PodChecker: An Interpretable Fact-Checking Companion for Podcasts

Anna Irmetova[1], **Haoran Liu**[1], **Maria Teleki**[*1], **Peter Carragher**[*2], **Julie Zhang**[*3], **James Caverlee**[1]

[1]Texas A&M University
[2]Carnegie Mellon
[3]Independent
{annairmetova, liuhr99, mariateleki, caverlee}@tamu.edu, zhangjunyan0225@gmail.com, petercarragher@cmu.edu

Abstract

We present PodChecker, a user-facing system for automated, claim-level fact-checking of podcast content. PodChecker processes podcast audio or RSS feeds by transcribing episodes, extracting atomic factual claims, and assigning each claim one of four fine-grained labels – *true, false, misleading/partially true, or unverifiable* – using retrieval-augmented verification. The system presents fact-checking results at the level of individual claims, accompanied by simple visual indicators and links to supporting/conflicting sources. This design, implemented via an interactive web-based interface, enables users to inspect fact-checking outputs and underlying evidence directly, supporting interpretable and critical engagement with long-form audio content. By presenting claim-level evidence and labels, PodChecker assists both general listeners and professional fact-checkers in assessing podcast factuality.

Code — https://github.com/annatastic/PodChecker

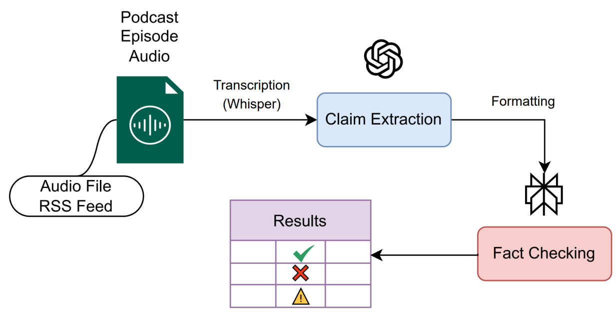

Figure 1: **PodChecker transforms podcast audio into interpretable, claim-level fact-checks** by performing (i) transcription via `whisper-small-en`, (ii) atomic factual claim extraction via `gpt-5-mini`, and (iii) retrieval-augmented verification of the extracted claims via `sonar`.

Introduction

Podcasts have become an important part of online information ecosystems, serving as a venue for news consumption, political discussion, and public commentary (Shearer et al. 2023; Litterer, Jurgens, and Card 2024). While audience shares for traditional information media such as radio are in decline, audience shares for podcasts are on the rise (Aubin 2023), increasing their role in shaping public understanding of current events. The rise of podcasts as a dominant information medium brings with it the challenge of maintaining trustworthiness and credibility.

Compared to text-based media, podcasts pose distinct challenges for assessing information quality. Episodes are often long and conversational, with factual claims embedded alongside opinions and anecdotes (Setty and Becker 2025). Podcasts also tend to elicit relatively high levels of listener trust, often attributed to parasocial relationships between hosts and audiences (Johnson and McCall 2025; Schlütz and Hedder 2022). However, podcast production and distribution generally involve limited editorial oversight. Together, these characteristics amplify the potential impact of unverified or misleading information in podcasts and underscore the necessity of effective fact-checking mechanisms tailored to this medium.

Fact-checking initiatives have traditionally relied on crowdsourced efforts or specialized organizations such as *Snopes* and *PolitiFact* (Pilarski, Solovev, and Pröllochs 2024; Lee et al. 2025). Rather than replacing these important actors, recent work has explored how *automated systems can scale and support fact-checking efforts.* In particular, advances in large language models (LLMs) have enabled automated fact-checking and claim verification at scale (Kim et al. 2024; Vykopal et al. 2024). LLM-based approaches typically decompose content into atomic factual claims (Min et al. 2023; Tang, Laban, and Durrett 2024) and verify each claim using retrieved external evidence (Lewis et al. 2020; Li et al. 2024). However, these methods have largely been developed and evaluated on written text, and have not been fully adapted to long-form spoken content such as podcasts. Meanwhile, existing computational work on podcast fact-checking has primarily focused on benchmark creation and annotation (Chun et al. 2025; Setty and Becker 2025) rather than developing end-to-end verification tools.

To bridge this gap, we present ***PodChecker***, an LLM-powered fact checking tool for podcasts and similar audio

[*]These authors contributed equally.

```
You are a fact-checker. Please check the
accuracy of the provided claim, taking
into account any previous claims provided
as context. Return one of these ratings:
true, false, misleading/partially true,
unverifiable.
Do not add any explanations or extraneous
formatting.
Here are 4 examples:
Input: "Antonio Guterres is the UN
Secretary General"
Output: "true"
...
```

Figure 2: k**-Shot Prompt for Fact Verification:** To verify claims generated in the claim extraction step, we instruct the LLM to act as a fact checker and assign a label to each one. Examples written by the authors are provided for guidance.

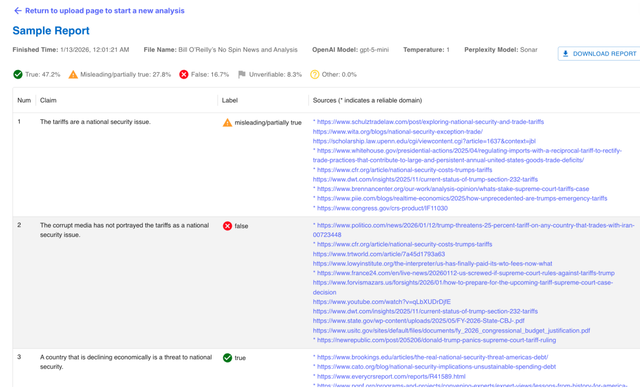

Figure 3: **PodChecker User Interface**: PodChecker ingests podcast audio or RSS feeds and outputs fact-checking results in a table with visual indicators and relevant links.

content. This end-to-end system leverages the strengths of existing LLMs for claim-level, retrieval-augmented verification. It allows users to upload samples of their choice and performs claim extraction and verification, detecting true as well as misleading and inaccurate claims and displaying them in an intuitive user interface. PodChecker emphasizes epistemic agency by enabling users to inspect the evidence underlying each assessment, providing links for each claim. In addition, the system presents an aggregated factuality overview for each episode, supporting both high-level assessment and deeper claim-level exploration.

PodChecker: An Interpretable Tool for Automated Fact Checking

Podcast Location & Transcription. PodChecker takes an input from the user in the form of an English language audio file or an RSS feed link to the podcast of their choice. If the RSS link is provided, the most recent episode is selected. Users must also provide their API keys for OpenAI and Perplexity, which are used to make calls but are not stored. The audio is then transcribed using Whisper's `small.en` model.

Claim Extraction. Once a transcript has been generated, an API call is made to `gpt-5-mini` to perform atomic claim extraction. Here, `gpt-5-mini` is given a 1-shot prompt to break down the transcript into independent factual claims, building on the Min et al. (2023) prompt.

Fact Verification. Perplexity's `Sonar` model is used to check each claim one by one and assign one of four labels: TRUE, FALSE, MISLEADING/PARTIALLY TRUE, and UNVERIFIABLE. As shown in Figure 2, we implement a 4-shot prompt with one example per label category, and included at most five most recent previous claims in the prompt for context. This model was chosen because of its real-time web search capabilities; we found that its retrieval approach helps improve speed and mitigate hallucination of source URLs.[1]

[1]We found that `gpt` and `gemini` models tended to hallucinate non-existent URLs.

We then collect both the generated label and the search result URLs that were gathered by `Sonar`. Using a dataset of source credibility ratings for news domains from Carragher, Williams, and Carley (2024), we identify search result URLs from domains that have a reliability score of $>= 5$ (on a scale of 1-6) and annotate them by stars on the UI to signal when the factuality label is backed by a credible source.

User Interface. The front end of this application provides an intuitive interface for users to submit podcast audios[2], view factuality analysis results, and download a report. PodChecker's landing screen features options to provide a podcast episode audio and API keys. We also provide three pre-processed examples drawn from short podcast excerpts (Clifton et al. 2020; Reuters 2025; O'Reilly 2025) for users to select and view.

After running the system on a specific podcast, the results are displayed in a table format with **each claim** that was extracted from the podcast sample, **its factuality label** accompanied by an intuitive symbol, and **the linked sources** used to support the verification. A panel at the top of the screen displays the percentage of claims falling into each label category, allowing for a quick high-level judgment of the episode's factuality. Metadata – such as processing time – is also displayed for reference. Users can opt to download the results as a report in `json` format.

Example Users

PodChecker is primarily useful for assisting with the fact checking process for podcast episodes and other spoken content discussing politics and current events. Fact checking is often time consuming and laborious – our tool speeds up this process. We view it as a starting point for two key users: general podcast listeners and professional fact-checkers.

[2]Although the system is API-based, it operates only on user-supplied inputs and does not ingest, store, or redistribute third-party datasets. Users may submit podcast audio or excerpts that they are authorized to use in accordance with the original data licenses.

General Podcast Listeners. It is not always clear how factual the information in a podcast is: PodChecker helps address this concern for general listeners, who can use the tool to gauge the approximate accuracy of the host's claims, with a quick visual gauge, and the sources allow listeners to further explore topics, supporting their epistemic agency. Listeners can also use PodChecker to assist with the process of choosing a podcast to listen to.

Professional Fact-Checkers. Rather than conducting the entire fact-checking process manually, fact-checkers and journalists who work with podcasts can use this tool to assist them in their work. One approach could be to use PodChecker results as a draft or reference to catch 'obvious' cases, and then manually refine the more ambiguous cases. We hope that using PodChecker as a companion can help professionals make their work more scalable.

Conclusion

With podcast popularity showing no signs of slowing down, the question of factuality and misinformation in this unique medium is highly relevant. The advances of LLMs have made it possible to automate the detection and verification of factual claims. PodChecker is an end-to-end tool that leverages these capabilities to fact-check podcast episode audio. PodChecker follows a three-stage pipeline of transcription, atomic claim extraction, and claim-level retrieval-augmented verification, and presents the results via an intuitive UI. Future work may systematically evaluate LLM strengths and limitations in fact-checking long-form conversational audio. Our system serves as a move toward combating misinformation in the podcast medium, encouraging accountability for creators and empowering listeners to take control of their information diets.

Limitations and Future Work

As with prior automated fact-checking systems, PodChecker's assessments may reflect errors arising from transcription noise, claim decomposition, or limitations of the retrieval-based verification model, and should be interpreted as probabilistic signals rather than definitive judgments. At the moment, the tool is only available for audio in English. In future iterations, we aim to add new features, such as feedback functionality for users and agentic sourcing of RSS feeds based on podcast titles.

A natural next direction for PodChecker is a performance analysis: future work can compare how human annotators and LLMs understand podcast episode content, extract factual claims, and assign source-supported labels. Work like this can lead to empirical system improvements as well as a greater understanding of LLMs and fact-checking in the context of podcast media.

Acknowledgments

We thank Majid Alfifi and Xiangjue Dong for the discussion. AI tools were used to assist with portions of the coding and to support the organization, editing, and refinement of the manuscript. All generations were carefully reviewed, verified, and, where necessary, modified by the authors, who take full responsibility.

References

Aubin, C. S. 2023. Audio and Podcasting Fact Sheet. Fact sheet, Pew Research Center. Accessed 2026-01-05.

Carragher, P.; Williams, E. M.; and Carley, K. M. 2024. Detection and discovery of misinformation sources using attributed webgraphs. In *ICWSM*.

Chun, C.; et al. 2025. MAD: A Benchmark for Multi-Turn Audio Dialogue Fact-Checking. In *arXiv*.

Clifton, A.; et al. 2020. The spotify podcast dataset. *arXiv preprint arXiv:2004.04270*.

Johnson, K.; and McCall, M. 2025. Trust in Pod: Listener Trust of News Content Heard on Different Genre Podcasts. *Media and Communication*, 13(0).

Kim, K.; et al. 2024. Can LLMs produce faithful explanations for fact-checking? Towards faithful explainable fact-checking via multi-agent debate. *arXiv*.

Lee, S.; Seo, H.; Xiong, A.; and Lee, D. 2025. Partisan Fact-Checkers' Warnings Can Effectively Correct Individuals' Misbeliefs About Political Misinformation. *ICWSM*.

Lewis, P.; et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *NeurIPS*.

Li, Z.; et al. 2024. Retrieval augmented generation or long-context llms? a comprehensive study and hybrid approach. *arXiv preprint arXiv:2407.16833*.

Litterer, B.; Jurgens, D.; and Card, D. 2024. Mapping the podcast ecosystem with the structured podcast research corpus. *arXiv preprint arXiv:2411.07892*.

Min, S.; et al. 2023. FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation. In *EMNLP*. Association for Computational Linguistics.

O'Reilly, B. 2025. Bill O'Reilly's No Spin News and Analysis. https://www.billoreilly.com/.

Pilarski, M.; Solovev, K. O.; and Pröllochs, N. 2024. Community Notes vs. Snoping: How the Crowd Selects Fact-Checking Targets on Social Media. *ICWSM*.

Reuters. 2025. Reuters World News Podcast. https://www.reuters.com/podcasts/.

Schlütz, D.; and Hedder, I. 2022. Aural Parasocial Relations: Host–Listener Relationships in Podcasts. *Journal of Radio & Audio Media*.

Setty, V.; and Becker, A. J. 2025. Annotation Tool and Dataset for Fact-Checking Podcasts. In *Companion Proceedings of WWW 2025*. ACM.

Shearer, E.; et al. 2023. Podcasts as a Source of News and Information. Research Report April 18, 2023, Pew Research Center. Accessed 2026-01-05.

Tang, L.; Laban, P.; and Durrett, G. 2024. Minicheck: Efficient fact-checking of llms on grounding documents. *arXiv*.

Vykopal, I.; et al. 2024. Generative large language models in automated fact-checking: A survey. *arXiv*.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes; we discuss the methodology of the system to back up claims about its usage.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes; we acknowledge that podcast inputs reflect only English language audio on politics and current events topics.
   (e) Did you describe the limitations of your work? Yes, the limitations are described in the section "Limitations and Future Work."
   (f) Did you discuss any potential negative societal impacts of your work? No, because this is a demo paper with limited space.
   (g) Did you discuss any potential misuse of your work? No, because this is a demo paper with limited space. However, we mention that PodChecker is a companion in the fact-checking process that produces probabilistic signals to discourage user overreliance on its automated assessments.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, we mention that the system does not store users' personal information (API keys) or third-party datasets
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes. Creators of datasets and podcasts are cited as references; AI/ML tools used in the system are cited via URLs in Table 1
   (b) Did you mention the license of the assets? Yes, where applicable. The license for the Spotify Podcast Dataset is discussed in Footnote 2 and Table 1
   (c) Did you include any new assets in the supplemental material or as a URL? Yes, they are included as a GitHub repository link.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, where applicable. The Spotify Podcast Dataset is used in accordance with its publicly documented license. The remaining podcast content consists of accessible episodes produced for public dissemination, and was limited to short excerpts and cited.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? No; the assets used do not contain personal or private information or offensive content
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see **?**)? NA

Table 1: **External Components Used in PodChecker**

Tool	Purpose	Link
Whisper `small.en`	A speech recognition model used to transcribe audio inputs.	https://huggingface.co/openai/whisper-small.en
`gpt-5 mini`	A LLM from OpenAI used to extract atomic claims from transcripts.	https://platform.openai.com/docs/models/gpt-5-mini
`Sonar`	A search LLM from Perplexity with real time web search capabilities used to fact-check each claim.	https://docs.perplexity.ai/getting-started/models/models/sonar
Spotify Podcast Dataset	A corpus of 100,000 podcasts, one of which was sampled for demonstration of PodChecker, licensed under a Creative Commons Attribution license.	https://research.atspotify.com/publications/100000-podcasts-a-spoken-english-document-corpus https://creativecommons.org/licenses/by/4.0/

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA