What's The Hot Topic in Topic Town?

Credible News Reporting using Natural Language Processing and Hugging Face Transformers

Kelvin K. Ahiakpor

Ashesi University, kelvin.ahiakpor@ashesi.edu.gh

Emmanuel A. Acquaye

Ashesi University, emmanuel.acquaye@ashesi.edu.gh

CCS CONCEPTS • Information systems ~ Information retrieval • Information systems ~ World Wide Web ~ Web services • Computing methodologies ~ Machine learning; Artificial intelligence ~ Natural language processing

Keywords: Trending topics, Fake News, Artificial Intelligence, Natural Language Processing

1 INTRODUCTION

Misinformation is a critical issue that affects individuals and communities worldwide, particularly in Africa [1]. The rampant spread of false information, whether through misinformation or disinformation [2] in the modern day, leads to widespread ignorance, wars, and mislead opinions about current affairs both locally and globally. Imagine opening social media or online articles only to be bombarded with various pieces of information, leaving one unable to deduce the authenticity, meaning, and biases. For instance, recently in Ghana, a chef lied about winning a Guinness Book of Records award, and before the lie was uncovered, TV3, the most notable news station in the country had already reported the hoax. This might seem trivial but we can also consider cases where Russia purposefully disinforms African countries to proliferate wars to advance their agenda [1]. If institutions like TV3 and countries receive false information, then many other Africans receive fake news on an individual level, making it difficult for information to be trusted.

Our project addresses the issue of disinformation, misinformation and misunderstanding of current affairs through artificial intelligence. This approach leads to the provision of verified, insightful and tailored information. It integrates

machine learning to analyze news, understand the underlying sentiments and verify the news. By doing so, the project impacts society by empowering individuals and corporations with reliable information to curb misinformation and improve decision making.

This project's significance lies in its potential to transform the consumption of information in Africa. We envision an Africa where individuals have accurate and improved public awareness. We also aim for a well-informed society capable of making data-driven decisions. Our project also provides individuals with accessible intelligence tools to verify and analyze news data, promote critical thinking, and develop media literacy.

2 PROBLEM STATEMENT

Misinformation and disinformation across the internet's global village is at undesirable levels, particularly in Africa [1]. This issue has led to widespread ignorance and misinformed opinions on current affairs. In 2021, China suffered from disinformation all over the globe when COVID struck as false conspiracies were spread to defame the country. Today, Africa still experiences misinformation. For example, in Africa's politically unstable regions, disinformation is used to proliferate wars and civil unrest. There is a need for a solution that verifies information and offers a comprehensive understanding of news to empower individuals and corporations with factual information. Although fact-checking websites and AIs like AfricaCheck, Snopes, FactCheck.org, and FullFact, exist to verify information [6,7,9], they do not have real-time verification, sentiment analysis, and educational components. This project seeks to fill this gap by leveraging artificial intelligence and machine learning to provide a platform that offers that verifies the hottest news in real-time, performs sentiment analysis and personalizes information needs for specific purposes.

3 PROJECT OBJECTIVES

1. Scrape news from at least 1 credible source: AllAfrica, BBC ,The New York Times, Reuters, etc. [11].
2. Fine-tune a pre-trained Hugging Face Transformer with labeled news for 1-paragraph summary generation.
3. Perform sentiment analysis on the top 10 news articles.
4. Deploy a functional web app that provides real-time news with sentiment analysis and educates users on issues surrounding misinformation in Africa.
5. Prepare and deliver a comprehensive presentation of the project by the deadline.

Extras

6. Scrape news from up to 5 heavily biased sources: Fox News, MSNBC, InfoWars, Sky News, and Breitbart News [10].
7. Scrape news from up to 50 untrustworthy sources: Individuals, disinformation campaigns, and hate groups [12].
8. Train a deep neural network NLP model accuracy to classify credible, biased news, and untrustworthy news.
9. Evaluate the accuracy of the NLP model in classifying news types: achieve 80% overall accuracy.
10. Consider future integration by developing an API for company and individual use 6 months after deadline.

4 PROJECT SCOPE

Boundaries

This project has the following boundaries:

- The project focuses on addressing misinformation and disinformation prevalent in Africa.
- It involves real-time news scraping, sentiment analysis, and fact-checking functionalities.
- The scope encompasses both credible and biased news sources but excludes false and hate-driven content.
- News will be harvested from at least one the following sites: AllAfrica, BBC, The New York Times.

Technology Involved

This project involves the following technology:

- Web Scraping
- Natural Language Processing
- Pre-trained NLP models for sentiment analysis
- Web Deployment Interface

Limitations

The project has the following limitations:

- The intensity of web scraping may be flagged for suspicious activity.
- Accuracy of sentiment analysis and news classification may vary based on the quality and bias of news sources.
- Real-time data processing may face challenges related to web scraping rate limits and processing speed.
- The initial implementation will be limited to English language news articles.
- Personalization may not be achieved if a news classifier is not trained within the time scope.

.

Domains Involved

This project will potentially impact the following domains:

- Artificial Intelligence
- Machine Learning
- Natural Language Processing
- Web Development

5 TECHNICAL REQUIREMENTS

Table 1: Technical Requirements

Requirement	Purpose	Usage Method
News APIs (RapidAI)	Enable real-time data retrieval and processing	Select the best relevant APIs from RapidAI and integrate these to fetch live news feeds
NLP Libraries (NLTK)	Analyze sentiment in news articles	Apply VADER sentiment analysis tool from NLTK to understand the emotional tone of news articles
Hugging Face Transformers (BART)	Generate concise summaries of news articles	Fine-tune the pre-trained BART model from Hugging Face Transformers on a labeled dataset of news articles and utilize it.
User Interface Frameworks	Develop intuitive and user-friendly interfaces	Use React to build an interactive and responsive web application to contain project functionality.

REFERENCES

[1] Africa Center for Strategic Studies. 2024. Mapping a Surge of Disinformation in Africa. *Africa Center for Strategic Studies*. Retrieved July 14, 2024 from https://africacenter.org/spotlight/mapping-a-surge-of-disinformation-in-africa/

[2] BBC. 2024. Misinformation v disinformation: What's the difference? *BBC Bitesize*. Retrieved July 14, 2024 from https://www.bbc.co.uk/bitesize/articles/z3hhvj6

[3] Shevon Desai. 2018. Research Guides: News Sources: International News Sources. *guides.lib.umich.edu*. Retrieved July 14, 2024 from https://guides.lib.umich.edu/c.php?g=283063&p=1885956

[4] Greg Gondwe. 2023. *Misinformation in countries with limited technological literacies: How individuals in sub-Sahara Africa engage with fake news Permanent link Terms of Use*. Retrieved July 14, 2024 from https://dash.harvard.edu/bitstream/handle/1/37375267/Misinformation%20and%20Digital%20Literacy.pdf?sequence=1&isAllowed=y

[5] Ansong Joana. 2024. On the frontlines in the fight against dangerous misinformation. *WHO | Regional Office for Africa*. Retrieved July 14, 2024 from https://www.afro.who.int/news/frontlines-fight-against-dangerous-misinformation

[6] Gretel Kahn. 2024. Generative AI is already helping fact-checkers. But it's proving less useful in small languages and outside the West | Reuters Institute for the Study of Journalism. *reutersinstitute.politics.ox.ac.uk*. Retrieved July 14, 2024 from https://reutersinstitute.politics.ox.ac.uk/news/generative-ai-already-helping-fact-checkers-its-proving-less-useful-small-languages-and

[7] Rick Lyman. 2013. Nonpartisan Fact-Checking Comes to South Africa. *The New York Times*. Retrieved July 14, 2024 from https://www.nytimes.com/2013/07/24/world/africa/nonpartisan-fact-checking-comes-to-south-africa.html?pagewanted=all

[8] Dani Madrid-Morales and Herman Wasserman. 2021. Social media users in Kenya and South Africa trust science, but still share COVID-19 hoaxes. *The Conversation*. Retrieved July 14, 2024 from https://theconversation.com/social-media-users-in-kenya-and-south-africa-trust-science-but-still-share-covid-19-hoaxes-157894

[9] Mark Aaron Polger. 2022. CSI Library: Misinformation and Disinformation: Thinking Critically about Information Sources: Web Sites for Fact Checking. *library.csi.cuny.edu*. Retrieved July 14, 2024 from https://library.csi.cuny.edu/c.php?g=619342&p=4310783

[10] PureVPN. 2022. 10 Most Unbiased News Sources in 2022 - PureVPN Blog. *PUREVPN*. Retrieved July 14, 2024 from https://www.purevpn.com/blog/unbiased-news-sources/

[11] The Berlin School Of Creative Leadership. 2017. 10 Journalism Brands Where You Find Real Facts Rather Than Alternative Facts. *Forbes*. Retrieved July 14, 2024 from https://www.forbes.com/sites/berlinschoolofcreativeleadership/2017/02/01/10-journalism-brands-where-you-will-find-real-facts-rather-than-alternative-facts/

[12] Wikipedia Contributors. 2019. List of fake news websites. *Wikipedia*. Retrieved July 14, 2024 from https://en.wikipedia.org/wiki/List_of_fake_news_websites