Ethics in AI-Powered Automation of Fact-Checking

Abstract

We present a discourse on the current state of AI-powered fact-checking, the current codes of ethics and principles in place to guide the fact-checkers and programmers who develop AI systems, and the myriad of pitfalls that may besiege their work and research. We propose some suggestions to help alleviate these ethics-related issues and begin the discussion on how these should be approached. Our findings, gathered from observations of state-of-the-art research in this area and the first-hand experience gained from our own projects, show that there are still large gaps we have to bridge before we can create systems that are robust, transparent, and trustworthy in the eyes of the public.

Primary discipline: Computer Science

Introduction

Our society is struggling with an unprecedented amount of falsehoods which do harm to wealth, democracy, health, and national security. In domestic political controversies, politicians repeat false claims even after they are debunked. “Fake news” divides the society and causes turmoil such as the “Pizzagate”. [1] Clickbaits and fake reviews damage the online ecosystem. Misinformation and disinformation also present serious national security threats, of which the most prominent example is the ongoing investigation of the Russian meddling with the 2016 election.

Mitigating false information calls for interdisciplinary collaboration of and advancement in multiple areas, including journalism, communication studies, law and public policy, psychology, and political science. Nevertheless, computing technology plays a crucial role in it, since the modern-day falsehoods are largely spread online.

Lately there has been a great leap in the capability of artificial intelligence systems, driven by compounded breakthrough in our means of storing and analyzing data, novel computing architecture, and machine learning algorithms. What follows is an explosion of interests in exploiting AI in almost every type of application reaching every corner of our world. It is thus only natural to observe a substantial growth in efforts for AI-powered fact-checking (Shu et al. 2017). These efforts tackle various fronts, such as the detection of unverified or deliberately made false information on Twitter using neural networks and machine learning algorithms (Ma, Gao, and Wong 2018; Zhao, Resnick, and Mei 2015), an end-to-end fact-checking system that uses SVM and neural networks for discovering checkworthy factual claims (Jimenez and Li 2018; Hassan et al. 2017a; 2017b), flagging clickbait articles (Chakraborty et al. 2016; Rony, Hassan, and Yousuf 2018), and detecting social media bot accounts which makes false impression of popularity and alters the perception of reality regardless of veracity, using a combination of graph-based, crowd-sourcing, and feature-based approaches (Ferrara et al. 2016).

As AI brings fundamental transformations to our society, it has also raised significant ethics concerns regarding fairness, transparency, trust, and misuse in AI-powered systems. By now we are familiar with a variety of conversations centered around the concerns regarding AI: how algorithms may discriminate against women in job applications and against African American inmates; how robots are exploited and how humans are replaced by robots; weaponization of AI technology; moral conundrum faced by autonomous vehicles; and so on. In the wake of these important issues, for the first time since 1992, the Association for Computing Machineary (ACM) has recently updated its Code of Ethics and Professional Conduct. [2]

The ethics concerns are particularly pertinent to fact-checking. The risks of misuse of AI are already manifested in this arena. Some of the social media bots spreading falsehoods may use advanced machine learning algorithms to decide who to follow and which topics to engage with and to tailor their program-generated posts for specific conversations. Clickbaits can be optimized for financial gains in advertisement revenue through sophisticated orchestration of titles, keywords, images, and target audience, and such can be steered by AI algorithms. We have also seen AI-generated videos that mimic real life in very interesting, and perhaps disturbing ways given the potential uses (Suwajanakorn, Seitz, and Kemelmacher-Shlizerman 2017). Such technology can pose future challenges to fact-checking re-

[1] https://www.washingtonpost.com/local/pizzagate-from-rumor-to-hashtag-to-gunfire-in-dc/2016/12/06/4c7def50-bbd4-11e6-94ac-3d324840106c_story.html

[2] https://www.acm.org/code-of-ethics

lated efforts as we try to tackle discerning whether something that seemingly came out of an individual's mouth is real or fake.

Ethics concerns arise regarding not only malicious adversaries but also unintended consequences of acts by samaritans—while fact-checkers and fact-checking tools discern truth from false, who is there to check them? Trust, transparency, and fairness are of paramount importance in fact-checking. Violating ethics in fact-checking defies its purpose. The list of AI-themed research projects and innovations for fact-checking is ever-expanding and in many ways unchecked. With no real precedents to look at, many have thrown themselves at the challenge at hand, pioneering techniques to help pave a way forward. While there is self-awareness about the ethical considerations and there have been some discussions on different issues, no real guidelines have truly been established which are enforced or the de facto standard. The International Fact-Checking Network (IFCN) brings together fact-checkers in the world to commit to its code of principles. [3] However, its code of principles is for professionals at fact-checking organizations instead of AI algorithms.

In this paper, we aim at presenting a discourse that tries to tackle the ethics in fact-checking assisted by AI technologies, particularly the ethical pitfalls and potential solutions in the automation of fact-checking steps. This essay can then engage more in the discussion of how we can build more ethics-compliant fact-checking tools. By far some of the biggest threats to fact-checking, fake-news detection, and related research are bias, transparency, and explainability. Focusing on these issues, our discussion looks at how we have spotted them in different projects including our own efforts and areas where they may easily creep into related research as well. This discussion is pertinent to various steps in the fact-checking workflow, including data collection, claim spotting, claim matching, fact verification, and verdict generation stages. For results to be produced from this discourse We expect ensuing, thorough investigations from not only computer scientists working in this area but also interdisciplinary efforts involving all relevant fields and especially the fact-checking community.

This paper is organized as follows. We will begin by presenting a brief background on fact-checking and fake-news detection, then proceed to address relevant standards that exist today, after which we explain automation in fact-checking and the role AI has played in it. We will then analyze ethical pitfalls in AI-powered fact-checking, and finally we will briefly discuss potential solutions to mitigate ethical concerns in this process before we conclude the paper.

Background: Fact-Checking and "Fake News"

Fact-checking has long been a staple of journalism, and the act of "fake news" perhaps can be dated back to the age when "yellow journalism" started to appear. These two became household terminologies since the 2016 U.S. election, probably due to several prominent reasons, among others. First, while politicians may refrain themselves from making outright false claims to avoid being fact-checked, oftentimes they even double down after their false claims are debunked. Second, some politicians and public figures use the term "fake news" to label personals and organizations holding opposing opinions or facts that will not back their agenda. Third, the ongoing investigation of interference with election using disinformation further raises the stakes.

In countering false information, the number of active fact-checking organizations, including the likes of The Washington Post, New York Times, and FactCheck.org, has grown from 44 in 2014 to 161 in November 2018, according the Duke Reporters's Lab. [4] Fact-checkers vet claims by presenting relevant data and documents and publishing their verdicts. For instance, PolitiFact.com, one of the earliest and most popular fact-checking projects, gives factual claims truthfulness ratings such as true, half true, false, and even "pants on fire". These fact-checking organizations are traditionally politics-oriented, focusing on vetting he said, she said, with a few (e.g., Snopes.com) also checking rumors, urban legends, and gossips.

A massive number of websites and social media users have been discovered to engage in spreading misinformation, disinformation, and propaganda. While a lot of false and exaggerated information exists in the form of misleading, captivating headlines, e.g., clickbaits created to deceive web users into clicking at those headlines only to read less appealing articles, fake news was also created to spread derogatory rumors, promote societal and political tensions, manipulate public opinion, and influence national election outcomes. The problem is exacerbated by bots that spread and amplified falsehoods published on dedicated fake news websites. These are social media accounts run by computer programs that automatically publish and forward content, follow other accounts, and leave comments, creating networks of seemingly legitimate news sites and user accounts. Algorithms employed by Facebook, Google, and Twitter heavily rely on popularity measures, failed to discern the bots, and promoted false information as trending topics and top search results, exposing it to an audience of millions. [5] The situation is worsen by the filter-bubble and echo-chamber phenomena—social media users tend to trust their own social groups over conventional news media. A recent study, offering evidence that bots may have influenced the election outcome, reports that a sample of 140,000 Twitter users in the battleground state of Michigan shared as many junk news items as professional news during the final ten days of the 2016 election, each constituting 23% of the web links they shared on Twitter in that period. [6]

Current Standards

A code of ethics provides guidance and establishes common ethical standards to promote fairness, consistency, and transparency. Currently, we have the International Fact-Checking Network (IFCN) Code of Principles [3] which is dedicated to

[3] https://ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles

[4] https://reporterslab.org/fact-checking/

[5] https://nbcnews.to/2DqeAgx

[6] http://politicalbots.org/?p=1064

fact-checking organizations and professional fact-checkers, and the ACM Code of Ethics and Professional Conduct [2] which is for all computing professions, including those developing AI backed systems and algorithms.

Biased or unsourced fact-checking taints public perception of the enterprise and leads to a general lack of trust in the media and in professional fact-checkers. A committee of fact-checkers under the supervision of IFCN designed the code of principles. It consists of a series of commitments, in order to assure fairness, transparency, and unbiasedness in their fact-checking. As of November 4th, 2018, 53 verified fact-checking organizations from 32 countries have signed onto this code of principles.[7] The commitments these signatory organizations adhere to are (1) applying the same standard on every fact-check, (2) providing all sources in enough detail so that readers can verify verdicts themselves, (3) assuring that funders have no impact on the outcomes of fact-checks, (4) explaining the procedures they follow to select, examine, write, edit, publish and correct their fact checks, and (5) announcing their corrections policy and correcting assuredly and transparently in conformance with the corrections policy.

The ACM Code of Ethics and Professional Conduct expresses the guiding principles of the computing profession. It encompasses the recommendations for several facets of computational professionals, including general ethical requirements, professional responsibilities, leadership responsibilities, and code compliance guidelines. Though not a tool for solving ethical problems, it offers principles for making ethical decisions in the field of computing. The underlying essence of the code is to establish public good as the primary consideration for all computation related activities while the guidelines serve professionals in promoting accountability and transparency to all stakeholders.

Artificial Intelligence in the Automation of Fact-Checking

The amount of misinformation and the speed at which they spread is way beyond the capability of current fact-checkers, since fact-checking is an intellectually demanding and laborious process. For instance, it takes about one day to research and write a typical article about a factual claim (Hassan et al. 2015). This difficulty leaves many harmful claims unchecked, particularly at the local level. Even if the fact-check has already been published, the public must undertake research to look it up. This gap in time and availability limits the effectiveness of fact-checking.

These challenges create an opportunity for automated fact-checking assistive systems. Fact-checking can be defined as the task of assessing the truthfulness of a claim made in a written or spoken modality. A typical fact-checking process is comprised of several stages, including claim monitoring, claim spotting, claim matching, claim checking, and verdict presentation. The goal of *claim monitoring* is to monitor live discourses (e.g., interviews, speeches and debates), social media, and news to extract content. This might require extracting text from speech audio. The step of *claim spotting* aims to detect factual claims that are worth checking. Given a factual claim, the step of *claim matching* detects matching claims from a repository of existing fact-checks. In the simplest case, the verdict of such an existing fact-check can be presented to readers and viewers if a match exists. *Claim checking* is the stage where research on the veracity of a claim is conducted to reach a verdict on the claim. In *verdict presentation*, a report is created to explain the findings used to reach the verdict.

Recently, there have been efforts to deploy AI technologies to aid the steps in fact-checking. A white paper from FullFact.org (Babakar and Moy 2016) surveys existing tools that can be put together for some of the aforementioned steps. Moreover, there are several research projects on algorithmic innovations in machine learning and natural language processing for fact-checking. ClaimBuster (Hassan, Li, and Tremayne 2015; Hassan et al. 2017a; 2017b; Jimenez and Li 2018) is the first end-to-end fact-checking system that automates all the aforementioned stages of fact-checking for certain types of claims. Two other similar projects are Chequeabot [8] and FactStream. [9]

Particularly, ClaimBuster automates claim spotting by training a machine learning model on human-labeled sentences from past presidential debates. ClaimBuster's model produces a score that indicates how likely a sentence contains an important factual claim that should be checked. Another claim detection work is ClaimRank (Gencheva et al. 2017) which is trained by using SVM and FNN to rank claims. It is trained on a dataset of fact-checked statements from political debates, using a wide variety of features including the speaker of the statement.

With respect to claim matching, it is not always as straightforward as finding existing fact-checks on identical claims. Instead, oftentimes claims are rephrased, a fact-check partially supporting or refuting a claim at hand is still insightful in vetting the claim, and even a related fact-check can be helpful. Hence, a more general approach to claim matching can benefit from coreference resolution (Clark and Manning 2015), entity matching (Mudgal et al. 2018), paraphrase detection (Socher et al. 2011), semantic similarity (Ji and Eisenstein 2013), and textual entailment (Padó et al. 2009). For instance, consider the following two statements, both from Hillary Clinton: "African-Americans are more likely to be arrested by police and sentenced to longer prison terms for doing the same thing that whites do." and "... if you're a young African-American man and you do the same thing as a young white man, you are more likely to be arrested, charged, convicted, and incarcerated." Albeit lacking an exact match or even a full entailment, these two statements express the same idea and thus the verdict on one can effectively help fact-check the other.

Such non-trivial claim matching can potentially exploit FrameNet (Baker 2014) and its many extensions, [10] as they provide means for structured and semantic modeling of fac-

[7] https://ifcncodeofprinciples.poynter.org/signatories

[8] https://chequeado.com/automatizacion/

[9] https://bit.ly/2DSBInN

[10] https://framenet.icsi.berkeley.edu/fndrupal/fnbibliography?page=1

tual claims. For example, consider the following factual claim, annotated based on FrameNet's *Taking_sides* frame: "[Sen. Kamala Harris]$_{COGNIZER}$ is supporting [the animals of MS-13]$_{ISSUE}$." With this modeling, we can exploit the key components of claims (e.g., *COGNIZER* and *ISSUE*) in claim matching. Detecting frames in sentences and extracting the components are accomplished through a supervised learning approach. [11]

Another approach to claim matching is to identify relevant articles for a given claim (Wang et al. 2018). They crawled the web and found fact-checking articles by utilizing the claim review markup [12] in web pages. For every fact-checking article, they extracted the fact-checks embedded in it and they also discovered a list of relevant, supporting web pages. The outcome of this work was a repository of fact-checking articles and their supporting articles, ready to be matched against a given factual claim.

There are also efforts in verifying claims about simple statistics by translating them into aggregate queries over database tables (Jo et al. 2018). The process roughly involves decomposing a claim into relevant parts that can be used to construct a query. These parts can be entities, relations between entities, and how these are grouped, e.g., aggregation, ranges, and so on. Given the nontrivial goal, for more general type of claims, it is feasible to use machine learning approaches such as neural networks for the translation.

Another way to assess the veracity of a claim is to validate it against a knowledge base or a knowledge graph (KG). KGs store facts about real-world entities in triples in the form of (head entity, relation, tail entity), e.g., (Microsoft, founded-by, Bill Gates). There are many efforts in recent years to create large-scale KGs, which have become an important resource for AI-related applications, including fact-checking. Validating a claim using a KG can entail employing natural language processing techniques to convert the claim to a query over the KG. Another approach is to find triples relevant to the claim and use the connectivity and distance between these relevant triples to assess the claim's truthfulness (Thorne and Vlachos 2017). Their proposed method focuses only on simple statistical claims such as "The population of Germany in 2015 was 80 million."

Ethical Pitfalls in AI-Powered Fact-Checking

There are many areas, as we have mentioned, that are susceptible to falling prey to ethically questionable practices. In this section we focus on bias within fact-checking, transparency of methods, and explainability of results. Some of what we talk about, has been addressed in other places (e.g., transparency of methodologies from the IFCN code of principles). However, we wish to consolidate all the myriad of ethical pitfalls and biases in one place.

Bias in the Stages of Fact-Checking Let us start with a look at bias in data collection by taking the ClaimBuster project (Hassan et al. 2015), as an example, which required a decent amount of data to produce their results. A large portion of the efforts for this project went into collecting and labelling data (as is evident from their developing of a data collection website[13]) which they could use to demonstrate that their methods and algorithms worked. Data labelling for any task is monotonous, so one challenge this project, like many others, would have faced is finding labellers willing to do the work. Even after finding labellers, researchers are still left with vetting and filtering out bad data.

In the ClaimBuster paper, the authors state that they used some sentences to vet labellers and then introduced more of these throughout the labelling process to score labellers and only used the responses from the best ones. That still leaves the question of whether by doing so they introduced some bias, towards labellers that are aligned with their own preconceived notions and if it is enough to use a rather short vetting procedure in a process that allows anyone to become a labeller. Similar labelling efforts have popped up. FullFact for one has received funding from various sources to help them collect data and train models[14] for their end-to-end fact checking system that is in the works. Bias can also be introduced in the features themselves, as can be seen from the ClaimRank project we mentioned previously. One detractor from their approach is that they use the name of the speaker which introduces bias into their model.

One can even conceive scenarios in which those that are coordinating the data collection might have workshops that are aimed at getting labellers started in the labelling process. The reasons for this might be to overcome language barriers or cultural ones that might prevent some labellers from understanding what it is that makes a statement check-worthy. Doing so, however, introduces the workshop conductor's bias into the labellers themselves. Thus the result is one in which the labellers' answers now reflect what would have been the researchers, so some of that variation in the data is lost and new biases introduced. The hardest part to consolidate is that during this process fact-checkers are trying to extract the collective idea of what is check-worthy, and certainly there must be something that can be considered the average check-worthy sentence with respect to a populace, but reaching that ideal is difficult without having some notion of what it is already (i.e., having a bias for what they consider check-worthy). For efforts like these to succeed, we need to collectively establish, what is and is not important through more empirical methods than a general check for the current temperament of the populace. Bias in data collection is one that should be avoided at all costs, because it then propagates to any process that follows it.

Transparency of Methods Aside from bias, we can also look at the transparency of fact-checking and fact-finding systems that are presented and the results of these (i.e., verdicts, claims found, etc.) While there are some established question answering systems, arguably some of the best[15] are not open source and their methods are a black box. How-

[11] https://github.com/swabhs/open-sesame

[12] https://schema.org/ClaimReview

[13] https://idir-server2.uta.edu/classifyfact_survey/

[14] https://fullfact.org/blog/2016/nov/automated-factchecking-hub/

[15] https://products.wolframalpha.com/api/explorer/

ever, even within academia there is still a big challenge in creating systems that are transparent in their methods and results. Having so many systems based on deep neural networks presents the ever present challenge of thoroughly explaining exactly what is happening within the model without some hand-waving and referencing to abstractions of the core concepts that underlie the networks themselves. The field has become somewhat empirical in nature, in the sense that in order to create the best model many different configurations are tested but at the end we are only left with the process and no real sense regarding why those particular choices were the best for the task at hand. For this we can join the efforts that have begun to make machine learning models more transparent in nature[16](Ribeiro, Singh, and Guestrin 2016), and we can also focus on what takes place before and after. As we have mentioned, having unbiased data that resonates with some pre-established standards in terms of how it was collected is paramount. Similarly, how we interpret the model should be as transparent as possible. If for example, the result of a model simply outputs a label for a given input, then it is likely that in generating that condensed result we omit some extra information regarding the other labels or aspects of that model. We might do this for simplicity, as it allows users to focus on the relevant piece of information our model is producing. However, the best approach would be to include the results in their entirety with some brief explanation of any assumptions or liberties that are taken with respect to these. Things like assuming that the class with the highest probability is the correct one, or applying some extra steps to refine the final label selection are all essential.

Explainability of Results We also ought to consider the explainability of fact-checking results, e.g., how easy it is to explain a given verdict. PolitiFact[17] provides evidence along with any verdict, as does the ClaimBuster project,[18] "Share the Facts",[24] and surely many others. These projects do so, if not for the underlying ethical reasons, to be seen as credible and a reliable source of information. It is apparent, however, that simply including some evidence is not enough. Any assumptions or shortcuts that took them from evidence to verdict must be included with the results. The role of fact-checkers should not be to influence the opinion of users, but to present them with the tools and sources to form their own opinions. They need to detach themselves a bit from the sensationalism that is ever too present in the media of today and strive for the stoic and mundane in order to capture the trust of those who seek, not an opinion, but a reliable source of unbiased information. Toward this goal, automated fact-checking tools can benefit from some standard format in which to present results, that bases itself in objectivity and not in the pursuit of garnering as much attention as possible. Such a format might include the evidence and verdict, as most fact-checking websites do now, but also include assumptions about the evidence that might influence the verdict. Such assumptions could include the threshold that was considered to ascertain a claim as true, or assumptions on what was meant in something that was said that might be ambiguous. For example, when dealing with numerical claims, a naive approach would be to apply a simple threshold of a few significant digits, but such a naive approach has too many drawbacks to become the go-to approach for fact-checkers and researchers. A simple census of the claims that are fact-checked and the context of these can show that the best approach would be to have different rules for different types or categories of claims that are best suited to deal with them. That is to say that leniency might be more important when dealing with things like trade deficits than when dealing with dates or times, it is all context dependent. A simple case of this might be Trump claiming the trade deficit with Japan is between "$69 to $100 billion a year",[19] and an investigation into this finds that by some standards it's true but by others it's less so.[20]

A Miscellany of Pitfalls Finally, we discuss various other pitfalls that we should strive to avoid in fact-checking. When collecting data, training models, or displaying results we should avoid cherry-picking at all costs. While this might seem obvious, it is easy to omit something by justifying that it may be an outlier or perhaps makes the results more digestible to the end user. We should take care to handle datasets that have lurking variables with care. This helps us avoid situations like the Simpson's paradox (Pearl 2014), in which we observe some properties in some given groups of data but have these then disappear when the smaller groups are consolidated into larger groups. Properly understanding the domain that one is working in and what the goal is, can help avoid these issues.

Sensationalism, or exaggeration, within our work should be avoided. Although, we have touched on this, it is worth reiterating as it seems to be a logical fallback for projects that wish to garner attention. Fact-checkers should represent the most objective sensibilities and avoid succumbing to trends that now plague media and have caused the divide between it and the populace. We should take care when selectively presenting results, and only do so when it absolutely makes sense and even then explain thoroughly what is being done. For example, the Maverick project (Zhang, Jimenez, and Li 2018) aims to find interesting facts in a knowledge graph and so scours the graph for these based on algorithms. When presenting these perhaps only the top-k are displayed, but from the interface it is hard to interpret the results for one and also to discern how these were chosen, i.e., based on what criteria. As it stands, the presentation in that project could use improvement and the results are displayed as a matter-of-fact without really trying to reason with the user as to why they were chosen and the limitations of the system. One can hypothesize a similar system aimed at displaying political facts could sway people's opinions depending on what facts are displayed and how they are displayed. Efforts to add more transparent context to current events, like congressional votes, have been implemented.[21] There are also

[16] https://github.com/marcotcr/lime

[17] https://www.politifact.com/

[18] https://idir-server2.uta.edu/factchecker/

[19] https://bit.ly/2qKwHFU

[20] https://bit.ly/2AOTSEE

[21] https://today.duke.edu/2016/09/icheck

projects that aim to add narratives to facts (Hassan et al. 2014), but such efforts should tread more carefully as bias can creep in the narrative generation portion of these types of efforts.

As mentioned earlier, large-scale knowledge graphs (KG) are an important resource for fact checking. Despite their large sizes, KGs are usually far from complete. In fact-checking, using an incomplete KG may lead to flagging a true claim as false due to missing triples that otherwise should be present in the KG. The challenges manifested by incomplete KGs have motivated many researchers to propose methods to link prediction, i.e., finding missing triples. Numerous machine learning models were proposed for link prediction, e.g., TransE (Bordes et al. 2013) and DistMult (Yang et al. 2015). A benchmark dataset, called Freebase15K, has been widely employed to train and evaluate these models. Unfortunately, a major problem exists in the dataset. For a majority of triples in the test set, their inverse triples also exist in the training set, e.g., (Microsoft, founded-by, Bill Gates) in training set and (Bill Gates, founded, Microsoft) in test set. These inverse triples present a serious selection bias, as the link prediction task on these triples may degenerate into trivially looking up their inverse triples in the training set. A few groups have investigated the impact of this bias (Toutanova and Chen 2015; Dettmers et al. 2018; Akrami et al. 2018). Their experiment results reveal that, for many papers in the literature that reported outstanding link prediction accuracy of their models, the models' performance became substantially worse after the inverse triples are taken out. For that reason, many models that supposedly improve upon TransE actually perform worse. This instance, highly related to fact-checking, serves as an excellent reminder for researchers to construct truly realistic datasets in building and evaluating AI models.

Possible Solutions

The one key theme throughout our own work, and one we have reiterated here as well, has been that we need more collective efforts to establish standards and use the ones that are around currently and are sensible (e.g., IFCN Code of Principles). Furthermore, adopting tools like the ClaimReview schema would benefit our community by providing a publicly available, structured dataset of fact-checked claims. For example, if a user posts a question of a fact-checked claim, previously encoded with ClaimReview schema, the Google search engine [22] will return a summarized version of the claim. ClaimReview [23] is a structured markup standard that promotes data uniformity across the web content. Ideally, these standards and tools won't just be de-facto baselines, but ones that are enforced or at least used to gauge the quality of work that is presented, with projects lacking their use suffering a credibility or usability penalty. We need to have ways to measure the viability and reliability of what is being generated in this domain, and to this end we need to have a joint discussion on what we ourselves expect from our peers and what the public expects from us. There is no point in creating all these tools, if in the end we lose the battle for trust in our systems. We cannot afford to mistakenly measure our success in echo-chambers of our peers, so we must be inclusive in how we grade ourselves with respect the public perception of the work that we produce.

The work that is coming out of "Share the Facts",[24] is a good start with respect to fact-checking related sharing of information. This template could certainly still be improved, but that should come once more people take up the joint effort. Some other issues that need addressing are reaching some consensus on what the average opinion on what is check-worthy and how this type of data should be reliably collected and labeled. Perhaps a simple census for the opinion of users would suffice for the first, but the latter issue still poses some challenges which might only be addressed by accepting the bias but handling it responsibly. By responsibly we mean that if there is some bias towards some specific type of claim then this should be clearly stated, and ideally a gamut of models, each addressing the deficits/biases in the others, would be generated. Projects should also try to start incorporating tools like LIME[16] to enhance the explainability of their methods and results. Other efforts that require the collective efforts of our community, are establishing a central repository for datasets that can be used to train models, producing highly reproducible results through, and establishing a complete pipeline for the data collection process.

Conclusion

Throughout this paper we have presented some of the key pitfalls that can plague fact-checking projects in all their forms. We have also presented some preliminary suggestions for solutions and shared our own experiences when dealing with these issues. With regards to our own research, we are going to continue to strive to produce work that is as unbiased as possible, but also begin to take up more standards. Currently we have a few projects that make use of some well established frame-works and knowledge sources with the hope that, by adopting standards rather than choosing to create ones tailored for our projects, we are joining a larger effort in consolidating the methods and techniques used in fact-checking and related tasks. Our goal is to work towards a better environment for fact-checking that can benefit all who do work or conduct research in the domain.

References

Akrami, F.; Guo, L.; Hu, W.; and Li, C. 2018. Re-evaluating embedding-based knowledge graph completion methods. In *CIKM*, 1779–1782.

Babakar, M., and Moy, W. 2016. The state of automated factchecking. *Full Fact.*

Baker, C. 2014. Framenet: a knowledge base for natural language processing. In *Proceedings of Frame Semantics in NLP: A workshop in honor of Chuck Fillmore*, 1–5.

[22] https://developers.google.com/search/docs/data-types/factcheck

[23] https://schema.org/ClaimReview

[24] http://www.sharethefacts.org/

Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*, 2787–2795.

Chakraborty, A.; Paranjape, B.; Kakarla, S.; and Ganguly, N. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *ASONAM*, 9–16.

Clark, K., and Manning, C. D. 2015. Entity-centric coreference resolution with model stacking. In *ACL*, 1405–1415.

Dettmers, T.; Pasquale, M.; Pontus, S.; and Riedel, S. 2018. Convolutional 2d knowledge graph embeddings. In *AAAI*.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; and Flammini, A. 2016. The rise of social bots. *Commun. ACM* 59(7):96–104.

Gencheva, P.; Nakov, P.; Màrquez, L.; Barrón-Cedeño, A.; and Koychev, I. 2017. A context-aware approach for detecting worth-checking claims in political debates. In *Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP)*, 267–276.

Hassan, N.; Sultana, A.; Wu, Y.; Zhang, G.; Li, C.; Yang, J.; and Yu, C. 2014. Data in, fact out: Automated monitoring of facts by FactWatcher. *Proceedings of the VLDB Endowment (PVLDB)* 7(13):1557–1560.

Hassan, N.; Adair, B.; Hamilton, J. T.; Li, C.; Tremayne, M.; Yang, J.; and Yu, C. 2015. The quest to automate fact-checking. In *Proceedings of the 2015 Computation+Journalism Symposium*.

Hassan, N.; Arslan, F.; Li, C.; and Tremayne, M. 2017a. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 1803–1812.

Hassan, N.; Zhang, G.; Arslan, F.; Caraballo, J.; Jimenez, D.; Gawsane, S.; Hasan, S.; Joseph, M.; Kulkarni, A.; Nayak, A. K.; Sable, V.; Li, C.; and Tremayne, M. 2017b. Claimbuster: The first-ever end-to-end fact-checking system. *Proceedings of the VLDB Endowment (PVLDB)* 10(12):1945–1948.

Hassan, N.; Li, C.; and Tremayne, M. 2015. Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th ACM conference on Information and knowledge management (CIKM)*, 1835–1838.

Ji, Y., and Eisenstein, J. 2013. Discriminative improvements to distributional sentence similarity. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 891–896.

Jimenez, D., and Li, C. 2018. An empirical study on identifying sentences with salient factual statements. In *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*.

Jo, S.; Trummer, I.; Yu, W.; Liu, D.; and Mehta, N. 2018. The factchecker: Verifying text summaries of relational data sets. *arXiv preprint arXiv:1804.07686*.

Ma, J.; Gao, W.; and Wong, K.-F. 2018. Rumor detection on twitter with tree-structured recursive neural networks. In *ACL*, 1980–1989.

Mudgal, S.; Li, H.; Rekatsinas, T.; Doan, A.; Park, Y.; Krishnan, G.; Deep, R.; Arcaute, E.; and Raghavendra, V. 2018. Deep learning for entity matching: A design space exploration. In *SIGMOD*, 19–34.

Padó, S.; Galley, M.; Jurafsky, D.; and Manning, C. D. 2009. Textual entailment features for machine translation evaluation. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, StatMT, 37–41.

Pearl, J. 2014. Comment: Understanding simpsons paradox. *The American Statistician* 68(1):8–13.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. "why should I trust you?": Explaining the predictions of any classifier. In *KDD*, 1135–1144.

Rony, M. M. U.; Hassan, N.; and Yousuf, M. 2018. Baitbuster: A clickbait identification framework. In *AAAI*, 8216–8217.

Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.* 19(1):22–36.

Socher, R.; Huang, E. H.; Pennin, J.; Manning, C. D.; and Ng, A. Y. 2011. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Advances in neural information processing systems*, 801–809.

Suwajanakorn, S.; Seitz, S. M.; and Kemelmacher-Shlizerman, I. 2017. Synthesizing obama: Learning lip sync from audio. *ACM Trans. Graph.* 36(4):95:1–95:13.

Thorne, J., and Vlachos, A. 2017. An extensible framework for verification of numerical claims. In *EACL*, 37–40.

Toutanova, K., and Chen, D. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, 57–66.

Wang, X.; Yu, C.; Baumgartner, S.; and Korn, F. 2018. Relevant document discovery for fact-checking articles. In *Companion Proceedings of the The Web Conference*, 525–533.

Yang, B.; Yih, W.-t.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*.

Zhang, G.; Jimenez, D.; and Li, C. 2018. Maverick: Discovering exceptional facts from knowledge graphs. In *Proceedings of the 2018 ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 1317–1332.

Zhao, Z.; Resnick, P.; and Mei, Q. 2015. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *WWW*, 1395–1405.