Unmasking "Fake News": A Pipeline for Fact-Checking Radio Shows via AI-Driven Claim Extraction

Paul Okewunmi [1] Bunmi Akinremi [1] Caroline Anipah [1]

[1]Centre for Journalism Innovation and Development

Abstract

Addressing the broader challenge of information disorder, commonly called 'fake news' and which encompasses misinformation and disinformation, remains a critical concern for governments and organisations worldwide. Aside from traditional media (television, radio and newspapers), social media and Artificial Intelligence have enabled the speedy spread of information disorder, necessitating innovative solutions to the problem. As technology advances, Artificial Intelligence (AI) has risen to the challenge and is used to fact-check and combat misinformation and disinformation on social media, news websites, and blogs[4]. However, a largely untapped domain is radio, one of the most relied on sources of information in West Africa.

Our work focuses on advancing the frontier of radio fact-checking with AI by targeting specific talk shows in Nigeria and Ghana. Leveraging the Ajala.ai[1] tool, we transcribe multiple radio shows into text, combining this dataset with the established 'ClaimBusters'[2] benchmark dataset. Through training on Open AI's Ada model[3], DistilBERT, and BERT, we achieve impressive results, with the Open AI Ada model attaining an accuracy of 94%, while DistilBERT and BERT achieve accuracies of 77.8% and 81.1%, respectively. To operationalise our findings, we integrate our methodology into a software tool that autonomously monitors radio broadcasts, converts spoken content to text, extracts claims from the textual data, and subsequently presents these claims to human fact-checkers for meticulous review.

Introduction

Radio, a vital information source in West Africa, is notorious for spreading 'fake news', posing challenges for timely fact-checking due to its transient nature and the sheer number of stations in the sub-region. Automated fact-checking, a rapidly growing field within the expansive domain of Natural Language Processing (NLP), is revolutionising information fact-checking and verification. Yet, radio content remains largely uncharted territory, unlike text-focused pioneers like FullFact. It is in response to this challenge and gap identified by the Centre for Journalism Innovation and Development's (CJID) fact-checking practice, Dubawa (dubawa.org), that this project was developed.

The project introduces an automated radio fact-checker tool for intelligently monitoring Nigerian and Ghanaian stations for misinformation and disinformation. It employs AI to convert audio to text, identify potential falsehoods, and prioritise claims for verification by human fact-checkers. Given radio's role in disseminating 'fake news' with potential impacts on conflicts, public health, and democratic processes, this project effectively tackles the critical task of fact-checking radio claims. As pioneers of this innovative idea, CJID's Dubawa continues to lead in countering information disorder and facilitating accurate information dissemination.

Objectives

- **Objective 1:** Develop an automated radio fact-checker application to monitor radio stations in Nigeria and Ghana.
- **Objective 2:** Implement a working pipeline to convert audio content into text and identify potential false or misleading claims on the radio for human fact-checkers to verify, helping combat the spread of disinformation and misinformation.

Methodology

Our automated radio claim extraction process employs a scheduled script to capture diverse radio programs within set timeframes. It also allows audio file uploads for analysis, ensuring versatility in handling content sources. Recorded audio is transcribed using Ajala.ai, a voice automation solution for African languages, ensuring precise representation across diverse accents. Central to our pipeline is the claim extraction model, fine-tuned from OpenAI's GPT-3 base model, Ada. The model identifies fact-checkable claims within radio transcripts.

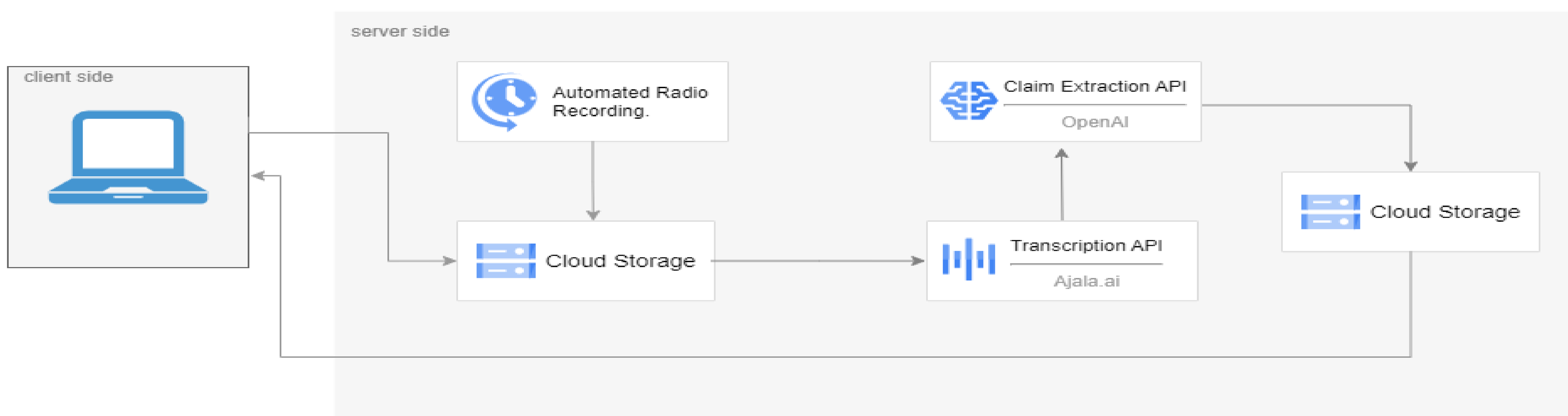

Figure 1. Architecture diagram

Data Collection

The data collection process involved carefully curating a comprehensive dataset for training the claim extraction model. We combined two distinct datasets to enable us to build a versatile and domain-aware model primed for detecting fact-checkable claims in radio programs with enhanced accuracy.

Dataset 1: Radio Transcriptions

We sourced transcriptions of diverse radio programs, capturing a wide range of content from various broadcasting stations. The radio transcriptions encompassed colloquial expressions, informal language, and diverse accents typical of spoken communication. This dataset provided valuable real-world examples of claims found in the context of radio broadcasts, aligning directly with our project's objectives.

Dataset 2: A Benchmark Dataset of Check-worthy Factual Claims

To supplement our radio-focused data, we integrated the ClaimBusters dataset—containing 23,533 sentences from conventional media, including claims from debates, categorized for fact-checking relevance into non-factual, unimportant factual, and check-worthy factual statements.

Figure 2. Radio Schedule Sample Data

Country	Radio	Recorded Hours
NGA	Brekete (Human Right Radio)	135
	State Affairs (Splash FM)	96
GHA	Press Pass (Asaase FM)	59
	Sunrise (3FM)	100
	Total	390

Figure 3. Summary of Claim Detection Dataset

Dataset	Claims	Non-Claims
ClaimBusters	1994	5982
Radio Extracted Claims	250	-
Combined Total	2244	5982

Results

During the experimentation phase, we explored several popular language models, including BERT and DistilBERT, to address the task of claim extraction. While these models showed promising results, we eventually decided to fine-tune OpenAI's GPT-3 base model, Ada. The decision was based on the understanding that the Large Language Model (LLM) capabilities of GPT-3, combined with the flexibility of fine-tuning Ada, would likely improve performance compared to the other models. Using the fine-tuned Ada model, we could avoid the additional effort and complexity of deploying and productionalising the model ourselves. OpenAI's infrastructure covers these aspects, allowing us to focus on the core task of claim extraction and maximise efficiency in our development process.

After training the fine-tuned Ada model, we evaluated its performance on a separate test dataset to assess its accuracy in identifying fact-checkable claims. The results demonstrated an impressive accuracy of 95%. Additionally, the model's precision, recall, and F1 score were measured at 88.4%, 87.2%, and 87.7%, respectively, indicating its effectiveness in accurately identifying and extracting claims for further fact-checking.

Table 1. Model performance

Model	Accuracy (%)	Recall (%)	Precision (%)
DistilBERT	77.8	57.5	40.3
BERT	81.1	66.4	42.0
Fine-tuned ADA	94.3	87.2	88.3

Figure 4 shows sample claims extracted using our tool's software interface. The interface includes tabs for scheduling radio programs for monitoring, uploading user audio files, and a dashboard offering a file overview.

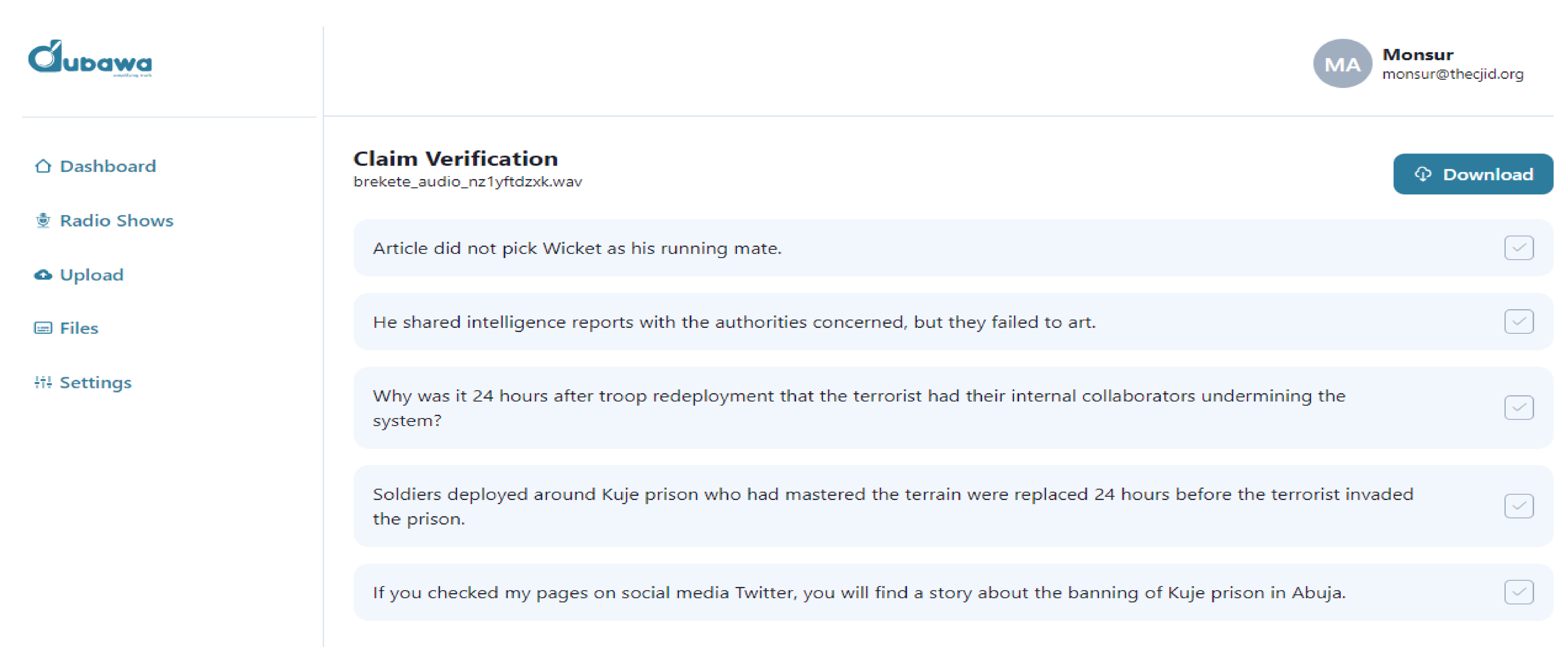

Figure 4. Sample claims

Limitations

- The transcription service, Ajala.ai, is tailored for African languages, yet it encounters difficulties in accurately identifying specific audio elements such as names and terms. For instance, in the first claim shown in Figure 4, it initially states, **"Article did not pick Wicket as his running mate"** when it should have read as **"Atiku did not pick Wike as his running mate"**, where 'Atiku' and 'Wike' denote names of Nigerian politicians.
- Also, the quality of the transcription is influenced by the audio input quality, and variations in audio quality can affect the accuracy of the transcriptions and subsequently impact the claims extracted from them.
- Diverse accents in radio broadcasts and other sources, leading to variations in speech patterns and pronunciation, pose challenges for the claim extraction model.
- Errors or inaccuracies at any stage of the transcription pipeline can have a cascading effect, affecting the overall accuracy and reliability of the final transcriptions and claim extraction results.

Conclusions

Our project significantly advances the domain of automated fact-checking in traditional media, addressing the challenges posed by radio. Through the integration of state-of-the-art AI tools, we have successfully crafted a robust claim extraction pipeline tailored to the unique characteristics of audio content. This pipeline demonstrates promising potential in enhancing media credibility by providing accurate and timely fact-checking.

Looking ahead, several areas for further work, such as exploring the integration of audio, video, and textual data sources to create a more comprehensive and robust fact-checking ecosystem; and addressing transcription challenges in the context of African languages and accents to improve accuracy further, emerge.

Moving forward, our project promises to empower audiences with verified information, reinforcing the critical role of technology in promoting informed and trustworthy journalism.

Acknowledgments

We extend our heartfelt gratitude to Google for their invaluable support through the Google News Initiative, which funded and propelled this project to fruition. We are also indebted to Ajala.ai for their dedicated provision of a transcription API tailored to diverse African accents, significantly enhancing the accuracy and scope of our research. This project owes its success to the unwavering dedication of our skilled developers and machine learning engineers, whose expertise and relentless efforts brought our vision to life.

References

[1] Ajala studios: Voice automation solutions designed for african languages. `https://www.ajala.ai/`.

[2] Fatma Arslan, Naeemul Hassan, Chengkai Li, and Mark Tremayne. A benchmark dataset of check-worthy factual claims. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, pages 358–361, June 2020.

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, and ... Language models are few-shot learners. *CoRR*, abs/2005.14165, 2020.

[4] Yudhanjaya Wijeratne and Dimuthu C. Attanayake. Artificial intelligence for factchecking: Observations on the state and practicality of the art. 2021.