Automated Fact-Checking for Assisting Human Fact-Checkers

Automated Fact-Checking for Assisting Human Fact-Checkers

Preslav Nakov[1*], **David Corney**[2], **Maram Hasanain**[3], **Firoj Alam**[1], **Tamer Elsayed**[3], **Alberto Barrón-Cedeño**[4], **Paolo Papotti**[5], **Shaden Shaar**[1], **Giovanni Da San Martino**[6]

[1]Qatar Computing Research Institute, HBKU, Qatar, [2]Full Fact, UK, [3]Qatar University, Qatar,
[4]Università di Bologna, Italy, [5]EURECOM, France, [6]University of Padova, Italy
{pnakov, fialam, sshaar}@hbku.edu.qa, david.corney@fullfact.org, a.barron@unibo.it,
{maram.hasanain, telsayed}@qu.edu.qa, papotti@eurecom.fr, dasan@math.unipd.it

Abstract

The reporting and analysis of current events around the globe has expanded from professional, editor-lead journalism all the way to citizen journalism. Politicians and other key players enjoy direct access to their audiences through social media, bypassing the filters of official cables or traditional media. However, the multiple advantages of free speech and direct communication are dimmed by the misuse of the media to spread inaccurate or misleading claims. These phenomena have led to the modern incarnation of the *fact-checker* — a professional whose main aim is to examine claims using available evidence to assess their veracity. As in other text forensics tasks, the amount of information available makes the work of the fact-checker more difficult. With this in mind, starting from the perspective of the professional fact-checker, we survey the available intelligent technologies that can support the human expert in the different steps of her fact-checking endeavor. These include identifying claims worth fact-checking; detecting relevant previously fact-checked claims; retrieving relevant evidence to fact-check a claim; and actually verifying a claim. In each case, we pay attention to the challenges in future work and the potential impact on real-world fact-checking.

1 Introduction

The spread of fake news and misinformation on the web and in social media has become an urgent social and political issue. In the web sphere, social media have been widely used not only for social good but also to mislead entire communities. To fight against such bad information there has been initiatives for manual and automated fact-checking. Notable fact-checking organizations include FactCheck.org,[1] Snopes,[2] PolitiFact,[3] and FullFact.[4]

Such organizations are also potential beneficiaries of or leaders in automated fact-checking research. As misinformation became a major concern globally, tech companies, national and international agencies began work in the area. Recently, several international initiatives have also emerged such as the *Credibility Coalition*[5] and *EUfactcheck*,[6]. Along side some tools have also been made available such as Google Factcheck[7] and Hoaxy[8]. Moreover, fact-checking is a common task in settings that go beyond online misinformation, as the verification of content accuracy is a priority for many organizations [Karagiannis *et al.*, 2020].

A large body of research is devoted to developing automatic systems for fact-checking [Vo and Lee, 2018; Shu *et al.*, 2017; Thorne and Vlachos, 2018; Li *et al.*, 2016; Lazer *et al.*, 2018; Vosoughi *et al.*, 2018]. The studies include the development of datasets [Hassan *et al.*, 2015; Augenstein *et al.*, 2019], systems, and evaluation campaigns [Barrón-Cedeño *et al.*, 2020b]. However, there are credibility issues with automated systems [Arnold, 2020]. Hence, a reasonable solution (i.e., human in the loop) is to facilitate human fact-checkers using the automated systems. Towards this direction, there has been limited work. The work on identifying previously fact checked claims is one such example [Shaar *et al.*, 2020].

To facilitate human fact-checkers, in this study we explore what fact-checkers want and what research has been done that can actually support them in their work. This is important because manual fact-checking is a time-consuming process, going through several manual steps. The study by [Vlachos and Riedel, 2014] describes the following typical sequence of fact-checking steps: *(i)* extracting statements that are to be fact-checked, *(ii)* constructing appropriate questions, *(iii)* obtaining the pieces of evidence from relevant sources, and *(iv)* reaching a verdict using that evidence. Typically, this process takes several hours or days, and by that time misleading statements may have spread out of control.

In the current information ecosystem (including web and social media), there is a large volume of false claims not only in textual form, but also misleading or manipulated images

[*] Contact Author

[1] http://www.factcheck.org/

[2] http://www.snopes.com/fact-check/

[3] http://www.politifact.com/

[4] http://fullfact.org/

[5] https://credibilitycoalition.org/

[6] https://eufactcheck.eu/

[7] https://toolbox.google.com/factcheck/explorer

[8] https://hoaxy.osome.iu.edu/

and videos, including “deepfakes”. In order to detect such content, there has been significant recent work. However in this study, we limit our focus to automated fact-checking on text, as this remains the focus of most professional fact-checkers.

In the literature, there have been a number of surveys on “fake news” [Zhou and Zafarani, 2020; Cardoso Durier da Silva *et al.*, 2019], misinformation [Islam *et al.*, 2020], fact-checking [Kotonya and Toni, 2020], truth discovery [Li *et al.*, 2016], and propaganda detection [Martino *et al.*, 2020]. Unlike that work, here we adopt a different point of view: we start with the desiderata of fact-checkers and then survey the research attempts that aim to meet them.

2 What Fact-Checkers Want

Recently, Full Fact carried out extensive interviews with professional fact-checkers from 24 organizations serving around 50 countries [Arnold, 2020]. The report discussed some key challenges they face where they believe technology can help. These include monitoring potentially harmful content, selecting claims to check, creating and distributing articles, and managing tips and suggestions from readers (such as tip lines serving WhatsApp or Signal).

The same report revealed that most fact-checkers do *not* believe that tools to automate the verification of facts, i.e., the last step of a typical fact-checking pipeline [Vlachos and Riedel, 2014], will be used in the foreseeable future. Some believe that the required intuition and creativity can never be automated, even if some parts of their work can be supported or automated.

This sets up a twin challenge for AI practitioners working in verification: *first*, to develop practical tools that solve the problems fact-checkers face; and *second* to demonstrate their value to fact-checkers in their day-to-day work. In the meantime, there is a recognised need for tools to help with finding claims, including previously fact-checked claims, and in finding relevant evidence to help write fact-checking articles.

2.1 Finding Claims Worth Fact-Checking

Choosing which claims to check is a complex process. Fact-checking takes time and it often takes effort to determine if a claim can even be checked, let alone whether it is misleading. Fact-checkers have to balance the potential harm that a misleading claim may cause (including risk to health; risk to democratic processes; and the risk of exacerbating emergency situations) against the effort required to check a claim, if it can be checked at all. In many countries, governments choose not to publish reliable official statistics, making related claims impossible to verify.

While simple algorithms can often decide whether content is viral, it is much harder to estimate the “checkworthiness” of a claim. For example, breaking news stories are often both popular and accurate. Given the limited resources of fact-checking organizations, many claims that are check-worthy nonetheless remain unchecked; thus using historic lists of claims that were or were not checked is *not* a reliable indication of whether similar claims are worth checking.

Claims may be found in many sources, including news websites, social media (text, audio, or video), and broadcast media. To monitor such a range of sources, fact-checkers often use a variety of technologies, such as news alerts, automatic speech recognition and translation tools, all of which typically depend on underlying AI technologies.

2.2 Detecting Previously Fact-Checked Claims

Misleading claims are often repeated in multiple channels, independent of any fact-checks or rebuttals.[9] Once a claim has been established as misleading, the ongoing spread of repeats or copies of the claim can be minimised by their rapid detection. In the simplest cases, the repeats may be simple “copy and paste” repeats that are relatively easy to detect, but more often they will be paraphrases of the original or endlessly evolving variations. Given the resources required to write fact-check articles, it is preferable to respond to multiple repeats of a claim with a single fact-checking article.

The number of fact-checking initiatives continues to grow. The Duke Reporters’ Lab lists 305 active fact-checking organizations.[10] While some organizations debunked just a couple of hundred claims, others such as Politifact, FactCheck.org, Snopes, and Full Fact have each fact-checked thousands or even tens of thousands of claims.

Also, manual fact-checking often comes too late. It has been shown that “fake news” spreads six times faster than real news [Vosoughi *et al.*, 2018], and that over half of the spread of some viral claims happens within the first ten minutes of their posting on social media [Zaman *et al.*, 2014]. To counter this, if detecting that a new viral claim has already been fact-checked can be done automatically and quickly, it allows for a timely action that can limit the spread and the potential harmful impact.

For a journalist, the ability to discover quickly whether a claim has been previously checked could be revolutionizing as it would allow them to put politicians on the spot during live events. In such a scenario, automatic fact-checking would be of limited utility as, given the current state of technology, it does not offer enough credibility in the eyes of a journalist.

Also, false claims often get made in one language but then translated to multiple other languages. Tools that can spot repeated claims across languages would be useful to address this. More generally, multi-lingual tools can help fact-checkers around the world, even those with limited resources.

2.3 Evidence Retrieval

The process of fact-checking is often limited by the time available: there are far more claims to be checked than there is time to check them. Even if the fully automatic verification of claims remains out of reach (see the next section), tools that support fact-checkers in their manual verification process are to be welcomed.

Tools that automatically retrieve relevant data from trusted sources may save the fact-checkers time. This is especially true if the evidence is contained in large text documents, audio-visual recordings or streams, or is in a language the

[9] President Trump repeated one claim over 80 times: http://tinyurl.com/yblcb5q5.

[10] http://reporterslab.org/fact-checking/

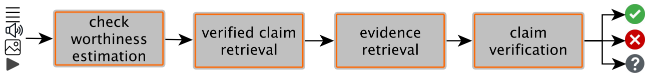

Figure 1: A fact-checking pipeline.

fact-checker is not familiar with. Combining automatic transcription, summarization, translation, and search can make sources of evidence available to fact-checkers that would be impossible or impractical to access otherwise.

2.4 Automated Verification

On first consideration, the automated verification of claims seems like the ultimate application of AI to fact-checking. And where such technologies can be developed and deployed, they will allow fact-checking organizations to be faster and to provide a more comprehensive coverage than manual checking could ever achieve. However, many claims are not simply 'correct' or 'incorrect', but may be 'partially correct', or 'correct but misleading' without extra context. One key role of professional fact-checkers is to help their audience gain a full understanding of a claim, with all its nuance and complexity, rather than simply apply a binary label.

Fact-checkers can only have an impact if they are trusted by their readers. They therefore take great care to only publish fact-checks after meticulous research, and adhere to strict editorial standards, as outlined, for example, in the International Fact-Checking Network (IFCN) fact-checkers' code of principles.[11] This leads to a major hurdle before adopting fully automated verification methods: such methods will inevitably be imperfect, and publishing incorrect fact-checks could seriously damage the reputation of the responsible fact-checking organization. They may be more valuable as internal tools by presenting the evidence, reasoning and conclusion regarding a claim, before the (human) fact-checker writes and publishes their fact-check article.

3 What Technology Currently Offers

Fact-checking is not a straightforward or routine process. It requires a chain of steps that go from sensing media to spotting check-worthy claims all the way through to concluding whether the claim is true, partially-true, false, misleading or perhaps impossible to judge. Figure 1 shows a prototypical fact-checking pipeline (partially derived from [Barrón-Cedeño *et al.*, 2020b]). Below, we discuss each of the boxes in that figure.

3.1 Finding Claims Worth Fact-Checking

In this modern age of multi-modal and multi-domain news propagation, fact-checkers are flooded with sentences that potentially include claims that are worth fact-checking. This has encouraged the development of AI solutions for this crucial component in the fact-checking pipeline inside research communities (e.g., CLEF CheckThat! lab [Nakov *et al.*, 2018; Elsayed *et al.*, 2019; Barrón-Cedeño *et al.*, 2020a; Nakov *et al.*, 2021]) and dedicated fact-checking organizations (e.g., Full Fact [Corney, 2019]).

The problem is widely tackled as a ranking problem where the system produces a ranked list of claims coupled with check-worthiness scores. Such a score is important to increase system's transparency and provide fact-checkers the ability to prioritize or filter claims. Fact-checkers can also provide feedback on how reflective this score is of actual check-worthiness of a claim, which can be later used to tune the system. One tool for check-worthy claims identification is ClaimBuster [Hassan *et al.*, 2017], which has been used by professional fact-checkers through Duke Reporters' Lab project.[12] It distinguishes between *non-factual sentences*, such as personal beliefs or opinions, *unimportant factual claims*, and *check-worthy factual claims*. By combining sentiment analysis, entity recognition, part-of-speech tags, tf-idf and claim length to train a support vector machine, they achieve usable levels of accuracy.

In a more recent version of ClaimBuster, a transformer (BERT) model coupled with gradient-based adversarial training was proposed, resulting in improved performance [Meng *et al.*, 2020]. We also see an increased adoption of transformer-based models in several systems attempting to identify check-worthy claims. Models like RoBERTa or BERT are usually used as part of the classification architecture itself (e.g., [Hasanain and Elsayed, 2020; Williams *et al.*, 2020; Nikolov *et al.*, 2020]) or sometimes as sources of content representation to be used as features in the classification system (e.g., [Kartal *et al.*, 2020]).

During a recent general election, Full Fact used a fine-tuned BERT model to classify claims made by each political party [Corney, 2019], according to whether they were *numerical claims*, *predictions*, *personal beliefs*, etc. This allowed the fact-checkers to rapidly identify the claims most likely to be checkable, hence focus their efforts in the limited time available while voters are making their final decisions.

Social media giants are also working on combating misinformation on their platforms. Facebook describes a proprietary tool to identify claims that should be fact-checked [Facebook, 2020]. The system leverages flags provided by users for a post indicating that it is potentially false. Additionally, features from the content of the replies to a post are exploited. These signals are part of a machine learning model that predicts whether a post contains misinformation. The model is updated using feedback from fact-checkers.

By rapidly surfacing important checkable claims, these tools can help fact-checkers respond quickly to misleading claims whenever they appear.

3.2 Detecting Previously Fact-Checked Claims

Interestingly, despite the importance of detecting whether a claim has been fact-checked in the past, this problem has only recently been explored by the research community. Most notably, we can point to the work of Shaar *et al.* (2020), who formulated the task, proposed learning-to-rank algorithms, and released two specialized datasets: one on tweets, which are to be fact-checked against the claims in Snopes, and another on claims in political debates, which are to be fact-checked against the claims in PolitiFact. The social me-

[11] http://www.poynter.org/ifcn-fact-checkers-code-of-principles/

[12] https://reporterslab.org/tech-and-check

dia version of the task was then featured at the CheckThat! Lab at CLEF-2020 and 2021 [Barrón-Cedeño *et al.*, 2020a; Nakov *et al.*, 2021]. It was also explored by Vo and Lee (2020) from a multi-modal perspective, where social media claims about images were matched against a previously fact-checked claims in the ”fauxtography” section of Snopes.

Full Fact is currently trialling a novel claim matching tool that uses machine learning to combine modern language models (BERT) with traditional information retrieval methods (BM25) and specially created claim-type classifiers, entity recognition and topic detection algorithms.

Recently, Google has released the *Fact Check Explorer*,[7] which is an exploration tool that allows users to search a number of fact-checking websites (those that use ClaimReview from `schema.org`[13]) for the mentions of a topic, a person, etc. However, the tool cannot handle complex claims, as it runs Google search, which is not optimized for semantic matching of long claims. While this might change in the future, as there have been reports that Google has started using BERT in its search, at the time of writing, the tool could not handle a long claim as input.

One tool developed by Logically compares incoming claims against a database of checked claims [Adler and Boscaini-Gilroy, 2019] to find repeats. They use word embeddings to represent the claims, then the DBScan clustering algorithm to find semantically similar groups of claims.

Automatically finding repeated instances of a misleading claim can help slow the spread of misinformation without requiring excess manual effort from fact-checkers.

3.3 Evidence Retrieval

Evidence retrieval aims to find external evidence to help decide on the factuality of a claim. This has been shown to be an effective approach when assessing claim veracity [Alhindi *et al.*, 2018]. Evidence retrieval is a component that can be composed of multiple optional steps. Whereas the input consists of a check-worthy claim and a (potentially closed) data collection, the process could finish in the production of a ranking of the relevant data —as in a *standard* retrieval scenario— or in the extraction of specific pieces of evidence; e.g., a text snippet, or a recording.

The CLEF 2013 INEX lab [Bellot *et al.*, 2013] included a shared task that required models to retrieve evidence snippets from a pool of $50k$ books to *confirm* or *refute* a claim. The main finding of INEX was that entity matching is among the most relevant pieces of information. Cartright *et al.,* [2011] paid more attention to the process of producing a sensitive query. Both the statement and its context are combined in the query to retrieve the potentially-supporting material on the basis of diverse search engines.

The Fact Extraction and Verification shared task (FEVER) [Thorne *et al.*, 2018] focused on extracting evidence sentence, related to a claim, from Wikipedia articles and determining whether such evidence *supports* the claim, *refutes* it, or does not provide enough information. As in INEX, named entities are among the key pieces of information, and they were often used to compose the queries to retrieve the most relevant articles and select the evidence sentences, usually exploiting neural-network architectures [Malon, 2018; Hanselowski *et al.*, 2018].

When dealing with a closed reference collection, the task can be addressed as a simple ranking computation on the basis of the similarity (e.g., cosine over some vectorial representation) of the claim against the document collection [Touahri and Mazroui, 2020]. Whereas most models intend to assess the relevance of full documents against the query, recent efforts try to capture relevance also at the sentence level and combine it to score full documents [Akkalyoncu Yilmaz *et al.*, 2019].

When a document is considered relevant to challenge a claim, its most relevant snippet for the fact-checking can be identified. Alshomary *et al.,* [2020] proposed a snippet generation model that produces snippets representing arguments in favour or against the claim.

By combining the earlier methods, Fan *et al.* [2020] provide snippets of background knowledge that brief the fact-checkers on the claim’s background knowledge. They do this by generating and retrieving relevant *passage briefs*, identifying and retrieving documents based on *entity briefs*, and generating and answering *question answering briefs* decomposed from the claim.

In some cases, evidence retrieval goes beyond text, e.g., when checking an image or a video, reverse image search allows fact-checkers to find other contexts where the multimedia content was used before. This allows to detect out-of-context content, e.g., an image or a video from one event portrayed as being from a different event, as well as potentially manipulated image/video. Popular tools include TinEye,[14] Google Image Search, and Yandex Image Search. Relevant research tools are being developed in two EU projects: WeVerify[15] and InVID.[16]

Presenting professional fact-checkers with a range of relevant evidence as they assess a claim can help save them time searching for evidence themselves, and also present evidence that they might otherwise have missed. Ideally, automated evidence retrieval can therefore help produce *better* fact-checks *faster*.

3.4 Automated Verification

Three broad approaches to automated verification can be identified in the literature [Babakar and Moy, 2016]: *reference approaches*, where the claim is checked against a trusted source, such as a database [Ahmadi *et al.*, 2019; Chen *et al.*, 2020]; *machine learning approaches*, where a probabilistic model is created to predict whether a claim or its contradiction is more likely [Thorne and Vlachos, 2018; Augenstein *et al.*, 2019]; and *contextual approaches* that use metadata to estimate the reliability of a claim, such as analysing who is spreading or repeating a claim in social media, the language used, and whether it is being already rebutted by other users [Shu *et al.*, 2017; Vosoughi *et al.*, 2018].

[13] http://schema.org/ClaimReview

[14] http://tineye.com/

[15] http://weverify.eu/tools/

[16] http://www.invid-project.eu

While automatic verification is hard, there are promising results for different kind of claims. For example, an explicit claim about a numerical value, such as "In 2017, global electricity demand grew by 3%", can be verified automatically with access to official statistics. Recent work has also shown that text reporting the results of complex formula can be automatically verified [Karagiannis *et al.*, 2020]. Success here also depends on the availability of reliable data, presented in a consistent format, and this varies widely between countries and between fields. Similarly, simple claims can be verified with promising accuracy results when good evidence is available, for example for entities popular on the Web [Augenstein *et al.*, 2019]. Most of these methods are based on machine learning algorithms that are trained on existing annotated corpora of claims, such as those in the FEVER challenges [Thorne *et al.*, 2018].

While new methods keep increasing the accuracy and the scope of the automated checking algorithms, two main problems prevent their adoption in fact-checking organizations. First, even on the corpora of claims that are available in the literature, their effectiveness is not high enough to allow automatic decisions. Second, most claims in the public realm are more complex, such as claims that COVID-19 vaccines have been developed too quickly and are still experimental[17]. To check these claims, fact-checkers might need to interview experts, collaborate with other fact-checkers, understand the context and framing of claims, track down and verify multiple sources and pieces of evidence — all of which require human levels of intelligence. The general verification of arbitrary claims requires a deep understanding of the real world that currently eludes AI. Indeed, most methods are designed to assist fact-checkers in their work with suggestions and assume that a human user will assess the verification output before assigning a true/false label to the given content.

3.5 Some Real-World Systems

In this section, we explore systems that cover multiple steps in the fact-checking pipeline. As was presented in the previous sections, many tools have been developed to automate fact-checking or at least some steps in the pipeline. However, majority of the existing systems are usually prototypes that do not offer suitable interfaces to allow users (especially fact-checkers) to interact with the system. Moreover, many systems only target one step in the pipeline. We present below a brief overview of notable systems that cover multiple steps of the fact-checking pipeline and also offer a user interface.

AFCNR: the system of Miranda *et al.* [2019] accepts a claim as input, searches over news articles, retrieves potential evidence and then presents to the user a judgment on stance of each evidence towards the claim and an overall rating of the claim veracity given the evidence. The system was extensively tested by eleven journalists from BBC.

BRENDA is a browser extension, which allows users to fact-check claims directly while reading news articles [Botnevik *et al.*, 2020]. It supports two types of input, either the full page opened in the browser, or a highlighted snippet inside the page. In the first scenario, the system applies check-worthiness identification in order to decide which sentences in a page to fact-check.

ClaimPortal:[18] this system includes a check-worthiness scoring component, which is only applicable to tweets with a focus on political content [Majithia *et al.*, 2019]. After retrieving tweets in response to an input search query, the system scores tweets by check-worthiness using the ClaimBuster model and also attempts to verify each tweet by retrieving similar claims previously fact-checked by fact-checking organizations (e.g., PolitiFact).

Squash: developed at Duke Reporters' lab, this system *(i)* listens to speech, debate and other events, *(ii)* transcribe them into text, *(iii)* identifies claims to check, and then *(iv)* fact-check them by finding matching claims already fact-checked by humans [Adair, 2020].

Full Fact's system is designed to support fact-checkers. It *(i)* follows news sites and social media, *(ii)* identifies and categorizes claims in the stream, *(iii)* checks whether a claim has been already verified, and then *(iv)* enrich claims with more supporting data to be presented to the fact-checker. It is in daily use in the UK and several countries in Africa [Dudfield, 2020].

We believe the prototypes presented above are good examples of steps being taken towards developing systems that cater to fact-checkers. More systems are now designed to *efficiently* identify claims originating from *various types* of sources (e.g., news articles, broadcast and social media). Moreover, the fact-checker is now becoming a part of the system by providing feedback to the system, rather than just being a consumer. Finally, we see an increase in systems' transparency by providing explainable decisions, therefore making them more an assistive tool rather than a replacement for the fact-checker. However, there are several challenges left to tackle, as we present in the next sections.

4 Lessons Learned

The main lesson from our analysis is that there is a partial disconnection between what fact-checlers want and what technology has to offer. A detailed analysis is reported in the following.

1. Over time, many tools have been developed either to automatically check claims or to provide facilities to the fact-checkers to support their manual fact-checking process. However, there are still limitations in both automated and manual processes: *(i)* the credibility issue in automated systems, as they do not provide supporting evidence; and *(ii)* the scalability issue of manual fact-checking.

2. Automated fact-checking systems can help fact-checkers in different ways: *(i)* finding claims from the large information ecosystem; *(ii)* finding previously fact-checked claims; *(iii)* finding supporting evidence (in a form of text, audio or video), translating (for multilingual content) and summarizing relevant posts, articles and documents if needed; and *(iv)* detecting claims that are spreading faster to slow them down.

[17] https://fullfact.org/online/covid-19-survival-rate-less-998/

[18] https://idir.uta.edu/claimportal/

3. There is a lack of collaboration between researchers and practitioners in terms of defining the tasks and developing datasets to train models and develop automated systems. In general, a human-in-the-loop can be an ideal setting for fact-checking, which is currently not being fully explored.

5 Challenges and Future Forecasting

Below we discuss some major challenges and we forecast some promising research directions:

5.1 Major Challenges

- **Leveraging multi-lingual resources:** The same claim, with slightly different variants, often spreads over different regions of the world at almost the same time or sometimes at different times. These may be "international claims" such as the medical claims around COVID-19, or stories that are presented as local, but with varied, false locations. Those claims might be fact-checked in one language, but not the others. Moreover, resources in English are abundant, but in low-resource languages, such as Arabic, they are clearly lacking. Aligning and coordinating the verification resources and leveraging those resources over different languages to improve fact-checking is a challenge.
- **Ambiguity in the claims:** another reason why automatic fact-checking is challenging is related to the fact that often a claim has multiple interpretations. An example is "The COVID death rate is rising." Is this about mortality or about fatality rate? Does it refer to today/yesterday or to the last week/month? Does it refer to the entire world or to a specific area? In such cases, knowledge about the context is necessary in order to properly frame the claim and to filter out unlikely interpretations. After that, all remaining interpretations should be analyzed, which would further slow down the work of fact-checkers. One system that proposes a solution to this problem is CoronaCheck[19].
- **System bias:** The majority of existing systems are trained using datasets curated by a small group of people and often annotated by non-experts. This in turn results in systems biased towards how the system developers perceive factuality and how the annotation task was described to annotators. The dangers of bias in large language models is becoming increasingly obvious [Bender *et al.*, 2021], and should not be ignored just because the purpose of the system is benevolent.
- **Contextual information:** The current state-of-art for automated fact-checking mostly makes a limited use of contextual information, for example comments of the readers, linked sources of news articles, social network data for social media posts. Such information can provide useful signals for enriching current models.
- **Multimodality:** Information is typically disseminated through multiple modalities such as text, image, speech, video, temporal, user profile, and network structure. Addressing the problem based on a single modality can be a step towards failure. For example, it might be difficult to detect fake news pieces that are automatically generated using deep fakes and/or GPT-3-style text generation. To avoid such issues, multimodal approaches would be one way to go, if evidence can be gathered from multiple types of source at the same time. This in turn requires multimodal datasets to develop suitable models.

5.2 Future Forecasting

- **Close collaboration between fact-checking platforms and researchers:** We envision closer collaboration between professionals from fact-checking platforms alongside researchers in the domain to discuss common interests, existing solutions, and future directions, has been a challenge.
- **Integrated solutions:** We further envision unified and open-source initiatives to develop resources for system development and benchmarking.
- **Usability:** We further forecast more research on the system interface design, which would facilitate the adoption of AI by fact-checkers. It is important to develop systems that require minimal technical knowledge and reduce cognitive load. Such systems can help a larger number of fact-checkers and journalists in the fact-checking process.
- **Interpretability and explainability:** Models should be designed in such a way that their outcomes are explainable, unbiased, and more accountable to ethical considerations.
- **Efficient and real-time solutions:** Finally, in order to tackle the velocity of the spread of fake news there is a need to develop systems that are efficient and scalable for real-time solution. To be effective, such systems would need to be embedded within, or accessible by, social networks and other big technology companies.

6 Conclusion

We have presented a survey of the available intelligent technologies that can support the human experts in the different steps of the manual process of fact-checking claims. These include tasks such as identifying claims worth fact-checking, detecting relevant previously fact-checked claims, retrieving relevant evidence to support the manual fact-check of a claim, and actually verifying a claim. In each case, we paid attention to the challenges in future work and the potential impact on real-world fact-checking.

We argued that there is currently only a partial overlap between what fact-checkers want and what the research community considers as a priority. We then discussed lessons learned and major challenges that need to be overcome. Finally, we suggested several research directions, which we forecast will emerge in the near future.

[19] https://coronacheck.eurecom.fr

References

[Adair, 2020] Bill Adair. Squash report card: Improvements during State of the Union … and how humans will make our AI smarter. https://reporterslab.org/squash-report-card-improvements-during-state-of-the-union-and-how-humans-will-make-our-ai-smarter/, 2020.

[Adler and Boscaini-Gilroy, 2019] Ben Adler and Giacomo Boscaini-Gilroy. Real-time Claim Detection from News Articles and Retrieval of Semantically-Similar Factchecks. In *NewsIR'19 Workshop at SIGIR*, 2019.

[Ahmadi *et al.*, 2019] Naser Ahmadi, Joohyung Lee, Paolo Papotti, and Mohammed Saeed. Explainable fact checking with probabilistic answer set programming. In *TTO*, 2019.

[Akkalyoncu Yilmaz *et al.*, 2019] Zeynep Akkalyoncu Yilmaz, Wei Yang, Haotian Zhang, and Jimmy Lin. Cross-domain modeling of sentence-level evidence for document retrieval. In *EMNLP*, 2019.

[Alhindi *et al.*, 2018] Tariq Alhindi, Savvas Petridis, and Smaranda Muresan. Where is your evidence: Improving fact-checking by justification modeling. In *FEVER*, 2018.

[Alshomary *et al.*, 2020] Milad Alshomary, Nick Düsterhus, and Henning Wachsmuth. *Extractive Snippet Generation for Arguments*. ACM, 2020.

[Arnold, 2020] Phoebe Arnold. The challenges of online fact checking. Technical report, Full Fact, 2020.

[Augenstein *et al.*, 2019] Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims. In *EMNLP*, 2019.

[Babakar and Moy, 2016] Mevan Babakar and Will Moy. The state of automated factchecking. Technical report, Full Fact, 2016.

[Barrón-Cedeño *et al.*, 2020a] Alberto Barrón-Cedeño, Tamer Elsayed, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari, and Preslav Nakov. CheckThat! at CLEF 2020: Enabling the automatic identification and verification of claims on social media. In *ECIR*, 2020.

[Barrón-Cedeño *et al.*, 2020b] Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Nikolay Babulkov, Bayan Hamdan, Alex Nikolov, Shaden Shaar, and Zien Sheikh Ali. Overview of CheckThat! 2020: Automatic identification and verification of claims in social media. In *CLEF*, 2020.

[Bellot *et al.*, 2013] Patrice Bellot, Antoine Doucet, Shlomo Geva, Sairam Gurajada, Jaap Kamps, Gabriella Kazai, Marijn Koolen, Arunav Mishra, Véronique Moriceau, Josiane Mothe, Michael Preminger, Eric Sanjuan, Ralf Schenkel, Xavier Tannier, Martin Theobald, Matthew Trappett, and Qiuyue Wang. Overview of INEX 2013. In *CLEF*, volume 8138, 2013.

[Bender *et al.*, 2021] Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *ACM Conference on Fairness, Accountability, and Transparency*, 2021.

[Botnevik *et al.*, 2020] Bjarte Botnevik, Eirik Sakariassen, and Vinay Setty. BRENDA: Browser extension for fake news detection. In *SIGIR*, 2020.

[Cappellato *et al.*, 2020] Linda Cappellato, Carsten Eickhoff, Nicola Ferro, and Aurélie Névéol, editors. *CLEF 2020 Working Notes*, CEUR Workshop Proceedings. CEUR-WS.org, 2020.

[Cardoso Durier da Silva *et al.*, 2019] Fernando Cardoso Durier da Silva, Rafael Vieira, and Ana Cristina Garcia. Can machines learn to detect fake news? a survey focused on social media. In *HICSS*, 2019.

[Cartright *et al.*, 2011] Marc-Allen Cartright, Henry A. Feild, and James Allan. Evidence finding using a collection of books. In *CIKM*. ACM, 2011.

[Chen *et al.*, 2020] Wenhu Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyou Zhou, and William Yang Wang. TabFact: A large-scale dataset for table-based fact verification. In *ICLR*, 2020.

[Corney, 2019] David Corney. How we use AI to help fact check party manifestos. https://fullfact.org/blog/2019/dec/how-we-use-ai-help-fact-check-party-manifestos/, 2019.

[Dudfield, 2020] Andrew Dudfield. How we're using AI to scale up global fact checking. https://fullfact.org/blog/2020/jul/afc-global/, 2020.

[Elsayed *et al.*, 2019] Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeño, Maram Hasanain, Reem Suwaileh, Pepa Atanasova, and Giovanni Da San Martino. CheckThat! at CLEF 2019: Automatic identification and verification of claims. In *ECIR*, 2019.

[Facebook, 2020] Facebook. How our fact-checking program works. https://www.facebook.com/journalismproject/programs/third-party-fact-checking/how-it-works, 2020.

[Fan *et al.*, 2020] Angela Fan, Aleksandra Piktus, Fabio Petroni, Guillaume Wenzek, Marzieh Saeidi, Andreas Vlachos, Antoine Bordes, and Sebastian Riedel. Generating fact checking briefs. In *EMNLP*, 2020.

[Hanselowski *et al.*, 2018] Andreas Hanselowski, Hao Zhang, Zile Li, Daniil Sorokin, Benjamin Schiller, Claudia Schulz, and Iryna Gurevych. UKP-athene: Multi-sentence textual entailment for claim verification. In *FEVER*, 2018.

[Hasanain and Elsayed, 2020] Maram Hasanain and Tamer Elsayed. bigIR at CheckThat! 2020: Multilingual BERT for ranking Arabic tweets by check-worthiness. In Cappellato et al. [2020].

[Hassan *et al.*, 2015] Naeemul Hassan, Chengkai Li, and Mark Tremayne. Detecting check-worthy factual claims in presidential debates. In *CIKM*, 2015.

[Hassan *et al.*, 2017] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *SIGKDD*, 2017.

[Islam *et al.*, 2020] Md Rafiqul Islam, Shaowu Liu, Xianzhi Wang, and Guandong Xu. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *SNAM*, 10(1), 2020.

[Karagiannis *et al.*, 2020] Georgios Karagiannis, Mohammed Saeed, Paolo Papotti, and Immanuel Trummer. Scrutinizer: A mixed-initiative approach to large-scale, data-driven claim verification. *VLDB*, 13(11), 2020.

[Kartal *et al.*, 2020] Yavuz Selim Kartal, Busra Guvenen, and Mucahid Kutlu. Too many claims to fact-check: Prioritizing political claims based on check-worthiness. In *CIKM 2020 Workshops*, 2020.

[Kotonya and Toni, 2020] Neema Kotonya and Francesca Toni. Explainable automated fact-checking: A survey. In *COLING*, 2020.

[Lazer *et al.*, 2018] 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), 2018.

[Li *et al.*, 2016] Yaliang Li, Jing Gao, Chuishi Meng, Qi Li, Lu Su, Bo Zhao, Wei Fan, and Jiawei Han. A Survey on Truth Discovery. *ACM SIGKDD Explorations Newsletter*, 17(2), 2016.

[Majithia *et al.*, 2019] Sarthak Majithia, Fatma Arslan, Sumeet Lubal, Damian Jimenez, Priyank Arora, Josue Caraballo, and Chengkai Li. Claimportal: Integrated monitoring, searching, checking, and analytics of factual claims on twitter. In *ACL*, 2019.

[Malon, 2018] Christopher Malon. Team papelo: Transformer networks at FEVER. In *FEVER*, 2018.

[Martino *et al.*, 2020] Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. A survey on computational propaganda detection. In *IJCAI*, 2020.

[Meng *et al.*, 2020] Kevin Meng, Damian Jimenez, Fatma Arslan, Jacob Daniel Devasier, Daniel Obembe, and Chengkai Li. Gradient-based adversarial training on transformer networks for detecting check-worthy factual claims. *arXiv:2002.07725*, 2020.

[Miranda *et al.*, 2019] Sebastião Miranda, David Nogueira, Afonso Mendes, Andreas Vlachos, Andrew Secker, Rebecca Garrett, Jeff Mitchel, and Zita Marinho. Automated fact checking in the news room. In *WWW*, 2019.

[Nakov *et al.*, 2018] Preslav Nakov, Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Lluís Màrquez, Wajdi Zaghouani, Pepa Atanasova, Spas Kyuchukov, and Giovanni Da San Martino. Overview of the clef-2018 checkthat! lab on automatic identification andverification of political claims. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, 2018.

[Nakov *et al.*, 2021] Preslav Nakov, Giovanni Da San Martino, Tamer Elsayed, Alberto Barrón-Cedeño, Rubén Míguez, Shaden Shaar, Firoj Alam, Fatima Haouari, Maram Hasanain, Nikolay Babulkov, Alex Nikolov, Gautam Kishore Shahi, Julia Maria Struß, and Thomas Mandl. The CLEF-2021 CheckThat! Lab on detecting check-worthy claims, previously fact-checked claims, and fake news. In *ECIR*, 2021.

[Nikolov *et al.*, 2020] Alex Nikolov, Giovanni Da San Martino, Ivan Koychev, and Preslav Nakov. Team_Alex at CheckThat! 2020: Identifying check-worthy tweets with transformer models. In Cappellato et al. [2020].

[Shaar *et al.*, 2020] Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, and Preslav Nakov. That is a known lie: Detecting previously fact-checked claims. In *ACL*, 2020.

[Shu *et al.*, 2017] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *SIGKDD*, 19(1), 2017.

[Thorne and Vlachos, 2018] James Thorne and Andreas Vlachos. Automated fact checking: Task formulations, methods and future directions. In *COLING*, 2018.

[Thorne *et al.*, 2018] James Thorne, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, and Arpit Mittal. The fact extraction and VERification (FEVER) shared task. In *FEVER*, 2018.

[Touahri and Mazroui, 2020] Ibtissam Touahri and Azzeddine Mazroui. EvolutionTeam at CheckThat! 2020: Integration of linguistic and sentimental features in a fake news detection approach. In Cappellato et al. [2020].

[Vlachos and Riedel, 2014] Andreas Vlachos and Sebastian Riedel. Fact checking: Task definition and dataset construction. In *Workshop on LT and CSS*. ACL, 2014.

[Vo and Lee, 2018] Nguyen Vo and Kyumin Lee. The rise of guardians: Fact-checking url recommendation to combat fake news. In *SIGIR*, 2018.

[Vo and Lee, 2020] Nguyen Vo and Kyumin Lee. Where are the facts? searching for fact-checked information to alleviate the spread of fake news. In *EMNLP*, 2020.

[Vosoughi *et al.*, 2018] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380), 2018.

[Williams *et al.*, 2020] Evan Williams, Paul Rodrigues, and Valerie Novak. Accenture at CheckThat! 2020: If you say so: Post-hoc fact-checking of claims using transformer-based models. In Cappellato et al. [2020].

[Zaman *et al.*, 2014] Tauhid Zaman, Emily B. Fox, and Eric T. Bradlow. A Bayesian approach for predicting the popularity of tweets. *Ann. Appl. Stat.*, 8(3), 2014.

[Zhou and Zafarani, 2020] Xinyi Zhou and Reza Zafarani. A survey of fake news: Fundamental theories, detection methods, and opportunities. *CSUR*, 2020.