Fakipedia: Building and exploiting an AI model for detecting online fake news in Burkina Faso

Angelique Sidbewendin Yameogo, Aminata Sabané, Tégawendé F. Bissyandé, Rodrique Kafando, Abdoul-Kader Kabore

Fakipedia: Building and exploiting an AI model for detecting online fake news in Burkina Faso

Angelique Sidbewendin Yameogo[1], **Aminata Sabané**[2], **Tégawendé F. Bissyandé**[1], **Rodrique Kafando**[1], **Abdoul-Kader Kabore**[12]

[1]Centre d'Excellence CITADEL, Université Virtuelle du Burkina Faso
[2]Département d'Informatique, UFR/SEA, Université Joseph KI-ZERBO, Burkina Faso

*E-mail : yameogoa94@gmail.com

Abstract
Misinformation poses a significant challenge, especially in developing countries with low literacy rates. The rapid spread on social media, coupled with their lack of robust verification mechanisms, makes distinguishing between credible and false information increasingly difficult. This document outlines our efforts to address this challenge in Burkina Faso using the pre-trained linguistic model CAMEMBERT. Our preliminary results show high performance (up to a 99.34% F1 score) on Fakipedia, suggesting that certain blatant instances of false information can be effectively identified and stopped by automated tools.

Keywords
Fake News ; Machine Learning ; Text Classification ; camemBERT ; Artificial Intelligence

I INTRODUCTION

In recent years, Fake News has garnered significant attention, particularly during key events such as the 2016 US presidential elections, the COVID-19 pandemic, and the Russia-Ukraine war[2, 12]. Fake news, often disseminated to deceive or influence opinions, has had notable consequences, including contributing to the spread of diseases such as COVID-19 in Burkina Faso[9].

The challenge intensifies during high-profile events, where the rush for breaking news risks amplifying false information and fueling hysteria. Brainard and Hunter[8] proposed an agent-based model, suggesting that reducing harmful online advice or limiting the ability of a portion of the population to share false information can mitigate the severity of an epidemic.

Despite emerging detection techniques, the sheer volume of information on social networks and online media makes manual verification nearly impossible. Detecting fake news remains a formidable classification task, emphasizing the need for tools, especially Artificial Intelligence, to protect against misinformation. Hence, we pose the following questions:

1. Question n°1: Can a centralized database address data availability issues regarding fake news in Burkina Faso?
2. Question n°2: How effective are artificial intelligence techniques in detecting fake news?

Our research aims to create a centralized repository akin to Wikipedia to enhance fake news accessibility and improve automatic detection using advanced technologies, particularly artificial intelligence. The article comprises five sections: introduction, state of the art in fake news detection, proposed approach, technical implementation, and results.

II STATE OF THE ART

2.1 Fake news in Burkina Faso

Misinformation is prevalent in Burkina Faso's information landscape due to the rapid growth of digital networks and online media. Traditional media, now operating largely online, often disseminate false information from government sources. Influencers and whistleblowers also contribute to the challenge of controlling misinformation, affecting both public discourse and private life [3] .

To address this issue, the Burkinabe government has implemented measures ranging from criminal provisions against fake news to journalist training in fact-checking. Dr. Sampala Balima suggests a comprehensive awareness program for the public and providing ministers with advisors during press conferences to handle technical questions.

Private initiatives, like FasoCheck, led by Burkinabe journalists trained by Deutsche Welle Akademie, focus on fact-checking to ensure the accuracy of information disseminated in public debates.

2.2 Fake News Detection

During your research, it was challenging to find direct works related to your topic. However, you discovered relevant works in the realm of artificial intelligence, particularly Natural Language Processing (NLP), for automatic fake news detection.

- **Early Detection Approaches [5]:** Graves and Amazeen propose early detection of fake news, relying on a large quantity of observed fake news over an extended period. However, their methods exhibit a significant delay of at least 12 hours in detecting fake news, which could be problematic given the rapid spread of misinformation.
- **Comparative Study on COVID-19 Fake News Detection [10]:** Joy, Dofadar, et al. conduct a comparative analysis using five transformer-based models (BERT, BERT without LSTM, ALBERT, RoBERTa, and a BERT-ALBERT hybrid) to detect COVID-19 fake news. The RoBERTa model outperforms others with an impressive F1 score of 0.98.
- **exBAKE: Automatic Fake News Detection Model [4]:** Ongsuk, Park, et al. propose exBAKE, an automatic fake news detection model enhancing BERT by addressing data imbalance. Their approach leverages BERT's deep and contextual nature, achieving a 0.14 improvement in F-score compared to previous state-of-the-art models.

While showing promising results, most studies have used data in English. Implementing an English-based model in a French language would be difficult.

III PROPOSED APPROACH

3.1 Architecture of Our System

The figure 1 below illustrates the comprehensive architecture of our system, which is structured into four sequential stages: Step 1 involves the archiving of fake news, followed by Step 2 where data extraction and pre-processing take place. Step 3 is dedicated to the learning phase, and the process culminates in Step 4, which focuses on the utilization of the learned models for fake news detection.

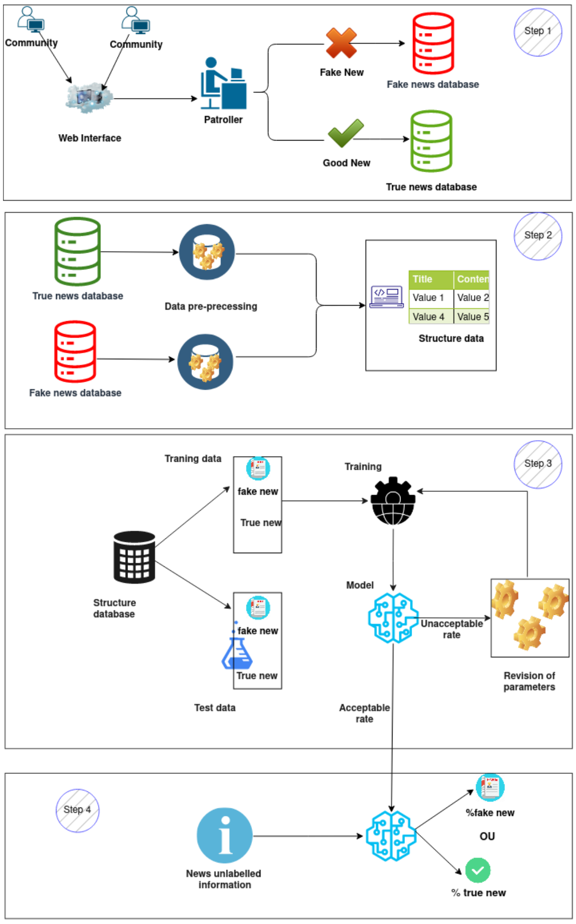

Figure 1: Architecture of our system

1. Step 1: Archiving Fake News

   In this stage, we focus on the architecture of news archiving in a scalable database known as Fakipedia. The procedure is carried out in the following steps:

   - Initially, Members of the Fakipedia community draft and contribute news articles rich in metadata, including title, content, authorship, original publication date, and the associated URL.
   - Designated Fakipedia patrollers are responsible for verifying the influx of data, having exclusive access to raw, unverified data. They utilize digital tools for fact-checking and validation, with TinEye facilitating image source validation. [7] and Google Reverse [7], which trace the origins of images across the web.
   - If a source URL is no longer active, patrollers turn to the Internet Archive [1], a digital library that archives various websites over time, to fetch the historical content.

   Upon successful verification, patrollers mark the information as 'reviewed,' thus signaling to other patrollers that the particular entry is validated and needs no further checking.
   In the case of text-only articles, we utilize conventional methods for verification based on the subject matter. This is achieved either through targeted keyword searches or information triangulation, which involves corroborating details from at least three different

reliable sources. Specific indicators of content reliability are extracted, such as:

- Source Credibility: The article's sources are scrutinized for their reliability and abundance.
- Title Examination: The article's title and any subtitles are assessed as they serve to engage the reader.
- Author Identification: The credibility of the author is considered as part of the content evaluation.

If the information is proven to be false, it is then cataloged in Fakipedia's designated fake news database. Conversely, verified information is stored in a separate database. Sites such as faso-check and lefaso.net are also used to retrieve known false news articles.
Due to the dynamic nature of news content, regular database updates are carried out to ensure that all information remains current.

2. Step 2: Data Extraction and Preprocessing
   - Objective: To extract and preprocess pertinent features from Fakipedia for model training.
   - Process: Features relevant to fake news detection are extracted, cleaned, and structured for further use.

   Firstly, raw textual data was retrieved from Fakipedia. Next, a cleaning process was implemented to eliminate redundancies, inconsistencies and any superfluous information. Textual data was standardised to ensure consistency of format, and duplicates were removed. Natural language processing (NLP) techniques were used to extract relevant information from the texts. This included syntactic analysis to understand sentence structure, named entity extraction to identify key elements, and lemmatisation to normalise words. Finally, the extracted information was structured in a pre-defined format. As a result, the textual data is now structured and ready to be used.
3. Step 3: Learning
   - **Training :** To train our model, we employ an algorithm previously described in the preceding chapter. This algorithm excels in its ability to identify and characterize linguistic elements such as proper nouns, verbs, adverbs, and adjectives, while also accurately capturing the nuances of French grammar and syntax [11].
   - **Testing :** Testing evaluates the model's capability to generalize to new data. Initially, we partition the dataset into training and testing subsets to mitigate the risk of overfitting, which would occur if the model were tested on the same set it was trained on.

4. Step 4: Usage
   This is the culminating and most critical stage of our system. Having optimized our model for the highest recognition rate in the prior steps, we now deploy it on unlabeled news data. The model provides a classification of the news as either 'fake' or 'real,' accompanied by a confidence score."

3.2 Chatbots

Recognizing the press's importance for global awareness, we aim to combat the spread of fake news by creating software. Using open-source conversational AI, we'll develop a chatbot. Users input an article's title, and the chatbot swiftly provides the probability of it being fake, simplifying the process for wider adoption.

3.3 Creation of the Dataset

To train our machine learning system, a dataset is essential, allowing the algorithms to predict the class of test data. However, no public dataset meeting our criteria for Fake News, containing both fake and real news with sufficient features, was found. Consequently, we created our dataset, focusing on Burkina Faso through the Fakipedia database. Since we only had 43 data points specific to Burkina Faso from fact-checkers at FasoCheck, we augmented our dataset by combining this information with data from recognized French sources that propagate both fake and real news. These sources include reputable French websites known for publishing accurate information, as well as sites disseminating false information. The dataset was collected using scrapy tools, and the source code is accessible at [6]. Further details about the dataset are provided below:

- The dataset for real articles was obtained from https://webhose.io/free-datasets/ and includes reputable and widely recognized sources, such as europe1.fr, 20minutes.fr, lexpress.fr and others.
- The corpus of fake news consists of various websites, including legorafi.fr, secretnews.fr, mespropresrecherches.com and others.
  The data has been divided into two files named real news.csv and fake news.csv depending on whether they are true or false for reliable and unreliable information.

We analyzed the two files, real news.csv and fake news.csv, from our dataset and retrieved a total of 16,383 fake data and 11,968 real data. We performed data preprocessing by removing stop words, special characters, and duplicates. Next, we tokenized our articles (split them into tokens). Figure 2 shows, for example, the number of articles retrieved per site.

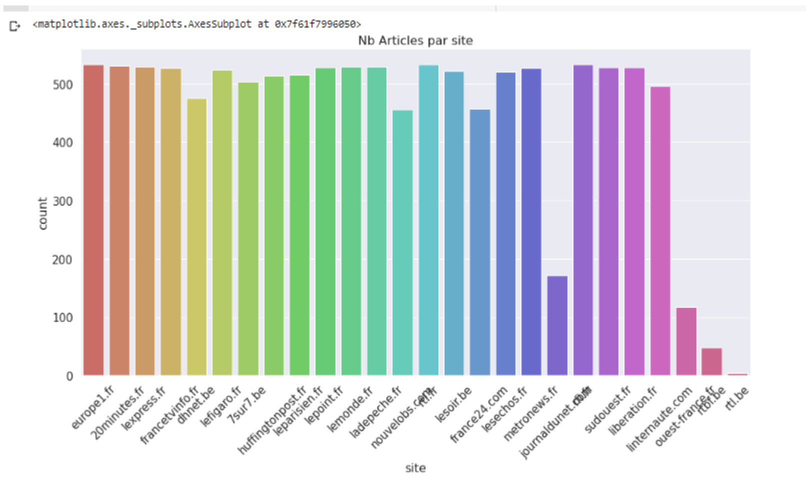

Figure 2: Number of Articles per Site

IV IMPLEMENTATION

In the quest for a robust and reliable fake news detection model, our study capitalizes on the strengths of neural network architectures and transfer learning techniques. We specifically adopt the CamemBERT model, fine-tuned for the particular nuances of fake news categorization. The subsequent sections elucidate the architecture of CamemBERT, the pre-training methodologies employed, and how transfer learning is integrated into our approach.

4.0.1 camemBERT

CamemBERT is a leading neural network model rooted in the Transformer's Encoder-Decoder with Attention architecture. As a French-specific iteration of the BERT family, it stands out for its bidirectional capabilities. Shared across CamemBERT, RoBERTa, and BERT, its architecture includes 12 layers, 768 hidden dimensions, 12 attention heads, and 110 million parameters. The BERT architecture's advantage lies in parallel processing, making CamemBERT the first BERT-type model in French.

4.0.2 Pre-training

The pre-training phase readies the model for the main task using various methodologies. Key differences from the BERT model include:

- Different batch size: 8000 compared to 256 in BERT.
- Masked Language Model (MLM) predicting a masked word in a sentence.
- Next sentence prediction.

In comparison to BERT, roBERTa adopts different training methods and hyperparameters:

- Longer duration and more data, from 16GB to 160GB of text.
- No next sentence prediction training.

CamemBERT diverges from roBERTa primarily in its use of a different tokenizer. Additionally, camemBERT is trained using the new French sub-corpus from OSCAR.

4.0.3 Transfer Learning

Our approach employs transfer learning for a binary classification task: "fake" or "not fake." Transfer learning, widely used in machine learning, involves leveraging a model's previous training to perform a similar task. Deep neural networks, prevalent in image recognition or natural language processing (NLP), undergo resource-intensive training. Transfer learning allows us to reuse one model's training phase in another. To achieve this, we append a classifier to CamemBERT's output, including:

- Input: n hidden states from the first token.
- Dropout intermediate layer (to prevent overfitting).
- Dense layer with n inputs/n outputs.
- Activation function: tanh.
- Final linear layer: n inputs/1 output.

4.0.4 Model training environment

All transformer-based experiments were conducted in the Google Colab GPU environment, while other experiments were carried out on local machines. The model training was repeated 5 times (starting from scratch) to achieve better results. The technical specifications provided by Colab are as follows: a GPU processor, 25 GB of RAM, and a 166 GB disk. As training parameters for our model, we used: train_percent : 0.7 ; test_percent : 0.3 ; nb_epochs : 25 ; random_state : 0

V EXPERIMENTS, RESULTS, AND ANALYSIS

5.1 Experiments

To evaluate our approach, we considered two (02) research questions. To answer each of these research questions, we established an experimental procedure that we present in this section.

5.1.1 Research Questions

- Question n°1: Could a database solve the problem of data availability regarding fake news in Burkina Faso?
- Question n°2: Can artificial intelligence techniques help us detect fake news?

5.1.2 Experimentations

For each of the research questions, we have defined the following experiments:

- Experimentation for Question n°1 To address the lack of fake news data from Burkina Faso, we've established a web interface and a database for easy data collection. This platform allows individuals to contribute suspected false news articles along with metadata such as titles, content, sources, authors, and timestamps. Fakipedia patrollers review, verify, and record the submitted information if its false nature is confirmed. Importantly, only patrollers have access to articles submitted by contributors.
- Experimentation for Question n°2 The proliferation of fake news through the Internet, especially via social media, necessitates effective control measures. This misinformation's widespread impact on individuals, organizations, and businesses underscores the urgency in combatting it, given that many people accept such information without thorough verification.
  The solution lies in leveraging machine learning, enabling quick and automatic detection of fake news upon publication. Machine learning algorithms analyze message content, identifying erroneous information. Researchers are actively developing classification models using various machine learning classifiers to enhance fake news detection.
  In our experiment, we utilize the data collected in response to the first research question (1) to train our fake news detection model.

5.2 Results

At the end of our work, we successfully established an interface for collecting data, which we named Fakipedia. Next, we preprocessed this data to obtain suitable training data. Finally, we were able to train our model (transfer learning), achieving an accuracy of 99.17%.

We will first introduce Fakipedia, followed by the results of fake news detection with CamemBERT.

5.2.1 Fakipedia

As mentioned in the previous chapter, we have implemented a tool for collecting fake news, and anyone can contribute to it. To register information in the Fakipedia database, it must be validated by patrollers to confirm that the information is indeed false. Figures-3 and -4 respectively provide an overview of the account creation and login pages on Fakipedia. Fakipedia's target audience is everyone, and they can log in anonymously. Only patrollers are required to have non-anonymous accounts.

Figures-6 and -7 respectively show an overview of the Fakipedia homepage and an article written by a contributor.

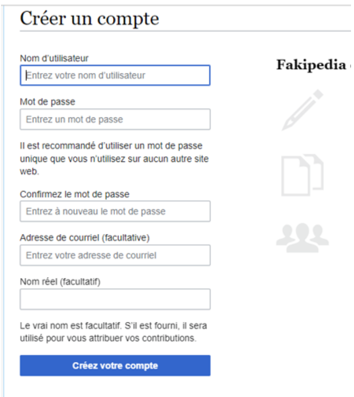

Figure 3: Account Creation Form

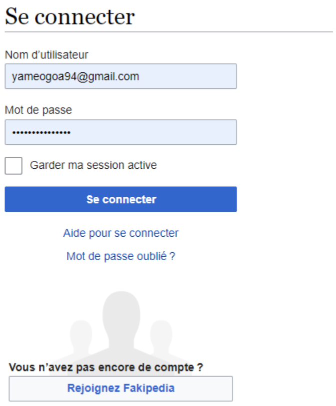

Figure 4: Connection

Figure 5: Account Creation Page

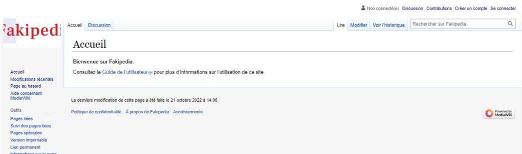

Figure 6: Homepage of Fakipedia

Figure 7: Article Written by a Contributor

5.2.2 *Training Results*

As mentioned in the previous chapter, we retrieved and preprocessed the data for training. Once the data preprocessing was completed, we fed it into our model for training.

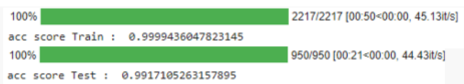

Figure 8: Training Results

We can observe in Figure-8 that we achieved a training accuracy of 99.99% and a test accuracy of 99.17%.

VI CONCLUSION

The application of artificial intelligence in fake news detection is a rapidly evolving research frontier. Our work contributes significantly to this field. We began with a comprehensive review of state-of-the-art techniques, guiding our subsequent research. Using insights gained, we created the Fakipedia database and enhanced a pre-trained French language model through transfer learning.

Addressing the challenges of AI-based fake news detection was a substantial undertaking, yielding encouraging and knowledge-advancing results. Our research opens avenues for further exploration, including implementing an automated verification bot via the Twilio WhatsApp interface, expanding Fakipedia for comprehensive training data, and potentially integrating our model into an online platform as an automated patroller for fake news.

ACKNOWLEDGEMENT

This work was conducted as part of the Artificial Intelligence for Development in Africa (AI4D Africa) program, with the financial support of Canada's International Development Research Centre (IDRC) and the Swedish International Development Cooperation Agency (Sida).

References

[1] T. Schwarz, M. Baker, S. Bassi, B. Baumgart, W. Flagg, C. van Ingen, K. Joste, M. Manasse, and M. Shah. "Disk failure investigations at the internet archive". In: *Work-in-Progess session, NASA/IEEE Conference on Mass Storage Systems and Technologies (MSST2006)*. 2006.

[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu. "Fake news detection on social media: A data mining perspective". In: *ACM SIGKDD explorations newsletter* 19.1 (2017), pages 22–36.

[3] *Fake news : Dr Sampala Balima analyse l'écosystème des fausses informations au Burkina*. 2019.

[4] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim. "exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert)". In: *Applied Sciences* 9.19 (2019), page 4062.

[5] L. G. et M. A. Amazeen. *Fact-Checking as Idea and Practice in Journalism*. Feb. 2019.

[6] Akachbat. *FrenchFakeNews/scraper*. 2020.

[7] M. Kondal and V. Singh. "Comparative analysis of tineye and google reverse image search engines". In: (2020).

[8] Y. B. et P. R. Hunter. "Misinformation making a disease outbreak worse: outcomes compared for influenza, monkeypox, and norovirus". In: *SIMULATION* 96, no 4, p. 365-374 (Apr. 2020).

[9] B. Seytre, C. Barros, P. Bona, K. Blahima, A. Rodrigues, O. Varela, B. Yoro, and B. Fall. "Une enquête socio-anthropologique à l'appui de la communication sur le Covid-19 en Afrique de l'Ouest". In: *Médecine Tropicale et Santé Internationale* 1.3 (2021).

[10] S. K. S. Joy, D. F. Dofadar, R. H. Khan, M. S. Ahmed, and R. Rahman. "A Comparative Study on COVID-19 Fake News Detection Using Different Transformer Based Models". In: *2022 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*. IEEE. 2022, pages 1–5.

[11] C. Langevin. "Les technologies de l'intelligence artificielle au service des médias et des éditeurs de contenus: traitement du langage naturel (TAL)". In: *I2D-Information, donnees documents* 1.1 (2022), pages 30–37.

[12] H. A. et M. Gentzkow. "Social Media and Fake News in the 2016 Election". In: *J. Econ. Perspect* 31, no 2 (20017).