Barriers to the Introduction of Artificial Intelligence to Support Communication Experts in Media and the Public Sector to Combat Fake News and Misinformation

Walter Seböck, Bettina Biron, and Thomas J. Lampoltshammer(✉)

Department for E-Governance and Administration, University for Continuing Education Krems,
Dr.-Karl-Dorrek-Str. 30, 3500 Krems an der Donau, Austria
{walter.seboeck,bettina.biron,
thomas.lampoltshammer}@donau-uni.ac.at

Abstract. Public trust represents a cornerstone of today's democracies, their media, and institutions and in the search for consensus among different actors. However, the deliberate and non-deliberate spreading of misinformation and fake news severely damages the cohesion of our societies. This effect is intensified by the ease and speed of information creation and distribution that today's social media offers. In addition, the current state-of-the-art for artificial intelligence available to everybody at their fingertips to create ultra-realistic fake multimedia news is unprecedented. This situation challenges professionals within the communication sphere, i.e., media professionals and public servants, to counter this flood of misinformation. While these professionals can also use artificial intelligence to combat fake news, introducing this technology into the working environment and work processes often meets a wide variety of resistance. Hence, this paper investigates what barriers but also chances these communication experts identify from their professional point of view. For this purpose, we have conducted a quantitative study with more than 100 participants, including journalists, press officers, experts from different ministries, and scientists. We analyzed the results with a particular focus on the types of fake news and in which capacity they were encountered, the experts' general attitude towards artificial intelligence, as well as the perceived most pressing barriers concerning its use. The results are then discussed, and propositions are made concerning actions for the most pressing issues with a broad societal impact.

Keywords: Fake News · Artificial Intelligence · Media Forensic · Journalism · Social Media · Public Sector

1 Introduction

The advent of the Internet has fundamentally changed how information is spread and perceived within societies. Not only are the entrance barriers much lower than in classical media, but the speed at which information can be shared worldwide is unrivaled.

Published by Springer Nature Switzerland AG 2023
N. Edelmann et al. (Eds.): ePart 2023, LNCS 14153, pp. 67–81, 2023.
https://doi.org/10.1007/978-3-031-41617-0_5

The “post-factual” [1], also called “post-truth” [2], society is amid an “information war” and poses immense challenges for the media and the public sector within democracies [3, 4]. During the early stage of this revolution of interpersonal and mass communication via digital technologies, this paradigm change was perceived as a huge chance to reduce inequality by providing increased access to the public discourse and hence give a voice to virtually everybody, which in turn would ultimately support democracy within our societies [5, 6]. An assessment that still holds today. But the downside is that the easier access opens the door to disinformation from various (dangerous) sources [7]. In an age of innovation through knowledge for a sustainable, cohesive society [8], misinformation and fake news have a direct negative impact on public value creation through falsified or misleading information [9]. In this context, media [10] and public administrations [11] have a shared responsibility as gatekeepers to ensure the accuracy of public information. Due to this shared responsibility, we decided to focus our study on both parties from a combined point of view.

Following the argument of shared responsibility, journalists and the public sector are in a difficult situation. Trying to resolve misinformation and inform the public often results in the original misinformation being distributed even more intensively. This circumstance is partly due to the backfire effect [12]. This effect relates to potential cognitive biases within individuals and will cause feelings in cases the deepest beliefs or world views are “violated” by information that would contradict them. Consequently, the affected individuals will try to protect their beliefs even more vehemently and hence, render entirely the original intention of correction counterproductive. Also, studies have demonstrated that negative news is often more likely to be picked up and spread among the general public than positive news [13, 14].

Thus, the media and the public sector are in a problematic discrepancy between protecting free expression and disseminating information versus distorting democratic elections through massive disinformation campaigns. Moreover, in this tension range, they must deal with distrust and attacks often determined by prejudice, fear, and hate [15, 16]. This problematic situation is additionally pushed by social bots, which can massively spread whole global disinformation campaigns [17–19]. In addition, continuous development in artificial intelligence (AI), especially deep fakes, makes it increasingly challenging, even for experienced communication experts, to distinguish information from disinformation [20, 21].

But AI can also be a potent solution for identifying and fighting fake news. However, many barriers impede the implementation and use of tools by the leading media and the public sector to detect disinformation [20, 21]. To get a deeper understanding of those barriers, we conducted a quantitative survey with more than 100 experts from the field of leading media and the public sector, with a particular focus on the use of AI to fight disinformation.

The remainder of this paper is structured as follows: Sect. 2 provides a short discourse about state-of-the-art solutions using AI to combat fake news. In Sect. 3, we present the underlying methodology of this study and the collected data, including an overall profile of the participants. Section 4 then continues with the analysis of the results of the survey. After that, Sect. 5 discusses key learnings and practical implications. Section 6 then closes the paper with the conclusions and outlook for future work.

2 Related Work

A growing body of literature exists concerning technical solutions for using AI to combat fake news and misinformation. In this section, we provide a short discourse along the work of Shahid et al. [22] to inform the reader about state-of-the-art solutions currently used within available tools to the media and the public sector. Based on their analysis, current research streams can be separated into the following categories (ibid.):

- Automatic detection: the idea behind this approach is to extract features of fake news within deep learning models to be used for the automated classification of news items. Examples of this approach include the research of Ozbay and Alatas [22], who developed a solution to detect fake news in social media via a transfer process of unstructured data toward structured data, combined with a multi-algorithm analysis.
- Language-specific detection: this approach targets the development of a language-specific model beyond the limitation of English as the primary language. Studies that have used this approach, including the work of Faustini and Covões [23], build upon textual features and are not bound to a specific language, significantly increasing the overall usability, especially in an international context.
- Dataset-based detection: the main goal is to develop highly specialized datasets to test and challenge existing and newly developed algorithms. Examples include Neves et al. [24], who developed a method of removing fingerprints of algorithms (i.e., Generative Adversarial Networks) in face manipulation of images to challenge existing detection tools.
- Early detection: focuses on detecting fake news to limit its propagation at the earliest stage possible. Studies following this direction include Zhou et al. [25], who targeted the prevention of spreading fake news on social media via a supervised classification approach, building on social sciences and psychology theories.
- Stance detection: the idea behind this approach is not only to detect fake news but to deepen the underlying understanding of it. This is achieved by also including the stance of the reporting news outlets toward the reported event or incident. Research following this idea includes the work of Xu et al. [26], who integrated the reputational factors of news distributors, such as registration behavior, timing, ranking of domains, and their popularity.
- Feature-based detection: while this approach is similar to the automatic detection described before, it goes beyond classical textual features and includes topological and semantical features to improve the overall classification. Studies that have followed this idea include de Oliveira et al. [27], who incorporated stylistic information of social media posts, i.e., tweets, to improve the accuracy of fake news detection.
- Ensemble learning: the concept behind this approach is to use not one but a combination of multiple algorithms to identify and classify fake news. Examples of such combined approaches include Elhadad et al. [28], who addressed the issue of misleading information in the context of the COVID-19 pandemic, combining ten machine-learning algorithms with several feature extraction approaches.

3 Methodology and Data

In order to derive recommendations on how AI tools can be used for disinformation detection for leading media and the public sector, it is crucial to consider several factors, motivations, and potential barriers. These include challenges with implementation and the working environment, technological maturity, data protection, uncertainty about AI, and advancing technological progress in general. To address this challenging domain rigorously, a questionnaire was created during the applied research project *defalsif-AI (Detection of Disinformation via Artificial Intelligence)* aimed at communication experts. The questionnaire was created based on literature around dimensions of fake news, misinformation, and information disorder [5, 29, 30], with a particular focus on professionals and their perspectives on i) the types of media to be confronted with, ii) individual detection approaches, iii) types of fake news encountered, iv) attitudes toward AI technological progress, as well as v) experience on currently used AI tools in the respective working environments.

In the first step, this questionnaire was circulated among the consortium partners, and in the second step, a snowball-based system was applied to other related areas. This approach helped us to significantly increase the overall coverage of experts that deal with fake news within their professional environment.

These experts included journalists, press officers, experts from different ministries, and scientists dealing with the topic of fighting fake news and disinformation. Overall, we collected n = 106 completed surveys. Since the population in these sectors is unknown, it is seldom possible for expert surveys to be representative. Nevertheless, they allow profound assessments to be made of trends among professionals.

In addition to demographic data and questions about media genres and usage behavior, the questionnaire focused on the frequency and risk of dealing with fake news and misinformation in everyday work, intuitive and technological detection, research activities, and, last but not least, the desire for or possible rejection of AI-based software. The survey was conducted from May to June 2021.

49% of the respondents work in the media domain (journalists, press officers, PR professionals, etc.), while 31% in communication and security, including fields such as the police or the Ministry of the Interior. 7% of the respondents work in the field of diplomacy or the Ministry of Foreign Affairs, and 13% in the field of research.

Concerning the age distribution of our participants, about 85% of them resided within the mid-career and late-career levels.

Regarding professional experience, about 62% had more than ten years of professional experience. Hence, a high level of insight and proficiency is represented among the study participants.

4 Analysis and Results

4.1 Fake News and Misinformation Within Working Environments

Only 23.6% of the respondents see little or no threat to democracy in disinformation, and 76.4% of the respondents consider fake news to be a high or very high risk for democracy. In the context of AI-based media forensics, it is necessary to understand the medium through which experts often come into contact with disinformation (see Fig. 1).

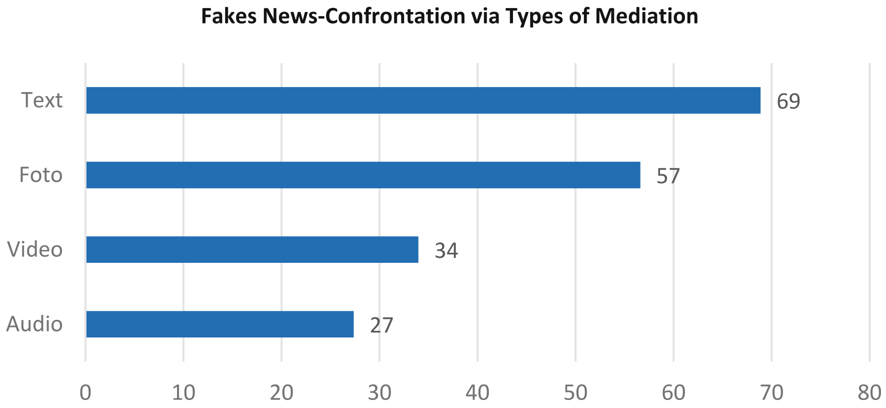

Fig. 1. Types of Mediation (n = 106); agreement high and very high (in %, n = 106; 6-point Likert scale; multiple answers possible).

Most subjects are confronted with disinformation via text, followed by manipulated photos often or very often. Text and photos are also the favored means of communication in traditional media, although video and audio are becoming increasingly popular, primarily through social media. In this context, this also raises the question of whether even experts can recognize manipulation, given the rapid technological development of deep fakes by video and audio. Studies also indicate that time of day, emotional state, fatigue, or age can significantly detect deepfakes [31, 32]. Concerning the odds of sharing misinformation, such as deep fakes, between individuals with a high interest in politics and those without, they later seem more prone to forwarding such misinformation [33]. In addition, personality traits such as optimism, especially for social media, can also play a role in classifying and spreading [34]. Ahmed points out that there is still limited knowledge about how social media users deal with this newer form of disinformation [33]. Our survey reveals a similar picture asking about the experts' strategies (see Fig. 2) in case of suspicion of fake news, and the following picture emerges.

Research whether and how other media report on it (78.5%), a critical look at the imprint of the medium (64,5%); checking the background of the author (54,2%), research how coherent contextual information such as geographic data, weather data, etc. are (39,3%), using fact-checking services like Mimikama, Correctiv, Hoaxsearch, etc. (28.9%), reverse image search on the Internet to check the actual origin of an image like Reverse Google Image Research, tineye.com Yandex (24.3%), and checking the metadata of an image (14,9%).

Since technology is advancing increasingly in mass manipulation, the results could indicate that training and AI-based tools will be increasingly necessary, especially for detecting deep fakes [35]. Especially since the sinister combination of manipulated videos and WhatsApp, e.g., in India [36], has already led to lynching mobs with innocent deaths, a change of modalities and further studies, training, and detection tools seem to be necessary for the context of security.

Continuing our analysis, we asked the participants to name the type(s) of misinformation and fake news most relevant to them in their daily business (see Fig. 3).

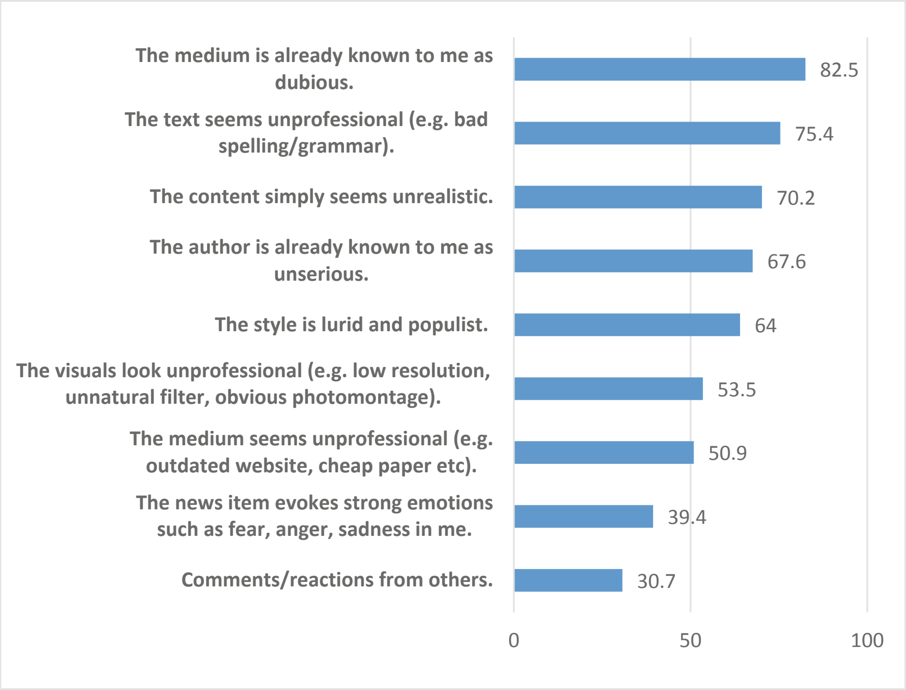

Fig. 2. Intuitive detection: Question: Based on which indications do you intuitively suspect whether it could be Fake News? (n = 106; agreement high and very high (in %, n = 106; 6-point Likert scale; multiple answers possible)).

The respondents stated that they were mainly involved in news fabrication in their professional life. Fabrication in this context implies that the generated news items are not based on facts. However, due to their style and presentation, they create the impression with readers that they are real. Similar to fabricated news is propaganda, usually originating from a political motivation to either praise or discredit an individual or entity. Examples of such approaches, besides others, can be found within official Russian news channels, deliberately using narratives to convey a particular image to their audience [3]. Similarly, tear-jerking misleading headlines were used to create click-bait and were frequently named by our respondents as a challenge they have to cope with within their own professional routine.

In the second place, however, are already photo and video manipulation. This observation is only, at first glance, contradictory to Fig. 1, in which photo and video manipulation are not classified as particularly frequent. It seems reasonable to assume that these manipulations are challenging to recognize precisely because of the technical know-how and effort; thus, the motives behind them must be exceptionally high. The respondents least frequently mentioned mouse-to-mouse propaganda, i.e., paid customer reviews, which are popular with large online retailers.

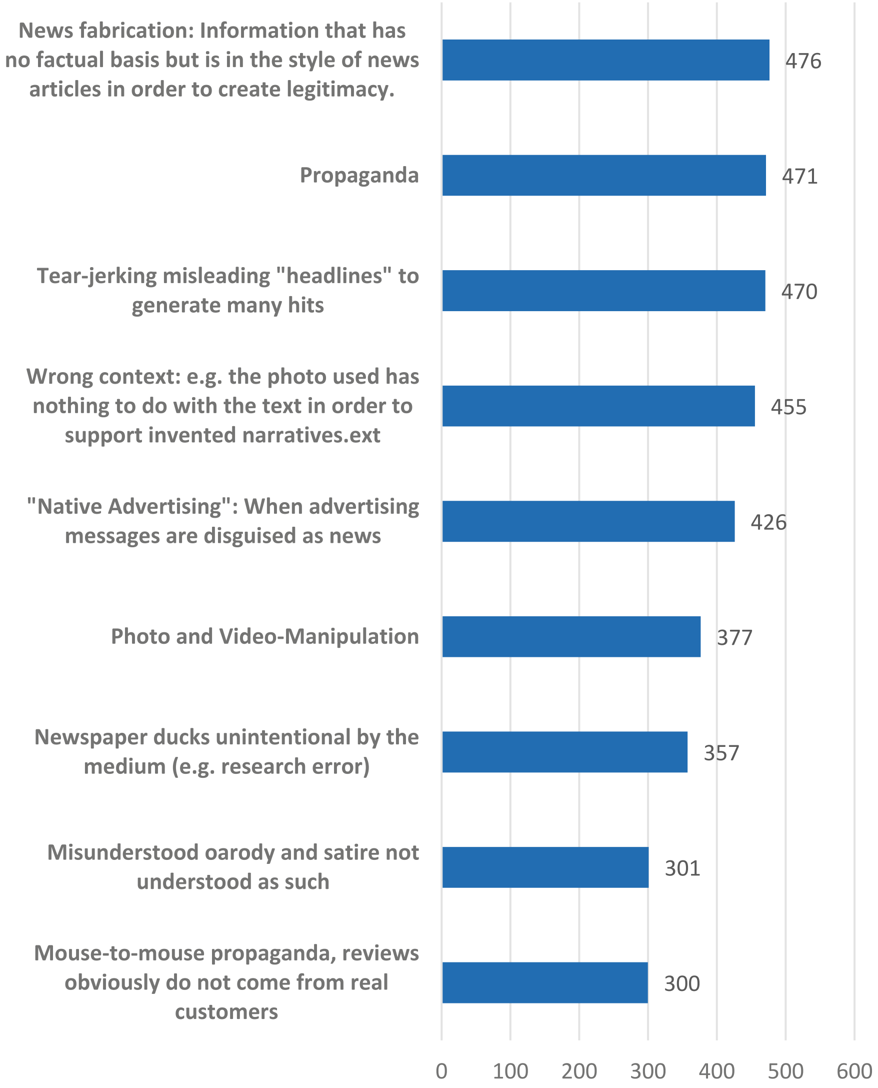

Fig. 3. Question: Which types of fake news are particularly relevant to you professionally? (n = 106; agreement high and very high (in %, n = 106; 6-point Likert scale; multiple answers possible)).

4.2 Barriers and Trust in AI

The application of technologies in the context of decision-making in the public sector always impacts the lives of citizens. Reasons for the introduction of these technologies often include cost-saving, increased efficiency, and improved ‘objectivity’ due to ‘fair’ algorithms [27]. Yet these technologies can also trigger unintended side-effects, which bear risks that are hard to foresee, measure, and thus be prepared for. When these risks come into force, the negative consequences affect the citizens and public administration [27].

In this tense field, it is decided if citizens gain trust due to better decisions or increase their distrust in government decision-making due to the perception of the underlying algorithms as 'black boxes, which in both ways will impact all aspects of daily life and social cohesion. Research has shown that the increased automation of decisions and centralization of those decisions will likely motivate distrust among citizens [29].

Especially in content filtering, e.g., to fight the spreading of fake news, the removal or restriction of content might be perceived as censorship [30]. Hence, it is essential to consider the ethical aspects of data-driven algorithms from the beginning of designing and implementing such systems.

Figure 4 shows the experts' attitudes to technological progress and AI within their professional environment. In general, the respondents see technological progress as more problematic than positive. The majority fear difficulties with data protection, ethical problems, and significant dangers such as cyber-attacks and blackouts. Effects on leisure time are viewed in a balanced way. The most positive expectation of 46.8% of respondents was new opportunities for creativity and innovation.

The impact of disruptive technologies on the work environment should not be underestimated. It is often the case that decisions to implement such technologies are made with little prior knowledge of possible limitations or potentials. This lack of knowledge, in turn, can directly impact the work itself, its results, and its quality, both positively and negatively [37].

It is crucial to understand the processes and activities of potential users to include them in technology development. Only in this way is there a possibility that the new technologies can cover the functions necessary for the users [37]. Grabowski et al. speak of a technology being used when it is accepted. This, in turn, is related to the trust in the technology, whether it can reliably fulfill the desired functions and means more efficient work [38].

The topics addressed, among other things, around the basic skepticism based on experience that new technologies do not necessarily mean a simplification of everyday work but can sometimes even lead to more work without a recognizable improvement in quality. However, it must be noted that the target groups are, by and large, technology-savvy and technology-friendly groups of people who rarely tend to be overwhelmed by new software solutions in this context.

Turning to our last part of the survey, we asked the participants to express their opinion concerning barriers to using a fake news detection software tool in their workplace (see Fig. 5).

In addition to a lack of application options, the respondents see unclear or non-transparent strategies, high time and cost expenditure, and a lack of customized solutions as obstacles to using software solutions for fake news detection. Lack of acceptance by the workforce and high demands on data protection and security is mentioned the least, but more than a third of the respondents still cite them as possible obstacles. Winning the acceptance of employees should therefore be considered in training courses.

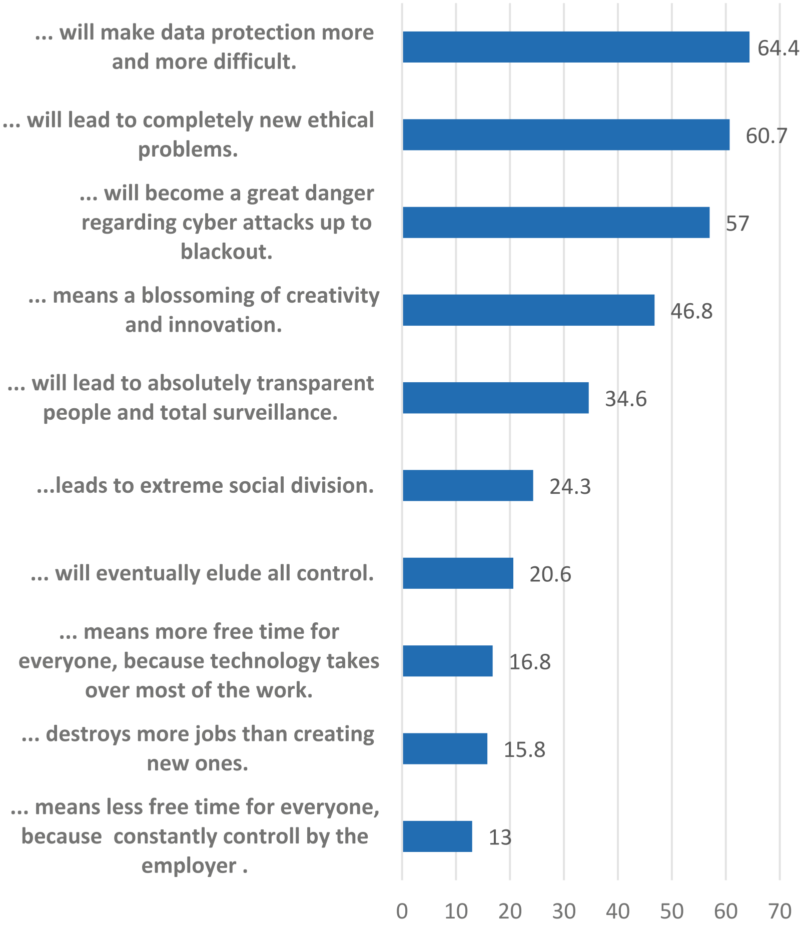

Fig. 4. Attitudes to technological progress and AI; agreement high and very high (in %, n = 106; 6-point Likert scale).

5 Lessons Learned and Propositions

The analysis of our survey has demonstrated the most pressing barriers that experts from the media and the public sector currently see in using AI to fight fake news and disinformation. Amongst the top-ranking results were: i) lack of trust in the technology, ii) in-transparent organizational strategies, and iii) ethical and privacy concerns. Hence, in the following, we provide some selected propositions and discussion points of lessons learned and what needs to be addressed to overcome the identified barriers.

In tools and data, we trust – attitudes towards AI as a 'Colleague'. Using AI to identify and communicate fake news to the general public is not without criticism, and trust in the technology is one of the key issues to ensure acceptance [39]. The literature shows that the same norms often come into play here as in interpersonal interaction [40]. In this context, it is also essential to consider that people tend to perceive AI as a

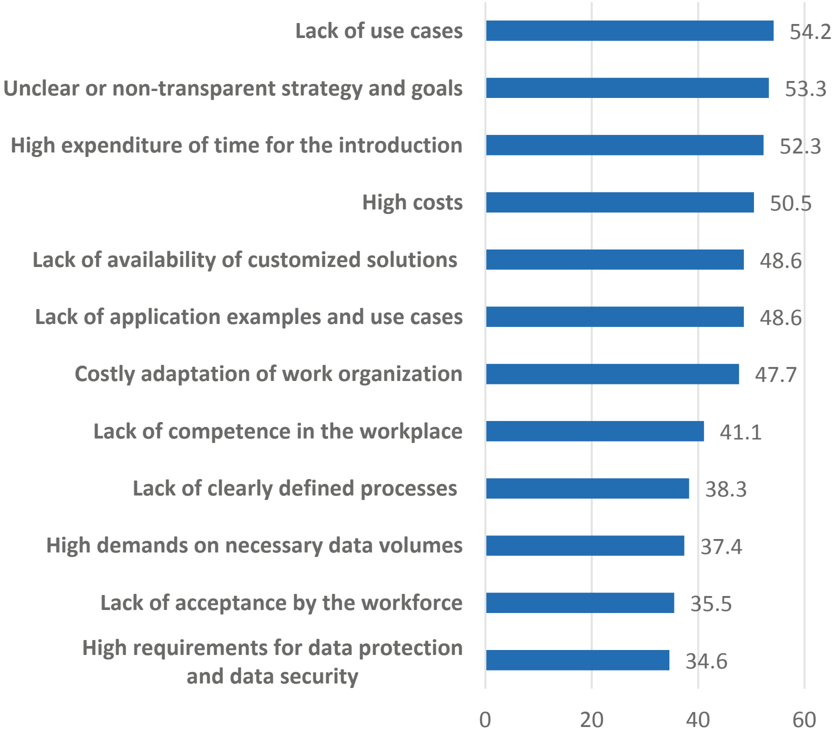

Fig. 5. What do you think would be barriers to using a fake news detection software tool in your workplace? (Agreement high and very high; in %, n = 106; 6-point Likert scale).

"counterpart" and not as a tool [41]. AI and its results must also be trustworthy in times of personal uncertainty [42]. In-depth research into the influence of perception and trust in the context of AI is, therefore, necessary [43].

I know it as well as the back of my hand – the importance of personal experience with AI. Many users have considerable reservations about AI-based fact-checking tools [44]. Overcoming these reservations is an open challenge due to such tools' increasing distribution and use [45]. The accuracy of the analysis results is not always the decisive aspect of whether users trust the tools [43]. The users' understanding of how to use the tools and how they work can have a lasting influence on their trust in the technology [46]. Personal experience in dealing with these tools [47] can also lead to realistic expectations of the tool itself [48] and, thus, to a more positive attitude toward AI [49]. It is, therefore, essential to define solutions that embed the presentation of results and the handling of the AI tool in the user's experience. If this succeeds, it could lead to greater self-reflection and a more critical approach to news and information through fact-checkers and evaluation tools [43].

Digital ethics – the importance of societal consensus and consent. Following the paradigm of digital humanism as a mindset of understanding the highly entangled and complex relationship of humans and technology [50], ultimately, technology should foster the free development of the individual to their full potential, but at the same time, not negatively impact others. This view also implies that tendencies towards anti-humanism through technology, e.g., artificial intelligence, should be identified and questioned [51]. This demand necessitates the fundamental need for ethical considerations embedded in all organizational processes. The essential question at hand: where to start? A plethora of frameworks is targeted at the ethical aspects of AI, where interested individuals can quickly lose oversight [52].

Furthermore, many of these frameworks are either on a high meta-level and thus hard to operationalize or on the opposite side, i.e., specific for a particular field or domain; hence, transferability is often limited [53]. Consequently, an approach needs to be selected that allows experts in communication to map common principles of digital ethics and the use of AI into their domain. Becker et al. have developed a three-step approach, i.e., analysis of principles, mapping the derived principles, and deriving an individual code of digital ethics [54]. Adopting this or similar frameworks can support communication experts in building their respective codes of conduct and guidelines for using AI. This adoption would ease internal barriers, as most refer to the missing knowledge and transparent and understandable guidelines.

6 Conclusions

Our study among the professionals has demonstrated that the situation is critical and that although AI can be a significant support within the daily work of communication experts, it is a blessing and a curse simultaneously. While the technology enables them to identify potentially fake news and misinformation, they struggle to communicate the results quickly and reach the necessary target audience. They are also facing fears and rejection concerning the use of AI by the general public. Censorship, violation of the free press, and intended overblocking are only a selection of accusations they are confronted with. This backlash leads to the build-up of internal barriers to adopting artificial intelligence within their organization. One of the biggest challenges comes in the lack of internal knowledge and capacity, which is also reflected in many follow-up barriers, such as fear of data privacy violation, mass surveillance, societal dived, or personal liability. What would be required is sophisticated training and proper adaptions to existing processes and work routines.

Consequently, this would lead to a deeper understanding of the underlying technology, its capabilities, and its limitations. In this context, the transparency of the use of algorithms and tools and the underlying decision process of these tools would be increased. Consequently, the responsible use would be strengthened, as well as the overall accountability for the application, interpretation, and dissemination of results. This overall increased knowledge would also become beneficial in terms of privacy protection while working with various sources of data and information.

For future work, several paths opened up based on our study results. The discussion around the regulation of AI within the EU is currently omnipresent. Thus, an examination of to what extent the handling of disinformation is regulated on a national level in the DACH countries and on an EU level (e.g., GDPR, Digital Services Act) or which initiatives exist in this regard in order to develop a well-founded recommendation for the future regulation of disinformation will be of interest. For a responsible approach to AI-based disinformation detection, the significance of the EU's AI Act is of high importance to the research community and the community of practitioners, and also, what consequences are to be drawn from a legal perspective. In addition to the provisions of the AI Act, national developments should also be considered to develop a framework for the legal, ethical, and transparent use of AI systems to detect disinformation. The aim is to shed light on the legal framework for designing AI systems to detect disinformation and to make recommendations based on a comprehensive consideration of the fundamental rights of the citizens affected.

Another interesting aspect for future research comes from the ever-increasing flood of disinformation, not least multiplied by bots, trolls, and generative AI, which raises concerns about the destabilization of society and a post-factual future. Technological development enables the massive increase of disinformation in quantity and quality while, at the same time, also providing solutions in the area of detection. However, paying particular attention to this ambivalent relationship to AI is vital, especially in the context of information dissemination in society. A representative survey of the Austrian population will empirically record the rejection, fears, and hopes regarding various aspects such as data protection, freedom of opinion, "overblocking" and transparency. From this data material, concrete recommendations for action are derived from promoting the acceptance of a broad population and taking ethics and diversity into special consideration.

Acknowledgments. The work described in this paper was funded in the context of the defalsif-AI project (FFG project number 879670, funded by the Austrian security research program KIRAS of the Federal Ministry of Agriculture, Regions, and Tourism BMLRT).

References

1. Hossová, M.: Fake news and disinformation: phenomenons of post-factual society. Media Literacy Acad. Res. **1**, 27–35 (2018)
2. Bybee, C.: Can democracy survive in the post-factual age?: A return to the Lippmann-Dewey debate about the politics of news. Journal. Commun. Monographs **1**, 28–66 (1999)
3. Khaldarova, I., Pantti, M.: Fake news: the narrative battle over the Ukrainian conflict. Journal. Pract. **10**, 891–901 (2016). https://doi.org/10.1080/17512786.2016.1163237
4. Seboeck, W., Biron, B., Lampoltshammer, T.J., Scheichenbauer, H., Tschohl, C., Seidl, L.: Disinformation and fake news. In: Masys, A.J. (ed.) Handbook of Security Science, pp. 1–22. Springer, Cham (2020). https://doi.org/10.1007/978-3-319-51761-2_3-1
5. Fraas, C., Klemm, M., Gesellschaft für Angewandte Linguistik (eds.) Mediendiskurse: Bestandsaufnahme und Perspektiven. P. Lang, Frankfurt am Main ; New York (2005)
6. Kriesi, H., Lavenex, S., Esser, F., Matthes, J., Bühlmann, M., Bochsler, D.: Democracy in the Age of Globalization and Mediatization. Palgrave Macmillan UK, London (2013). https://doi.org/10.1057/9781137299871
7. Bennett, W.L., Livingston, S.: The disinformation order: disruptive communication and the decline of democratic institutions. Eur. J. Commun. **33**, 122–139 (2018). https://doi.org/10.1177/0267323118760317

8. Carayannis, E.G., Barth, T.D., Campbell, D.F.: The Quintuple Helix innovation model: global warming as a challenge and driver for innovation. J. Innov. Entrepreneurship. **1**, 1–12 (2012)
9. Van Meter, H.J.: Revising the DIKW pyramid and the real relationship between data, information, knowledge, and wisdom. Law Technol. Hum. **2**, 69–80 (2020)
10. Guo, L.: China's "fake news" problem: exploring the spread of online rumors in the government-controlled news media. Digit. Journal. **8**, 992–1010 (2020)
11. Ninkov, I.: Separating truth from fiction: legal aspects of "fake news." Biztonságtudományi Szemle. **2**, 51–64 (2020)
12. Wood, T.J., Porter, E.: The elusive backfire effect: mass attitude' steadfast factual adherence. Polit. Behav. **41**, 135–163 (2019)
13. Huijstee, D., Vermeulen, I., Kerkhof, P., Droog, E.: Continued influence of misinformation in times of COVID-19. Int. J. Psychol. ijop.12805 (2021). https://doi.org/10.1002/ijop.12805
14. Jacobson, N.G., Thacker, I., Sinatra, G.M.: Here's hoping it's not just text structure: the role of emotions in knowledge revision and the backfire effect. Discourse Process. 1–23 (2021). https://doi.org/10.1080/0163853X.2021.1925059
15. Appel, M. (ed.): Die Psychologie des Postfaktischen: über Fake News, "Lügenpresse" Clickbait & Co. Springer, Heidelberg (2020). https://doi.org/10.1007/978-3-662-58695-2
16. Hagen, L.: Nachrichtenjournalismus in der Vertrauenskrise. "Lügenpresse" wissenschaftlich betrachtet: Journalismus zwischen Ressourcenkrise und entfesseltem Publikum. ComSoz. **48**, 152–163 (2015). https://doi.org/10.5771/0010-3497-2015-2-152
17. Hajli, N., Saeed, U., Tajvidi, M., Shirazi, F.: Social bots and the spread of disinformation in social media: the challenges of artificial intelligence. Brit. J. Manag. 1467–8551.12554 (2021). https://doi.org/10.1111/1467-8551.12554
18. Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A., Menczer, F.: The spread of fake news by social bots. **96**, 104. arXiv preprint arXiv:1707.07592 (2017)
19. Wang, P., Angarita, R., Renna, I.: Is this the era of misinformation yet: combining social bots and fake news to deceive the masses. Presented at the Companion Proceedings of the Web Conference 2018 (2018)
20. Zhang, T.: Deepfake generation and detection, a survey. Multimedia Tools Appl. **81**, 6259–6276 (2021). https://doi.org/10.1007/s11042-021-11733-y
21. Mirsky, Y., Lee, W.: The creation and detection of deepfakes: a survey. ACM Comput. Surv. **54**, 1–41 (2022). https://doi.org/10.1145/3425780
22. Ozbay, F.A., Alatas, B.: Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Stat. Mech. Appl. **540**, 123174 (2020)
23. Faustini, P.H.A., Covoes, T.F.: Fake news detection in multiple platforms and languages. Expert Syst. Appl. **158**, 113503 (2020)
24. Neves, J.C., Tolosana, R., Vera-Rodriguez, R., Lopes, V., Proença, H., Fierrez, J.: Ganprintr: improved fakes and evaluation of the state of the art in face manipulation detection. IEEE J. Sel. Top. Sig. Process. **14**, 1038–1048 (2020)
25. Zhou, X., Jain, A., Phoha, V.V., Zafarani, R.: Fake news early detection: a theory-driven model. Digit. Threats Res. Pract. **1**, 1–25 (2020)
26. Xu, K., Wang, F., Wang, H., Yang, B.: Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Sci. Technol. **25**, 20–27 (2019)
27. de Oliveira, N.R., Medeiros, D.S., Mattos, D.M.: A sensitive stylistic approach to identify fake news on social networking. IEEE Sig. Process. Lett. **27**, 1250–1254 (2020)
28. Elhadad, M.K., Li, K.F., Gebali, F.: Detecting misleading information on COVID-19. IEEE Access **8**, 165201–165215 (2020)
29. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. J. Econ. Perspect. **31**, 211–236 (2017). https://doi.org/10.1257/jep.31.2.211
30. Wardle, C., Derakhshan, H.: Information disorder: toward an interdisciplinary framework for research and policymaking. Council of Europe Strasbourg (2017)

31. Jung, T., Kim, S., Kim, K.: Deepvision: deepfakes detection using human eye blinking pattern. IEEE Access **8**, 83144–83154 (2020)
32. Müller, N.M., Pizzi, K., Williams, J.: Human perception of audio deepfakes. Presented at the Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia (2022)
33. Ahmed, S.: Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size. Telematics Inform. **57**, 101508 (2021)
34. Valenzuela, S., Halpern, D., Katz, J.E., Miranda, J.P.: The paradox of participation versus misinformation: social media, political engagement, and the spread of misinformation. Digit. Journal. **7**, 802–823 (2019). https://doi.org/10.1080/21670811.2019.1623701
35. Weerawardana, M., Fernando, T.: Deepfakes detection methods: a literature survey. In: 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), pp. 76–81 (2021). https://doi.org/10.1109/ICIAfS52090.2021.9606067
36. Sundar, S.S., Molina, M.D., Cho, E.: Seeing is believing: is video modality more powerful in spreading fake news via online messaging apps? J. Comput.-Mediat. Commun. **26**, 301–319 (2021). https://doi.org/10.1093/jcmc/zmab010
37. Pennathur, P.R., Bisantz, A.M., Fairbanks, R.J., Perry, S.J., Zwemer, F., Wears, R.L.: Assessing the impact of computerization on work practice: information technology in emergency departments. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 51, pp. 377–381 (2007). https://doi.org/10.1177/154193120705100448
38. Grabowski, M., Rowen, A., Rancy, J.-P.: Evaluation of wearable immersive augmented reality technology in safety-critical systems. Saf. Sci. **103**, 23–32 (2018). https://doi.org/10.1016/j.ssci.2017.11.013
39. Gillath, O., Ai, T., Branicky, M.S., Keshmiri, S., Davison, R.B., Spaulding, R.: Attachment and trust in artificial intelligence. Comput. Hum. Behav. **115**, 106607 (2021). https://doi.org/10.1016/j.chb.2020.106607
40. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. J. Soc. Isssues **56**, 81–103 (2000). https://doi.org/10.1111/0022-4537.00153
41. Seeber, I., et al.: Machines as teammates: a research agenda on AI in team collaboration. Inf. Manag. **57**, 103174 (2020). https://doi.org/10.1016/j.im.2019.103174
42. Okamura, K., Yamada, S.: Adaptive trust calibration for human-AI collaboration. PLoS ONE **15**, e0229132 (2020). https://doi.org/10.1371/journal.pone.0229132
43. Shin, J., Chan-Olmsted, S.: User perceptions and trust of explainable machine learning fake news detectors. Int. J. Commun. **17**, 23 (2022)
44. Brandtzaeg, P.B., Følstad, A.: Trust and distrust in online fact-checking services. Commun. ACM. **60**, 65–71 (2017). https://doi.org/10.1145/3122803
45. Zhou, X., Zafarani, R.: A survey of fake news: fundamental theories, detection methods, and opportunities. ACM Comput. Surv. **53**, 1–40 (2021). https://doi.org/10.1145/3395046
46. Siau, K., Wang, W.: Building trust in artificial intelligence, machine learning, and robotics. Cutter Bus. Technol. J. **31**, 47–53 (2018)
47. Mohseni, S., Zarei, N., Ragan, E.D.: A Multidisciplinary survey and framework for design and evaluation of explainable AI systems. ACM Trans. Interact. Intell. Syst. **11**, 1–45 (2021). https://doi.org/10.1145/3387166
48. Matthews, G., Lin, J., Panganiban, A.R., Long, M.D.: Individual differences in trust in autonomous robots: implications for transparency. IEEE Trans. Human-Mach. Syst. **50**, 234–244 (2020). https://doi.org/10.1109/THMS.2019.2947592
49. Araujo, T., Helberger, N., Kruikemeier, S., de Vreese, C.H.: In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc. **35**(3), 611–623 (2020). https://doi.org/10.1007/s00146-019-00931-w

50. Hofkirchner, W., Kreowski, H.-J.: Digital humanism: how to shape digitalisation in the age of global challenges? In: IS4SI 2021, p. 4. MDPI (2022). https://doi.org/10.3390/proceedings2022081004
51. Schmölz, A.: Die Conditio Humana im digitalen Zeitalter: Zur Grundlegung des Digitalen Humanismus und des Wiener Manifests. MedienPädagogik. 208–234 (2020). https://doi.org/10.21240/mpaed/00/2020.11.13.X
52. Floridi, L., Cowls, J.: A unified framework of five principles for AI in society. Harvard Data Sci. Rev. (2019). https://doi.org/10.1162/99608f92.8cd550d1
53. Hickok, M.: Lessons learned from AI ethics principles for future actions. AI Ethics **1**(1), 41–47 (2020). https://doi.org/10.1007/s43681-020-00008-1
54. Becker, S.J., Nemat, A.T., Lucas, S., Heinitz, R.M., Klevesath, M., Charton, J.E.: A code of digital ethics: laying the foundation for digital ethics in a science and technology company. AI Soc. (2022). https://doi.org/10.1007/s00146-021-01376-w