Claim Check-worthiness in Podcasts: Challenges and Opportunities for Human-AI Collaboration to Tackle Misinformation

Ujwal Gadiraju,[1] **Vinay Setty,** [2,3] **Stefan Buijsman** [1]

[1] Delft University of Technology
[2] Factiverse AS, [3] University of Stavanger
u.k.gadiraju@tudelft.nl, vinay@factiverse.no, s.n.r.buijsman@tudelft.nl

Abstract

The rapid proliferation and adoption of large language models (LLMs) and generative AI across various domains have pivotal implications for the diffusion and propagation of misinformation. The last decade has showcased numerous harmful consequences of widespread misinformation. In this context, fact-checking is essential to assessing the factual accuracy of different content and mitigating false information. Several automated fact-checking systems and pipelines have been developed to this end, albeit with limited success. Challenges pertaining to different stages of such pipelines remain unsolved — from claim detection to evidence retrieval, verdict prediction, and production of justifications. In this work, we consider the context of *audio podcasts* as a medium that poses unique challenges and opportunities for building automated fact-checking systems. We focus on claim check-worthiness, and propose a system to elicit check-worthiness annotations from non-expert crowd workers. Finally, we discuss open research questions and future directions to facilitate the development of multi-modal fact-checking systems by using the lens of audio podcasts.

1 Introduction and Background

We are now gripped by the extreme narratives surrounding AI boons and banes, in a reality that lies somewhere in between (Shneiderman 2020). While striving to build human-centered AI systems that can better augment human experiences and assist us in completing various tasks in our everyday lives, it is important to consider the unintended impact of technological advances on society. A consequence of advances in generative AI and the democratization of large language models (LLMs) is the growing ease with which more information can be generated, diffused, and consumed. This has critical implications for the propagation of misinformation with potentially damaging ramifications (Bergstrom and West 2023). One such medium in which information is increasingly being consumed is podcasts. Audio/video podcasts can be thematically diverse and vary significantly in their formats (Tian, Hauff, and Chandar 2022). Recent work has also identified a growing trend across the world to consume news via short online videos (Newman et al. 2024). What is arguably common to most podcasts, especially those that have the potential to inform or otherwise mislead listeners, is the need to fact-check utterances by speakers on the podcasts. It is tedious to listen to a podcast and manually identify potentially controversial and check-worthy claims made in the podcast. Automating this step could save a significant amount of time for fact-checkers, journalists, and podcast platforms to quickly assess whether the podcast is a candidate for a fact-check. In this work, we consider this unique context of building automated fact-checking systems for audio podcasts and explore the first step of assessing the check-worthiness of claims in podcasts.

Claim Check-worthiness

To ensure the veracity of the information created, generated, propagated, and/or consumed by users, automatic fact-checking systems have been developed in recent years (Zeng, Abumansour, and Zubiaga 2021; Guo, Schlichtkrull, and Vlachos 2022). Claim check-worthiness is an important component of such systems to reduce costs and optimize the use of resources. For instance, it would be computationally expensive to fact-check every single claim in a sea of information on one hand, and it would be rather rudimentary to squander the time of expert fact-checkers to serve this purpose on the other hand. As a result, understanding which claims are worth checking is the first step that is necessary, and this is a task that has gained prominence in recent years. Researchers and practitioners in the natural language processing (NLP) and machine learning communities continue to build systems capable of automatically detecting checkable and check-worthy claims that warrant further inspection for their factual accuracy (Hassan et al. 2015, 2017; Kotonya and Toni 2020). A vital ingredient in this process is the availability of labeled data in the context of interest – news articles, social media posts, or discussions on forums. Crowdsourcing has emerged as a scalable means to acquire labeled data by leveraging human input through existing marketplaces such as Amazon Mechanical Turk, Prolific, Toloka AI, or other means (Demartini et al. 2017; Pinto et al. 2019; Godel et al. 2021; Allen, Martel, and Rand 2022).

Fact-checking Podcasts — A New Frontier

Podcasts are becoming increasingly popular and are estimated to become a 4 billion dollar industry by 2024 (Shapiro

2023). Considering the widespread popularity of podcasting, the growth in platforms, and the increasing amount of content available on many topics, fact-checking podcast content is extremely important (Cherumanal, Gadiraju, and Spina 2024). Limiting the propagation of misinformation via podcasts, safeguarding listeners, and empowering them to make informed decisions by providing factually accurate information is a valuable goal to strive for. In this work-in-progress, we describe the task design for claim check-worthiness of podcasts that can be deployed on a crowdsourcing platform. Next, we present a synthesis of some key challenges and open research questions for the benefit of the community.

2 Claim Check-worthiness Task Design

We built a web application, with a React.js front-end to facilitate the claim check-worthiness annotation process, catering to the first step in the effective identification of check-worthy claims. The back-end of the application uses an API developed with Python and the Django REST framework.

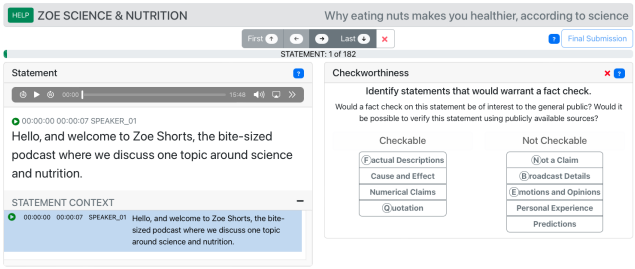

Figure 1: The annotation interface that workers can use to assess the check-worthiness of statements in the podcasts.

On starting the task and after reading through the instructions, workers are presented with the annotation interface (see Figure 1). A panel on the left-hand side of the screen presents the "statement" that needs to be assessed for check-worthiness. Workers can listen to the audio snippet of the statement from the podcast in addition to reading it if they wish to do so. At the bottom of the panel, workers can scroll through the preceding context (when it exists) to help obtain more context and make an informed assessment when in doubt. The right-hand side of the screen presents the check-worthiness assessment panel. The podcast is presented to workers as a chronologically ordered series of utterances (one statement at a time).

Checkable and Not Checkable Statements

The first decision that a worker has to make is whether or not the statement is verifiable using publicly available sources (i.e., determine whether the statement is '`CHECKABLE`' or '`NOT CHECKABLE`'). Following this decision, the worker needs to provide a rationale for their assessment of check-worthiness. `CHECKABLE` statements match one of the four following characterizations:

- **Factual Descriptions**: Claims about the existence or characteristics of notable people, places, things, events, or actions, and which are possible to verify with public sources. For example, "*She won the London Marathon last year.*" — "*Rival groups were involved in a gunfight on the outskirts of the city.*"
- **Cause and Effect**: Claims asserting one thing is caused by or linked with another, which can be checked against reputable sources. For example, "*The company collapsed after a rogue employee was discovered to be embezzling funds.*" — "*Smoking causes cancer.*"
- **Numerical Claims**: Claims which involve specific statistics or would require counting or analysis of numerical data to verify. For example, "*The average Mexican consumes more sugar per day than the average American.*" — "*The latest poll shows that 80% of people are unhappy with the current government.*"
- **Quotation**: Repeating the words of another notable person or entity which can be verified in public sources. For example, "*The mayor was clear when he said, 'All flooded households will receive emergency assistance after a damage assessment.'*" — "*President Roosevelt famously said, 'Ich bin ein Berliner.'*"

Statements that are labeled `NOT CHECKABLE` match one of the following five characterizations:

- **Not a Claim**: Not making any sort of claim, including questions not including some factual assertion. For eaxmple, "*Hello, how are you?*" — "*How old are you?*" — "*Thanks for chatting with us today.*"
- **Broadcast Details**: Introducing the speakers, describing the program, or giving details related to the episode contents. For eaxmple, "*Welcome to the show, I'm your host, John Smith.*" — "*Today we're going to be talking about the history of the internet.*"
- **Emotions and Opinions**: An emotion that is being felt or expressed, or an opinion that doesn't contain a checkable factual assertion. For example, "*I love how the tulips look early on a spring morning.*" — "*He's really upset about the way things are going at school.*"
- **Personal Experience**: Claims a person makes about their own experience, but which cannot be verified in public sources. For example, "*I passed four empty buses on my way to work yesterday.*" — "*My grandmother used lard in her pie crusts.*"
- **Predictions**: Claims and predictions about future events or plans that can't be confirmed at present. For example, "*Elon Musk will visit Mars.*" — "*New car sales will increase every month going forward.*"

The distinctions between `CHECKABLE` and `NOT CHECKABLE` statements stem from the practices employed by professional fact-checkers. In this endeavor, we collaborated closely with fact-checkers from Faktisk.no,[1] who regard this classification as the initial step in selecting claims for fact-checking. The rationale for this approach is to effectively sift through sentences and identify those that may not qualify as claims or present difficulties in verification, such as opinions and predictions made by the speakers.

[1] `Faktisk.no AS` is a non-profit organization and independent newsroom for fact-checking the social debate and public discourse in Norway.

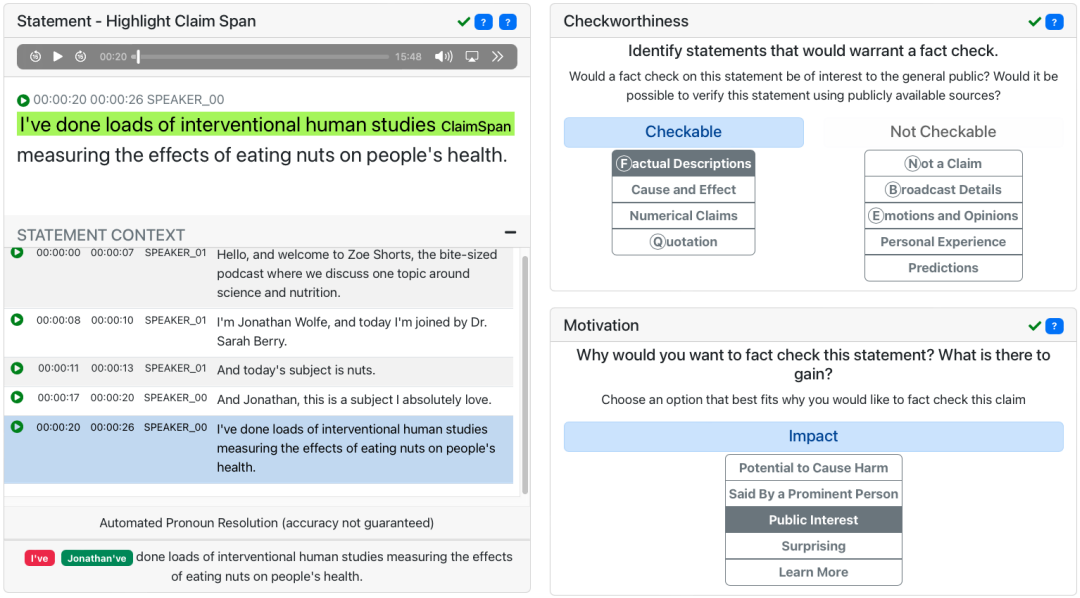

Figure 2: The annotation interface that workers can use to assess the check-worthiness of statements in the podcasts.

By doing so, we ensure that the fact-checking process focuses on statements that can be objectively validated by providing evidence. When workers decide that a particular statement is CHECKABLE, they are asked to complete two steps. As shown in Figure 2, workers first highlight the part of the podcast statement that is worth fact checking. Next, they explain their motivation for fact-checking the statement by choosing one of the following five options:

1. **Potential to Cause Harm**: I think this statement could cause harm if false.
2. **Said by a Prominent Person**: I want to check if this prominent person actually said this.
3. **Public Interest**: I believe the fact checking of this claim is for the public interest.
4. **Surprising**: I find this statement surprising, shocking, or otherwise hard to believe.
5. **Learn More**: I would gain new knowledge about this topic by fact checking this statement.

Annotations gathered in this manner can play a crucial role in efforts to enhance and refine check-worthy claim detection models in automated fact-checking pipelines within podcasts and other media. The ultimate goal is to quickly identify podcasts that might contain potentially inaccurate information, along with the claims that are inaccurate with their supporting rationales. This will also solve challenges which fact-checkers such as Faktisk.no and many other media companies face when having to fact-check lengthy podcasts and other media.

These fine-grained annotations alongside the rationale for considering a claim as check-worthy, can also serve as valuable resources for understanding model behavior from an XAI perspective. Through the creation of diagnostic datasets, we can gain insights into the performance of production models and take actions to enhance the models' robustness, improving automated fact-checking systems.

3 Open Research Questions

1. How can we use crowdsourced input from non-experts to build effective **multimodal** and **multilingual** automated or hybrid human-AI fact-checking systems? Although recent work has begun addressing challenges in this realm (Kazemi et al. 2021; Yao et al. 2023; Mubashara et al. 2023), there is plenty of ground to cover before we can fully solving this spectrum of problems.
2. How can we foster **appropriate trust** and **reliance** on automated fact-checking systems among fact-checkers and other stakeholders for such systems? Fact-checking can serve as a unique context where users need to be supported with XAI tools and techniques that can prevent under-reliance and reduce over-reliance on automated systems (Robbemond, Inel, and Gadiraju 2022).
3. How can we best **engage listeners** in fact-checking podcasts? Unlike other social media platforms, podcast listeners often have a limited means to challenge or respond to contentious statements. On the one hand, there is a need for real-time, interactive inspection from listeners that can help develop robust and automated fact-checking systems for podcasts. On the other, there are several intriguing questions surrounding how we can augment information about the factual accuracy of claims made in podcasts without hampering users' listening experience.
4. How can we **prioritize** between different statements that can be fact-checked? The capacity of fact-checkers is limited and in a growing podcast landscape, we have to choose. In addition, listeners should not be overloaded with fact-checks. Choosing which statements are the most crucial to check is value-laden, as exemplified by the question of how to weigh the five different motivations for fact-checking against each other.

Acknowledgements

We thank Adam Becker from the University of Stavanger for his substantial contributions in developing the web application. This work was partially supported by the Digital Ethics Center, the TU Delft AI Initiative, and Toloka AI. The first author was partially affiliated with Toloka AI in Amsterdam, NL at the time of contributing to this work.

References

Allen, J.; Martel, C.; and Rand, D. G. 2022. Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–19.

Bergstrom, C. T.; and West, J. D. 2023. How publishers can fight misinformation in and about science and medicine. *Nature Medicine*, 1–3.

Cherumanal, S. P.; Gadiraju, U.; and Spina, D. 2024. Everything We Hear: Towards Tackling Misinformation in Podcasts. *In International Conference On Multimodal Interaction (ICMI '24)*.

Demartini, G.; Difallah, D. E.; Gadiraju, U.; Catasta, M.; et al. 2017. An introduction to hybrid human-machine information systems. *Foundations and Trends® in Web Science*, 7(1): 1–87.

Godel, W.; Sanderson, Z.; Aslett, K.; Nagler, J.; Bonneau, R.; Persily, N.; and Tucker, J. A. 2021. Moderating with the mob: Evaluating the efficacy of real-time crowdsourced fact-checking. *Journal of Online Trust and Safety*, 1(1).

Guo, Z.; Schlichtkrull, M.; and Vlachos, A. 2022. A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, 10: 178–206.

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*. Citeseer.

Hassan, N.; Arslan, F.; Li, C.; and Tremayne, M. 2017. 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*, 1803–1812.

Kazemi, A.; Garimella, K.; Gaffney, D.; and Hale, S. A. 2021. Claim matching beyond English to scale global fact-checking. *arXiv preprint arXiv:2106.00853*.

Kotonya, N.; and Toni, F. 2020. Explainable automated fact-checking: A survey. *arXiv preprint arXiv:2011.03870*.

Mubashara, A.; Michael, S.; Zhijiang, G.; Oana, C.; Elena, S.; and Andreas, V. 2023. Multimodal Automated Fact-Checking: A Survey. *arXiv preprint arXiv:2305.13507*.

Newman, N.; Fletcher, R.; Robertson, C.; Arguedas, A.; and Nielsen, R. 2024. Reuters Institute Digital News Report 2024. *Reuters*.

Pinto, M. R.; de Lima, Y. O.; Barbosa, C. E.; and de Souza, J. M. 2019. Towards fact-checking through crowdsourcing. In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 494–499. IEEE.

Robbemond, V.; Inel, O.; and Gadiraju, U. 2022. Understanding the Role of Explanation Modality in AI-assisted Decision-making. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 223–233.

Shapiro, A. 2023. Podcasting will be a 4 billion industry by 2024. https://www.theverge.com/2022/5/10/23065056/podcasting-industry-iab-report-audacy-earnings-patreon-pulitzer. Accessed: 2023-08-15.

Shneiderman, B. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6): 495–504.

Tian, M.; Hauff, C.; and Chandar, P. 2022. On the Challenges of Podcast Search at Spotify. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 5098–5099.

Yao, B. M.; Shah, A.; Sun, L.; Cho, J.-H.; and Huang, L. 2023. End-to-end multimodal fact-checking and explanation generation: A challenging dataset and models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2733–2743.

Zeng, X.; Abumansour, A. S.; and Zubiaga, A. 2021. Automated fact-checking: A survey. *Language and Linguistics Compass*, 15(10): e12438.