Automated Fact-Checking to Support Professional Practices: Systematic Literature Review and Meta-Analysis

LAURENCE DIERICKX¹
CARL-GUSTAV LINDÉN
ANDREAS LOTHE OPDAHL
University of Bergen, Norway

Fact-checking is a time-consuming process that automation can potentially make more efficient. This study provides a comprehensive, multidisciplinary state of the art that considers a holistic and sociotechnical approach to studying automated fact-checking (AFC) from a journalistic perspective. It identifies how AFC tools, as boundary objects, connect with their end users. The findings highlight that most research in AFC focuses on providing technological solutions to solve the multidimensional social problem of information disorder, and they also indicate the promise of human-machine teaming. This suggests that the cursor can be moved from a technological point of view toward a social one, provided that a relationship of confidence is established between the communities—developers and fact-checkers/journalists—involved from either side of the tool. Although the conditions of use of a technological artifact are multifaceted, the reliability of the results that it provides participates in building such a trust-based relationship.

Keywords: fact-checking, journalism, artificial intelligence, systematic literature review, meta-analysis

The prevalence of disinformation, misinformation, and propaganda online has been on the rise, accompanied by multiple political, societal, and cultural implications (Kalsnes, 2018). As such, automated fact-checking (AFC) has increasingly garnered interest in research over the past years (Guo et al., 2022). These different forms of information disorder are not new to the world’s history (Kalsnes, 2018; Posetti & Matthews, 2018), but online social interactions contribute to their large-scale accessibility and visibility, considering the central role played by social media and platforms (Guess et al., 2020).

Even though the verification of information has always been a part of editorial processes, it was popularized as a prominent subgenre in recent years (Singer, 2021), referring to “good journalism” (Singer,

Laurence Dierickx: l.dierickx@uib.no
Carl-Gustav Lindén: carl-gustav.linden@uib.no
Andreas Lothe Opdahl: andreas.opdahl@uib.no
Date submitted: 2023-01-25

¹ This research was funded by EU CEF Grant No. 2394203.

Copyright © 2023 (Laurence Dierickx, Carl-Gustav Lindén, and Andreas Lothe Opdahl). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at http://ijoc.org.
2018) or “unbiased journalism” (Graves, 2018). At the same time, the fact-checking community also encompasses political activists or fact-checkers acting on behalf of social progress (Mena, 2019). Although they share the same idea of defending and practicing a form of accountability reporting, these actors may view this ideal differently (Graves, 2018). Such heterogeneity is challenging for AFC tool designers, particularly in defining the requirements or needs of potential end users whose various organizational contexts and professional standards contribute to shaping their practices.

In this article, we define the end user as a fact-checker employed either in a fact-checking organization in a newsroom model or in a newsroom alongside or not alongside other journalistic activities. These professionals assimilate their identity and professional practices into journalism, considering that fact-checking is a genre of journalism (Cavaliere, 2020; Graves & Cherubini, 2016). Therefore, the professional values that frame their practices are linked to the traditional ethical values of journalism—accuracy, fairness, and objectivity (Frost, 2015)—that support the social responsibility of the journalists regarding the audiences to whom they are accountable (Bardoel & d’Haenens, 2004). These principles are strongly connected to the transparency defended in the ethical principles of the International Fact-Checkers Network (IFCN; Mena, 2019) and the European Fact-Checking Standards Network, which bring together fact-checkers to promote excellence in fact-checking.

Considering the speed at which misinformation and disinformation spread, AFC can be fast and effective tools not only to find claims worth fact-checking, relevant previously fact-checked claims, or supporting evidence but also to translate or summarize content (Nakov et al., 2021). The four stages of manual fact-checking are extracting statements, constructing appropriate questions, obtaining answers from relevant sources, and reaching a verdict using these answers (Vlachos & Riedel, 2014). Concretely, they encompass four primary tasks: monitoring media and capturing content, detecting claims, checking claims, and publishing content (Konstantinovskiy, Price, Babakar, & Zubiaga, 2021). These activities require the conjugation of critical thinking and know-how related to online investigative journalism and multimedia forensics.

The boundary object (BO) theory allows us to approach AFC tools as technological artifacts that exist between adjacent communities of designers and end users. According to this theory, as boundary objects, AFC tools “negotiate meaning to understand and articulate connections and disconnections between communities, cultures, and information infrastructures” (Huvila, Anderson, Jansen, Mckenzie, & Worrall, 2017, p. 1812). The BO theory can be seen as a valuable instrument to highlight the issues that researchers face when technology and journalism intertwine because it involves distinct occupational and social worlds that do not necessarily share the same values or visions of journalism (Lewis & Usher, 2016; Sirén-Heikel, Kjellman, & Lindén, 2022). From the point of view of the approach to the relationship between design and use (or between developers and fact-checkers), these issues are potentially linked to the work of interpreting what fact-checking is, that is, in terms of objectives or in terms of process. Therefore, it would involve more than an accurate understanding or transfer of knowledge from fact-checkers to designers because interdisciplinary interactions transcend differences in meaning between disciplines (Akkerman & Bakker, 2011; Fox, 2011; Trompette & Vinck, 2009).

This systematic literature review maps the field of AFC and its progress over the past five years to follow the recent developments in artificial intelligence (AI) in this field. It identifies how AFC tools, as
boundary objects, connect or disconnect adjacent communities on either side of the technological artifact. In other words, how do they adapt to journalism and infuse professional practices? The objective is to provide a comprehensive multidisciplinary state of the art that considers a holistic and sociotechnical approach to studying AFC from a journalistic perspective to nourish future works. Therefore, two research questions are investigated:

**RQ1:**  To what extent does research on AFC tools consider end users, and how?

**RQ2:**  As boundary objects, how do AFC tools connect and disconnect with the social world of journalism?

### Method

A systematic literature review is a means of collecting and synthesizing previous research, providing an overview of areas covered by the research, and demonstrating evidence on a meta-level (Snyder, 2019). Its main objective is to answer specific questions by relying on rigorous and explicit methods to ensure the work’s transparency, transferability, and replicability (Thomas & Harden, 2008). Although systematic literature reviews are commonly used in the medical and computer sciences, they can also be utilized in social sciences to provide an overall picture of the evidence in a topic area to guide future research (Petticrew & Robert, 2006, p. 21). Meta-analysis refers to statistical techniques to obtain overall estimations or a synthesis of published research works (Çoğaltay & Karadağ, 2015; Shelby & Vaske, 2008). In this study, the meta-analysis was mobilized through descriptive statistics to highlight the corpus characteristics and provide quantified evidence to complement the systematic literature review (Davis, Mengersen, Bennett, & Mazerolle, 2014; Mengist, Soromessa, & Legese, 2020).

From a methodological perspective, this systematic literature review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure transparent and complete reporting (Liberati et al., 2009). It consists of a checklist of 27 items that frame the method, the writing of a systematic review report, and a flow diagram that shapes information retrieval and selection (Page et al., 2021). The tool Parsifal, available online (Durier da Silva, Bicharra Garcia, & Matsui Siqueira, 2022), was used to define the research planning (objectives, research questions, keywords, search strings, sources, and selection criteria) as well as to identify the articles published in Scopus. The tool was also used to import all the collected data as spreadsheets containing the articles’ metadata to identify duplicate entries. The data collection was carried out via Google Scholar, Semantic Scholar, and Scopus between March 14 and 31, 2022. The search terms were approached through three complementary queries to refine the quality of the corpus and limit the phenomena of noise (linked to unexpected results) and silence (referring to the absence of expected results). This iterative process followed an inverted pyramid model, that is, going from the general to the specific, considering that Google Scholar was the most general database used and Scopus was the most specific:

- Fact-checking AND "machine learning" (Google Scholar)
- Automated AND fact-checking AND journalism (Semantic Scholar)
- ("machine learning" OR automated) AND fact-checking (Scopus)

These three queries returned 918 results. Inclusion and exclusion criteria were applied to refine the corpus selection. Published peer-reviewed articles; open-access articles published on arXiv, which are
published after moderation but are not peer-reviewed; book chapters; and proceedings related to AFC were included because the first objective of this review was to get a broad overview of the research works. Duplicate entries were removed along with all the articles that fell into the scope of the exclusion criteria: non-English texts, articles dated before 2017, undated articles, articles unrelated to AFC, and articles not available either in a PDF format or online. Dissertations were excluded from the review due to their extensive length, which is not comparable to research articles, and their limited accessibility and availability. Moreover, because their authors could publish their work in peer-reviewed journals, this exclusion criterion was also intended to limit the number of duplicate research. After the processes, the main corpus consisted of 267 articles used for a state of the art (Figure 1).

<table>
<thead>
<tr>
<th>Retrieving (918)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar: 205</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of paper + close and distant reading (267, main corpus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar: 101</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Users OR journalism (72, subcorpus 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar: 42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Users AND journalism (21, subcorpus 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar: 11</td>
</tr>
</tbody>
</table>

**Figure 1. Corpus selection process.**

To fit the scope of this study, the selection criteria were addressed using two complementary methods: (1) close reading of all of the collected abstracts and, in cases of uncertainty, additional close reading of the full texts, which concerned half of the collected corpus, and (2) distant reading of the abstracts through text mining and text analysis techniques using the programming language R (Ramage, Rosen, Chuang, Manning, & Mcfarland, 2009; Welbers, Van Atteveldt, & Benoit, 2017) and dedicated packages to proceed a meta-analysis and n-grams frequencies (tidyverse, tidytext, TM, quanteda, highcharter), topic modeling (LDA) and clustering (textmineR). This process allowed us to define a first subcorpus of 72 articles that considered users or journalism, as well as a second subcorpus of 21 articles that referred to users and journalism (Figure 1).

The R packages previously mentioned were also used for the meta-analysis of the main corpus. The meta-analysis combined a deductive and an inductive approach (Grimmer, Roberts, & Stewart, 2021; Molina & Garip, 2019) to support discoveries considering research questions that globally refer to the challenges of AFC in the social world of journalism. In addition, we have created a database including a unique identifier, the title of the article, the abstract, the field of research, the type of article, the year of publication, and the number of citations. We added two columns to identify articles related to users and journalism or newsrooms. The datasets and the source code of all these operations are available on GitHub ([https://github.com/laurence001/AFC_SLR](https://github.com/laurence001/AFC_SLR)).
The limitations of the research strategy are related to the level of accuracy provided by distant reading. Indeed, it is recognized that topic modeling is not suitable for advanced data relationships and performs poorly when documents do not have a sufficient length (Vayansky & Kumar, 2020). Clustering is also challenging for finding similarities between data points and grouping similar ones into the same cluster (Qaddoura, Faris, & Aljarah, 2020). Consequently, these results were primarily used to support human analysis. In addition, it is challenging to claim the completeness of a corpus queried through databases during a short period. Nevertheless, the examination of the references in the articles collected indicates the representativeness of the corpus on recent research works on AFC.

**Results**

The presentation of the results is divided into three parts: (1) a meta-analysis of the corpus to identify the state of the art of AFC technologies as well as the general challenges of AFC in journalism, (2) a review of the first subcorpus related to the uses and users of AFC tools to answer RQ1, and (3) an analysis of the second subcorpus identifying the challenges of developing AFC tools in a journalistic context to address RQ2. This three-level analysis aims to provide evidence to fuel the discussion about how research should consider end users more when designing AFC tools.

**Meta-Analysis and Challenges for Journalism**

Most of the articles collected were published in 2019, 2020, and 2021 (respectively 22.47%, 22.10%, and 29.96%). However, it does not mean that AFC gained particular interest over this period, considering the time required for the research and the time needed for reviewing and publishing, although preprint articles represent 20.52% of the main corpus. This corpus also includes 51.49% of articles, 46.64% of proceedings, and 1.87% of book chapters. Computer science was the main research area covered (80.90%), followed by social computing (9.74%), and the articles related to social science or journalism studies represented less than 7% of the corpus. It is not surprising that computer science is the most represented field given that the development of AFC tools involves using specific technologies such as machine learning, natural language processing, and knowledge graphs (Gallofré Ocaña & Opdahl, 2020; Lakshmanan, Simpson, & Thirumuruganathan, 2019). In addition, social computing and information science are closely related to computer science, and journalism studies and social science intersect.

AFC covers four main functionalities that assist and support human fact-checkers: finding claims, detecting already fact-checked claims, evidence retrieval, and verification (Nakov et al., 2021). Based on this typology, automated analysis of the abstracts of the main corpus demonstrated that the current state of the art in AFC technologies primarily covers claim detection (34.33%), followed by claim verification (14.18%) and evidence retrieval (8.20%). The technologies developed in AFC are most frequently oriented to machine learning (supervised and unsupervised), followed by natural language processing and knowledge graphs. Blockchain technologies are marginally represented (0.75%).

AFC technologies are essentially text focused. Only 14 articles were dedicated to images and/or videos (5.24%). They included datasets of images and/or videos (e.g., Papadopoulou, Zampoglou, Papadopoulos, & Kompatsiaris, 2019; Zlatkova, Nakov, & Koychev, 2019), supervised machine learning to
detect deceptive images or for image classification (Boididou et al., 2018; Reis & Benevenuto, 2021), the assessment of image forensics services (Katsaounidou, Gardikiotis, Tsipas, & Dimoulas, 2020; Nakov et al., 2021), social-computing solutions for user-generated content verification (Middleton, Papadopoulos, & Kompatsiaris, 2018), and deepfake detection (Hoque, Ferdous, Khan, & Tarkoma, 2021). Four articles focused on multimodal solutions to detect misleading texts and images (Abdelnabi, Hasan, & Fritz, 2021; Dhankar, Zaïane, & Bolduc, 2022; Gao, Hoffmann, Oikonomou, Kiskovski, & Bandhakavi, 2022; Yang et al., 2018). One article was dedicated to verifying audio content associated with video (Vryzas, Katsaounidou, Vrysis, Kotsakis, & Dimoulas, 2022). The full-text search results indicated that 129 articles were about experimental-stage systems (35%), and 179 articles were about frameworks (48.5%). In addition, the authors of 45 articles provide online the source code of their fact-checking system (60%) and/or the datasets they built (62.2%) through a GitHub project page.

Based on the abstracts’ content, datasets are the most prominent topic because it was found in 40% of the corpus. This is not surprising because AFC tools rely on data for classification and prediction purposes: it is how machine-learning works. Several articles focused more specifically on creating training datasets that are specific to one particular news context (e.g., on the Syrian war, U.S. politics, or COVID-19). They are mainly published in English, except for one multilingual article and three others in Spanish, Arabic, and Czech. They rely on several different classification systems such as “True,” “False,” or “Half-true” (Wang, 2017); “Unproven” (Kotonya & Toni, 2020); and “Contradiction,” “Compatible,” or “Unrelated” (Sepúlveda-Torres, Bonet-Jover, & Saquete, 2021). They also encompass variable amounts of data; for instance, the FEVER dataset—which consists of 185,445 claims generated by altering sentences extracted from Wikipedia—comprises 185,000 rows (Thorne, Vlachos, Christodoulopoulos, & Mittal, 2018), whereas AraStance has more than 4,000 rows. It consists of a multicountry and multidomain dataset of Arabic stance detection for fact-checking, based on claim-article pairs from a diverse set of sources comprising three fact-checking websites and one news website (Alhindi, Alabdulkarim, Alshehri, Abdul-Mageed, & Nakov, 2021).

The performance of an AFC system can also be tackled through the detection of overfitting, which occurs when the model is trained for too long and, as a result, comes to reflect the specifics of the training data rather than
the general characteristics of the underlying domain. Consequently, the model fails to find a general predictive rule (Dietterich, 1995). In this corpus, the overfitting of the machine-learning model was pointed out in 24 articles. This does not mean that the research has failed, but it illustrates the challenges in developing efficient and scalable AFC tools.

As boundary objects, AFC systems represent a clear division between their designers and fact-checkers. This finding is confirmed in the second part of the analysis, which focuses on the relationship between AFC systems and the social world of journalism. However, some attempts to connect technologies to work environments and practices have also been observed, particularly when considering keeping the human fact-checker in the loop.

**Uses and End Users of AFC Systems**

The proportion of articles considering end users and/or journalism is relatively low, representing 20.22% and 14.61% of the corpus, respectively. This highlights that attempts by developers in AFC are driven by the goal to provide a technical solution to the social problem of information disorder. According to this approach, the growing importance of information disorder creates a need for new tools to speed up the process of detecting and verifying claims (e.g., Azevedo, 2018; Çarik & Yeniterzi, 2021; Du, Bosselut, & Manning, 2022; Kar, 2020). However, we found 72 articles directly or indirectly related to the sociotechnical dimension of AFC technologies, that is, 26% of the main corpus, comprising the first subcorpus devoted to the uses and end users.

Several articles highlighted the time-consuming aspects of human fact-checking activities and the challenge that human fact-checkers cannot keep up with the amount of misinformation and the speed at which it spreads (e.g., Adair, Li, Yang, & Yu, 2017; Gencheva, Koychev, Marquez, Barron-Cedeño, & Nakov, 2019; Hanselowski & Gurevych, 2017; Jiang, Baumgartner, Ittycheriah, & Yu, 2020). Scholars also emphasized that human fact-checking requires expertise (Jaroucheh, Alissa, Buchanan, & Liu, 2020) and, in a certain way, human intuition and creativity that can't be automated (Nakov et al., 2021). Furthermore, the relationship between AFC tools and human users remains problematic (Borges, Martins, & Calado, 2019) due to limitations related to credibility issues for automated systems and scalability issues for human fact-checking (Nakov et al., 2021). When researchers consider uses or end users, the focus is mostly on technological solutions to speed up the AFC process or assist human fact-checkers (e.g., Brand, Roitero, Soprano, & Demartini, 2021; Majithia et al., 2019). However, the fact-checkers toolbox is also explored (Bañon Castellón, 2021; Nygren, Guath, Axelsson, & Frau-Meigs, 2021; Svahn & Perfumi, 2021), compared (Školkay & Filin, 2019), classified (Nakov et al., 2021), and evaluated (Komendantova et al., 2021; Picha Edwardsson, Al-Saqaf, & Nygren, 2021). Another perspective is related to the benefits of human-computer interactions (Miranda et al., 2019; Shi, Bhattacharya, Das, Lease, & Gwizdka, 2022; Yang et al., 2019). These approaches can be understood in terms of challenges and opportunities that concern both users and developers (e.g., Demartini, Mizzaro, & Spina, 2020; Gallofré Ocaña & Opdahl, 2020). They encourage a hybrid process where humans are augmented by the use of AI, broadening a technical solution into a sociotechnical one, which is likely to result in the reconfiguration of professional practices (Diakopoulos, Trielli, & Lee, 2021). This concerns not only the AFC system’s functionalities but also how the tool was
designed, considering the users’ requirements (Nguyễn et al., 2018) or the transparency of the model in facilitating human interactions (Rony, Hoque, & Hassan, 2020).

Making an AFC system transparent depends on its explicability. Transparency concerns both the upstream components of the tool (Reis, Correia, Murai, Veloso, & Benevenuto, 2019)—for example, the explanation of the model and the information sources used (Gencheva et al., 2019) or why the expert assigned that label to a training dataset (Berendt et al., 2021)—its interface (Katsaounidou et al., 2020); and the results it provides (Denaux & Gomez-Perez, 2020; Middleton et al., 2018). Therefore, the use of AI tools is first and foremost a matter of trust (Demartini et al., 2020). Providing explanations not only helps humans perform fact-checking but also can foster trust toward the tool or improve AFC systems (Gad-Elrab, Stepanova, Urbani, & Weikum, 2019). However, this can be difficult when the system relies on black-box methods, such as deep neural networks (Saeed & Papotti, 2021).

Contextual information about the application domain is another topic that has characterized this first subcorpus. It implies some shared expertise between coders and end users (Berendt et al., 2021). Providing context with the results contributes to a better understanding of a claim (Vizoso, Vaz-Álvarez, & López-García, 2021; Vryzas et al., 2022) and suggests trustworthiness in the tool and its results (Middleton et al., 2018). In addition, the lack of knowledge of the application domain may interfere with the correctness and the quality of a training dataset’s annotations (Singh, Das, Li, & Lease, 2021). It was also pinpointed that human fact-checking also requires human judgment and sensitivity (Graves, 2018). Therefore, research in AFC also considers context-based approaches (Boididou et al., 2018; Fairbanks, Fitch, Bradfield, & Briscoe, 2020; Gencheva et al., 2019), which can be interpreted as another form of an attempt to link boundary objects with the knowledge and practices of a given professional community (Trompette & Vinck, 2009).

**AFC and Journalism**

The analysis of the second subcorpus of 21 articles (7.86% of the main corpus) demonstrated that the most prominent perspective is about developing tools (42.85%) or supporting journalistic practices (52.38%), mainly for claims or stance detection (33.33%). Nearly half of the articles (42.86%) presented prototypes (e.g., Adair et al., 2017; Horne, Dron, Khedr, & Adali, 2018), of which five consisted of experiments (Adler & Boscaini-Gilroy, 2019; Boididou et al., 2018; Jiang et al., 2020; Masood & Aker, 2018; Vryzas et al., 2022). In addition, one article focused on fact-checkers’ needs by exploring the tools they use and how AFC tools can support them (Nakov et al., 2021), and four articles—all related to journalism studies—were dedicated to journalism practices within newsrooms (Bañón Castellón, 2021; de Haan, Van Den Berg, Goutier, Kruikemeier, & Lecheler, 2022; Diakopoulos, 2020; Picha Edwardsson et al., 2021).

The world of journalism is less considered in research on AFC. However, when it is considered, the complementarity between the journalist and the tool is the most frequently underlined aspect. That is mainly because fact-checking is a time-consuming process that still requires human input either to assess the validity of a claim or to adapt to dynamic news contexts by extending the daily collection of the source, for instance, as news information collection is a continuous process that occurs in a moving reality (Berendt et al., 2021; Hassan et al., 2017). Several articles also stressed that human-machine complementarity is still
needed because AFC systems are not strong enough to give rise to fully automated solutions (Komendantova et al., 2021).

The question of the fact-checkers’ user needs was tackled either through a sociotechnical prism (Diakopoulos et al., 2021) or by considering the cultural background and attitudes of journalists who may be skeptical toward technology and who generally lack algorithmic culture (de Haan et al., 2022). Still, from a journalistic point of view, other challenges are related to the “lack of collaboration between researchers and practitioners in terms of defining tasks and developing datasets for automated systems” (Nakov et al., 2021, p. 5), the lack of time to learn how to use digital tools (Picha Edwardsson et al., 2021), the lack of consideration of the specific needs that can be expressed by fact-checkers, and the lack of transparency of some fact-checking services (Komendantova et al., 2021). Although addressing some of these issues might be seen as a fruitful path to better intersect AFC systems’ users and designers, all cannot be technically solved, mainly because they also rely on social and organizational variables that involve human decisions.

Designing an AFC system is also challenging because of the complexity of defining central concepts such as “claim” and “verification” (e.g., Konstantinovskiy et al., 2021). Other issues can also be related to the limited availability of training datasets or to the overall quality of some existing ones, including the aspects related to the quality of their annotations (e.g., de Haan et al., 2022). Systems must also be adapted to the complexity of the newsroom’s workflows, where the multiplication of information channels is required to speed up the processes and provide newsworthy insights (Gallofré Ocaña & Opdahl, 2020).

**Discussion and Conclusion**

Regarding RQ1, related to the integration of end-user perspectives, the findings indicated that most research focuses on providing technological solutions to a social problem without embedding end users’ views or needs. As machine-learning systems rely on training data, quality issues appear crucial, specifically when annotations are based on crowdsourcing by nonexperts. Annotations are inherently prone to errors even when control procedures are set for correction (Northcutt, Athalye, & Mueller, 2021). They are also language, domain, and context dependent, making them less adaptive or reusable. In addition, the datasets used for machine learning have various sizes, rely on binary classes for the labeling, and tend to be unbalanced, which is not ideal for training (Zeng, Abumansour, & Zubiaga, 2021).

Because they are likely to rely on various annotation process types, the question is also that of developing a standardized annotation system (Zeng et al., 2021). Such diversity also complicates both comparing approaches and gauging their improvement over time. At the same time, there is no consensus on the best classification strategies and sets of features for AFC tools developed for detection (Silva, Santos, Almeida, & Pardo, 2020). Nonetheless, researchers have underlined that the datasets developed for natural language processing purposes are more valuable, cover different domains, and help progress research in automated claim validation (Zeng et al., 2021). Furthermore, automating detection and selection processes is also challenging because the criteria are likely to vary from one fact-checking organization to another, although common patterns are observed, such as prominence within the opinion or debate, virality, and measures of social engagement (Micallef, Armacost, Memon, & Patil, 2022).
Working with data from datasets based on Wikipedia raises the same risk of potential quality issues as any other user-generated content in terms of accuracy or reliability because there are no guarantees about the users’ expertise. In this regard, quality control of the datasets and their maintenance over time also appeared as two other obstacles to efficiency, and data quality has “a huge impact on the efficiency, accuracy and complexity of machine learning tasks” (Gupta et al., 2021, p. 1). In addition, the use of crowd workers raises “practical and ethical issues, such as funding and remuneration” (Berendt et al., 2021, p. 11). Although tremendous efforts are dedicated to gathering and annotating datasets for machine-learning tasks, the lack of relevant ones could be a constraint on developing AFC tools (Pathak & Srihari, 2019).

At the same time, the lack of vision about the fact-checkers’ needs and requirements appears problematic as several fact-checking tools are developed without considering them (Komendantova et al., 2021). The explicability of the system contributes to building a relationship of confidence between the user and the tool, and it plays a central role in using AI tools (Stray, 2019). In a context where fact-checkers vigorously promote values of transparency, being able to explain algorithms at work and the results they provide Katsaounidou et al., 2020) can be considered a lever for trust, or an instrument to demystify tools that might be considered black boxes (e.g., Bartneck, Lutge, Wagner, & Welsh, 2021; de Haan et al., 2022; Zhou, Hu, Li, Yu, & Chen, 2019).

From the BO theory perspective addressed in RQ2, AFC tools might exist between the adjacent communities of researchers and fact-checkers by involving the latter in the process when it comes to assessing the performance and the usability of the tool, for instance because this is likely to help improve it (Miranda et al., 2019). This implies a transfer of knowledge from one professional community to another, although the computational process results from a work of interpretation (Huvila et al., 2017). The social aspects of fact-checking also play a role in the adoption of the technological artifact, including in terms of its perceived advantages (Fox, 2011). Specific training in identifying manipulated images and deep fakes, along with the ease of use and integration of AFC tools into editorial systems, are additional outcomes from journalism studies to enhance the integration of AFC tools in newsrooms (Katsaounidou et al., 2020; Picha Edwardsson et al., 2021; Vizoso et al., 2021).

AFC tools are helpful to detect falsehoods, but they do not eliminate the need for human intervention (Bañón Castellón, 2021). Moreover, AFC tools should meet the users’ trust (Nakov et al., 2021), which can be achieved only if both the system and its outcomes are trustworthy too. When professional fact-checkers use their human expertise and take charge of the labeling of training datasets for machine-learning-based systems (Berendt et al., 2021), this “human in the loop” perspective is a means to achieve connection between the communities surrounding the BO. Integrating the users’ knowledge also improves a system’s transparency and enhances its trustworthiness (Nguyen et al., 2018) because it also emphasizes on the social components of technology. Ultimately, “automated fact-checking works well in some cases,” but it “still needs improvement prior to widespread use” (Lazarski, Al-Khassaweneh, & Scotts Howard, 2021, p. 1).

Reconnecting adjacent professional communities from either part of the BO may also be achieved by paying acute attention to the users’ beliefs and the possibility “to infuse their views and knowledge into the system” (Shi et al., 2022, p. 315). Furthermore, the necessary compatibility of a technological artifact with journalistic ideals and values was also stressed in the context of the diffusion of news automation within
newsrooms (Diakopoulos, 2019). However, in journalism, psychological and cultural barriers challenge the effectiveness of AFC tools, although technology can make fact-checking easier and faster (Cazalens, Lamarre, Leblay, Manolescu, & Tannier, 2018). We can draw a parallel with previous qualitative research on the possibilities of AI technologies within newsrooms where human journalists intend to keep the lead in these new forms of human-machine collaborations (Dierickx, 2020; Gutierrez Lopez et al., 2022).

Effective AFC tools require efficient technology. Nevertheless, no technology can be used sufficiently on its own, despite various technical challenges. Therefore, (re)connecting communities means responding to both technological and social issues that arise upstream and downstream of the automated process. Our observations provide useful reflections for further research in AFC, whether working on improving data quality, models for standardizing annotations, the explainability of processes and results, or improving human-computer interaction and the overall user experience to better support professional practices. Drawing links between the adjacent communities surrounding the BO also implies that it connects research fields because AFC involves different but complementary scientific disciplines.

References


