Journal of Information Technology & Politics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/witp20

Political-RAG: using generative AI to extract political information from media content

Muhammad Arslan, Saba Munawar & Christophe Cruz

To cite this article: Muhammad Arslan, Saba Munawar & Christophe Cruz (23 Oct 2024): Political-RAG: using generative AI to extract political information from media content, Journal of Information Technology & Politics, DOI: 10.1080/19331681.2024.2417263

To link to this article: https://doi.org/10.1080/19331681.2024.2417263

Published online: 23 Oct 2024.

JOURNAL OF INFORMATION TECHNOLOGY & POLITICS
https://doi.org/10.1080/19331681.2024.2417263

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Political-RAG: using generative AI to extract political information from media content

Muhammad Arslan, Saba Munawar, and Christophe Cruz

ABSTRACT

In the digital era, media content is crucial for political analysis, providing valuable insights through news articles, social media posts, speeches, and reports. Natural Language Processing (NLP) has transformed Political Information Extraction (IE), automating tasks such as event extraction and sentiment analysis. Traditional NLP methods, while effective, are often task-specific and require specialized expertise. In contrast, Large Language Models (LLMs) powered by Generative Artificial Intelligence (GenAI) offer a more integrated solution. However, domain-specific challenges persist, which led to the development of the Retrieval-Augmented Generation (RAG) framework. RAG enhances LLMs by incorporating external data retrieval, addressing issues related to data availability. To demonstrate RAG's capabilities, we introduce the Political-RAG system, designed to extract political event information from media content, including Twitter data and news articles. Initially developed for event extraction, the Political-RAG system lays the foundation for developing various complex Political IE tasks. These include detecting hate speech, analyzing conflicts, assessing political bias, and evaluating social trends, sentiment, and opinions.

KEYWORDS

Natural Language Processing; Information Extraction (IE); Large Language Models (LLMs); Retrieval-Augmented Generation (RAG); Political events extraction

Introduction

In today's digital age, media content serves as a rich source of information for political scientists, offering insights into various aspects of political landscapes, events, and sentiments (Barberá & Steinert-Threlkeld, 2020; Demiros et al., 2008). Media content includes news articles, social media posts, speeches, reports, and more. Understanding and analyzing media content is crucial for political scientists to gain valuable insights into public opinion, political trends, and policy implications (Barberá & Steinert-Threlkeld, 2020). While manual methods for extracting political information from media content have long been employed (Piskorski & Yangarber, 2013; Small & Medsker, 2014), they are often time-consuming, labor-intensive, and prone to human error. With the exponential growth of digital data, manual methods have become increasingly inadequate for handling the sheer volume of information available.

The advent of NLP has revolutionized the process of Information Extraction (IE) from media content (Zhou et al., 2020). NLP techniques enable automated analysis of text data, allowing political scientists to derive meaning from large collections of textual information. These techniques have significantly enhanced the efficiency and accuracy of IE tasks in various domains, including political science. There is a wide array of NLP methods designed specifically for extracting different types of information within the realm of political science (Hobbs & Riloff, 2010). These include event detection, analysis, and prediction; hate speech and conflict detection; political bias, profiling, and social analytics; sentiment, opinion, and trend analysis; political polarization, and risk management, among others (see Figure 1). Each of these tasks plays a vital role in understanding political dynamics, shaping public discourse, and informing policy decisions.

The recent development of LLMs based on GenAI principles is a major advancement in extracting information from media content (J. Yang et al., 2023a). LLMs can produce text that closely resembles human language and exhibit an impressive capacity for contextual understanding (Touvron et al., 2023). These models have showcased remarkable performance across a spectrum of NLP tasks, including language translation, text summarization, and sentiment analysis (Touvron

CONTACT Muhammad Arslan muhammad.arslan@u-bourgogne.fr; muhammad.arslan@uwe.ac.uk School of Architecture and Environment, University of the West of England, Bristol, United Kingdom; Laboratoire Interdisciplinaire Carnot de Bourgogne (ICB), Université de Bourgogne, Dijon, France

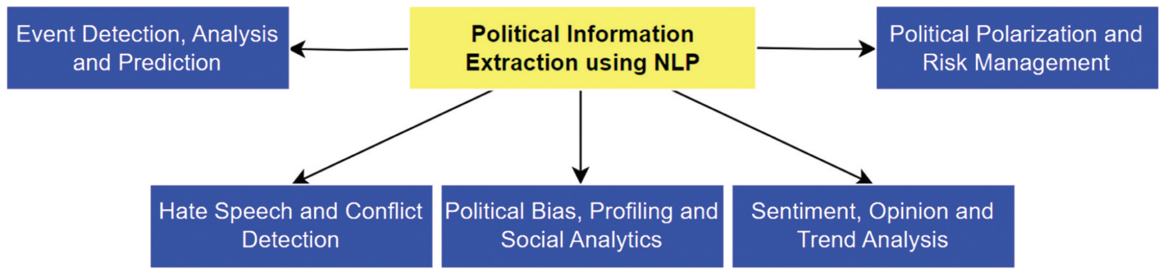

Figure 1. NLP-based political IE tasks.

et al., 2023). However, while LLMs hold great promise for IE tasks, they often encounter difficulties when confronted with domain-specific inquiries, leading to inaccuracies and irrelevant outputs, particularly in scenarios where data availability is limited.

In response to this challenge, the Retrieval-Augmented Generation (RAG) framework was introduced (Lewis et al., 2020), aiming to augment the capabilities of LLMs by integrating external data retrieval into the generative process. Through the utilization of external knowledge sources during inference, RAG mitigates the risk of unreliable information generation, thereby supporting the suitability of LLMs for practical applications in real-world settings (Lewis et al., 2020).

This paper introduces the Political-RAG system, a framework for retrieving political event updates from tweets and news articles using the RAG framework integrated with LLMs, specifically the Llama2 model. This approach combines the generative capabilities of LLMs with the retrieval capabilities of RAG, resulting in a deeper contextual understanding of political events and more accurate extraction results. The system also allows for updates to news datasets (Misra, 2022), keeping analysts informed about the latest political developments. By monitoring tweets and news articles, it supports timely decision-making and analysis for researchers, policymakers, and other stakeholders.

Background

The literature relevant to this research falls into two main categories. The first segment introduces Political IE and underscores its significance in informing political decisions. It also reviews prior studies utilizing diverse NLP techniques for Political IE tasks. The second segment introduces applications of RAG with LLMs. Ultimately, this research identifies a gap by underscoring the lack of studies focusing on RAG with LLMs tailored specifically for Political IE.

Political IE

The literature on Political IE spans various applications, including event detection, sentiment analysis, hate speech detection, political bias analysis, and political polarization studies (Haq et al., 2020). However, these represent only a subset of the diverse applications within Political IE. Event detection research focuses on identifying political events like protests and conflicts, employing techniques such as syntactic and semantic analysis (Lorenzini et al., 2022). Sentiment analysis explores public sentiment toward political figures and policies (Pair et al., 2021), while hate speech detection aims to check harmful speech using multimodal detection methods (Aziz et al., 2023). Political bias analysis investigates affiliations and ideologies (Demidov, 2023), while political polarization studies seek to understand and address polarization across different societal groups (Varma & Harsha, 2022). Across these Political IE applications, a variety of NLP techniques are employed, tailored to different datasets such as news articles, Twitter data, debate speeches, blogs, etc. These techniques facilitate the extraction of diverse political information for tasks such as event detection, sentiment analysis, hate speech detection, political bias analysis, and political polarization studies, as outlined in Table 1.

The benefits of these systems are well documented in their respective studies. With numerous systems available, selecting and validating the

Table 1. Exploring existing political information extraction (IE) studies.

IE Type	No.	Use case	Methods	Dataset used
Event Detection, Analysis and Prediction	1	Protest event analysis over time and space (Lorenzini et al., 2022)	Keyword search, location-based filter, document classifier, etc.	News articles
	2	Systematic analysis of triggers of state-led mass killings (Burley et al., 2020)	Support Vector Machines (SVMs) and Recurrent Neural Networks (RNNs)	News articles
	3	Extracting political events using syntax and semantics (Halterman, 2020)	Convolutional Neural Network (CNN), and Latent Dirichlet Allocation (LDA)	News articles
	4	Event detection in smart cities (Hodorog et al., 2022)	Multinomial Naive Bayes Classifier, Complement Naive Bayes Classifier and Random Forest Classifier	Twitter data
	5	Linking events and locations in political text (Halterman, 2018)	Neural-net based model	News articles and Wikipedia
	6	Automated acquisition of patterns for coding political event data (Makarov, 2018)	Stanford CoreNLP toolkit, and Cosine similarity	LexisNexis data service, English Gigaword documents, ACE 2005, etc.
	7	Coding political events (Liang et al., 2018)	NER model and Wiki-bio	Wikipedia data and Gigaword text corpus
	8	Detecting battle events (Tanev et al., 2023)	XLM-RoBERTa, NEROne, NEXUS and Mordecai3	Telegram data
	9	Event prediction by few-shot abductive reasoning (Shi et al., 2024)	GPT-3	Political events, interactions between socialpolitical entities and user reviews.
	10	Event detection from real-time streaming data (Singh et al., 2023a)	Conditional Term Frequency-Average Inverse Window Frequency (CTF-AIWF) based on TF-IDF	Twitter data
	11	Multilingual protest events detection (Hürriyetoğlu et al., 2021)	SVM, Big-Bird-RoBERTa and XLM-RoBERTa	News articles
	12	Protest event identification (Suri et al., 2022)	Multilingual BERT	News articles
	13	Multilingual socio-political and crisis event detection (Hettiarachchi et al., 2021)	BERT-based variants	News articles
	14	Civil unrest event detection (Delucia et al., 2023)	Multi-Instance Learning (MIL) approach	Twitter data
	15	Protest event detection (Won et al., 2017)	CNN	UCLA protest image dataset
	16	Approximate pattern matching for event detection (Tanev, 2024)	BERT	News snippets
Hate Speech and Conflict Detection	17	Multimodal hate speech detection (K. Singh et al., 2023b)	Xlm-Roberta-base	Text-embedded images
	18	Multimodal hate speech detection (Aziz et al., 2023)	BiLSTM module	Text-embedded images
	19	Stance and hate event detection related to climate activism (Thapa et al., 2024)	BERT and its variants	Twitter data
	20	Detection of hate speech and targets (Avanthika et al., 2023)	Logistic regression and SVM classifiers	Text-embedded images
	21	Hate speech detection (Esackimuthu & Balasundaram, 2023)	ALBERT (A Lite BERT) Base v1 and Artificial Neural Network (ANN)	Text-embedded images
	22	Multimodal hate speech detection (Yamagishi, 2024)	LLaVA-1.5	Text-embedded images
	23	Hate speech and text-image correlation detection in real life memes (Armenta-Segura et al., 2023)	Pre-trained BERT	Text-image memes
Political Bias, Profiling and Social Analytics	24	Political profiling (Mallavarapu et al., 2018)	Term Frequency (TF) and Term Frequency - Inverse Document Frequency (TF-IDF), and Random Forest model	Googles Knowledge graph, Wikipedia, and news articles
	25	Social analytics on Russia–Ukraine cyber war (Sufi, 2023)	TF-IDF and LDA	Twitter data
	26	Categorizing statements to values on the political compass (Hey, 2023)	LinearSVM, BERT, and OpenAI's ChatGPT3.5	Occupy Democrats data, Cato Institute data, and news articles.
	27	Analyzing political party manifestos (Orellana & Bisgin, 2023)	BERTopic	Manifesto Project database
	28	NLP for policymaking (Jin & Mihalcea, 2022)	Text classifier, topic modelling, event extraction, and text scaling.	Twitter and Facebook data, survey responses, and news articles.
	29	Socio-political news classification (Büyüköz et al., 2020)	ELMo and DistilBERT	News articles

(*Continued*)

Table 1. (Continued).

IE Type	No.	Use case	Methods	Dataset used
	30	Analysis of Portuguese political parties' communication (Costa et al., 2021)	Natural Language Toolkit (NLTK), TF-IDF, and neural network.	Twitter data
	31	Detection of criminal organizations (Osorio & Beltran, 2020)	Extreme Gradient Boosting (XGB) model, ALBERT, Multinomial Naive Bayes (NB), Support Vector Machine (SVM) model and CNN models.	News articles
	32	ConfliBERT-Arabic: A model for politics, conflicts and violence (Alsarra et al., 2023)	BERT	News articles
	33	ConfliBERT-Spanish: A model specialized in political conflict and violence (W. Yang et al., 2023b)	BERT	News articles
	34	ConfliBERT: A model for conflict and political violence (Hu et al., 2022)	BERT	United Nations' websites and databases, government sources, Wikipedia, etc.
	35	Semantic political knowledge inference (Doumit & Minai, 2011b)	LDA	News articles
	36	Analyzing parliamentary debates and political cohesion (Sawhney et al., 2020)	GPolS: Graph Political Sentiment analyzer: a neural framework	Debate speeches
	37	Identify viral moments in the U.S. presidential debate (Lukito et al., 2019)	ARIMAX model, time series analysis and domain adapted word embeddings.	Twitter data
	38	Trace socio-political issues (Katre, 2019)	LDA and Comparative Visual Analytics	Political speech transcripts
	39	Online news media bias analysis (Doumit & Minai, 2011a)	LDA and Antelope	News articles
	40	Identifying stance by analyzing political discourse (Johnson & Goldwasser, 2016)	Weakly supervised Machine Learning (ML) model	Twitter data
	41	Event-level moral opinions (Lei et al., 2024)	Longformer and Bi-LSTM	News Articles
	42	Political bias of news content (Demidov, 2023)	BERT	News articles
	43	Political polarity classification (Varma & Harsha, 2022)	Word2Vec, Naïve Bayes classifier and Random Forest Classifier.	Twitter data
	44	Political bias in the media (Ali et al., 2024)	ConvBERT	News articles
	45	Analyzing political content (Kawintiranon & Singh, 2022)	PoliBERTweet	Twitter data
Sentiment, Opinion and Trend Analysis	46	Automatic detection of political opinions (Maynard & Funk, 2012)	GATE and NER	Twitter data and WordNet
	47	Context and emotion classification of political speeches (Efat et al., 2023)	XGB Classifier, Cat Boost Classifier and Linear Discriminant Analysis	Political speeches
	48	Quantification of gender bias and sentiment toward political leaders (Pair et al., 2021)	GloVe (Pennington et al., 2014) and Word2Vec (Mikolov et al., 2013).	News articles
	49	Analyzing political sentiment (Bose et al., 2019)	NRC Emotion Lexicon, ParallelDots AI APIs by ParallelDots, Inc.	Twitter data
	50	Sentiment analysis of Nigerian presidential election (Oyewola et al., 2023)	Long Short-Term Memory (LSTM), Peephole Long Short-Term Memory (PLSTM), and Two-Stage Residual Long Short-Term Memory (TSRLSTM) models	Twitter data
	51	Target-aware opinion classification (Bestvater & Monroe, 2023)	SVM and BERT	Twitter data
	52	Identifying trends towards Indian general elections (Almatrafi et al., 2015)	Naive Bayes algorithm	Twitter data
	53	Real-time sentiment analysis of election tweets (Hitesh et al., 2019)	Word2vec and Random Forest Model	Twitter data
	54	Understanding people emotions towards national parties (Kuamri & Babu, 2017)	Unsupervised lexicon-based approach	Twitter data
	55	Detecting political leniency (Kowsik et al., 2024)	NLTK and VADER	Twitter data
	56	Opinion analysis of bi-lingual event data (Javed & Saeed, 2023)	LSTM classifier	Twitter data
	57	Fine-tuning language models on Dutch protest event tweets (Loerakker et al., 2024)	Bernice, TwHIN-BERT and Sentence Transformers	Twitter data

(*Continued*)

Table 1. (Continued).

IE Type	No.	Use case	Methods	Dataset used
Political Polarization and Risk Management	58	Analyzing political polarization (Sharber, 2020)	Ensemble learner	Blogs and political websites
	59	Political risk management (W. Wang et al., 2023)	LDA	News articles
	60	Understanding political polarization (Gode et al., 2023)	Word2Vec, Doc2Vec and Longformer	Individual pages of politicians
	61	Quantifying polarization across political groups on key policy issues (Bor et al., 2023)	Valence Aware Dictionary for sEntiment Reasoning (VADER)	Twitter data
	62	Predicting content-based political inclinations (Rahmati et al., 2023)	LSTM and CNN with BERT	Twitter data

accuracy of each one according to specific organizational needs becomes daunting. Moreover, most systems are not openly accessible, and organizations may lack sufficient proprietary data to train models from scratch. These factors demand considerable time and resources, which small- to medium-sized organizations often lack for dedicated Research and Development (R&D) efforts. Given these challenges, we sought to leverage the potential of existing LLMs, which are readily accessible and pre-trained on extensive datasets. However, while these LLMs excel in general applications, they often fall short in specialized tasks like Political IE. To address this limitation, we turned to RAG to enhance the capabilities of LLMs specifically for Political IE tasks. The subsequent section elaborates on the existing literature in this domain.

Retrieval-Augmented Generation (RAG) with LLMs, driven by GenAI

LLMs, based on GenAI, are advanced machine learning models that generate human-like text (Raiaan et al., 2024). Despite their proficiency in NLP tasks related to IE, they often encounter challenges with domain-specific queries, leading to inaccuracies or irrelevant outputs. To address this, RAG enhances the generative capabilities of LLMs by incorporating an initial step where pertinent information is retrieved from external sources before text generation (Lewis et al., 2020). This integration notably enhances the accuracy and relevance of the generated content, thereby minimizing errors and elevating overall quality (Lewis et al., 2020). Consequently, this augmentation renders LLMs more applicable for real-world scenarios, ensuring that the generated content is well grounded in evidence and reliable.

The literature on the integration of LLMs with RAG spans various domains and applications. For instance, MIRAGE, a medical information LLM, utilizes RAG to furnish medical information (Xiong et al., 2024). In financial report chunking, RAG is employed to enhance context and information accuracy (Jimeno Yepes et al., 2024). Additionally, RAG finds applications in Electrocardiography (ECG) analysis (Yu et al., 2023) and Representative Vector Summarization (RVS) with RAG-assisted abstractive-extractive workflows (Manathunga & Illangasekara, 2023). Beyond these, RAG is employed in various contexts such as retrieval-augmented controllable review generation (Kim et al., 2020), knowledge graph reasoning (Sha et al., 2023), and extracting answers from table corpora (Pan et al., 2022).

In the medical field, RAG aids in developing liver disease-specific LLMs (Ge et al., 2023) and clinical language models (Zakka et al., 2024). Furthermore, RAG supports practices like tutoring assessment (Han et al., 2024) and frontline health worker capacity building (Al Ghadban et al., 2023). In biomedical text processing, RAG is utilized for frameworks like self-BioRAG (Jeong et al., 2024) and real-time composition assistance (Xia et al., 2023). Moreover, RAG finds applications in product information retrieval (Rackauckas, 2024), disaster reporting such as flood disasters (Colverd et al., 2023), and a range of NLP tasks including retrieval-augmented text-to-image generation (Chen et al., 2022) and patch generation for automatic program repair (W. Wang et al., 2023).

Upon examining the aforementioned studies, it becomes apparent that the integration of RAG with LLMs yields promising results in extracting structured information across diverse fields such as medicine, biomedical research, and software engineering. This integration enhances contextual understanding and ensures information accuracy, providing valuable insights for informed decision-making. However, despite its success in various domains, its application in the political realm remains relatively unexplored. Particularly, its potential to optimize and refine the IE process, thereby enhancing the efficacy of political information systems (Wang, 2023), warrants further exploration. To address this gap, we propose exploring the integration of RAG with LLMs to develop a Political-RAG system for more effective event extraction.

Extracting political events with GenAI

This section starts by defining political events and examining the properties that improve their contextual understanding. It then describes two datasets used for political event extraction: the first contains uncategorized COVID-19-related tweets focused on politics (short texts), and the second includes news articles (long texts) covering various topics like Politics, Health, and Business from the same period. Next, the Political-RAG system, built on RAG and LLM, is introduced. The section concludes with an evaluation of the proposed system's effectiveness in extracting political events from these datasets.

Defining political events

Political events, as defined by Davenport and Ball (2002), encompass actions undertaken by political actors within a specific temporal and spatial context. While various properties of political events are discussed in the literature, this proof-of-concept demonstration adopts the schema of event properties outlined by Halterman (2021). This schema outlines eight potential properties for defining a political event (see Figure 2). The "agent" property identifies the actor performing the action, typically represented by subject nouns. The "action" property describes the event's action, which includes at least one verb. The "recipient" property provides details about the actor receiving the action. The "instrument" property covers objects used by the actor or how the action is executed. The "reason" property explains the cause behind the reported event in political text. The "time" property indicates when the event occurred, while the "location" property specifies where it took place. Finally, the "reporter" property identifies the source reporting the event, such as "according to local sources." While an event

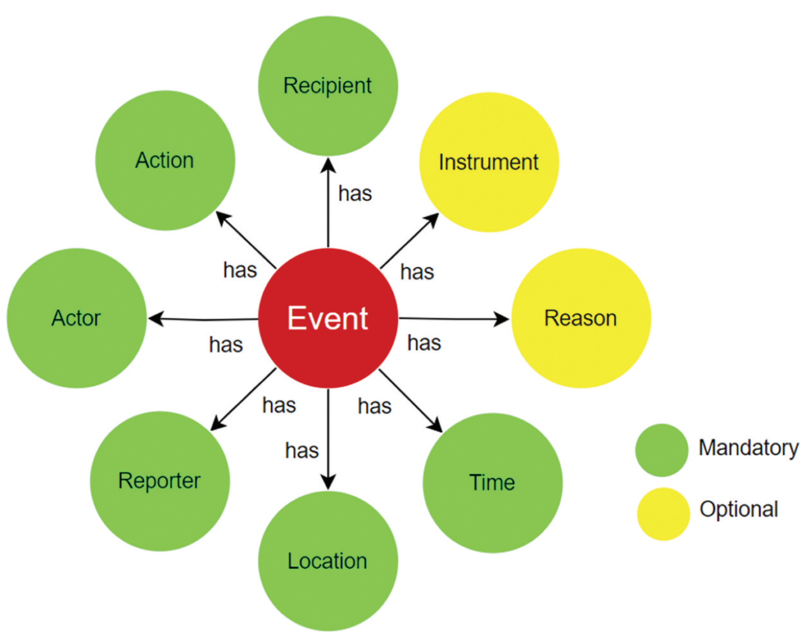

Figure 2. Event properties: mandatory and optional.

must include at least one action and one actor or recipient, other properties are optional, and sentences providing all eight pieces of information are rare (Halterman, 2021).

Datasets for political event extraction

We selected two open datasets for political event extraction. The first, "Coronavirus Tweets," includes approximately 56,000 tweets from March 29, 2020, in various languages. We filtered this to 10,000 English-language tweets because they are short and numerous. The second dataset, "Coronavirus News (COVID-19)," consists of around 1,000 categorized news articles from 2020. While the tweet dataset lacks topic categorization, the news articles are categorized into topics such as Politics, Business, and Health. Both datasets were collected during the pandemic. We chose these two datasets to evaluate the model's performance on short texts like tweets and long, detailed texts like news articles. Additionally, this approach allows us to assess the system's capability to extract events related to different topics beyond Politics, such as Business and Health. In the next section, we will describe the proposed Political-RAG system and its application to these datasets.

Political-RAG system

After identifying the key properties for constructing political events and selecting the datasets for event extraction, we utilized the RAG framework with the LLM, powered by GenAI, to develop the Political-RAG system. This approach automates the extraction of political events, eliminating the need for manual extraction of each event property. To facilitate interactive exploration of political events, we developed a chatbot assistant based on insights from recent literature (Jo et al., 2023; Linegar et al., 2023), enhancing seamless interaction with analysts for effective information exploration. We reviewed several publicly available LLMs, including BLOOM (Le Scao et al., 2022), Falcon (Almazrouei et al., 2023), GPT-4 (Achiam et al., 2023), Llama2 (Touvron et al., 2023), and Chinchilla (Hoffmann et al., 2022). We chose Llama2 for its superior performance and diverse training data, making it the most suitable for our needs (Touvron et al., 2023). To customize it for political-specific tasks and effectively utilize Political IE from news articles and Twitter datasets, we integrated RAG technology into our solution. This integration allows the chatbot assistant to better understand and respond to political-related queries with extracted information. The chatbot's

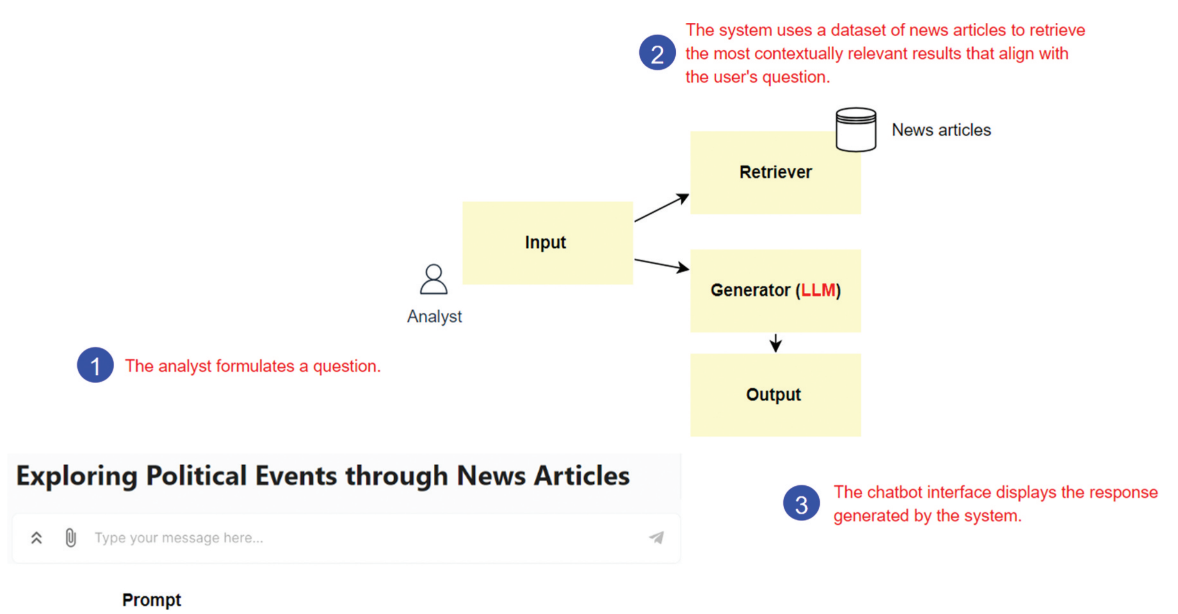

Figure 3. A scenario for developing a RAG with LLM system for exploring political events involves a user entering a query through the interface. The system leverages RAG combined with LLMs to analyze political datasets and generate a response based on the user's question.

effectiveness depends on the quality and comprehensiveness of the provided news dataset.

To create the chatbot assistant, we used the scenario in Figure 3 as a foundation, connecting Llama2 with a dataset to develop a RAG application for extracting political events and responding to queries. The process involves installing packages, importing modules, logging into Hugging Face (https://huggingface.co/), and initializing HuggingFaceLLM and LangchainEmbedding (https://docs.llamaindex.ai/en/stable/examples/embeddings/Langchain/) with ServiceContext. Datasets are loaded, embeddings are generated, and an index is constructed. Queries are executed using the query engine, leveraging Llama2 to retrieve relevant answers. The responses, containing extracted political events, are processed and displayed as text-based answers on the chatbot assistant terminal (see Figure 4), built using the Chainlit library (https://chainlit.io/). This terminal remains active for continuous interaction with analysts.

System evaluation

The Political-RAG system for Political IE, designed to enhance efficiency using the RAG with LLM model (see Figure 4), was evaluated to determine its effectiveness in extracting political events. The evaluation focused on identifying named entities (e.g., individuals and organizations) and extracting properties related to political events (see Figure 2). Two datasets were used for the system evaluation: Twitter dataset and news articles. This approach assessed the system's performance across different types (short texts and long texts) of political data and varying data volumes: 3,000, 6,000, and 10,000 tweets, and 300, 600, and 1,000 news articles. Additionally, the system was evaluated on Business and Health-related events using the news articles dataset to determine its effectiveness across topics beyond Politics.

For the evaluation of the Twitter dataset, we created a set of 150 questions focused solely on Politics. In contrast, the news articles dataset was evaluated using a total of 150 questions, with 50 questions each for the topics of Politics, Business, and Health. Since the Twitter dataset lacks topic categorization, it was not evaluated by topic. However, the news articles dataset includes topic-based categorization, which was used for its evaluation. A few sample questions (user queries), system responses, and ground truth answers (the correct responses provided by us, serving as the benchmark) are presented in Tables 2 and 3.

The evaluation, based on the data in Tables 2 and 3, was conducted using Precision and Recall metrics (Miao & Zhu, 2022). Precision measures the relevance of the retrieved context to the given question, while Recall assesses whether the retriever captures all relevant components necessary for answering the question. This ensures no pertinent information is missed during retrieval. By

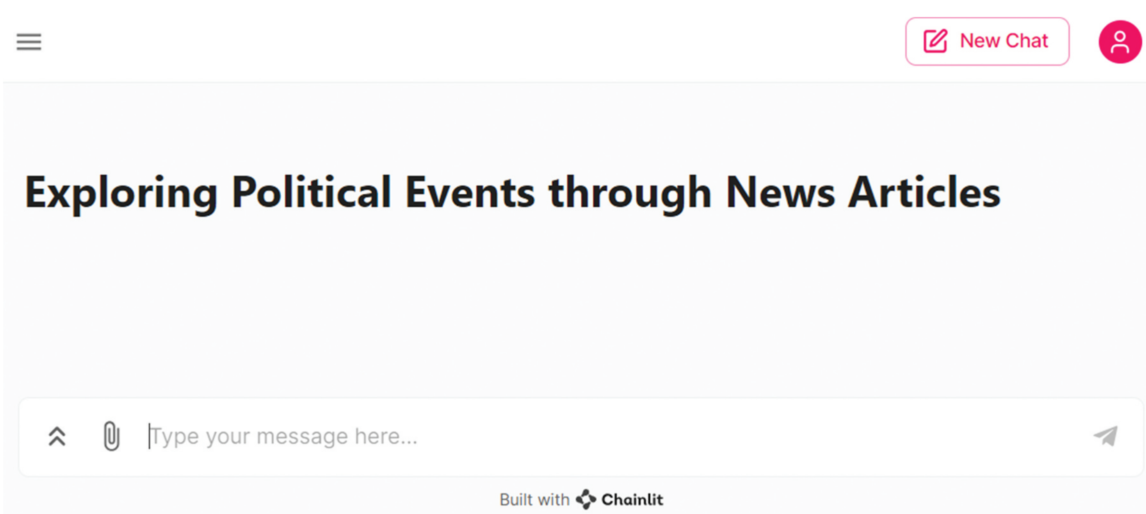

Figure 4. Designing an interface for the political-rag system to explore political events allows users to input questions related to politics.

Table 2. Extraction of events from tweets.

No.	Question	System Response	Ground Truth Answer
1	What measures are hospitals implementing to ensure continuity of medical care during the coronavirus lockdown?	Hospitals are implementing video consultations, tele-ICUs, and virtual in-house meetings to ensure continuity of medical care and minimize the spread of infection.	Hospitals are putting in place video consultations, tele-ICUs, and virtual in-house meetings to ensure continuity of medical care and minimize the spread of infection.
2	What is the call to action regarding Trump's accountability during the coronavirus pandemic?	The call to action is to "organize to ensure Trump is held accountable for all powers of the presidency he may try to exploit."	There is a call to organize to ensure Trump is held accountable for all powers of the presidency he may try to exploit.
3	What steps has the Saudi Ministry of Education taken to combat the coronavirus?	The Saudi Ministry of Education is contributing to the Health Ministry's efforts to combat the coronavirus by equipping six university hospitals to be on the frontline.	The Saudi Ministry of Education is contributing to the Health Ministry's efforts with six university hospitals equipped to be on the frontline.

Table 3. Extraction of events from news articles.

Topic	No.	Question	System Response	Ground Truth Answer
Politics	1	What was the nature of the interaction between China and the United States?	The nature of the interaction between China and the United States was a "war of words" with no signs of letting up on Monday.	There was a war of words between China and the United States, showing no signs of letting up.
	2	What is the nationwide death toll from the novel coronavirus according to the latest report?	According to the latest report, the nationwide death toll from the novel coronavirus was 19.	The nationwide death toll has reached 19.
Business	1	Which company launched a website to screen people for COVID-19?	Google sister company Verily launched a website to screen people who think they might have COVID-19 and point them to testing sites.	Google sister company Verily launched the website.
	2	How much financing did United Airlines raise?	United Airlines Holdings Inc raised an extra $2 billion in financing.	United Airlines raised an extra $2 billion in financing.
Health	1	What action has China taken in response to the suspicion that exotic animals infected humans with the coronavirus?	China has banned the trade of wildlife, suspecting that exotic animals infected humans with the coronavirus.	China has banned the trade of wildlife.
	2	What is a significant issue in the United States related to the coronavirus?	A significant issue in the United States related to the coronavirus is the small number of tests performed, which is a concern.	The significant issue is the small number of tests performed in the United States.

leveraging Precision (see Formula 1) and Recall (see Formula 2), the F1 score was calculated (see Formula 3). The F1 score is the harmonic mean of Precision and Recall, giving a single metric that balances both aspects of the system's retrieval performance (Miao & Zhu, 2022). Using these metrics, Table 4 evaluates the system's performance on the Twitter dataset across three volumes: 3,000, 6,000, and 10,000 tweets. As the dataset size increased, all performance metrics: Precision, Recall, and F1-score declined. Precision dropped from 0.85 to 0.75, indicating reduced accuracy, while Recall decreased from 0.88 to 0.78, showing a reduced ability to capture relevant information. The F1-score fell from 0.86 to 0.76. This trend highlights that the system performs well with smaller tweet volumes but struggles as the dataset grows, presenting challenges for handling larger datasets.

Table 5 assesses the system's performance on news articles by topic (Politics, Health, Business) and volume (300, 600, 1000 articles). Across all topics, Precision, Recall, and F1-score generally decreased with larger volumes. For Politics, Precision fell from 0.78 to 0.74, Recall from 0.82 to 0.78, and F1-score from 0.80 to 0.76. In Health, Precision dropped from 0.82 to 0.78, Recall from 0.75 to 0.71, and F1-score from 0.78 to 0.74. For Business, Precision decreased from 0.75 to 0.71, Recall from 0.78 to 0.74, and F1-score from 0.76 to 0.72. This trend indicates a challenge in maintaining system performance as the volume of news articles increases.

$$Precision = \frac{Number\ of\ relevant\ items\ retrieved}{Total\ number\ of\ items\ retrieved} \tag{1}$$

Table 4. Evaluation using Twitter dataset by volume.

No.	Evaluation metrics	3,000 Tweets	6,000 Tweets	10,000 Tweets
1	Precision	0.85	0.80	0.75
2	Recall	0.88	0.83	0.78
3	F1-Score	0.86	0.81	0.76

Table 5. Evaluation using news articles dataset by topic and volume.

Topic	No.	Evaluation Metric	300 News Articles	600 News Articles	1000 News Articles
Politics	1	Precision	0.78	0.76	0.74
	2	Recall	0.82	0.80	0.78
	3	F1-Score	0.80	0.78	0.76
Health	1	Precision	0.82	0.80	0.78
	2	Recall	0.75	0.73	0.71
	3	F1-Score	0.78	0.76	0.74
Business	1	Precision	0.75	0.73	0.71
	2	Recall	0.78	0.76	0.74
	3	F1-Score	0.76	0.74	0.72

$$Recall = \frac{Number\ of\ relevant\ items\ retrieved}{Total\ number\ of\ relevant\ items} \quad (2)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

During the evaluation, several instances were identified where the proposed system underperformed in terms of Precision and Recall. Table 6 highlights some of these examples, showcasing where the system struggled and produced lower quality results. For instance, when asked about the measures taken by African governments to contain the coronavirus, the system's response was vague, mentioning only limited travel and new rules, whereas the accurate answer included specific actions like closing borders and imposing quarantine requirements, leading to low precision but high recall. Similarly, for the question about which countries confirmed their first cases of the virus, the system correctly indicated that five African countries were involved but failed to name them, again showing low precision with high recall. Finally, regarding Detroit's plan to prevent virus spread, the system mentioned water restoration but omitted the reason behind it, resulting in low Precision and Recall.

Discussion

The Political-RAG system, designed to gather updates on political events from tweets and news articles by integrating RAG with LLMs like the Llama2 model, offers several key advantages. First, it accelerates development by avoiding the trial-and-error process often associated with traditional NLP techniques (see Table 1) for building Political IE tasks. By using pre-trained, publicly available LLMs such as Llama2, developers can quickly create customized Political IE solutions with minimal costs and resources, making it especially beneficial for organizations with limited R&D budgets aiming to initiate digital transformations. Additionally, the system supports easy and dynamic updates to the domain dataset, which includes tweets and news articles. By simply replacing outdated data files with new ones, the system can adapt to the latest information, facilitating the exploration of new political events.

The system's functionality is demonstrated by its ability to extract political events, providing a foundation for further analysis by analysts. Beyond this, the proposed system sets the foundation for future integration with various political datasets (as mentioned in Table 1) to extract information on hate speech, conflicts, political bias, profiling, social analytics, sentiment analysis,

Table 6. Examples of questions where the system underperformed.

No.	Question	System Response	Ground Truth Answer	Precision	Recall
1	What measures have several African governments taken to contain the spread of the coronavirus?	They limited travel and introduced new rules.	They closed borders, canceled flights, and imposed strict entry and quarantine requirements.	Low	High
2	Which countries confirmed their first cases of the new coronavirus on Friday?	Five African countries confirmed their first cases.	Kenya, Ethiopia, Sudan, Guinea, and Mauritania all confirmed their first cases.	Low	High
3	What plan was implemented in Detroit to help prevent the spread of the coronavirus?	Many Detroit residents who had their water shut off can now have it restored.	Thousands of Detroit residents who had their water service shut off due to nonpayment of bills can have it restored to allow them to wash their hands at home.	Low	Low

opinion mining, and trend analysis. This expanded capability offers valuable insights into public sentiment and prevailing opinions on political matters. Additionally, researchers can leverage the system to develop advanced tools for detecting political polarization and implementing risk management strategies, thereby enhancing their understanding of socio-political dynamics and mitigating potential risks. The quality and comprehensiveness of the extracted information depend on the diversity and accuracy of the datasets used. Future research is needed to evaluate the performance of the RAG with LLM integrated system for the aforementioned political IE tasks using different political datasets.

This study has notable limitations despite its benefits. While the Llama2 model was selected for its superior performance as reported in existing research, it is important to recognize that most LLMs are trained on general datasets. Future research could benefit from exploring and evaluating various LLMs specifically for the political domain. Additionally, querying the LLM with large and complex datasets can lead to delays in response times, affecting scalability and real-time performance. In our case, with a small dataset, query response times ranged from 5 to 9 seconds, with some exceeding a minute. The system used Google Colab's free GPUs to run the RAG with LLM framework, but scaling up to larger datasets increases complexity and response time. This limitation might be addressed by enhancing computational resources, such as increasing the number of GPUs, based on user needs and available resources.

Additionally, the timeliness of the information about political events that analysts can explore depends significantly on the system's ability to access and integrate updated news articles in real-time. In our current approach, the datasets used for analysis were initially downloaded and incorporated into the system. However, this method does not support real-time updates, as it relies on periodically updating the dataset manually. Developing a real-time information system presents several challenges. First, the system would need to continuously fetch and integrate data from multiple websites and sources where political events are reported. This requires robust data acquisition mechanisms capable of handling diverse formats and sources, such as news websites, social media platforms, government publications, and more. Ensuring the reliability and consistency of data across these sources poses another challenge, as different websites may report on the same event with varying details or perspectives.

Real-time systems must efficiently handle the volume and speed of incoming data, particularly with the rapidly changing and diverse reporting of political events. This requires scalable infrastructure and algorithms that can process large datasets quickly to maintain responsiveness and ensure that analysts receive the latest information promptly. Integrating real-time capabilities into the RAG with LLM architecture demands careful management of computational resources. Continuous querying and updating of datasets can strain resources, leading to longer response times or potential system failures if not managed effectively. Scaling resources, such as through cloud-based solutions, can address these issues but introduces additional complexity and cost. To overcome these challenges, advancements in data acquisition, real-time processing algorithms, and infrastructure scalability are needed. Future research should focus on developing efficient and reliable methods for real-time data integration and processing within the RAG with LLM framework, enhancing the ability of researchers and analysts to make timely and informed decisions regarding political events.

Conclusion

IE is essential for political scientists to make informed decisions. Historically, IE has evolved from simple keyword matching to sophisticated machine learning techniques. This advancement has been significantly driven by third-generation methods that employ NLP to analyze documents in context, leading to more nuanced and effective IE. LLMs have been essential in this progress, enabling rapid and accurate extraction of political events from news articles. In this context, our research introduces a chatbot designed to assist analysts with political event extraction. The Political-RAG system integrates RAG with LLMs, streamlining the development process and overcoming the trial-and-error phase typical of classical NLP approaches. By leveraging pre-trained LLMs, organizations can swiftly develop customized Political IE solutions, even with limited resources. Additionally, the system supports dynamic updates to news datasets, facilitating continuous exploration of new political events.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Agence Nationale de la Recherche, France and the University of the West of England, Bristol, UK.

Notes on contributors

Muhammad Arslan earned his Ph.D. in Computer Science in 2020 from the Université de Bourgogne, Dijon, France. He is currently working at the University of the West of England, Bristol. Prior to this, he held a position at the Université de Bourgogne starting in December 2021. Arslan's research interests include Digital Construction, Generative AI, Business Informatics, Open Data, and Knowledge Graphs.

Saba Munawar received her BS in Telecommunication Engineering in 2009 from the National University of Computer and Emerging Sciences (NUCES), Islamabad, Pakistan. She specializes in evaluating computer systems, focusing on designing and implementing test cases to ensure optimal performance and reliability.

Christophe Cruz is a Professor of Computer Science and currently serves as Head of the Computer Science Department at IUT Dijon-Auxerre, Université de Bourgogne, Dijon, France. Alongside his teaching duties, he is an active researcher affiliated with the Laboratoire Interdisciplinaire Carnot de Bourgogne (ICB). His research spans Applied Ontology, Data Science, Artificial Intelligence, and Machine Learning.

ORCID

Muhammad Arslan http://orcid.org/0000-0003-3682-7002

Dataset availability

The datasets utilized in this research can be accessed through the following links:

Twitter Data: https://www.kaggle.com/datasets/smid80/coronavirus-covid19-tweets-early-april.

News Articles Data: https://www.kaggle.com/datasets/saadaljebreen/corona-various-news-covid19.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L. McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Al Ghadban, Y., Lu, H. Y., Adavi, U., Sharma, A., Gara, S., Das, N., Hirst, J. E. (2023). Transforming healthcare education: Harnessing large language models for frontline health worker capacity building using retrieval-augmented generation. *medRxiv*, 2023-12.https://doi.org/10.1101/2023.12.15.23300009

Ali, O., Zaland, Z., Bazai, S. U., Ghafoor, M. I., Hussain, L., & Haider, A. (2024). Neural transformers for bias detection: Assessing Pakistani News. *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan (pp. 1–7). IEEE.

Almatrafi, O., Parack, S., & Chavan, B. (2015). Application of location-based sentiment analysis using twitter for identifying trends towards Indian general elections 2014. *IMCOM '15: The 9th international conference on ubiquitous information management and communication*, Bali, Indonesia (pp. 1–5).

Almazrouei, E., Alobeidli, H., Alshamsi, A., Cappelli, A., Cojocaru, R., Debbah, M., & Penedo, G. (2023). *The falcon series of open language models*. arXiv preprint arXiv:2311.16867. https://arxiv.org/pdf/2311.16867

Alsarra, S., Abdeljaber, L., Yang, W., Zawad, N., Khan, L., Brandt, P., D'Orazio, V. (2023). Conflibert-Arabic: A pre-trained Arabic language model for politics, conflicts and violence. *Proceedings of the 14th International conference on recent advances in natural language processing*, Varna, Bulgaria (pp. 98–108).

Armenta-Segura, J., Núnez-Prado, C. J., Sidorov, G. O., Gelbukh, A., & Román-Godínez, R. F. (2023). Ometeotl@ multimodal hate speech event detection 2023: Hate speech and text-image correlation detection in real life memes using pre-trained Bert models over text. *Proceedings of the 6th workshop on challenges and applications of automated extraction of socio-political events from text*, Varna, Bulgaria (pp. 53–59). https://aclanthology.org/2023.case-1.7

Avanthika, K., Kl, M., & Thenmozhi, D. (2023). Ssn-nlp-ace@ multimodal hate speech event detection 2023: Detection of hate speech and targets using logistic regression and svm. *Proceedings of the 6th workshop on challenges and applications of automated extraction of socio-political events from text*, Varna, Bulgaria (pp. 66–70).

Aziz, A., Hossain, M. A., & Chy, A. N. (2023). Csecu-dsg@ multimodal hate speech event detection 2023: Transformer-based multimodal hierarchical fusion model for multimodal hate speech detection. *Proceedings of the 6th workshop on challenges and applications of automated extraction of socio-political events from text*, Varna, Bulgaria (pp. 101–107). https://doi.org/10.26615/978-954-452-089-2_014

Barberá, P., & Steinert-Threlkeld, Z. C. (2020). How to use social media data for political science research. *The Sage Handbook of Research Methods in Political Science and*

International Relations, 404–423. https://doi.org/10.4135/9781526486387.n26

Bestvater, S. E., & Monroe, B. L. (2023). Sentiment is not stance: Target-aware opinion classification for political text analysis. *Political Analysis*, *31*(2), 235–256. https://doi.org/10.1017/pan.2022.10

Bor, D., Lee, B. S., & Oughton, E. J. (2023). Quantifying polarization across political groups on key policy issues using sentiment analysis. *arXiv preprint arXiv* 2302, 07775. https://doi.org/10.48550/arXiv.2302.07775

Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2019). Analyzing political sentiment using twitter data. Information and Communication Technology for Intelligent Systems: *Proceedings of ICTIS 2018*, California, USA (Vol. 2., pp. 427–436). Springer Singapore.

Burley, T., Humble, L., Sleeper, C., Sticha, A., Chesler, A., Regan, P., Brenner, P. (2020). Nlp workflows for computational social science: Understanding triggers of state-led mass killings. *PEARC '20: Practice and Experience in Advanced Research Computing 2020: Catch the Wave*, 152–159. https://doi.org/10.1145/3311790.3397343

Büyüköz, B., Hürriyetoğlu, A., & Özgür, A. (2020). Analyzing ELMo and DistilBERT on socio-political news classification. *Proceedings of the workshop on automated extraction of socio-political events from news 2020*, Marseille, France (pp. 9–18).

Chen, W., Hu, H., Saharia, C., & Cohen, W. W. (2022). *Reimagen: Retrieval-augmented text-to-image generator. arXiv preprint arXiv* 2209, 14491.

Colverd, G., Darm, P., Silverberg, L., & Kasmanoff, N. (2023). *FloodBrain: Flood disaster reporting by Web-based retrieval augmented generation with an LLM. arXiv preprint arXiv* 2311, 02597.

Costa, C., Aparicio, M., & Aparicio, J. (2021). Sentiment analysis of Portuguese political parties communication. *SIGDOC '21: The 39th ACM international conference on design of communication*, Virtual Event USA (pp. 63–69). https://doi.org/10.1145/3472714.34736

Davenport, C., & Ball, P. (2002). Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977-1995. *The Journal of Conflict Resolution*, *46*(3), 427–450. https://doi.org/10.1177/0022002702046003005

Delucia, A., Dredze, M., & Buczak, A. L. (2023). A multi-instance learning approach to civil unrest event detection on twitter. *Proceedings of the 6th workshop on challenges and applications of automated extraction of socio-political events from text*, Varna, Bulgaria (pp. 18–33).

Demidov, D. (2023). Political bias of news content: Classification based on individual articles and media. https://www.researchgate.net/publication/377074277_Political_Bias_of_News_Content_Classification_based_on_Individual_Articles_and_Media

Demiros, I., Papageorgiou, H., Antonopoulos, V., Pipis, A., & Skoulariki, A. (2008). Media monitoring by means of speech and language indexing for political analysis. *Journal of Information Technology & Politics*, *5*(1), 133–146. https://doi.org/10.1080/19331680802149632

Doumit, S., & Minai, A. (2011a). Online news media bias analysis using an LDA-NLP approach. International Conference on Complex Systems. https://eecs.ceas.uc.edu/~minaiaa/papers/doumit_iccs2011.pdf

Doumit, S., & Minai, A. (2011b). Semantic knowledge inference from online news media using an LDA-NLP approach. The 2011 International Joint Conference on Neural Networks, San Jose, CA, USA (pp. 3068–3071). IEEE (Institute of Electrical and Electronics Engineers). https://doi.org/10.1109/IJCNN.2011.6033626

Efat, A. A., Atiq, A., Abeed, A. S., Momin, A., & Alam, M. G. R. (2023). *EMPOLITICON: NLP and ML based approach for context and emotion classification of political speeches from transcripts*. IEEE Access.

Esackimuthu, S., & Balasundaram, P. (2023). Verbavisor@ multimodal hate speech event detection 2023: Hate speech detection using transformer model. *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, Varna, Bulgaria (pp. 79–83).

Ge, J., Sun, S., Owens, J., Galvez, V., Gologorskaya, O., Lai, J. C., & Lai, K. (2023). Development of a liver disease-specific large language Model chat interface using retrieval augmented generation. medRxiv. https://doi.org/10.1101/2023.11.10.23298364

Gode, S., Bare, S., Raj, B., & Yoo, H. (2023). Understanding political polarization using language models: A dataset and method. *AI Magazine*, *44*(3), 248–254. https://doi.org/10.1002/aaai.12104

Halterman, A. (2018). *Linking events and locations in political text. MIT political science department research paper No. 2018-21*, https://doi.org/10.2139/ssrn.3267476

Halterman, A. (2020). *Extracting political events from text using syntax and semantics. Technical report MIT.*

Halterman, A. (2021). *Three essays on natural language processing and information extraction with applications to political violence and international security* [Doctoral dissertation]. Massachusetts Institute of Technology.

Han, Z. F., Lin, J., Gurung, A., Thomas, D. R., Chen, E., Borchers, C., Gupta, S., & Koedinger, K. R. (2024). Improving assessment of tutoring practices using retrieval-augmented generation. *arXiv preprint arXiv*, 240214594. https://doi.org/10.48550/arXiv.2402.14594

Haq, E. U., Braud, T., Kwon, Y. D., & Hui, P. (2020). A survey on computational politics. *Institute of Electrical and Electronics Engineers Access*, *8*, 197379–197406. https://doi.org/10.1109/ACCESS.2020.3034983

Hettiarachchi, H., Adedoyin-Olowe, M., Bhogal, J., & Gaber, M. M. (2021). DAAI at CASE 2021 task 1: Transformer-based multilingual socio-political and crisis event detection. Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021) (pp. 120–130). https://doi.org/10.18653/v1/2021.case-1.16

Hey, A. M. (2023). Using NLP analysis to categorise statements to values on the political compass. https://www.alexhey.com/resources/MSc-Paper-Alexander-Hey.pdf

Hitesh, M. S. R., Vaibhav, V., Kalki, Y. A., Kamtam, S. H., & Kumari, S. (2019). Real-time sentiment analysis of 2019 election tweets using word2vec and random forest model. *2019 2nd international conference on intelligent communication and computational techniques (ICCT)*, Jaipur, India (pp. 146–151). IEEE. https://doi.org/10.1109/ICCT46177.2019.8969049

Hobbs, J. R., & Riloff, E. (2010). Information extraction. *Handbook of Natural Language Processing*, *15*, 16. https://doi.org/10.1002/9781444324044.ch18

Hodorog, A., Petri, I., & Rezgui, Y. (2022). Machine learning and natural language processing of social media data for event detection in smart cities. *Sustainable Cities and Society*, *85*, 104026. https://doi.org/10.1016/j.scs.2022.104026

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Sifre, L. (2022). Training compute-optimal large language models. *arXiv preprint arXiv* 2203.1555. https://doi.org/10.48550/arXiv.2203.15556

Hu, Y., Hosseini, M., Skorupa Parolin, E., Osorio, J., Khan, L., Brandt, P., & D'Orazio, V. (2022). *Conflibert: A pre-trained language model for political conflict and violence*. Association for Computational Linguistics.

Hürriyetoğlu, A., Mutlu, O., Yörük, E., Liza, F. F., Kumar, R., & Ratan, S. (2021). Multilingual protest news detection-shared task 1, case 2021. *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)* (pp. 79–91). https://doi.org/10.18653/v1/2021.case-1.11

Javed, I., & Saeed, H. (2023). Opinion analysis of bi-lingual event data from social networks. *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Istanbul, Turkiye (pp. 1–6). IEEE. https://doi.org/10.1109/HORA58378.2023.10156766

Jeong, M., Sohn, J., Sung, M., & Kang, J. (2024). Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. *Bioinformatics*, *40*(Supplement_1), i119–i129. https://doi.org/10.1093/bioinformatics/btae238arXiv preprint arXiv:2401.15269.

Jimeno Yepes, A., You, Y., Milczek, J., Laverde, S., & Li, L. (2024). *Financial report chunking for effective retrieval augmented generation*. arXiv e-prints, arXiv-2402.

Jin, Z., & Mihalcea, R. (2022). Natural language processing for policymaking. In Eleonora Bertoni, Matteo Fontana, Lorenzo Gabrielli, Serena Signorelli, & Michele Vespe (Eds.), *Handbook of computational social science for policy* (pp. 141–162). Springer International Publishing.

Jo, E., Epstein, D. A., Jung, H., & Kim, Y. H. (2023). Understanding the benefits and challenges of deploying conversational AI leveraging large language models for public health intervention. *CHI '23: CHI Conference on Human Factors in Computing Systems*, Hamburg, Germany (pp. 1–16). https://doi.org/10.1145/3544548.358150

Johnson, K., & Goldwasser, D. (2016). Identifying stance by analyzing political discourse on twitter. Proceedings of the First Workshop on NLP and Computational Social Science (pp. 66–75).

Katre, P. D. (2019). Text mining and comparative visual analytics on large collection of speeches to trace socio-political issues. 2019 IEEE 9th International Conference on Advanced Computing (IACC) (pp. 108–114). IEEE.

Kawintiranon, K., & Singh, L. (2022). PoliBERTweet: A pre-trained language model for analyzing political content on twitter. Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France (pp. 7360–7367). https://aclanthology.org/2022.lrec-1.801

Kim, J., Choi, S., Amplayo, R. K., & Hwang, S. W. (2020). Retrieval-augmented controllable review generation. *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (pp. 2284–2295). https://doi.org/10.18653/v1/2020.coling-main.207

Kowsik, V. S., Yashwanth, L., Harish, S., Kishore, A., S, R., Jose, A. C., & V, D. M. (2024). Sentiment analysis of twitter data to detect and predict political leniency using natural language processing. *Journal of Intelligent Information Systems*, *62*, 765–785. https://doi.org/10.1007/s10844-024-00842-3

Kuamri, S., & Babu, C. N. (2017). Real time analysis of social media data to understand people emotions towards national parties. *2017 8th international conference on computing, communication and networking technologies (ICCCNT)*, Delhi, India (pp. 1–6). IEEE. https://doi.org/10.1109/ICCCNT.2017.8204059

Lei, Y., Miah, M. M. M., Qamar, A., Reddy, S. R., Tong, J., Xu, H., & Huang, R. (2024). EMONA: Event-level moral opinions in news articles. arXiv preprint arXiv:2404.01715.

Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., & Al-Shaibani, M. S. (2022). Bloom: A 176b-parameter open-access multilingual language model. https://doi.org/10.48550/arXiv.2211.05100

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., , and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, *33*, 9459–9474. https://doi.org/10.48550/arXiv.2005.11401

Liang, Y., Jabr, K., Grant, C., Irvine, J., & Halterman, A. (2018, July). New techniques for coding political events across languages. *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, Salt Lake City, UT, USA (pp. 88–93). IEEE.

Linegar, M., Kocielnik, R., & Alvarez, R. M. (2023). Large language models and political science. *Frontiers in Political Science*, *5*, 1257092. https://doi.org/10.3389/fpos.2023.1257092

Loerakker, M., Müter, L., & Schraagen, M. (2024). Fine-tuning language models on Dutch protest event tweets. *Proceedings of the 7th Workshop on Challenges and*

Applications of Automated Extraction of Socio-political Events from Text (CASE 2024), St. Julian's, Malta (pp. 6–23).

Lorenzini, J., Kriesi, H., Makarov, P., & Wüest, B. (2022). Protest event analysis: Developing a semiautomated NLP approach. *The American Behavioral Scientist*, *66*(5), 555–577. https://doi.org/10.1177/00027642211021650

Lukito, J., Sarma, P. K., Foley, J., & Abhishek, A. (2019). Using time series and natural language processing to identify viral moments in the 2016 US presidential debate. *Proceedings of the third workshop on natural language processing and computational social science*, Minneapolis, Minnesota (pp. 54–64). https://doi.org/10.18653/v1/W19-2107

Makarov, P. (2018, August). Automated acquisition of patterns for coding political event data: Two case studies. *Proceedings of the second joint SIGHUM workshop on computational linguistics for cultural heritage, social sciences, humanities and literature*, Santa Fe, New Mexico (pp. 103–112 https://aclanthology.org/W18-4512).

Mallavarapu, C., Mandava, R., & Kc, S. (2018). Political profiling using feature engineering and nlp. *SMU Data Science Review*, *1*(4), 10.

Manathunga, S. S., & Illangasekara, Y. A. (2023). A comparison of transmissibility of SARS-CoV-2 variants of concern. arXiv preprint arXiv:2308.00479. *20*(1). https://doi.org/10.1186/s12985-023-02018-x

Maynard, D., & Funk, A. (2012). Automatic detection of political opinions in tweets. In R. García-Castro, G. Antoniou, & D. Fensel (Eds.), *The semantic web: ESWC 2011 workshops: ESWC 2011 workshops*, Heraklion, Greece, May 29-30, 2011, Heraklion, Greece, May 29-30, 2011 (pp. 88–99). Springer.

Miao, J., & Zhu, W. (2022). Precision–recall curve (PRC) classification trees. *Evolutionary Intelligence*, *15*(3), 1545–1569. https://doi.org/10.1007/s12065-021-00565-2

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13) - Lake Tahoe, Curran Associates Inc.), Nevada, (Red Hook, NY).

Misra, R. (2022). News category dataset. arXiv preprint arXiv:2209.11429. https://doi.org/10.48550/arXiv.2209.11429

Orellana, S., & Bisgin, H. (2023). Using natural language processing to analyze political party manifestos from New Zealand. *Information*, *14*(3), 152. https://doi.org/10.3390/info14030152

Osorio, J., & Beltran, A. (2020). Enhancing the detection of criminal organizations in Mexico using ML and NLP. *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK (pp. 1–7). IEEE. https://doi.org/10.1109/IJCNN48605.2020.9207039

Oyewola, D. O., Oladimeji, L. A., Julius, S. O., Kachalla, L. B., & Dada, E. G. (2023). Optimizing sentiment analysis of Nigerian 2023 presidential election using two-stage residual long short term memory. *Heliyon*, *9*(4), e14836. https://doi.org/10.1016/j.heliyon.2023.e14836

Pair, E., Vicas, N., Weber, A. M., Meausoone, V., Zou, J., Njuguna, A., & Darmstadt, G. L. (2021). Quantification of gender bias and sentiment toward political leaders over 20 years of Kenyan news using natural language processing. *Frontiers in Psychology*, *12*, 712646. https://doi.org/10.3389/fpsyg.2021.712646

Pan, F., Canim, M., Glass, M., Gliozzo, A., & Hendler, J. (2022). End-to-end table question answering via retrieval-augmented generation. arXiv preprint arXiv:2203. 16714. https://doi.org/10.48550/arXiv.2203.16714

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation, eds A. Moschitti & B. Pang," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Doha, Qatar, Association for Computational Linguistics, Stroudsburg. 10.3115/v1/D14-1162.

Piskorski, J., & Yangarber, R. (2013). Information extraction: Past, present and future. *Multi-Source, Multilingual Information Extraction and Summarization*, 23–49. https://doi.org/10.1007/978-3-642-28569-1_2

Rackauckas, Z. (2024). RAG-Fusion: A New take on retrieval-augmented generation. arXiv preprint arXiv:2402.03367. *13*(1), 37–47. https://doi.org/10.5121/ijnlc.2024.13103

Rahmati, A., Tavan, E., & Keyvanrad, M. A. (2023). Predicting content-based political inclinations of Iranian twitter users using BERT and deep learning. *AUT Journal of Mathematics and Computing*, *4*(2), 145–154.

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J. . . . Azam, S. (2024). *A review on large language models: Architectures, applications, taxonomies, open issues and challenges*. IEEE Access.

Sawhney, R., Wadhwa, A., Agarwal, S., & Shah, R. (2020). GPolS: A contextual graph-based language model for analyzing parliamentary debates and political cohesion. *Proceedings of the 28th international conference on computational linguistics*, Barcelona, Spain (Online) (pp. 4847–4859). https://doi.org/10.18653/v1/2020.coling-main.426

Sha, Y., Feng, Y., He, M., Liu, S., & Ji, Y. (2023). Retrieval-augmented knowledge graph reasoning for commonsense question answering. *Mathematics*, *11*(15), 3269. https://doi.org/10.3390/math11153269

Sharber, B. (2020). *Analyzing political polarization in news media with natural language processing* [(Doctoral dissertation]. University Honors College Middle Tennessee State University).

Shi, X., Xue, S., Wang, K., Zhou, F., Zhang, J., Zhou, J. , and Mei, H. (2024). Language models can improve event prediction by few-shot abductive reasoning. *Advances in Neural Information Processing Systems*, *36*. https://doi.org/10.48550/arXiv.2305.16646

Singh, J., Pandey, D., & Singh, A. K. (2023a). Event detection from real-time twitter streaming data using community detection algorithm. *Multimedia Tools & Applications*, 1–28. https://doi.org/10.1007/s11042-023-16263-3

Singh, K., Vajrobol, V., & Aggarwal, N. (2023b). Iic_team@ multimodal hate speech event detection 2023: Detection of hate speech and targets using Xlm-Roberta-base. *Proceedings of the 6th workshop on challenges and applications of automated extraction of socio-political events from text*, Varna, Bulgaria (pp. 136–143). https://aclanthology.org/2023.case-1.18

Small, S. G., & Medsker, L. (2014). Review of information extraction technologies and applications. *Neural Computing & Applications*, *25*(3–4), 533–548. https://doi.org/10.1007/s00521-013-1516-6

Sufi, F. (2023). Social media analytics on Russia–Ukraine cyber war with natural language processing: Perspectives and challenges. *Information*, *14*(9), 485. https://doi.org/10.3390/info14090485

Suri, M., Chopra, K., & Arora, A. (2022). NSUT-NLP at CASE 2022 task 1: Multilingual protest event detection using transformer-based models. *Proceedings of the 5th workshop on challenges and applications of automated extraction of socio-political events from text (CASE)*, Abu Dhabi, United Arab Emirates (pp. 161–168). https://doi.org/10.18653/v1/2022.case-1.23

Tanev, H. (2024). Leveraging approximate pattern matching with Bert for event detection. *Proceedings of the 7th workshop on challenges and applications of automated extraction of socio-political events from text (CASE 2024)*, St. Julians, Malta (pp. 32–39). https://aclanthology.org/2024.case-1.4

Tanev, H., Stefanovitch, N., Halterman, A., Uca, O., Zavarella, V., Hürriyetoğlu, A. , and Della Rocca, L. (2023, September). Detecting and geocoding battle events from social media messages on the Russo-Ukrainian war: Shared task 2, case 2023. *Proceedings of the 6th workshop on challenges and applications of automated extraction of socio-political events from text*, Varna, Bulgaria (pp. 160–166). https://aclanthology.org/2023.case-1.21

Thapa, S., Rauniyar, K., Jafri, F., Shiwakoti, S., Veeramani, H., Jain, R., Naseem, U. (2024). Stance and hate event detection in tweets related to climate activism-shared task at case 2024. *Proceedings of the 7th workshop on challenges and applications of automated extraction of socio-political events from text (CASE 2024)*, St. Julians, Malta (pp. 234–247). https://aclanthology.org/2024.case-1.33

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y. , and Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288. https://doi.org/10.48550/arXiv.2307.09288

Varma, S. H., & Harsha, Y. V. S. (2022). Political polarity classification using NLP. Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2021 (pp. 19–33). Springer Nature Singapore., Singapore.

Wang, M. D. (2023). Artificial intelligence techniques for political risk management: An NLP analysis of the 2019 US-China trade war. In M. A. Aceves-Fernández (Ed.), *Machine learning and data mining annual volume 2023*. IntechOpen. https://doi.org/10.5772/intechopen.110794

Wang, W., Wang, Y., Joty, S., & Hoi, S. C. (2023). Rap-gen: Retrieval-augmented patch generation with codet5 for automatic program repair. *Proceedings of the 31st ACM joint European software engineering conference and symposium on the Foundations of Software Engineering*, San Francisco, CA, USA (pp. 146–158). https://doi.org/10.1145/3611643.3616256

Won, D., Steinert-Threlkeld, Z. C., & Joo, J. (2017). Protest activity detection and perceived violence estimation from social media images. *Proceedings of the 25th ACM international conference on Multimedia*, Mountain View, California, USA (pp. 786–794). https://doi.org/10.1145/3123266.3123282

Xia, M., Zhang, X., Couturier, C., Zheng, G., Rajmohan, S., & Ruhle, V. (2023). Hybrid retrieval-augmented generation for real-time composition assistance. arXiv preprint arXiv:2308.04215. https://doi.org/10.48550/arXiv.2308.04215

Xiong, G., Jin, Q., Lu, Z., & Zhang, A. (2024). Benchmarking retrieval-augmented generation for medicine. arXiv preprint arXiv:2402.13178. https://doi.org/10.48550/arXiv.2402.13178

Yamagishi, Y. (2024). Yyama@ multimodal hate speech event detection 2024: Simpler prompts, better results-enhancing zero-shot detection with a large multimodal Model. *Proceedings of the 7th workshop on challenges and applications of automated extraction of socio-political events from text (CASE 2024)*, St. Julians, Malta (pp. 60–66). https://aclanthology.org/2024.case-1.7

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H. , and Hu, X. (2023a). Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data*, *18*(6), 1–32.

Yang, W., Alsarra, S., Abdeljaber, L., Zawad, N., Delaram, Z., Osorio, J., D'Orazio, V. (2023b). Conflibert-Spanish: A pre-trained Spanish language model for political conflict and violence. *2023 7th IEEE Congress on Information Science and Technology (CiSt)*, Agadir - Essaouira, Morocco (pp. 287–292). IEEE. https://doi.org/10.1109/CiSt56084.2023.10409883

Yu, H., Guo, P., & Sano, A. (2023). Zero-shot ECG diagnosis with large language models and retrieval-augmented generation. In Stefan Hegselmann, Antonio Parziale, Divya Shanmugam, Shengpu Tang, Mercy Nyamewaa Asiedu, Serina Chang, Tom Hartvigsen, & Harvineet Singh (Eds.), *Machine learning for health (ML4H)*, New Orleans, Louisiana, USA (pp. 650–663). PMLR.

Zakka, C., Shad, R., Chaurasia, A., Dalal, A. R., Kim, J. L., Moor, M., Hiesinger, W. (2024). Almanac—retrieval-augmented language models for clinical medicine. *Nejm Ai*, *1*(2), AIoa2300068. https://doi.org/10.1056/AIoa2300068

Zhou, M., Duan, N., Liu, S., & Shum, H. Y. (2020). Progress in neural NLP: Modeling, learning, and reasoning. *Engineering*, *6*(3), 275–290. https://doi.org/10.1016/j.eng.2019.12.014