AI-driven Generation of News Summaries

Leveraging GPT and Pegasus Summarizer for Efficient Information Extraction

Issiaka Faissal Compaore[1,2], Rodrique Kafando[1], Aminata Sabané[1,2], Abdoul Kader Kabore[1], Tegawendé F. Bissyandé[1,2]

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

issiakafaissal2j@gmail.com, rodrique.kafando@citadel.bf, aminata.sabane@ujkz.bf, abdoulkader.kabore@citadel.bf, , tegawende.bissyande@citadel.bf,

Abstract. The surge of online information makes it challenging to access relevant news quickly. This necessitates automated methods to effectively extract and summarize information. Our research focuses on designing an online press synthesis tool using advanced AI models. We investigate the feasibility of employing two pre-trained models, GPT-3.5 Turbo 16k and Pegasus Summarizer, to generate high-quality summaries from scraped articles. Our methodology encompasses robust web scraping, model integration, and metric evaluation. Experimental results demonstrate that GPT-3.5 Turbo 16k outperforms in accuracy, achieving a BLEU score of 16.39% and a ROUGE score of 0.66%. The turner007/pegasus-summarizer model also performs well, with a BLEU score of 15.45% and a ROUGE score of 0.45%. We identify areas for improvement, such as enriching the database with expert-authored summaries and adopting a dynamic approach to news adaptation. Additionally, we explore a unified models approach for further refinement.

Keywords: Online press, artificial intelligence, web scraping, Transformers, Abstractive summarization.

1 Introduction

The surge in online information has created a challenge in promptly and effectively accessing relevant news amidst a deluge of articles from diverse sources. An analysis of French news sites underscores editorial variety, covering between 300 and 700 distinct news topics daily [1].

JRI 2023, December 18-20, Ouagadougou, Burkina Faso

DOI 10.4108/eai.18-12-2023.2348180

Technological advancements have influenced the quantity and quality of information on online press platforms [2]. The abundance of online information recognized by [2] poses challenges in sorting through and finding pertinent content, reflecting significant editorial diversity. In this context, the automation of press synthesis through artificial intelligence (AI) emerges as a promising solution to distill essential information concisely.

Confronted with this information abundance, our inquiry revolves around creating an automated system capable of extracting articles and generating high-quality summaries. We aim to ensure that these summaries encapsulate the essence of original articles while remaining comprehensible and informative. Additionally, we seek to develop a tool that prevents information duplication, ensuring diverse, relevant, and non-redundant content for users. The topic ”Implementation of an AI-based online press synthesis tool” was proposed to address these questions.

The primary objective of this work is to construct an online press synthesis tool employing advanced AI models to automate article summarization. This approach offers users an efficient and swift means to access the core of information. Specific objectives encompass ensuring tool accessibility to the general public, efficient collection of online news article data, centralization of information from various press sites, implementation of a similarity calculation mechanism to prevent information duplication, efficient automation of information synthesis through high-performance models, and the creation of a user-friendly interface.

2 Related Work

Text summarization is a constantly evolving field where researchers continuously explore new avenues to enhance the quality and relevance of automatically generated content. In Natural Language Processing (NLP), [3] proposes a feature extraction algorithm based on the maximum entropy reordering model in statistical machine translation within language learning machines. The algorithm aims to extract more precise sentence reordering information, especially regarding inverted sentence features, to address the imbalance in data features during maximum entropy training in the original algorithm and improve sentence reordering accuracy during translation. Experimental verification is performed on the Chinese-English translation dataset, showing that different posterior word probabilities significantly impact the classification error rate. Combining linguistic features based on posterior word probability can reduce the classification error rate and improve error prediction performance in translation. Therefore, the proposed feature extraction algorithm improves the accuracy of sentence reordering in translation, solving the problem of feature data imbalance during maximum entropy training. Additionally, the addition of features related to the first word and combination features further enhances translation performance.

According to [4], extractive text summarization is the approach of selecting salient sentences from a source document to create a summary. They propose a new neural model called N-GPETS for extractive text summarization. N-GPETS integrates cross-sentence associations by combining the Transformer-based BERT model and the Graph Attention Network (GAT). It combines a heterogeneous graph attention network with the BERT model and uses TF-IDF values to improve interaction between sentences. N-GPETS incorporates the graph layer into the BERT encoder during the graph initialization step, improving its performance compared to other models using different neural network encoders. Empirical results on reference news datasets demonstrate favorable outcomes for N-GPETS compared to other models using the BERT model and graph structures without BERT.

The intersection of text detection and automatic content generation is one of the most dynamic areas of artificial intelligence. According to [5], traditional methods based on machine learning rely on numerous artificial features with detection rules. Additionally, traditional methods may experience information loss during feature extraction, and [5] proposes a mechanism for compensating personal information to solve this problem. [6] addresses challenges in text summarization, such as eliminating redundancy and covering informative content. He emphasizes the need to develop methods to eliminate redundancy and solve the problem of different words used to describe the same object in a text. He focuses on text summarization, generating a condensed version of a document by removing redundant information and identifying important differences between documents. He introduces FrameNet, a lexical database based on frame semantics, as a valuable resource for understanding the meaning of words and syntactic structures. It contains semantic roles called frame elements (FeS) linked to words described in the context of events or objects. The proposed Semantic Graph Model (SGM) uses semantic information from FrameNet and word integration to calculate sentence similarity. Experimental results demonstrate the feasibility and efficiency of SGM in summarizing text. SGM considers internal and external information in ranking sentences, improving the quality of generated summaries.

Regarding extractive summarization, [7] focuses on applying advanced extractive summarization techniques to legal regulatory documents, aiming to democratize the understanding of regulations for non-lawyers. They create a corpus named EUR-Lexsum containing selected European regulatory documents and their corresponding summaries. Experiments show that transformer-based models, when applied to this corpus, surpass traditional reference bases for extractive summarization in terms of ROUGE metrics. They also mention that transformer-based models outperform TextRank, a traditional reference model, even with a moderate amount of data for refinement. They suggest exploring advanced hybrid approaches and alternative methods of supervised extractive summarization in future research. They also show that increasing the number of sentences from 10 to 30 per generated summary improves the F1 score for all models in terms of ROUGE-1, 2, and L. However, increasing the summary length to 50 sentences does not yield further gains. BERT and Distilbert achieve results almost as good as Roberta in terms of F1 scores for the three ROUGE indicators.

The greedy selection algorithm used to build oracle summaries produces a remarkable number of documents containing fewer than the desired 32 sentences, with the average oracle summary consisting of only 21 sentences. This limits the amount of data available to refine transformer-based models.

Several artificial intelligence (AI) text synthesis models have been developed. The MTL-DAS model is a unified model for adaptive multidomain text synthesis, combining multitask learning for multidomain adaptation synthesis with BART. It aims to improve generalization capacity in multidomain scenarios. The model adapts the ability to detect summary-worthy content from the source domain and gain knowledge and generation style in target domains through text reconstruction and classification tasks. The experience on the AdaptSum dataset, including six domains in low-resource scenarios, shows that the unified model outperforms separately trained models and requires fewer computational resources [8]. MTL-DAS achieves better overall inference accuracy in six domains compared to fine-tuning BART, even without using labeled data from target domains. The multitask learning strategy studied in MTL-DAS improves the model's ability to capture key content and adapt to multiple target domains with low resources [8].

Abstract text summarization aims to condense a text corpus while preserving its meaning and grammatical accuracy. The introduction of sequence-to-sequence models, based on RNNs, has revolutionized the use of deep learning in natural language processing. These models have demonstrated competitive performance in tasks such as automatic translation [9]. [9] shows that the Transformer model, based solely on the attention mechanism, has surpassed previous state-of-the-art models in translation tasks. It explores the application of the Transformer model in abstract text summarization and describes its advantages and disadvantages compared to other available models. Output summaries generated by the Transformer model have proven to be more accurate and retain the main content of the summary compared to the reference model.

Therefore, the performance of the Transformer model has been evaluated using Rouge scores, indicating its effectiveness in abstract text summarization. Experiments on the Amazon Reviews dataset show that the base model performs better for shorter sequences, while the abstract model faces challenges due to the unavailability of synthesis tokens in the source. Removing stop words has improved Amazon dataset results by reducing input noise and avoiding overfitting. It is also noteworthy that [9] emphasizes the need for a better metric capable of evaluating synthesis systems comprehensively, considering grammatical and semantic accuracy, as Rouge scores focus solely on token frequency.

[10] states that controllable text generation (CTG) is an emerging field of natural language generation (NLG) focusing on generating text with specific constraints for practical applications. Pre-trained transformer-based language models have become a popular approach for CTG, allowing the generation of diverse and fluid texts. It provides a systematic critical review of controllable text generation (CTG) techniques using pre-trained transformer-based language models (PLM) from the perspective of controllability and interpretability of deep neural networks. It also outlines common tasks, main approaches, and evaluation methods in the CTG domain using transformer-based PLMs. Text-to-speech, driven by major advances in deep learning, is a key pillar of artificial intelligence applied to natural language processing. This revolutionary approach enables texts to be generated autonomously, opening up new perspectives in fields as varied as machine translation, computer-assisted writing and dynamic content creation. [11] analyzes the development context and the state of research in the field of text classification, providing a theoretical reference for its further development.

[11] performs a statistical analysis of relevant papers published in the field of text classification over the last ten years, examining annual publication trends, subject distribution, journal distribution, institution distribution, author distribution, analysis of frequently cited literature and research hotspots. Based on this analysis, they propose a networked news text classification model based on deep learning that combines a convolutional neural network, a cyclic neural network and an attention mechanism. They highlight the role of keywords and demonstrate the superiority of their model through experiments. To do this, they use a Chinese word segmentation database to segment news text and convert it into word sequences. They filter out empty words and non-Chinese characters

to better express the characteristics of the news text. Next, they compare the classification effect of their deep learning-based model with traditional models using the spatial vector model, TF-IDF and the SVM classifier, demonstrating the validity of their approach.

[12] proposes a new news text classification method based on convolutional neural network and deep learning to solve the problems of low classification accuracy and efficiency in existing methods. The method involves determining the weight of news text data using the VSM spatial vector model, calculating the information gain from mutual information and determining the characteristics of news text data. It analyzes the basic structure of the deep learning convolutional neural network and uses the convolutional layer to train the news text data and create a news text classifier. Experimental results show that the deep learning-based convolutional neural network improves the accuracy and speed of news text classification. It also discusses the use of the deep hashing algorithm to handle noise in news text features and the computation of mutual information for feature extraction. The spatial vector model VSM is used to transform unstructured or structured news text into structured computer-intelligible text, and the TF-IDF method is used to compute feature word weights.We note that the proposed method, along with existing methods, has achieved a classification accuracy of over 90% in the classification of news texts.

Sentiment analysis, also known as sentiment analysis or opinion mining, is a natural language processing (NLP) technique that aims to determine and evaluate the emotions, opinions, feelings or attitudes expressed in a given text. [13] proposes a hybrid model combining machine learning and deep learning techniques for detecting emotions in text. Convolutional neural network (CNN) and bi-GRU are used as deep learning techniques, while support vector machine (SVM) is used as a machine learning approach. The proposed model achieves an accuracy of 80.11% when evaluated using a combination of three different types of datasets: sentences, tweets and dialogues. These datasets are used to train and test the model, enabling a comprehensive evaluation of its performance in detecting emotions from different forms of text. By including these various types of datasets, the model can capture variations in language use, context and communication styles, providing a more robust and accurate emotion detection capability. This approach also enables the model to be applicable in different scenarios, such as customer review analysis, social network user safety and other potential business applications. The overall accuracy of the hybrid model, which integrates the CNN and bi-GRU deep learning models, as well as the SVM model, is said to have identified emotions in text. It achieved an accuracy of 80.11% when evaluated using the combination of sentence, tweet and dialogue datasets. Abstract text summarization [14] is a method that reformulates the original text to generate a summary composed of new sentences. It is implemented using deep learning models, in particular recurrent neural networks equipped with an attention mechanism and an LSTM. Whereas extractive text summarization is a method that generates a summary by selecting words and phrases directly from the original text based on linguistic and statistical features. To this end, [14] reviews recent approaches to abstract text summarization using deep learning models and explores the datasets used for training and validation. It studies the Gigaword dataset, which is commonly used for one-sentence summarized approaches, while the CNN/Daily Mail dataset is commonly used for multi-sentence summarized approaches. He brings that recurrent neural networks with attention mechanism and LSTM have proven to be widespread techniques for abstract text summarization, and that experimental results show that text summarization using a pre-trained encoder model resulted in high ROUGE scores. [14] shows us that various methods have been developed to generate summaries on a single document, with machine learning being a predominant approach. These methods rely on numerical representations of the text and extracted features to produce high-quality summaries. Earlier text summarization techniques included superficial approaches such as term frequency highlighting and phrase positioning techniques. These approaches identified the importance of terms and phrases in the text document. [15] proposes a news text classification method based on the combination of deep learning (DL) algorithms, such as CNN, LSTM and MLP, to address the challenges of text length, feature extraction and classification in news text. The proposed model uses word vectors and word dispersion to represent the relationship between words and categories, enabling comprehensive learning of spatial and chronological information on time series features. Several experiments were carried out to evaluate the stability and performance of the proposed method, demonstrating its effectiveness in achieving better precision, recall and overall value than other models. The study highlights the importance of considering the relationship between words and categories in the classification of news texts, which is often neglected in previous approaches.

English-Chinese translation models based on neural networks have replaced traditional methods, and [16] focuses on attention mechanisms and grammatical knowledge in translation models. It proposes a translation model based on the integration of LSTM attention and the LSTM model combined with prior grammatical knowledge, to improve the representation of source language contextual information and enhance translation quality.

The proposed model is simulated on the IWSLT2019 dataset and shows improved representation of source language contextual information compared to the standard LSTM model. The IWSLT2019 Chinese and English datasets are processed separately using language-specific word segmentation methods. For Chinese, a statistics-based word segmentation method is adopted, while for English, words are divided based on empty spaces.

The dataset is used for training and performance evaluation of the translation model. The BLUE value is used as an indicator to assess translation quality, while a higher BLUE value indicates better translation quality. The results showed that the proposed model offered a better representation of source language contextual information than the existing translation model based on the standard LSTM model. The LSTM model with attention integration stabilized after 80 training cycles, indicating its high learning capacity and ability to learn the corresponding text expression in a short time.

However, the translation model failed to capture terminological information, as prior knowledge in the source language struggled to fully understand the corresponding relationship between terms, resulting in less effective translation for longer terms. Whereas the LSTM model combined with prior syntactic knowledge, using simple identifiers to identify target terms as a group, helped the translation model to better integrate terminological knowledge during training and learn the semantic relationship between target terms and source statements. [16] points out that integrating the neural machine translation model can improve the neural machine translation model and enhance translation quality. In [17], abstract summarization of Arabic texts is considered a difficult task due to the complexity of the language. Most existing research studies on Arabic text summarization focus on extractive summarization, while abstract summarization is relatively limited.The quality of the generated summary depends strongly on the quality of the words incorporated in the model. Pre-processing Arabic text for summarization poses challenges such as removing empty words, unwanted characters and Arabic additions to words.[17] proposes a system that uses a sequence-sequence model with encoder and decoder components. Various deep artificial neural networks, such as closed recurrent units (GRU), long-term memory (LSTM) and bidirectional long-term memory (BiLSTM), are used to develop the encoder and decoder. The global attention mechanism is also used in the model, as it has been shown to perform better than the local attention mechanism. We have the ArabErt preprocess which is applied during the data preprocessing phase to improve the model's understanding of Arabic words and achieve state-of-the-art results. It has been shown that the Word2vec model with gram skipping outperforms the continuous bag-of-words (CBOW) Word2vec model in terms of abstract synthesis. Experimental results show that three layers of BiLSTM hidden states at the encoder level offer the best performance in abstract Arabic text synthesis. [17] highlights the challenge of developing purely abstract synthesis methods, as some existing systems still rely on extractive synthesis techniques. [18] explores the impact of GPT-3, a large linguistic model, on text summarization, particularly in the field of news summarization. GPT-3, a large language model, has had a significant impact on text summarization research. The authors of [18] compare GPT-3 with refined models trained on large synthesis datasets and find that GPT-3 summaries requested using only a task description are preferred by humans and do not present problems such as lack of factuality. Evaluation of GPT-3 abstracts is also discussed, highlighting the limitations of automatic reference-based and reference-free metrics to reliably evaluate these abstracts.Refined models trained on large synthetic datasets are compared to GPT-3, and GPT-3 outperforms them in terms of human preferences. Evaluating GPT-3 summaries is difficult, as automatic reference-based and reference-free metrics cannot reliably assess them.

Ultimately, we find that transformer-based models perform best in generating relevant and informative text summaries. These models are able to process long, complex data sequences and focus on the most relevant information. In conclusion, the field of AI-assisted text summarization has undergone a significant evolution, marked by the emergence of transformer-based models.These models, such as BERT, DistilBERT and GPT-3, have revolutionized natural language processing, enabling deeper and more accurate understanding of texts. Considering the information gained from previous research, the text summarization model that has caught our attention the most is GPT-3, demonstrated by [18] to be the most successful model in terms of human preference and without problems such as lack of factuality. To this, we will add other text-to-speech models based on the transformers architecture such as: 'moussaKam/barthez-orangesum-abstract' and 'facebook/bart-large' developed by Facebook, 'turner007/pegasus-summarizer' from the Hugging Face Transformers platform, and for GPT-3, we will use the 'gpt-3.5-turbo-16k' version developed by OPENAI. To evaluate these models and select the best performer for our tool implementation, we have selected the ROUGE and BLUE metrics, which are frequently used in the majority of the above-mentioned studies to evaluate models.

Table 1: Summary table of selected information

Model	Summary Type	Data Set	Metrics	Result/Observation
N-GPETS [19]	Extractive	--	--	Favorable results obtained
Transformers [7]	Extractif	EUR-Lexsum	ROUGE	Transformer-based models show good results
MTL-DAS [8]	Abstract	AdaptSum	--	Unified models outperform separately trained models.
Transformers [9]	Abstractif	--	ROUGE	Transformer-based models outperform the others.
Transformers	Abstract	GIGAWORD	ROUGE	Very efficient
Transformers	Abstract	--	ROUGE	Good Result
Transformers	Abstract	Diversified	ROUGE/BLUE and others	Good Result especially in terms of human preference.

3 Materials and Methods

This section is devoted to presenting our approach to implementing our AI-based online press synthesis tool. Figure 1 illustrates the global overview of our approach. It consists of three parts: the first describes the extraction of our data and the web-scraping techniques used, the second describes the study and use of pre-trained text summarization models, notably ”gpt-3.5-turbo-16k” and ”turner007/pegasus-summarizer”, not forgetting the use of the gTTS (Google Text-to-Speech) Python library to convert our summaries into audio. Finally, we store the data in our database for display on the user interface. We'll go into more detail about these parts in this chapter, while justifying our various choices.

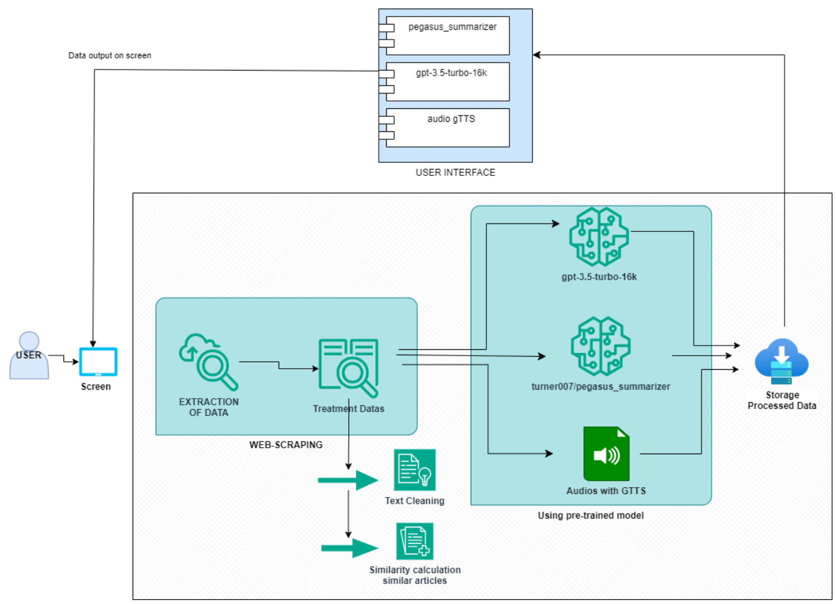

Fig. 1. Global architecture of the system.

3.1 Web scraping and data extraction

Web scraping is a fundamental technique in data science and computer engineering. It automates the retrieval of information from web pages, providing a rich and diverse source of data. Web scraping is the process of automatically collecting and extracting data from the Internet, in particular from websites. It involves the use of automated systems, known as web scrapers, to retrieve and analyze data from web pages.

For data collection, we have targeted online press sites :
Le Faso.net[1], Burkina24.com[2] and kaceto.net[3].
These sites were chosen for their popularity and as news sources with varied content. The data collection process was carried out using the Python Beautiful Soup library, a powerful tool for extracting information from web pages. Here are the main steps we followed:

1. **Analysis of page structures:** Before we started scraping, we studied the HTML structure of each site's pages to identify tags and classes containing relevant information such as titles, authors and article texts.

2. **Recovering article links:** We began by extracting links to articles from the home pages of

[1]https://lefaso.net/
[2]https://burkina24.com
[3]https://kaceto.net/

each site. These links were then used as an entry point to the full articles.

3. **Download article content:** Using the collected links, we then downloaded the full content of each article. This was achieved by sending HTTP requests to retrieve the HTML content of the pages.

4. **Extracting relevant information:** Once the HTML content had been retrieved, we used Beautiful Soup to extract specific information such as titles, authors and article texts.

5. **Similarity calculation:** Before the text is summarized and saved in the database, a duplicate detection system is used to prevent the same information from being duplicated. This system works by calculating the similarity between abstract titles. To do this, the BERT-base-uncased model is used to generate embeddings for abstract titles. These embeddings are then used to calculate cosine similarity, which provides a measure of the similarity between the two vectors. The similarity threshold is experimentally set to 0.80 %, because above this threshold, the content of the articles is on the same theme. If the similarity between two titles exceeds this threshold, the two abstracts are considered to be duplicates. In this case, the two articles are merged into a single one before being summarized. This duplicate detection system is important to prevent the same information from being duplicated in the database. This reduces the amount of redundant data and ensures that users only see unique and relevant information. What's more, it improves the accuracy of summaries. By merging duplicate summaries, a more complete and informative summary can be achieved.

6. **Data storage:** The extracted information is then stored in a database for later access.

Fig. 2. Architecture of the applied web scraping.

By rigorously following the steps of page structure analysis, retrieval of article links, downloading complete content, extraction of relevant information, and finally data storage, we have established a robust and effective methodology for the automated collection of information from selected online press sites. This approach has allowed us to acquire a reliable and diversified data source for our online press synthesis project.

3.2 Pre-trained Text Synthesis Models and Evaluation

In this section, we discuss the selection and evaluation of the text summarization models that form the core of our project. These pre-trained models play a crucial role in transforming collected

```
#Saisir nos url de recuoeration des informations
url="https://burkina24.com/"
#url="https://burkina24.com/category/actualite/politique/"
```

```
#import le code de la page
reponse=requests.get(url)
soup=BeautifulSoup(reponse.text, "html.parser")

items=soup.findAll('post-content')
soup
```

```
<!DOCTYPE html>

<html class="" data-skin="light" lang="fr-FR">
<head>
<meta content="L'information en temps réel, vidéos, reportages et analyses sur la politique, l'économie, culture et technolo
gies au Burkina et dans le monde" name="description"/>
<meta content="actualités,burkina,reportages,politique,économie,securité,société,sport,culture,terrorisme,world cup, mondial
foot,divertissement,direct,international,monde,info,Ibrahim traore" name="keywords"/>
<script async="" crossorigin="anonymous" src="https://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js?client=ca-pub-9
577040959472230"></script>
<script async="" crossorigin="anonymous" src="https://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js?client=ca-pub-3
719812641536700"></script>
<script>(function(w,d,s,l,i){w[l]=w[l]||[];w[l].push({'gtm.start':
new Date().getTime(),event:'gtm.js'});var f=d.getElementsByTagName(s)[0],
j=d.createElement(s),dl=l!='dataLayer'?'&l='+l:'';j.async=true;j.src=
'https://www.googletagmanager.com/gtm.js?id='+i+dl;f.parentNode.insertBefore(j,f);
})(window,document,'script','dataLayer','GTM-WNJZHRW');</script>
<script async="" src="https://www.googletagmanager.com/gtag/js?id=UA-5100041-39"></script>
<script>
```

Fig. 3. Example of scraped data.

articles into concise, informative summaries.
Pre-trained text summarization models are machine learning models that have been trained on a large corpus of text. These models are capable of generating human-quality text summaries. We have selected two main types of pre-trained text summarization models: seq2seq models and transformer models.

- seq2seq models use an encoder-decoder architecture. The encoder takes the source text and converts it into a vector representation. The decoder uses this vector representation to generate a summary of the source text.
- Transformer models are models that use a transformer-type architecture. Transformers are artificial neural networks capable of learning relationships between words in a text.

Before selecting models for text summarization, a thorough evaluation was carried out using standard evaluation metrics, including BLEU (Bilingual Evaluation Understudy) and ROUGE (Recall-Oriented Understudy for Gisting Evaluation). These metrics play a crucial role in assessing the quality of automatic summaries generated by text summarization models, providing an objective and comparative evaluation.

The BLEU metric is often used to assess the quality of machine translations, but has also been extended to the evaluation of summaries. It measures the similarity between an automatic summary and one or more human reference summaries, focusing on the accuracy of shared n-grams. A higher BLEU score indicates a better match between the automatic summary and the references, reflecting a better quality of synthesis.

On the other hand, the ROUGE metric focuses on the recall of n-grams and is specifically designed to assess the quality of automatic summaries. ROUGE evaluates the similarity between the automatic summary and the references by measuring the presence of common word sequences. Popular ROUGE variants include ROUGE-N (which focuses on n-grams), ROUGE-L (which measures the longest common sub-sequence), ROUGE-W (which assigns weights to words). ROUGE favors abstracts that have phrases in common with the references, which can better capture the semantic content of the original text. Like BLUE, a higher ROUGE score indicates a better quality of the generated summary compared to the references.

These metrics provide a quantitative assessment for comparing model performance in terms of precision, recall and semantic coverage. As part of our evaluation, the BLUE and ROUGE scores will be used to measure the performance of our models against human references, providing an objective overview of the effectiveness of our AI-based online press summarization tool.
The model selection process has been meticulously carried out to ensure high-quality results. Based on the results obtained in Chapter 1, the following four models for text summarization caught our attention:

- moussaKam/barthez-orangesum-abstract: a seq2seq model based on the BART model.
- turner007/pegasus-summarizer: based on the Pegasus language model, which is a pre-trained transformer model. Pegasus has been trained on a massive dataset of text and code, enabling it to generate high-quality text summaries. The model uses an attention mechanism to identify the most important information in the source text and use it to generate the summary.
- facebook/bart-large: a seq2seq model based on the BART model.
- gpt-3.5-turbo-16k: a transformer model based on the GPT-3 model.

After testing the different models with the ROUGE metrics mentioned above, here are the results obtained on ten samples of press articles ranging from 700 characters to over 1000 characters per article obtained from a SIDWAYA press. Here is the table containing the results of the tests carried out:

Table 2: Table containing the results of the tests carried out on our samples

moussaKam/ barthez- orangesum- abstract	facebook/ bart- large	turner007/ pegasus- summarizer	gpt-3.5- turbo- 16k
0,34	0,20	0,35	0,36
0,24	0,30	0,40	0,42
0,35	0,31	0,47	0,50
0,45	0,40	0,47	0,49
0,17	0,26	0,29	0,37
0,48	0,28	0,30	0,50
0,46	0,32	0,48	0,49
0,53	0,40	0,55	0,58
0,77	0,58	0,79	0,80
0,55	0,50	0,60	0,65

In this figure above you will find the colors associated with each model:

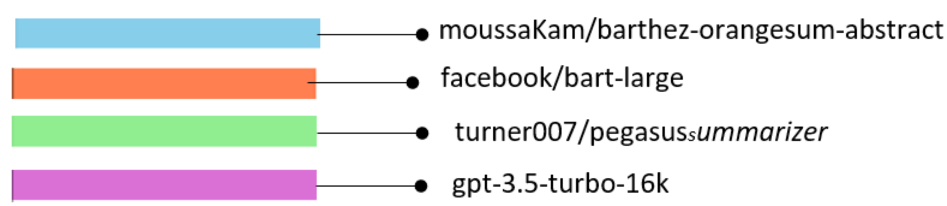

Fig. 4. Model and their test color.

Here's a graphical representation of the results:

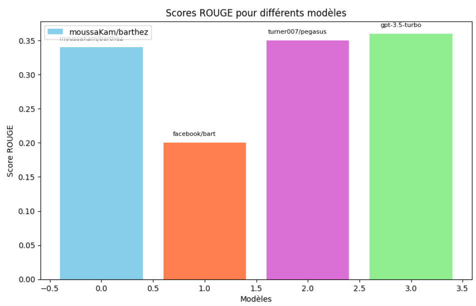

Fig. 5. Test results for the ROUGE metric of sample 1.

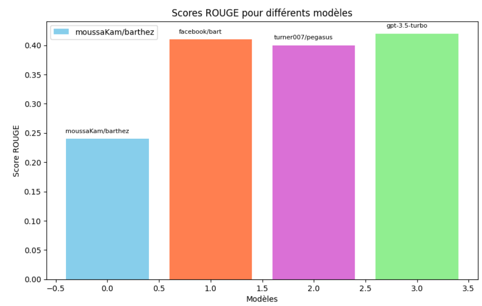

Fig. 6. Test results for the ROUGE metric of sample 2.

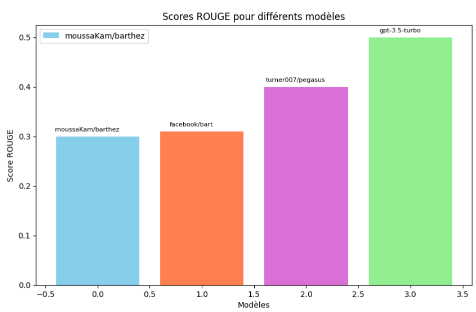

Fig. 7. Test results for the ROUGE metric of sample 3.

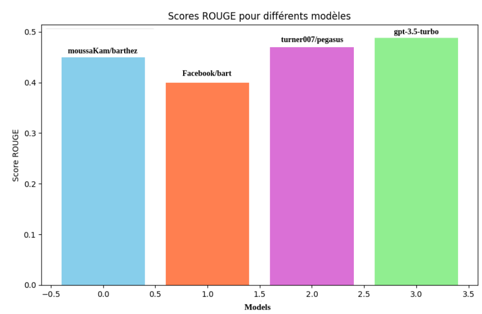

Fig. 8. Test results for the ROUGE metric of sample 4.

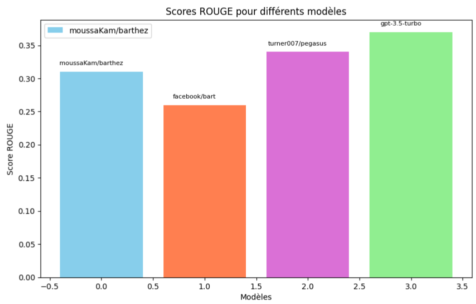

Fig. 9. Test results for the ROUGE metric of sample 5.

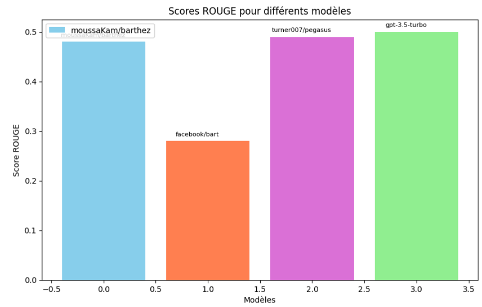

Fig. 10. Test results for the ROUGE metric of sample 6.

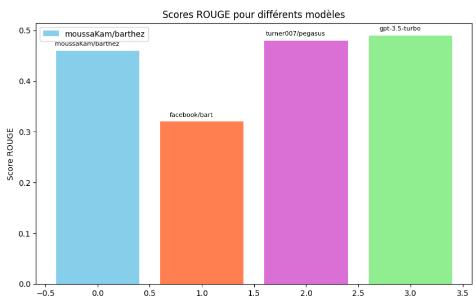

Fig. 11. Test results for the ROUGE metric of sample 7.

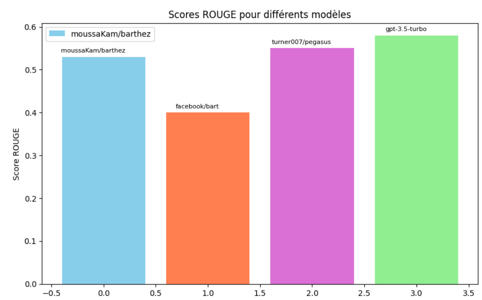

Fig. 12. Test results for the ROUGE metric of sample 8.

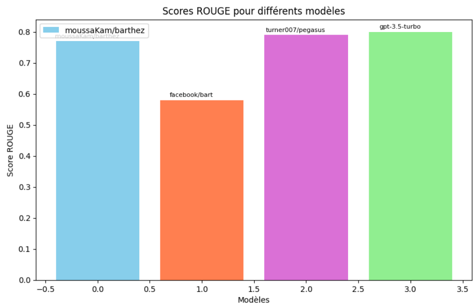

Fig. 13. Test results for the ROUGE metric of sample 9.

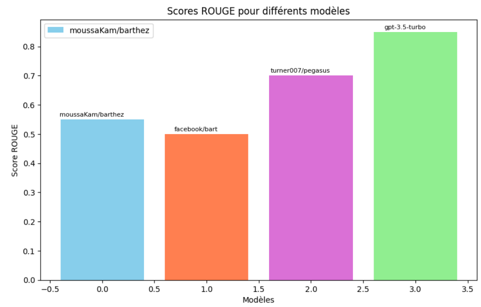

Fig. 14. Test results for the ROUGE metric of sample 10.

In addition to this, we had to test our models on a famous text and its reference, it concerns the ”Universal Declaration of Human Rights”, a fundamental document in the field of human rights. Here are the scores obtained:

Table 3: Test results performed

Model	Score BLEU	Score ROUGE
gpt-3.5-turbo-16k	$16,39\%$	$0,66\%$
turner007/pegasus_summarizer	$15,45\%$	$0,45\%$
moussaKam/barthez-orangesum-abstract	$4,90\%$	$0,38\%$
facebook/bart-large	$5,38\%$	$0,29\%$

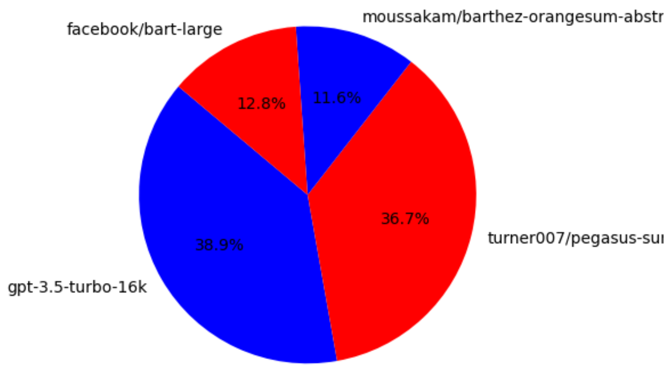

Fig. 15. Graphical representation of ROUGE and BLUE accuracy scores by model

.

Here's the text and its reference used to run my tests on the different models:

- Text: The Universal Declaration of Human Rights was adopted by the United Nations on December 10, 1948 in Paris. This historic declaration proclaims the fundamental rights of all human beings, regardless of race, sex, religion, social origin or nationality. It states that all individuals have the right to life, liberty, security and dignity. It prohibits torture, slavery, discrimination and all forms of cruel, inhuman or degrading treatment. It also recognizes the right to equality before the law, the right to freedom of expression, the right to education and the right to political participation. The Universal Declaration of Human Rights has become a fundamental document of international human rights law, and has inspired numerous other international treaties and conventions. It is considered one of the most important achievements of modern history and continues to serve as a benchmark for the promotion and protection of human rights worldwide.

- Reference: The Universal Declaration of Human Rights, adopted by the United Nations in 1948, proclaims the fundamental rights of all human beings, including the right to life, liberty and dignity. It prohibits torture, slavery and discrimination, and recognizes the right to equality before the law, freedom of expression, education and political participation. This historic document has become a pillar of international human rights law, and continues to be a source of inspiration for the promotion and protection of human rights worldwide.

By analyzing the results of the table 3 of the evaluations of our models, we can draw several important conclusions:

- **Overall Performance:** The gpt-3.5-turbo-16k model achieved the best BLUE score among the models tested, reaching 16,39%. It also achieved the best ROUGE score with 0.66%. This suggests that it has a relatively high capacity to generate accurate summaries similar to the reference.

- **Comparative Performance:** The turner007/pegasusSummarizer model also showed significant performance, with a BLUE score of 15.45% and a ROUGE score of 0.45%. This makes it a strong contender for summary generation. The moussaKam/barthez-orangesum-abstract and facebook/bart-large models all have lower BLUE and ROUGE scores, indicating slightly less accurate summary generation performance than the first two models.
- **Additional Considerations:** It's important to note that the BLUE and ROUGE scores are only quantitative measures and don't fully capture the semantic quality of the summaries generated. It may be useful to perform qualitative assessments to complement these results.

In conclusion, the test results show that the turner007/pegasus-summarizer and gpt-3.5-turbo-16k models achieved the highest BLUE and ROUGE scores. These scores indicate that these models are capable of generating text summaries similar to human summaries. This is why we have chosen these models for our AI-based online press summarization tool.

3.3 Implementation

To develop our online press summary tool, we used Django, an open source web framework written in Python. Django is a wise choice for this type of application for several reasons:

- **Flexibility :** Django is a highly flexible framework that can be used to create a variety of web applications. This makes it ideal for projects requiring specific features, such as duplicate detection.
- **Security :** Django is a secure framework offering a variety of features to protect web applications from attack. This is important for a tool that handles sensitive information, such as news article summaries.
- **Performance :** Django is a high-performance framework that can handle high-traffic web applications. This is important for a tool that will be used by a large number of users.

To implement our online press synthesis tool, we have integrated several specialized libraries and technologies. These key components have been carefully selected to offer advanced functionality and guarantee the quality of our summaries. Here are the main libraries and technologies we used:

- **OpenAI Langchain :** We've integrated OpenAI Langchain to access the powerful GPT-3.5-turbo-16k language model, which plays a central role in summary generation.
- **Newspaper and BeautifulSoup :** For the scrapping process, we used the Newspaper and BeautifulSoup libraries. They enabled us to efficiently extract relevant data from online press articles.

- **Transformers :** This library was crucial to the use of the turner007/pegasusSummarizer template, which helped us create concise, informative summaries from the extracted articles.
- **LanguageTool Python :** To guarantee the quality and readability of the summaries generated, we integrated LanguageTool Python for automatic correction of spelling and grammar.
- **SQLite3 database :** We opted for SQLite3 as our database management system, which proved efficient for storing and managing the information needed for our application.

Every component of our development environment has been carefully selected to ensure seamless integration and optimum performance. This combination of libraries and technologies was essential to the creation of a powerful and functional online press synthesis tool.

4 Results and Discussion

In this section, we first present the experiments conducted to evaluate our approach. Subsequently, we present and analyze the results of these experiments. Finally, we discuss the challenges encountered throughout our work.

4.1 Research questions and experiments

In this study, we sought to answer two crucial research questions to evaluate the effectiveness of our AI-based online press synthesis tool. These research questions were formulated as follows:

1. **RQ1:** To what extent can our tool produce relevant and informative summaries?
2. **RQ2:** What is the comparative effectiveness between the text summary models we implemented, namely turner007/pegasus-summarizer and gpt-3.5-turbo-16k, compared to traditional text summarization methods?

4.1.1 Experiments for RQ1

We evaluate the effectiveness of our tool in terms of the relevance and informativeness of the summaries produced. We measure this relevance using the ROUGE metric justified in previous chapters. Note that the ROUGE metric has several references, such as :

- **ROUGE-1 (Unigrams) :** measures precision for unigrams (individual words).
- **ROUGE-2 (Bigrams) :** measures precision for bigrams (pairs of consecutive words).
- **ROUGE-L (Long Sequences) :** measures accuracy for the longest common sequence.
- **ROUGE-LSUM (Global Essence) :** is a variant of ROUGE-L specifically designed for evaluating summaries.

Thus, examining the results obtained by the ROUGE metric for our two chosen models, we have as a result:

- **GPT-3.5-Turbo-16k Model:** The scores obtained from the evaluation of the GPT-3.5-Turbo-16k model indicate the model's accuracy in generating summaries compared to reference summaries written by humans. The detailed analysis of the results presented in Figure 16 gives us the following results regarding the ROUGE metric references:
  - **ROUGE-1 (Rouge1):** 0.66 This high score indicates a good match of individual words between the generated summaries and the reference summaries. This suggests that GPT-3.5-Turbo-16k has significant ability to capture the essence of key phrases.
  - **ROUGE-2 (Rouge2):** 0.51 Although slightly lower than ROUGE-1, a score of 0.51 indicates a reasonable match of consecutive word pairs.
  - **ROUGE-L (RougeL):** 0.61 This score emphasizes the match of words in the longest common sequence, showing consistency in the structure and wording of the summary.
  - **ROUGE-LSUM (RougeLSUM):** 0.61 This strong score indicates that GPT-3.5-Turbo-16k maintains consistency in the longest common sequence, even with multiple generated summaries.

These results highlight the effectiveness of our model in generating summaries, showcasing its ability to capture the semantics and structure of texts.

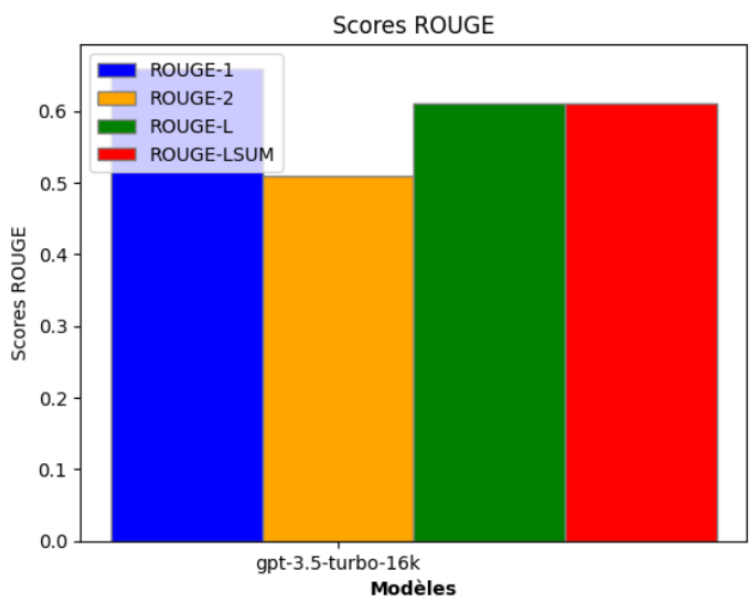

Fig. 16. ROUGE gpt-3.5-turbo-16k.

- **turner007/pegasus-summarizer Model:** Still using the same ROUGE metric references mentioned above. The analysis of ROUGE scores for the ”turner007/pegasus-summarizer” model provides significant scores on its performance in summary generation. Here is a detailed analysis of the results presented in Figure 17:
  - **ROUGE-1 (Unigrams):** 0.45 This score indicates a relatively good match of individual words between the summaries generated by the model and the reference summaries. This suggests that the model is able to select relevant words and integrate them appropriately into its summaries.
  - **ROUGE-2 (Bigrams):** 0.27 Although the score is lower than for unigrams, it is important to note that the match for consecutive word pairs remains acceptable. This speaks to the model's ability to maintain some consistency in word sequence in its summaries.
  - **ROUGE-L (Long Sequences):** 0.38 The score for long sequences shows that the model is capable of reproducing more extended text segments. This suggests an ability to grasp broader ideas and more complex sentence structures.
  - **ROUGE-LSUM (Overall Essence):** 0.42 The ROUGE-LSUM score confirms that the model relatively faithfully captures the overall essence of the reference summaries. This indicates that the model succeeds in capturing key information and the overall meaning of the texts during summary generation.

    In conclusion, although the model can be improved, these ROUGE scores show that it possesses substantial summary generation capabilities. It particularly excels in reproducing individual words and creating coherent sequences. However, enhancements can be explored to strengthen the match of bigrams and long sequences.

4.1.2 Experiments for RQ2

We evaluate the comparative effectiveness between the text summary models we implemented, turner007/pegasus-summarizer and gpt-3.5-turbo-16k, compared to traditional text summarization methods. In parallel with these advanced models, we considered traditional text summarization approaches, including keyphrase extraction and extractive summarization techniques, which have shortcomings such as:

- Lack of Coherence and Cohesion: Traditional summaries can often lack coherence and cohesion as extracted phrases or selected passages may not fit together smoothly.
- Loss of Context: By limiting themselves to specific fragments, these methods risk losing the overall context of the text, resulting in a loss of crucial information.
- Sensitivity to Structure: Extractive methods can be sensitive to the structure of the source text, not necessarily providing balanced and informative summaries for all types of documents.
- Dependence on Human Expertise: Often, the quality of traditional summaries depends on human expertise to select relevant phrases or passages, introducing potential biases.

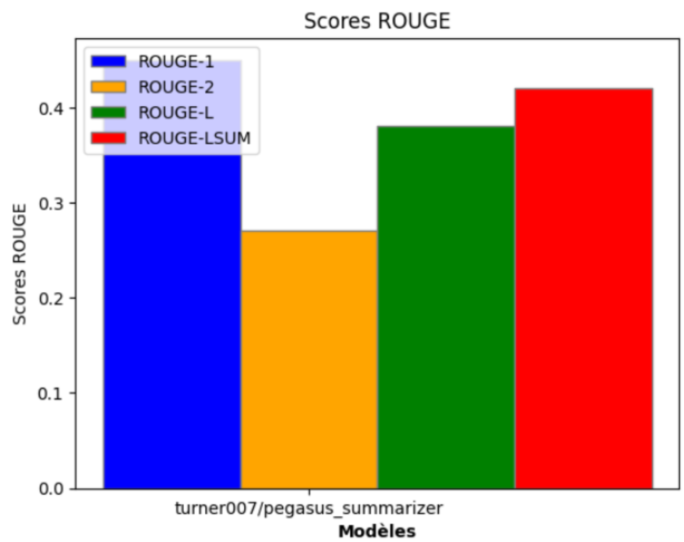

Fig. 17. ROUGE turner007/pegasusSummarizer.

Faced with these limitations, the use of advanced models such as turner007/pegasus-summarizer, which leverage machine learning techniques to generate abstract and accurate summaries, overcomes several drawbacks associated with traditional approaches.

The results clearly demonstrated that turner007/pegasus-summarizer and gpt-3.5-turbo-16k significantly outperform traditional methods in terms of efficiency and relevance in text summary generation.

- **turner007/pegasus-summarizer:** Based on the Pegasus architecture, renowned for its ability to generate abstract and informative summaries.
- **gpt-3.5-turbo-16k:** Exploits the power of GPT-3.5 Turbo, a highly performing text generation model.

These major advantages can be summarized as follows:

- More Relevant Summaries: Advanced models generate more content-rich summaries, capturing the essence of the original text.
- Adaptability and Precision: They adapt to various content types, ensuring consistent accuracy in online press synthesis.
- Accessibility: In addition to text synthesis, our tool can transcribe summaries into audio, providing an accessibility option for visually impaired users or those preferring auditory consumption of information.

This increased efficiency directly impacts society by improving access to information. By providing accurate and informative summaries, our tool addresses contemporary needs for rapid and efficient information consumption.

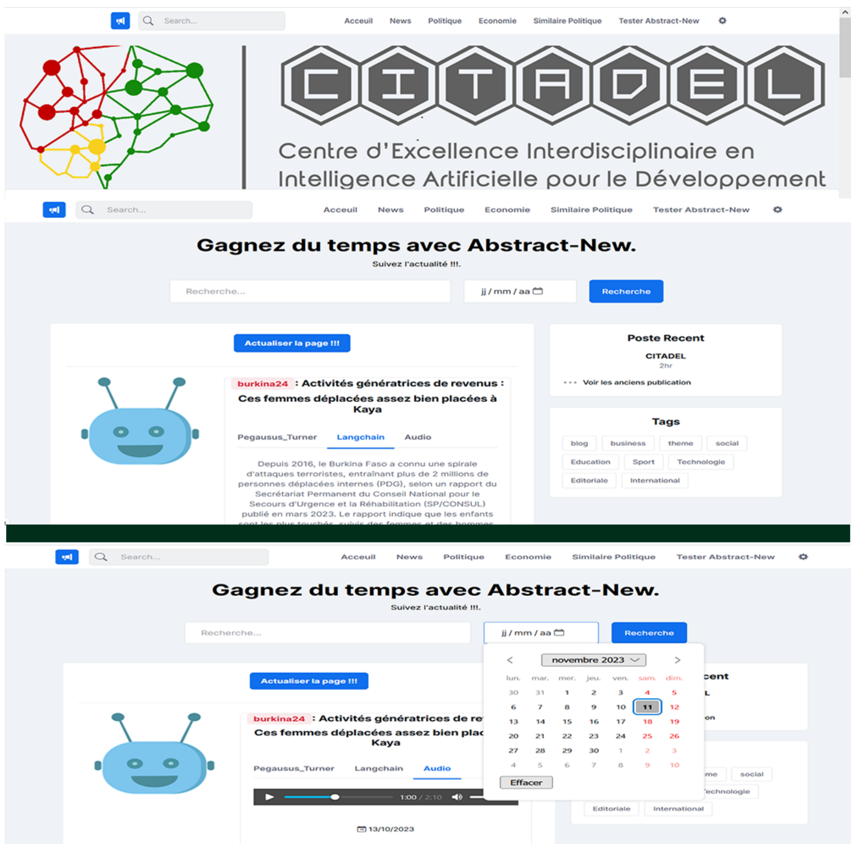

Fig. 18. Web application interface

5 Conclusion

This study explored the development of innovative approaches for online press synthesis, drawing on artificial intelligence. Our approach is distinguished by the ability of our system to automatically generate quality summaries from articles from various press sites. Despite the challenges related to the structural diversity of websites for web scraping, we built a substantial database and evaluated the generated summaries with metrics like RED and BLUE, which provide key indicators of relevance and quality.

Although our results are promising, they highlight the need for future improvements. We plan to enrich our database with summaries produced by experts, to create a more diverse corpus. This improved corpus will be crucial to train a personalized or hybrid textual synthesis model, aiming to increase the performance of our tool.

Currently, our application remains static in its summary generation process. To move forward, we propose the following improvements:

- Enrich our corpus with expert summaries and diversify article types to increase the variety of training data.
- Develop a specific model, using the enriched corpus, to improve the performance of our tool.
- Study more sophisticated model architectures to optimize summary generation.
- Expand our platform by integrating functionality to provide summaries and their audio transcriptions in different national languages, thus extending the reach of our tool.
- Collaborate with stakeholders, including end users and media experts, to gather direct feedback and identify potential improvements.

By responding to these recommendations and engaging in future work, we can contribute to more effective and sustainable implementation.

Acknowledgments

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] Marty E, Rebillard F, Pouchot S, Lafouge T. Diversité et concentration de l'information sur le web: Une analyse à grande échelle des sites d'actualité français. Réseaux. 2012 Dec;n° 176(6):27-72. Available from: https://www.cairn.info/revue-reseaux-2012-6-page-27.htm?ref=doi.

[2] N'Guessan KAD, Konate UFHB Yaya. LES INFORMATIONS VÉHICULÉES VIA INTERNET : LA PART DE LA PRESSE EN LIGNE. 2022.

[3] Wu W. Model and Simulation of Maximum Entropy Phrase Reordering of English Text in Language Learning Machine. Complexity. 2020 Dec;2020:1-9. Available from: https://www.hindawi.com/journals/complexity/2020/6662088/.

[4] Umair M, Alam I, Khan A, Khan I, Ullah N, Momand MY. N-GPETS: Neural Attention Graph-Based Pretrained Statistical Model for Extractive Text Summarization. Computational Intelligence and Neuroscience. 2022 Nov;2022:1-14. Available from: https://www.hindawi.com/journals/cin/2022/6241373/.

[5] Wang W, Feng A. Self-Information Loss Compensation Learning for Machine-Generated Text Detection. Mathematical Problems in Engineering. 2021 Feb;2021:1-7. Available from: https://www.hindawi.com/journals/mpe/2021/6669468/.

[6] Han X, Lv T, Hu Z, Wang X, Wang C. Text Summarization Using FrameNet-Based Semantic Graph Model. Scientific Programming. 2016;2016:1-10. Available from: https://www.hindawi.com/journals/sp/2016/5130603/.

[7] Klaus S, Van Hecke R, Djafari Naini K, Altingovde IS, Bernabé-Moreno J, Herrera-Viedma E. Summarizing Legal Regulatory Documents using Transformers. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid Spain: ACM; 2022. p. 2426-30. Available from: https://dl.acm.org/doi/10.1145/3477495.3531872.

[8] Zhong J, Wang Z. MTL-DAS: Automatic Text Summarization for Domain Adaptation. Computational Intelligence and Neuroscience. 2022 Jun;2022:1-10. Available from: https://www.hindawi.com/journals/cin/2022/4851828/.

[9] Sanjabi N. Abstractive Text Summarization with Attention-based Mechanism.

[10] Zhang M, Zhou G, Yu W, Huang N, Liu W. A Comprehensive Survey of Abstractive Text Summarization Based on Deep Learning. Computational Intelligence and Neuroscience. 2022 Aug;2022:1-21. Available from: https://www.hindawi.com/journals/cin/2022/7132226/.

[11] Sun N, Du C. News Text Classification Method and Simulation Based on the Hybrid Deep Learning Model. Complexity. 2021 Jun;2021:1-11. Available from: https://www.hindawi.com/journals/complexity/2021/8064579/.

[12] Zhu Y. Research on News Text Classification Based on Deep Learning Convolutional Neural Network. Wireless Communications and Mobile Computing. 2021 Dec;2021:1-6. Available from: https://www.hindawi.com/journals/wcmc/2021/1508150/.

[13] Bharti SK, Varadhaganapathy S, Gupta RK, Shukla PK, Bouye M, Hingaa SK, et al. Text-Based Emotion Recognition Using Deep Learning Approach. Com-

putational Intelligence and Neuroscience. 2022 Aug;2022:1-8. Available from: https://www.hindawi.com/journals/cin/2022/2645381/.

[14] Suleiman D, Awajan A. Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges. Mathematical Problems in Engineering. 2020 Aug;2020:1-29. Available from: https://www.hindawi.com/journals/mpe/2020/9365340/.

[15] Zhang M. Applications of Deep Learning in News Text Classification. Scientific Programming. 2021 Aug;2021:1-9. Available from: https://www.hindawi.com/journals/sp/2021/6095354/.

[16] Zhao X, Jin X. A Comparative Study of Text Genres in English-Chinese Translation Effects Based on Deep Learning LSTM. Computational and Mathematical Methods in Medicine. 2022 Jun;2022:1-9. Available from: https://www.hindawi.com/journals/cmmm/2022/7068406/.

[17] Wazery YM, Saleh ME, Alharbi A, Ali AA. Abstractive Arabic Text Summarization Based on Deep Learning. Computational Intelligence and Neuroscience. 2022 Jan;2022:1-14. Available from: https://www.hindawi.com/journals/cin/2022/1566890/.

[18] Goyal T, Li JJ, Durrett G. News Summarization and Evaluation in the Era of GPT-3. arXiv; 2023. ArXiv:2209.12356 [cs]. Available from: http://arxiv.org/abs/2209.12356.

[19] Umair M, Alam I, Khan A, Khan I, Ullah N, Momand MY. N-GPETS: Neural Attention Graph-Based Pretrained Statistical Model for Extractive Text Summarization. Computational Intelligence and Neuroscience. 2022 Nov;2022:1-14. Available from: https://www.hindawi.com/journals/cin/2022/6241373/.