17]. However, depending on the type and structure of the network, some SNA methods cannot be adopted or require additional adaptation. On X (Twitter), for example, single user profiling techniques are frequently practised [18], because metadata of user-accounts, like followers and followees, were easily accessible. As for Telegram, there are certain limitations in applying user-profiling methods due to different functionalities for securing the anonymity of user-accounts, which restricts the study of a user’s relationships and also the study of a post’s dissemination paths. Instead, other types of personas, like group-accounts and channel-accounts with different accessible information can be examined.

Dataset

The dataset analysed consists of three Telegram channels and three Telegram groups crawled via Telegram API in the period from 25-03-2022 to 31-07-2022, whereas the time span of the crawled messages varies, with the earliest post being from 20-09-2020 to the latest post being from 31-07-2022. The selected channels (Ch_1,Ch_2,Ch_3) and groups (Gr_1,Gr_2,Gr_3) are publicly accessible, well-known in German-speaking countries[1], and have a high number of activities. Due to data privacy concerns, we anonymised the names of the selected channels and groups. The dataset consists of 53,438 posts, where 7,211 posts were removed for our experiment because they do not contain any text

[1]From the selected channels and groups four of them (Ch_1,Ch_2,Ch_3,Gr_3) have official websites where the exact telegram addresses can be found. Gr_1 and Gr_2 have more members than Gr_3.

messages[2] or because the texts are too short (less or equal to 6 words)[3]. The final dataset consists of 46,227 posts, most of them identified as written in German. This dataset is part of a larger Telegram corpora introduced in the work by Schäfer and Choi [21].

Experimental Methodology

In our experiment, we selected and applied NLP methods for sentiment analysis, topic modeling, spam detection, and detecting scientific appearance. Additionally, we analysed the metadata using simple statistical methods.

Through **sentiment analysis**, texts can be classified by their polarity as positive, negative or neutral [22]. This classification task can be divided into supervised machine learning, unsupervised lexicon-based or concept/ontology-based approaches [23]. [23] compared different works and concluded that those based on supervised machine learning performed the best. We used XLM-T [24], a model based on the XLM-R language model [25] and pre-trained on 198 M multilingual Twitter posts (including German). This multilingual model achieved a F1 Score of 77.35% on German data (SB 10k dataset [26]) [24] and is also publicly available on Huggingface[4].

Another NLP technique applied is **topic modeling**. Besides two traditional topic modeling methods (LDA and NMF), there are more advanced methods such as BERTopic and Top2Vec [27, 28]. BERTopic uses a transformer-based language model for embedding. The algorithm for dimension reduction, clustering, and topic extraction can be chosen [29]. We used UMAP for dimension reduction, HDBSCAN for clustering, and TF-IDF for extracting topic descriptive words. The usage of transformer-based language models for embedding preserves semantic and syntactic information, unlike the LDA and NMF methods. BERTopic has achieved excellent results in previous work [27, 28, 30], therefore we used BERTopic for our experiment while using ‘paraphrase-multilingual-mpnet-base-v2’ for embedding. This model is based on Sentence BERT (SBERT), a work by [31], and has been trained on various (50+) languages.

Further content information can be obtained through **spam (advertisement) detection**. Spammers are people who send unwanted messages to people to either advertise a product or lure the victims into clicking malicious links [32]. In social media, one can distinguish spam detection by identifying spammer accounts [32] or by analysing the posted content. As no observable user profiles exist in Telegram, we focus here on identifying spam and advertisement by content. Spam can be recognised by certain words, [33] identified spam terms that appear in most spam E-mails, Twitter, or Facebook posts. Of these terms, we selected those that, we believe, are only applicable to social media. Our list of spam words included the terms: ‘Full refund’, ‘Get it Now’, ‘Order now’, ‘Order status’, ‘Make money’, ‘Earn extra cash’, ‘100% free’, ‘Apply now’, ‘Winner’, ‘Lose weight’, ‘Win’, ‘ipad’, ‘Mobi’, ‘paypal’, ‘shop’, ‘click’, ‘store’ (in English and German). Based on these spam words, we classified posts as spam or non-spam. First, we looked at the URLs in the messages separately, as we suspected that the URLs could be the best way to identify spam. Then, we looked at the entire posts and examined them for spam using the terms mentioned above.

In connection with the Covid-19 pandemic, false reports with a **scientific appearance** were spread on a large scale [34]. Scientific appearances can cause posts with disinformation to appear more true and genuine. Titles such as ’Doctor’ or ’Professor’ ensure a certain trustworthiness of readers towards the content. We aimed to identify such posts by searching for these titles and, if available, examine the content more closely. Therefore, we searched for the following terms in the posts: ’Prof.’, ’Professor’, ’Dr.’, ’Doctor’, ’Dipl.’, and ’Ph.D.’.

Results and Discussion

Four NLP methods were applied on the social media posts of six different Telegram personas. For all six personas, in the **sentiment analysis** mainly either negative or neutral sentiments were classified. An overview of the results from sentiment analysis, for the six personas, are presented in Figure 1. The highest percentage of negative contributions was observed in Ch_1 with 72.73% and the lowest in Gr_3 with 19.11%. The largest share of positive posts can be found at Gr_3 with 16.44%. Depending on the circumstances, the sentiment analysis can help to identify relevant personas for research. For example, since the negative sentiment is one of the significant indicators for disinformation [35], for those exploring disinformation, Ch_1 might be more relevant, followed by Gr_1 with 56.98% of the posts being negative. Thus, it allows primary selection of likely relevant pieces of information within a large dataset.

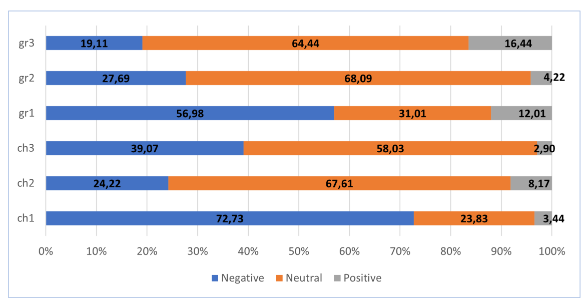

Figure 1. *Sentiment Analysis Results*

Topic modeling with BERTopic identified from a total of 3 (Gr_3) up to 38 (Gr_2) topics per persona. Table 1 provides an overview of the topics found in several personas. For better readability in Table 1, the topic name column was created by us based on the topic names provided by BERTopic, and the “original” topic names were additionally translated into English. Common topics across individual personas were the Ukraine war (Ch_1, Ch_3, Gr_2), the Covid-19 pandemic (Ch_1, Ch_3, Gr_2), vaccination (Ch_1, Ch_3, Gr_1), police and protests (Ch_1, Ch_3), the freedom of the press (Ch_1, Ch_3), the climate/-change (Ch_1, Ch_2) and China (Ch_2, Gr_2). It is noticeable that more comprehensible topic names were generated by BERTopic for the channels than for groups. In the groups, topics such as “”0_patrick_more_have_been_hello”, “7_clown_face_posts_groups” or “1_flyer_distribution_gives_theme” were formed, see Table 1. We suspect that this is due to discussions taking place in

[2]These are action posts like adding a user to a chat or they are posts with images only.

[3]In German Spotlight-DE corpus[19] consisting of different reading levels of articles of different topics, the lowest average length of short sentences is slightly above 6 words [20]

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

groups, which is functionally absent in channels, that leads to an increase in colloquial language. These results could be improved in future work through increased preprocessing of the colloquial language. In general, topic modeling with BERTopic worked well on social media posts and provides a good first impression of personas' topics.

Topic modeling results. In blue the topics of the persona "group".

Topic name	Persona	Topic name by BERTopic (translated from german)	Number posts
Ukraine War	Ch_1	0_putin_ukraine_war_russia	105
	Ch_3	3_ukraine_russia_war_nato	130
	Gr_2	1_ukraine_russia_putin_war	221
Covid-19 pandemic	Ch_1	6_corona_virus_measures_autumn	53
	Ch_3	7_corona_pandemic_measures_who	48
	Gr_2	4_china_shanghai_lockdown_covid	69
Vaccination	Ch_1	1_mandatory vaccination_biontech_vaccine_data	85
	Ch_3	0_mandatoryvaccination_vaccination_corona_covid	531
	Gr_1	1_vaccinated_vaccination_patrick	187
Police (protesters)	Ch_1	7_police_demonstrators_violence_berliner	52
	Ch_3	4_wien_today_demo_street	116
		8_police_police violence	34
Freedom of the press	Ch_1	9_federal press conference_journalists_press freedom_broadcasting	49
	Ch_3	6_media_censorship_telegram_facebook	76
Climate	Ch_1	2_climate_lengsfeld_vera_greens	78
	Ch_2	2_climate_climate change_climateprotection_glasgow	209
China	Ch_2	1_china_chinas_chinese_chinese	310
	Gr_2	4_china_shanghai_lockdown_covid	69

In **spam detection**, through keyword search, we classified posts as possibly spam. The idea behind spam detection was to identify personas with large amounts of advertisement. For the personas considered in our experiment, the proportion of spam was rather low. Ch_3 with 10.49% has the highest percentage of posts classified as spam. Figure 2 displays the distribution of detected spam posts and the respective classification method per persona.

We also used keyword search to examine the posts for **scientific appearance**. Posts with the keywords "professor" and "doctor" mostly contained news related to the Covid-19 pandemic (vaccination etc.), but also less frequently, energy and nuclear power. In Ch_1, 38 posts with a scientific appearance were found (2.08% of all posts). Figure 3 illustrates the percentage of posts with scientific appearance of each persona.

Examples of posts classified as containing scientific appearance can be found in Table 2. We believe that posts with a scientific appearance could create more trust in the content of the posts [5, 34]. In practice, since the mention of scientific titles does not

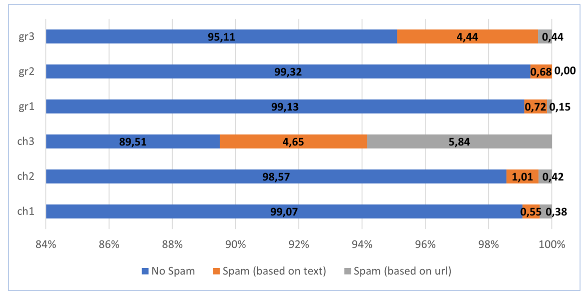

Figure 2. Percentage Spam Posts

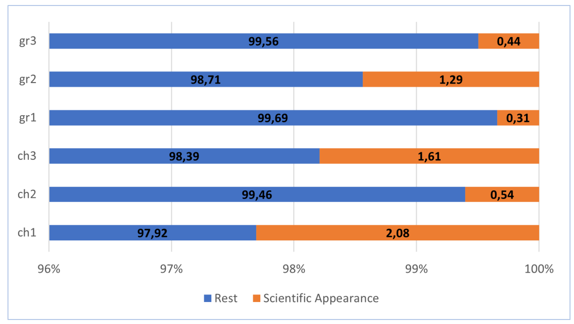

Figure 3. Percentage Scientific Appearance

Example results of scientific appearance (Ch_1); translated from German.

Example	Keyword
'"The next generation will not be so old". Prof. Dr. Dr. Christian Schubert and Michael Hüter in conversation. The collateral damage caused when dealing with children in the pandemic will run through the entire lives of the young generation, the two guests predict.'	Prof.
'Professors demand: Nuclear power plants must continue to run. Stuttgart Declaration: "Rising energy prices and falling security of supply endanger competitiveness and prosperity" - Professors of German universities rebel'	Professor

prove the factuality, users should be encouraged to critically scrutinize the posts. These posts can be also used to offer a first clue to identify relevant posts regarding identification of disinformation in connection with a scientific appearance. Furthermore, these posts can be primarily selected to be assessed in fact-checking systems, to clarify whether they are false information.

Besides the content analysis performed using NLP techniques, **metadata** provides interesting information about their general behaviours or characteristics. In the metadata analysis, the feature *percentage_of_activity_posts* displays the percentage of posts that do not have contents but shows activities within the platform such as adding a new user. The feature *percentage_of_original_posts* shows the percentage of posts that are not forwarded from others, i.e. original post produced by the channel or group. With *percentage_of_original_posts_w_forwards* the percentage of original posts that were forwarded to other channels at least once is shown. This reflects the dissemination power of the posts originally produced by the channel or group. *percentage_of_textual_content* is the percentage of posts that have textual contents. For example, a post containing only an image do not have any textual content. With *total_time_online* the total days from the day when the groups or channel is created until 31-07-2022 is shown.

Additionally, we generated our own parameters like *total_used_posts/total_posts* to show the percentage of posts used for the experiment, i.e. after posts are removed (cleaned) if they that have 1) no content, 2) no textual contents, and 3) too short textual contents that have less than or equal to 6 words. *total_languages_predicted* shows the total number of languages predicted using lingua language detection[5] for each persona after cleaning. *freq_language* shows the language with maximum number of posts in the final dataset. In *percentage_of_freq_language* the percentage of posts of *freq_language* is given.

Such metadata analysis can also help to reveal some undiscovered features of the observed social media platform. For example, according to the metadata only, the original posts in groups are not forwarded to other personas at all. Through further investigation, we found out that this is caused by the technical design of Telegram, so that the original posts in groups do not include information about whether they are forwarded or not. In other words, the users do not have the information about how many times the original posts in groups are forwarded, whereas in channels information about forwards for original posts are provided. In Table 3 the results of the metadata analysis are displayed.

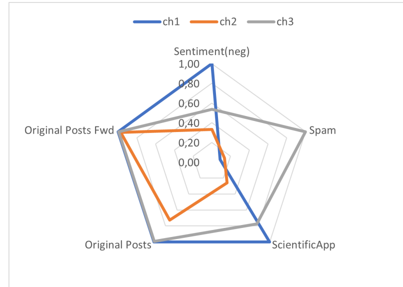

Figure 4. *Channel Profiling*

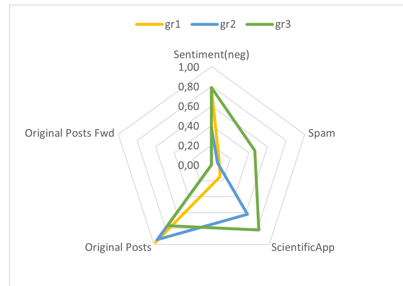

Figure 5. *Group Profiling*

Example Persona Profiling

We think that combining the insights gained with NLP with additional information from metadata provides helpful representation of a persona. For example, examining different features could allow a visualization of the differences between the two persona types, as shown in Figure 4 and 5. The different distributions of topics allow to predict the focus of the personas and how they might be topically related to each other. This can be beneficial for researchers dealing with massive amount of data. For example, Gr_3 does not cover any of the six identified common topics (see Figure 6), thus the behaviour of this group might be different compared to other personas. Ch_1 could be a good starting point for tracking the information flow as it covers most of the common topics and the percentage of original posts forwarded to other personas is very high (see Table 3), i.e. it could have more opinion-building contents than Ch_2 with noticeably lower negative sentiment. From these data, it is also possible to identify unique characteristics for each persona, such as Ch_3 having more spam than other personas and Ch_2 having relatively lower scientific appearances and original posts.

Such visualization can help users to gain quick general overview of personas which in turn encourages users to behave consciously in social media platforms. It allows them to question certain anomalies such as why a persona has much more advertisement or why a certain type of persona do not have any original forwards.

Researchers or explorers of social media platforms, can use such visualization to design hypothesis. For example, one could use them to identify the creators, the spreaders and the consumers of a social media platform. The creators are those who create content and are the starting point of information flow, here e.g. Ch_1. The spreaders are those who spread information and the consumers are the end points of the information flow, who mostly consume the information rather than spreading it. This idea is to be studied in our future work.

[5] https://github.com/pemistahl/lingua-py

Conclusion

Applying metadata and contextual analysis based on NLP and combining the analysis results into a visual comprehensive form are both essential elements for developing a tool for fostering transparency in social media. As social media becomes an essential part of human society, the study on Telegram (which is comparably less researched) and the study of visualizing analysis of social media data will become important not only for normal users but also other scientists of different disciplines and stakeholders from different sectors. Furthermore, this work tries to show the possibilities of conducting SMA in a less invasive way, while presenting the practice of handling large amounts of data by applying both metadata analysis and NLP technologies.

Topic modeling and sentiment analysis can provide helpful first impressions of the personas viewed. In our paper, especially the topics of the persona type “channel” could be determined well. Yet for the persona type “group” inferior results were achieved. The detection of scientific appearance should be further evaluated since in this paper we used a simple keyword search. In the future, we plan to evaluated further methods to detect not only “titles” but scientific descriptions and terminologies. By using spam detection with keywords, some posts were identified, especially an analysis of URLs appears promising. However, more advanced methods, based on machine learning, could be used in the future. Photos and videos are also often found in social media and thus, multimodal analysis approach is to be explored as well.

Dataset Analysis, divided in Activity and Content Analysis. In blue the smallest and in red the largest value.

Features	Ch_1	Ch_2	Ch_3	Gr_1	Gr_2	Gr_3
Activity Analysis						
avg_post_per_day (posts)	7.45	112.58	8.28	8.72	58.16	1.52
percentage_of_activity_posts (%)	0.001	0.005	0.001	0.000	0.017	0.505
percentage_of_original_posts (%)	1.000	0.728	0.996	0.974	0.941	0.767
percentage_of_original_posts_w_forwards(%)	0.999	0.972	0.992	0	0	0
avg_forwards_per_original_post (posts)	996.52	343.98	1074.68	0	0	0
percentage_of_textual_content (%)	0.997	0.900	0.901	0.940	0.988	0.761
total_time_online (days)	249	333	294	244	146	680
Content Analysis						
total_used_posts (posts)	1830	31781	2173	1948	8270	225
total_used_posts/total_posts (%)	0.987	0.848	0.892	0.915	0.974	0.216
max_length (words)	330	681	478	668	561	811
mean_length (words)	45.6	89.1	68.6	144.6	93.6	52.3
min_length (words)	8	7	7	7	7	7
total_languages_detected	2	16	1	1	1	2
freq_language	German	German	German	German	German	German
percentage_in_freq_ language (%)	0.999	0.995	1.000	1.000	1.000	0.996

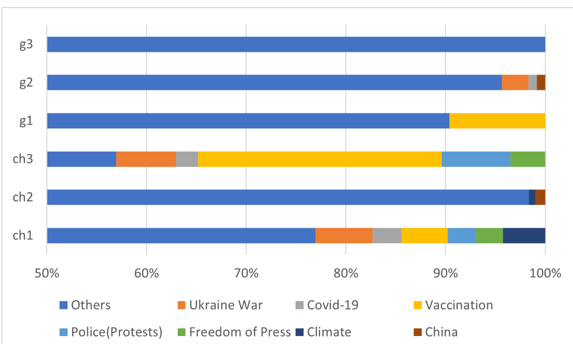

Figure 6. *Distribution of six common topics*

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

[1] Fabian Zimmermann and Matthias Kohring. Mistrust, disinforming news, and vote choice: A panel survey on the origins and consequences of believing disinformation in the 2017 german parliamentary election. *Political Communication*, 37(2):215–237, 2020.

[2] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on twitter during the 2016 u.s. presidential election. *Science*, 363(6425):374–378, 2019.

[3] Alexandre Bovet and Hernán A. Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature Communications*, 10(1), 2019.

[4] Sahil Loomba, Alexandre de Figueiredo, Simon J. Piatek, Kristen de Graaf, and Heidi J. Larson. Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. *Nature Human Behaviour*, 5(3):337–348, 2021.

[5] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[6] Andreas Jungherr and Ralph Schroeder. Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, 7(1):2056305121988928, 2021.

[7] Gordon Pennycook and David G. Rand. The psychology of fake news. *Trends in Cognitive Sciences*, 25(5):388–402, 2021.

[8] Liwei Hsu and Yen-Jung Chen. Neuromarketing, subliminal advertising, and hotel selection: An eeg study. *Australasian Marketing Journal (AMJ)*, 28(4):200–208, 2020.

[9] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49):12435–12440, 2018.

[10] Matthew Moore. Unilever dumps celebrities who pay for followers, Jun 2018.

[11] Todor Mihaylov, Ivan Koychev, Georgi Georgiev, and Preslav Nakov. Exposing paid opinion manipulation trolls. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 443–450, Hissar, Bulgaria, September 2015. INCOMA Ltd. Shoumen, BULGARIA.

[12] Megan McCluskey. Fake reviews: Inside the fight to stop them, Jul 2022.

[13] Liliana Bounegru, Melody Devries, and Esther Weltevrede. *The Research Persona Method: Figuring and Reconfiguring Personalised Information Flows*, pages 77–104. Springer Nature Singapore, Singapore, 2022.

[14] Hiba Sebei, Mohamed Ali Hadj Taieb, and Mohamed Ben Aouicha. Review of social media analytics process and big data pipeline. *Social Network Analysis and Mining*, 8, 04 2018.

[15] Stefan Stieglitz, Milad Mirbabaie, Björn Ross, and Christoph Neuberger. Social media analytics – challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39:156–168, 2018.

[16] Cécile Zachlod, Olga Samuel, Andrea Ochsner, and Sarah Werthmüller. Analytics of social media data – state of characteristics and application. *Journal of Business Research*, 144:1064–1076, 2022.

[17] Francesco Martino and Andrea Spoto. Social network analysis: A

brief theoretical review and further perspectives in the study of information technology. *PsychNology J.*, 4(1):53–86, 2006.

[18] Amandeep Singh, Malka N. Halgamuge, and Beulah Moses. An analysis of demographic and behavior trends using social media: Facebook, twitter, and instagram. *Social Network Analytics*, page 87–108, 2019.

[19] Zarah Weiss, Xiaobin Chen, and Detmar Meurers. Using broad linguistic complexity modeling for cross-lingual readability assessment. In *Proceedings of the 10th Workshop on NLP for Computer Assisted Language Learning*, pages 38–54, Online, May 2021. LiU Electronic Press.

[20] Zarah Weiss and Detmar Meurers. Assessing sentence readability for German language learners with broad linguistic modeling or readability formulas: When do linguistic insights make a difference? In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, pages 141–153, Seattle, Washington, July 2022. Association for Computational Linguistics.

[21] Karla Schäfer and Jeong-Eun Choi. Transparency in messengers: A metadata analysis based on the example of telegram. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*, pages 1–3, 2023.

[22] Amandeep Kaur and Vishal Gupta. A survey on sentiment analysis and opinion mining techniques. *Journal of Emerging Technologies in Web Intelligence*, 5(4):367–371, 2013.

[23] Chaima Messaoudi, Zahia Guessoum, and Lotfi Ben Romdhane. Opinion mining in online social media: a survey. *Social Network Analysis and Mining*, 12(1):1–18, 2022.

[24] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond. In *Proceedings of the Language Resources and Evaluation Conference*, pages 258–266, Marseille, France, June 2022. European Language Resources Association.

[25] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2019.

[26] Mark Cieliebak, Jan Milan Deriu, Dominic Egger, and Fatih Uzdilli. A twitter corpus and benchmark resources for german sentiment analysis. In *5th International Workshop on Natural Language Processing for Social Media, Boston MA, USA, 11 December 2017*, pages 45–51. Association for Computational Linguistics, 2017.

[27] Abeer Abuzayed and Hend Al-Khalifa. Bert for arabic topic modeling: An experimental study on bertopic technique. *Procedia Computer Science*, 189:191–194, 2021. AI in Computational Linguistics.

[28] Roman Egger and Joanne Yu. A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in Sociology*, 7, 05 2022.

[29] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[30] Darell Hendry, Fariz Darari, Raditya Nurfadillah, Gaurav Khanna, Meng Sun, Paul Constantine Condylis, and Natanael Taufik. Topic modeling for customer service chats. In *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 1–6, 2021.

[31] Nils Reimers and Iryna Gurevych. 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, 11 2019.

[32] B Mukunthan and M Arunkrishna. Spam detection and spammer behaviour analysis in twitter using content based filtering approach. In *Journal of Physics: Conference Series*, volume 1817, page 012014. IOP Publishing, 2021.

[33] Sanaa Kaddoura, Ganesh Chandrasekaran, Daniela Elena Popescu, and Jude Hemanth Duraisamy. A systematic literature review on spam content detection and classification. *PeerJ Computer Science*, 8:e830, 2022.

[34] José van Dijck and Donya Alinejad. Social media and trust in scientific expertise: Debating the covid-19 pandemic in the netherlands. *Social Media + Society*, 6(4):2056305120981057, 2020.

[35] Razieh Nokhbeh Zaeem, Chengjing Li, and K. Suzanne Barber. On sentiment of online fake news. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 760–767, 2020.

Author Biography

Jeong-Eun Choi and Karla Schäfer are researchers at the Media Security and IT Forensics division at Fraunhofer SIT. Both are working on research focusing on natural language processing and the detection of manipulations in audio data.

Prof. Dr. Martin Steinebach is the Manager of the Media Security and IT Forensics division at Fraunhofer SIT. In 2003 he received his PhD at the Technical University of Darmstadt for this work on digital audio watermarking. In 2016 he became honorary professor at the TU Darmstadt.