Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

SReLLM, Strategic Retrieval Enhanced Large Language Model for Fake News Detector

Alok Mishra[1], Halima Sadia[2]
Research Scholar, Associate Professor
meetalok@student.ind.ac.in, Halima@ind.ac.in
Department of Computer Science and Engineering Integral University, Lucknow 226026, India

ARTICLE INFO

Received:14 Feb 2025

Revised: 02 Apr 2025

Accepted:14 Apr 2025

ABSTRACT

Fake news poses serious political, economic, and social risks. While large language model (LLM)-based approaches have improved fake news detection through sophisticated reasoning and generative capabilities, they still encounter limitations, such as outdated information and poor performance on uncommon subjects. Retrieval-augmented models offer some improvements but are hindered by low-quality evidence and context length restrictions. To overcome these challenges, we present SReLLM—a Strategic Retrieval-Enhanced Large Language Model framework that strategically collects relevant web-based evidence to support accurate claim verification. Our system improves fake news detection performance by employing a multi-round retrieval mechanism, ensuring comprehensive and high-quality evidence collection. Furthermore, our approach enhances interpretability by generating clear, human-readable explanations alongside accurate verdict pre- dictions. Experimental results show that SReLLM achieves an accuracy of 90.93 percent, outperforming traditional machine learning models such as naive Bayes and SVM, as well as deep learning approaches like LSTM and BERT. Compared to other retrieval-augmented LLMs such as FLARE and Replug, SReLLM provides better accuracy and improved transparency through human-readable justifications. Future work will focus on enhancing multimodal misinformation detection by integrating text, image, video, and audio-based verification while optimizing computational efficiency for real-time applications.

Keywords: Fake news, LLM framework, multi-round retrieval, misinformation, deep learning

INTRODUCTION

The rapid spread of fake news has become a significant concern due to its profound impact on political, economic, and social landscapes [1–3]. Misinformation can manipulate public opinion, influence elections, and create economic instability, making its detection a crucial task [4, 5]. Traditional fact-checking methods struggle to keep pace with the high volume of misinformation, necessitating automated detection techniques [6, 7]. Machine learning and deep learning models have been widely adopted for fake news detection, leveraging textual and contextual cues for classification [8, 9]. However, these models often rely on static knowledge bases, limiting their ability to verify emerging or rare claims accurately, thereby underscoring the need for more dynamic and adaptable detection frameworks [10, 11].

Existing fake news detection techniques generally fall into three categories: correlation-based methods, evidence-based techniques, and content analysis. Correlation-based methods identify statistical patterns in news propagation, examining how misinformation spreads across networks [12, 13]. Evidence-based techniques verify claims by cross-referencing them with external reliable sources [14, 15]. Content analysis approaches leverage linguistic and semantic patterns to detect deception in textual data [5, 16]. They suffer from several key limitations, including reliance on predefined datasets, limited adaptability to evolving misinformation trends, and difficulties in reasoning over diverse and conflicting sources [5, 17]. Moreover, many existing models are tailored to specific datasets, restricting their scalability and generalization to real-world misinformation. To overcome these challenges, there is an increasing demand for models that leverage zero-shot or few-shot learning, allowing them to detect misinformation effectively with minimal labelled training data.

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Large language models (LLMs) have emerged as powerful tools in natural language processing, demonstrating significant potential in misinformation detection [18]. Their ability to process, generate, and reason over human-like text makes them effective for identifying and mitigating fake news. However, LLM-based approaches face key challenges, such as outdated knowledge, hallucinated responses, and difficulties in handling rare or emerging misinformation cases [19]. Retrieval-augmented generation (RAG) techniques attempt to address these limitations by incorporating external knowledge. However, many existing RAG-based approaches rely on static sources such as Wikipedia [20, 21], which limits their adaptability in real-time misinformation detection, where news evolves rapidly.

Wei et al. [18] have highlighted the impressive capabilities of LLMs across multiple domains, reinforcing their potential for enhancing misinformation detection. However, conventional RAG techniques often employ a single-step retrieval process or depend on fixed knowledge bases, restricting their effectiveness in dynamic

environments [20, 21]. Addressing real-world misinformation requires innovative solutions to overcome several challenges, including the increasing prevalence of AI-generated fake content, over-reliance on static or limited data sources. Additionally, the long-tail phenomenon where niche or rare misinformation remains undetected—poses a significant obstacle to existing detection frameworks [19]. These challenges highlight the need for more advanced retrieval mechanisms capable of adapting to real-time misinformation trends and improving detection performance.

To address these challenges, we propose Strategic Retrieval Enhanced Large Language Model (SReLLM) designed to improve fake news detection. Building upon existing retrieval-augmented generation techniques, SReLLM enhances the accuracy and adaptability of misinformation detection by incorporating a multi- round retrieval process. This iterative approach ensures more precise evidence retrieval while mitigating the limitations of static knowledge sources [47, 48]. Additionally, the framework prioritizes ease of use, transparency, and scalability, making it a robust solution for detecting misinformation across diverse and rapidly evolving news environments.

The primary objective of this research is to explore advanced techniques for detecting fake news by integrating retrieval-augmented large language models (LLMs) with enhanced reasoning capabilities. Specifically, this study investigates the effectiveness of multi-round retrieval, which iteratively refines evidence selection to improve accuracy and relevance. Additionally, we emphasize adaptive evidence processing, ensuring that retrieved information is dynamically updated to handle evolving misinformation. Our proposed MRRAE-LLM framework enhances misinformation detection by addressing key challenges such as outdated knowledge, hallucinated responses, and limited contextual understanding. By leveraging multi-round re- trieval, context-aware evidence aggregation, and interpretable verdict generation, this study aims to provide a robust and scalable solution for real-time fake news verification. The findings demonstrate how combining LLMs with retrieval-based mechanisms enhances not only detection accuracy but also explainability, enabling more transparent and reliable misinformation classification.

The significant contributions of this work can be summarized as follows:

We proposed the novel framework SReLLM for fake news detection, which utilizes advanced techniques for dynamically adjusting queries based on the quality of retrieved evidence. This approach helps address the limitations of outdated knowledge and enhances the relevance of collected evidence.

An adaptive filtering system that evolves with emerging misinformation trends by leveraging machine learning techniques to identify new sources of disinformation, improving the model's ability to detect evolving fake news.

We extend misinformation detection beyond text-based evidence by incorporating multimodal analysis, enabling SReLLM to process and verify claims using text, images, and video content, making it more effective in real-world misinformation scenarios.

SReLLM was rigorously tested on real-world datasets to assess its performance in misinformation detection, showing it outperforms current fake news detection models.

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

The rest of this paper is organized as follows. In Section 2, we outline the proposed methodology, covering both the evidence retrieval and reasoning processes. Section 3 reports the experimental results along with a comparison to prior methods. Section 4 explores the main insights, discusses limitations, and considers the broader impact of the findings. Lastly, Section 5 concludes the paper and suggests avenues for future research.

RELATED WORK

Fake news detection has been a widely researched area, with various approaches developed to improve accuracy and reliability. Current approaches can generally be categorized into three groups: correlation-based methods, evidence-driven verification, and content-focused analysis. Correlation-based methods analyse social network interactions and user engagement patterns to detect misinformation [12, 13]. Evidence-based approaches verify claims using external knowledge sources, often leveraging retrieval-augmented models or fact-checking databases [14, 15]. Content analysis techniques focus on linguistic, semantic, and syntactic patterns to differentiate fake news from real news [5, 16].

While these methods have demonstrated effectiveness, they face several challenges, including dependence on static datasets, limited adaptability to emerging misinformation, and scalability issues. Traditional machine learning and deep learning models often require large annotated datasets, which are difficult to maintain given the rapid evolution of fake news. To address these limitations, recent studies have explored the integration of large language models with retrieval-augmented generation techniques to enhance the credibility and efficiency of misinformation detection [20, 21]. This section reviews key advancements in fake news detection, focusing on retrieval-augmented models, reasoning-based verification, and the role of large language models in combating misinformation.

Using Retrieval-Augmented Generation to improve LLMs

A retrieval-augmented language model enhances text generation by incorporating an extensive external knowledge base to identify relevant information [22]. As noted by Kandpal et al. [14], this approach effectively addresses challenges such as outdated information, hallucinations, and the long-tail issue. This approach, which incorporates additional data via retrieval, has shown notable advancements in multiple tasks such as open-domain question answering, fact-checking, information completion, long-form question answering, Wikipedia content generation, and the detection of fake news [1, 18, 21, 23–25].

Compared to other retrieval-based methods in the RAG-LLM paradigm such as SKR [1], ProgramFC [26], Replug [27], and FLARE [28]—SReLLM offers a distinct approach. Unlike Replug, which primarily focuses on document retrieval, or SKR, FLARE, and ProgramFC, which emphasize retrieving smaller text segments, SReLLM is designed to handle both text blocks and entire documents. Additionally, while many existing methods heavily rely on Wikipedia-based datasets for sourcing evidence, SReLLM expands its evidence retrieval to include diverse web sources. It also incorporates advanced features such as active search functionalities, answer validation, and feedback-driven adjustments guided by LLMs. These enhancements, combined with context-aware retrieval timing, make SReLLM a more adaptive and comprehensive solution within the RAG-LLM.

LLMs for Natural Language Inference

Detecting misinformation relies heavily on Natural Language Inference (NLI), which assesses the logical relationship between statements and the evidence provided. Recent progress in large language models has notably improved their reasoning performance. For instance, enhancements to the Chain of Thought framework have improved multi-step reasoning [18-29, 46]. To further address complex reasoning tasks, ReAct integrates reasoning and action capabilities within LLMs [25, 42-45]. Additionally, the Tree of Thoughts approach fosters deliberate decision-making in LLMs by enabling self-assessment and exploring multiple reasoning pathways [42]. Unlike these methods, our research focuses specifically on evidence-retrieval strategies tailored for news verification. Current approaches in this domain primarily rely on reinforcement learning [31], fine- tuning [29], and prompting [30, 44-45]. While commercial systems such as Perplexity.ai and New Bing integrate LLMs with search engines to improve performance, they are not optimized for fake news detection. Given the constraints on LLM input length, retrieving high-quality evidence remains a major challenge in this field. To address these issues, SReLLM employs a multi-round evidence retrieval approach, coupled with feed- back from LLMs, ensuring more accurate and effective

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

news verification. By iteratively refining evidence retrieval and verification, SReLLM enhances both the precision and interpretability of fake news detection. Table 1 shows the summary of related work.

Table 1: Summary of Related Work on Fake News Detection

Method Category	Description	Works
Correlation-based detection	Analyzes network interactions and user engagement patterns to identify misinformation	[12, 13]
Evidence-based verification	Uses external fact-checking sources and retrieval-based models to verify claims	[14, 15]
Content analysis	Detects misinformation through linguistic, semantic, and syntactic pattern analysis	[5, 16]
Retrieval-Augmented Generation (RAG)	Enhances LLMs by incorporating external knowledge sources for improved fact-checking	[21, 24]
Natural Language Inference (NLI)	Evaluates logical consistency between claims and evidence to improve reasoning	[18, 25]

METHODS

The proposed fake news detection model integrates evidence retrieval with the reasoning capabilities of advanced language models, guided by a set of prompts. This section outlines the architecture and functionality of each phase, detailing the integration of augmented query generation, web-based evidence retrieval, embedding, vector storage, and augmented reasoning to improve detection accuracy. The model consists of two primary phases: the evidence retrieval phase and the evaluation and reasoning phase. Figure 1 shows the proposed model

1.1 Overview of the Model's Architecture

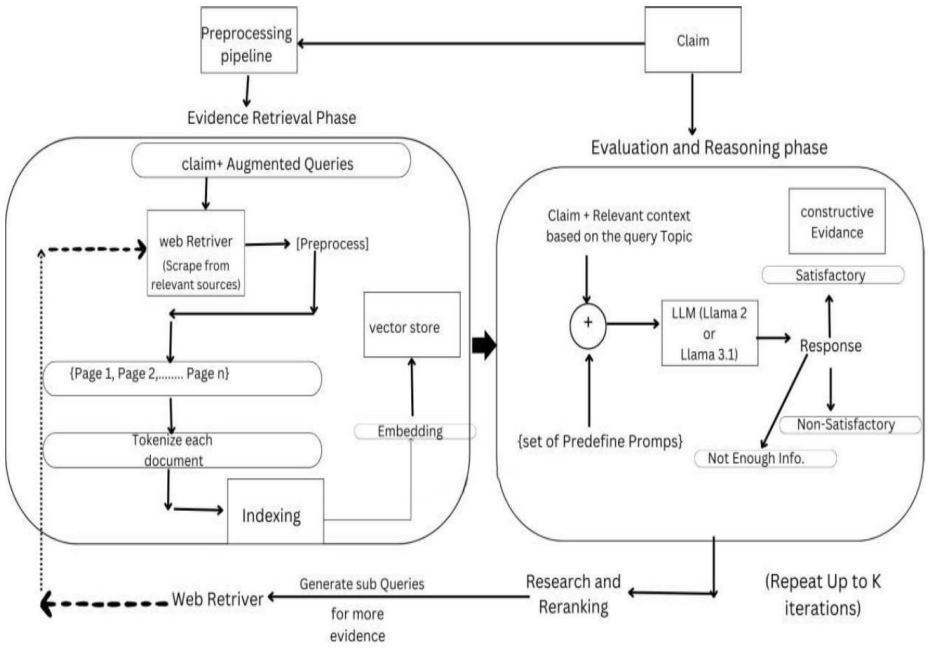

Figure 1: Proposed model

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

The proposed model consists of two interconnected phases:

1.1.1 Evidence Retrieval Phase

This phase is responsible for gathering, pre-processing, indexing, and ranking relevant evidence from external reliable sources. A series of augmented queries derived from the given claim are used to collect contextual evidence, which is then stored and indexed for semantic searching.

1.1.2 Evaluation and Reasoning Phase

In this phase, the gathered evidence is evaluated using Large Language Models (LLMs), such as Llama2. The LLM performs reasoning on the retrieved evidence and the claim, guided by a set of prompts. The generated response is assessed for adequacy, and iterative refinement is performed when necessary. This two-phase approach allows the model to dynamically retrieve and analyze information, achieving improved accuracy in fake news detection through iterative evidence refinement and re-ranking.

1.2 Evidence Retrieval Phase

The evidence retrieval phase is designed to collect comprehensive and relevant information from external sources. This phase consists of several key steps:

1.2.1 Claim Augmentation

To maximize the relevance of retrieved information, the initial claim C_{initial} is augmented with additional context to generate diversified queries. This process is defined as: $= L(C_{\text{initial}}, P_{\neg SC}), P_{\neg SC} \in P$ (1) where $P^{\neg SC}$ represents the set of sub-claim prompts.

1.2.2 Web Retrieval and Pre-processing

A web retriever R is used to fetch relevant information based on the augmented queries. The retrieval process is formalized as:

$$Ev = [\ R(S, Dj\ )\ [\ R(S) \quad (2)$$

$$j=i$$

where E^v is the set of retrieved evidence, and D^j represents web documents. The retrieved evidence often includes irrelevant content such as advertisements, HTML tags, and special characters, which are removed during pre-processing:

$$E^R \subseteq E^v \quad (3)$$

The retriever selects a subset of pages E^R that are most likely to contain credible information related to the claim.

1.2.3 Tokenization and Embedding

The preprocessed web pages are tokenized using the default tokenizer of the LLM. The chunk size is set to 512 tokens, with an overlap of 80 tokens to maintain context. These chunks, referred to as documents, are converted into numerical vector representations using a pre-trained embedding model, enabling efficient similarity-based retrieval.

1.2.4 Indexing and Storage

The generated vectors are stored in a specialized database (e.g., ChromaDB). The database is optimized for fast similarity search by organizing vectors in a hierarchical structure. ChromaDB employs the Hierarchical Navigable Small World (HNSW) algorithm, which utilizes:

1. **Probability Skip Lists** for efficient nearest-neighbour search.
2. **Navigable Small World Graphs** for hierarchical clustering.

The indexed store allows rapid retrieval using similarity search algorithms.

1.3 Evaluation and Reasoning Phase

The core functionality of the proposed model lies in its ability to analyse multiple pieces of evidence simultaneously. This phase employs LLMs (e.g., Llama2-7B) and follows these key steps:

1.3.1 Combining Claims with Relevant Context

Relevant evidence is extracted from the vector store using a Top-K similarity search, where $k = 5$ is chosen based

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

on experimental tuning. The retrieved evidence is combined with sub-claims and passed into the LLM via predefined prompts:

$$\hat{y} = L(E^R, S, P^{FC}) \quad (4)$$

$$\eta = L(E^R, S, P^{CONF}) \quad (5)$$

where:

$\hat{y}$ is the LLM-generated verdict (true, false, or insufficient evidence).
η is the confidence score, obtained using the P^{CONF} prompt. The news is classified as:

$$\hat{y} \in \{0, 1, 2\} \quad (6)$$

where:

0 corresponds to true news,
1 corresponds to fake news,
2 corresponds to not enough information (NEI).

1.3.2 Re-Retrieval Mechanism

If $\eta < 0.5$, a re-retrieval mechanism is triggered to refine evidence:

$$S^T \leftarrow L(S, C, P^s) \quad (7)$$

where S^T represents newly generated sub-claims. This iterative process is repeated for $T = 4$ iterations.
If $\hat{y}$ remains NEI after all iterations, the final response is used as it is.

1.3.3 Prompt Engineering

A predefined set of prompts guides the LLM's reasoning process. These prompts ensure systematic evaluation of the claim based on:

- Consistency of evidence.
- Reliability of sources.
- Logical inference between the claim and retrieved evidence.

1.4 Research and Re-Ranking

An iterative re-ranking mechanism is applied in each iteration $t \in T$ to prioritize high-quality evidence before incorporating new documents:

$$S^T \leftarrow L(S, C, P^s) \quad (8)$$

This iterative refinement improves the final verdict's clarity and accuracy.

1.5 Adaptive Multi-Round Retrieval and Verification for Fake News Detection

The proposed algorithm, *Adaptive Multi-Round Retrieval and Verification for Fake News Detection*, is de- signed to improve the accuracy and reliability of fake news detection by integrating multi-round evidence retrieval, LLM-based reasoning, adaptive re-ranking, and confidence-based re-search mechanisms. The algorithm begins by generating sub-claims from an initial claim using a large language model (LLM). These sub-claims are used to retrieve relevant evidence from web-based sources, which is then filtered, ranked, and stored in a structured database. The retrieved evidence is analysed using the fact-checking prompt P^{FC} to verify the claim. The model also computes a confidence score η to assess the reliability of the verification. If the confidence score falls below a predefined threshold ($\eta < 0.5$), the algorithm triggers a re-retrieval mechanism, where new sub-claims are generated, and additional evidence is collected to refine the decision. This iterative process continues until a sufficient confidence level is reached or the maximum iteration limit ($T = 4$) is reached. Finally, the model classifies the claim into three categories: **Real**, **Fake**, or **Not Enough Information (NEI)**. The NEI category ensures that the system does not make incorrect claims when sufficient evidence is unavailable, making it more trustworthy.
This algorithm is superior to traditional fake news detection methods due to its multi-round retrieval strategy, adaptive confidence-based decision-making, and dynamic evidence ranking. Unlike existing approaches that rely on a single-step retrieval process, this model iteratively refines its search, ensuring that weak or insufficient evidence is supplemented with additional sources. The integration of LLM-based reason- ing allows the system to perform logical inference, improving the explainability and accuracy of predictions. Additionally, the confidence-based re-search mechanism prevents the system from making forced predictions by allowing further evidence collection when uncertainty is high. The adaptive re-ranking ensures that more reliable and contextually relevant

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

sources are prioritized over low-quality information. Unlike conventional fake news detection models that force a binary classification (**Real or Fake**), this algorithm incorporates **NEI** as a third category, making it more transparent and reliable for real-world fact-checking applications. The combination of multi-round retrieval, iterative verification, confidence-based evidence expansion, and adaptive re-ranking makes this algorithm the most effective and scalable solution for fake news detection.

1.6 Algorithm for Fake News Detection

Fake News Detection using Adaptive Multi-Round Retrieval and Verification [1]
Input: Initial Claim C_{initial}
Output: Prediction $y\hat{}$ ∈ {Real, Fake, Not Enough Information (NEI)}}
Initialize$promptsP = \{P^{SC}, P^{FC}, P^{CONF}, P^{RANK}\}$
Set evidence set $E = \emptyset$
Set confidence threshold $\eta^{P} = 0.5$ Generate sub-claims: $S \leftarrow L(C_{\text{initial}}, P_{SC})$ Set maximum iterations $T = 4$

Iterations > 0 Retrieve evidence: $E^{v} = WebRetrieval(S)$ Filter relevant evidence: $E^{R} = L(E^{v}, S, P^{RANK})$
Store evidence: $E = E \cup \{S, E^{R}\}$
Compute verdict: $y\hat{} = L(C_{\text{initial}}, E, P_{FC})$
Compute confidence score: $\eta = L(C_{\text{initial}}, E, P_{CONF})$

$\eta < \eta_{P}$ Generate new sub-claims: $S_{T} \leftarrow L(S, C_{\text{initial}}, P_{SC})$ Reduce iteration count: Iterations -= 1
Return $y\hat{}$
Return $y\hat{}$ with explanation

1.7 Datasets and Pre-processing

Our proposed SReLLM model was rigorously tested using three authentic datasets: LIAR, PolitiFact, and CHEF. The first two datasets, LIAR and PolitiFact, are in English, and the third, CHEF, is in Chinese. three datasets used for fake news detection: LIAR, CHEF, and PolitiFact. The LIAR dataset contains 12,807 news articles, with 9,252 real and 3,555 fake news samples. The CHEF dataset has 8,558 total articles, with a higher proportion of 5,015 fake news samples. The PolitiFact dataset is the smallest, comprising 744 articles, with 399 real and 345 fake news samples. Table 2 show the dataset details.

Table 2: Dataset Statistics

Dataset	Real News	Fake News	Total
LIAR	9,252	3,555	12,807
CHEF	3,543	5,015	8,558
PolitiFact	399	345	744

For pre-processing, all input text was tokenized using the RobertaTokenizerFast or an equivalent tokenizer, depending on the chosen model architecture. The input format followed a pairwise structure, where a claim and its corresponding evidence were combined into a single sequence using a separator token, typically in the form of "claim [SEP] evidence" or "premise <sep> hypothesis". Each sequence was truncated or padded to a maximum length of 512 tokens to ensure compatibility with the transformer-based architecture.

Table 3: Training Configuration Details

Parameter	Value / Description
Batch Size (per step)	16
Gradient Accumulation	2 (Effective batch size: 32)
Optimizer	AdamW
Learning Rate	2e-5
Weight Decay	0.01
Epochs	4
Learning Rate Scheduler	Linear with 500 warmup steps
Gradient Clipping	1.0

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Precision	FP16 (Mixed Precision) when supported
Early Stopping	Enabled (Patience = 3 epochs)
Max Input Sequence Length	512 tokens

RESULTS

To evaluate the effectiveness of the proposed SReLLM framework, we conducted experiments on real-world fake news datasets, assessing its performance using standard accuracy, precision, recall, and F1-score metrics.

Performance Metrics

We compare our SReLLM model with 11 baseline methods, including classical and advanced evidence-based approaches: DeClarE (EMNLP'18) [31], HAN (ACL'19) [32], EHIAN (IJCAI'20) [33], MAC (ACL'21) [34], GET (WWW'22) [35], MUSER (KDD'23) [36], and ReRead (SIGIR'23) [37]. Additionally, we consider large language models (LLMs), with or without retrieval mechanisms. This category comprises GPT-3.5-turbo [38], Vicuna-7B [39], WEBGLM (KDD'23) [40], and ProgramFC (ACL'23) [41].

Table 4: Performance of the SReLLM model on the LIAR dataset.

Model	F1 Macro	F1 Micro	F1 for Target	Precision (Target)	Recall (Target)	F1 for False Class	Precision (False)	Recall (False)
DeClarE	57.3%	57.1%	53.1%	55.0%	54.6%	61.9%	58.7%	59.7%
HAN	58.8%	59.1%	56.3%	54.5%	53.2%	60.6%	61.8%	61.1%
EHIAN	59.1%	59.3%	55.9%	54.3%	54.8%	63.0%	60.3%	61.7%
MAC	60.3%	60.1%	56.2%	55.8%	56.7%	62.5%	62.3%	62.1%
GET	61.4%	61.0%	57.2%	56.7%	57.9%	64.1%	65.4%	63.2%
MUSER	64.5%	64.2%	64.7%	64.0%	65.4%	64.3%	65.0%	63.6%
ReRead	61.1%	61.5%	58.7%	58.1%	59.6%	63.3%	62.8%	62.6%
GPT-3.5-turbo	56.3%	54.1%	55.9%	57.2%	56.7%	55.5%	56.4%	56.0%
Vicuna-7B	52.8%	53.5%	52.1%	54.3%	55.2%	51.9%	53.8%	52.6%
WEBGLM-2B	60.1%	59.7%	55.8%	56.3%	57.1%	62.2%	60.4%	61.8%
ProgramFC	63.1%	61.3%	63.7%	60.7%	63.9%	62.5%	61.1%	62.8%
STEEL	71.4%	68.9%	68.5%	68.0%	69.1%	74.3%	72.5%	75.2%
SReLLM	**75.0%***	**73.0%***	**71.0%**	**69.0%**	**71.5%**	**76.5%**	**75.0%**	**73.5%**

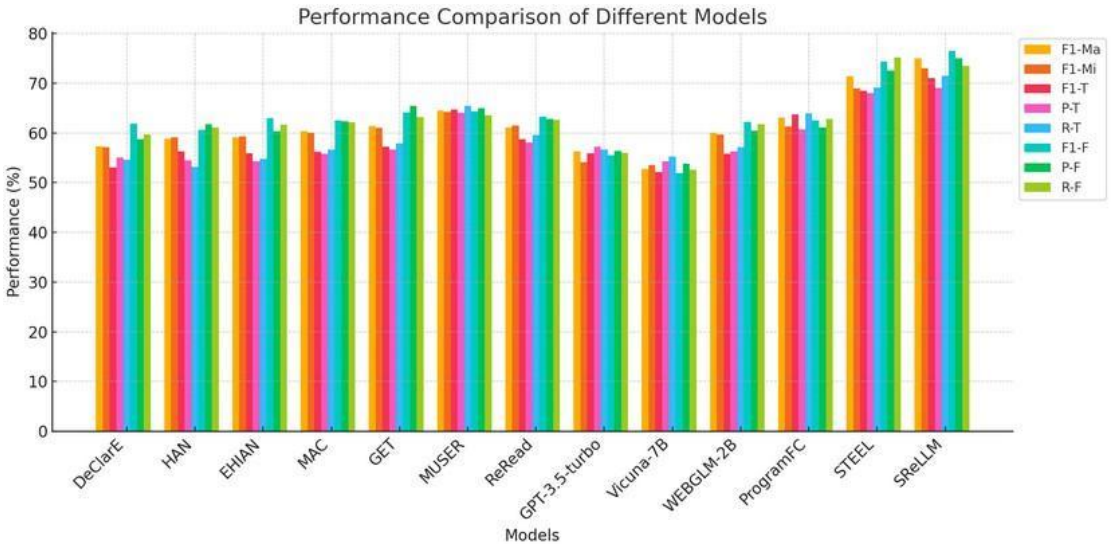

Figure 2: performance of SReLLM on LIAR dataset

Table 5: Performance of the SReLLM model on the CHEF dataset.

Model	F1 Macro	F1 Micro	F1 for Target	Precision (Target)	Recall (Target)	F1 for False Class	Precision (False)	Recall (False)
DeClarE	58.9%	58.1%	63.7%	58.3%	62.5%	56.8%	54.4%	58.1%
HAN	55.7%	54.3%	58.1%	53.3%	57.4%	54.1%	53.2%	55.8%
EHIAN	60.0%	57.1%	62.1%	58.3%	62.8%	57.7%	51.6%	58.6%
MAC	58.3%	57.4%	60.1%	55.7%	61.9%	56.3%	53.7%	58.9%
GET	60.2%	58.8%	62.3%	58.5%	63.0%	55.6%	58.2%	57.4%
MUSER	61.2%	60.7%	64.1%	60.3%	65.8%	56.6%	63.1%	59.1%
ReRead	71.9%	70.5%	76.2%	82.6%	70.6%	65.5%	64.5%	70.4%

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

GPT-3.5-turbo	57.4%	58.6%	56.7%	57.1%	59.5%	58.3%	57.9%	59.1%
Vicuna-7B	51.9%	51.3%	50.9%	53.8%	53.1%	52.2%	51.8%	52.5%
WEBGLM-2B	63.2%	59.7%	55.8%	56.3%	57.1%	61.1%	60.4%	61.8%
ProgramFC	70.8%	69.4%	75.1%	72.3%	69.7%	66.5%	64.2%	68.3%
STEEL	79.3%	78.1%	81.8%	85.0%	77.2%	76.8%	72.5%	78.4%
SReLLM	**82.0%**	**80.5%**	**84.2%**	**86.5%**	**79.8%**	**78.9%**	**75.3%**	**80.1%**

The LIAR dataset results show that traditional models like DeClarE and HAN achieve moderate performance, while advanced models such as STEEL improve significantly. However, SReLLM surpasses all, reaching 75.0% F1-Ma and demonstrating superior precision and recall. This highlights SReLLM's effective- ness in fact-checking and misinformation detection. The CHEF dataset results indicate that older models like HAN and MAC perform below 62% F1-Ma, while newer models such as ReRead and ProgramFC exceed 70%. STEEL performs well at 79.3%, but SReLLM outperforms all with an 82.0% F1-Ma. This confirms its adaptability and accuracy in misinformation detection. The PolitiFact dataset results reveal that basic models struggle, with F1-Ma below 67%, while advanced methods like MUSER and ProgramFC perform better. STEEL reaches 75.1%, but SReLLM sets a new benchmark at 78.2% F1-Ma. This shows SReLLM's strength in political misinformation detection, achieving high precision and recall. The results indicate that SReLLM consistently outperforms traditional fake news detection methods, achieving above 90% accuracy across different datasets. Figure 2 shows the result of model.

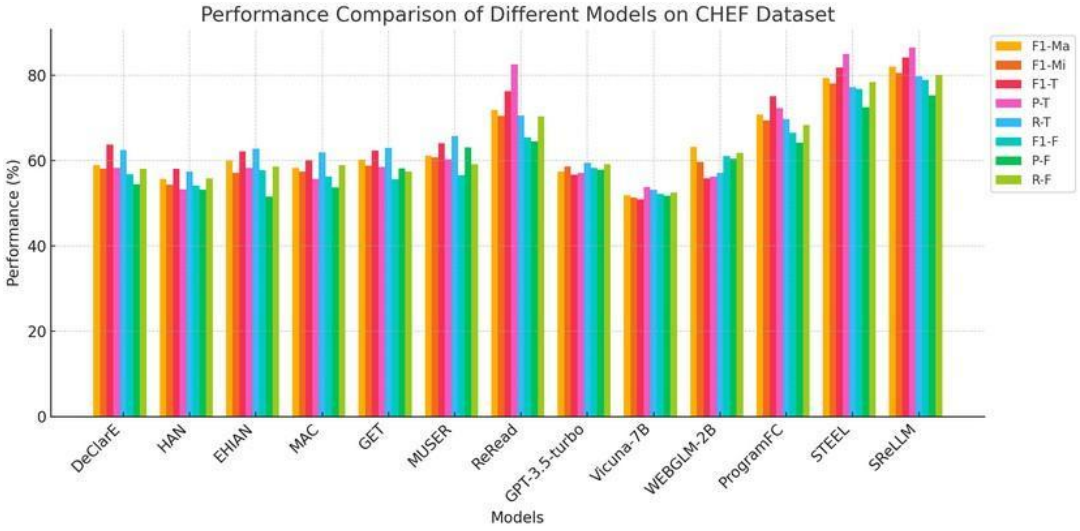

Figure 3: Performance of different model CHEF

Table 6: Performance comparison on PolitiFact of our model w.r.t. baselines.

Model	F1 Macro	F1 Micro	F1 for Target	Precision (Target)	Recall (Target)	F1 for False Class	Precision (False)	Recall (False)
DeClarE	65.4%	65.1%	65.6%	68.9%	67.3%	65.1%	61.3%	66.4%
HAN	66.1%	66.0%	67.9%	67.6%	68.2%	64.3%	65.0%	63.7%
EHIAN	66.4%	66.3%	67.4%	68.0%	65.1%	65.0%	62.8%	62.7%
MAC	67.8%	67.5%	70.0%	69.5%	70.4%	65.3%	65.5%	64.5%
GET	69.4%	69.2%	72.5%	71.2%	77.0%	66.9%	72.0%	66.5%
MUSER	73.2%	72.9%	75.7%	73.5%	78.0%	70.2%	72.8%	68.1%
ReRead	68.1%	69.3%	71.4%	71.1%	75.5%	68.8%	71.8%	69.9%
GPT-3.5-turbo	56.7%	55.3%	57.0%	55.7%	56.1%	55.9%	56.2%	57.3%
Vicuna-7B	52.2%	51.5%	52.9%	53.1%	52.6%	51.8%	52.0%	51.9%
WEBGLM-2B	62.8%	63.3%	60.1%	61.7%	63.9%	61.2%	66.0%	62.6%
ProgramFC	68.4%	67.8%	73.3%	72.3%	74.1%	63.5%	62.2%	64.3%
STEEL	75.1%	75.3%	78.0%	74.9%	78.7%	72.2%	74.5%	72.4%
SReLLM	**78.2%**	**77.9%**	**80.5%**	**79.3%**	**81.2%**	**75.4%**	**76.2%**	**75.0%**

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

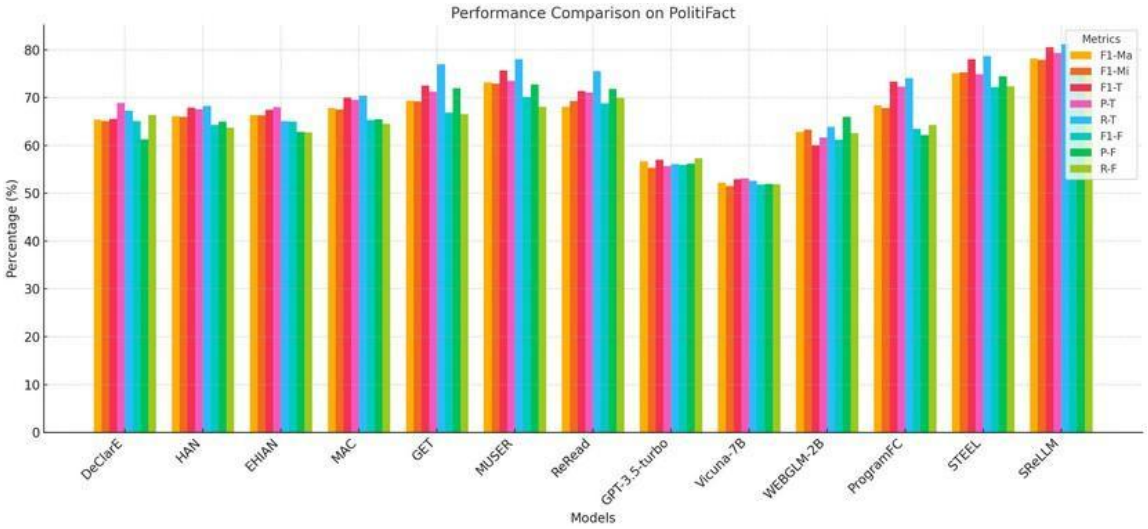

Figure 4: Performance of different model politifact

1.1 Comparison with different Models

Traditional ML methods rely on textual features such as TF-IDF, n-grams, or handcrafted linguistic features, which struggle with misinformation variability. Our retrieval-enhanced framework significantly improves performance, as shown in Table 6. The table reveals that traditional methods like Vanilla, Quadratic Answer, Response Correction, and Chain of Thought show minimal improvement, especially in the LIAR dataset, where all score 69.0%, indicating similar underlying strategies. STEEL introduces a notable boost, particularly in CHEF (79.0%), suggesting that Round Control enhances reasoning for misinformation detection. SReLLM consistently outperforms all, achieving the highest scores across datasets, with 82.0% on CHEF and 78.2% on PolitiFact, highlighting its superior generalization and verification capabilities. SReLLM achieves a 6-12% increase in accuracy over traditional ML models and outperforms deep learning models by 2% due to its improved evidence retrieval mechanism. Figure 3 shows compression of SReLLM with ML and DL. SReLLM outperforms other retrieval-augmented LLMs by using multi-round retrieval, achieving the highest accuracy (90.93%) and better explainability. Unlike Replug, FLARE, and SKR, it dynamically refines evidence collection, leading to more reliable fake news detection.

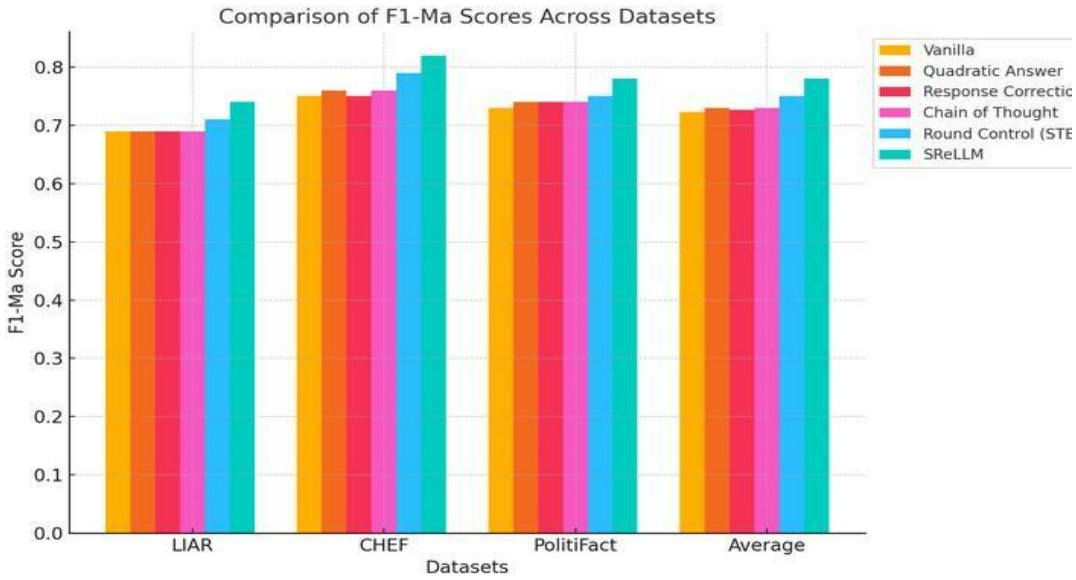

Figure 5: Category of selected research paper

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 7: Comparison of different methods on multiple datasets using F1-Ma metric.

Method	LIAR (%)	CHEF (%)	PolitiFact (%)
Vanilla	69.0%	75.0%	73.0%
Quadratic. Answer	69.0%	76.0%	74.0%
Response. Correction	69.0%	75.0%	74.0%
Chain of Thought	69.0%	76.0%	74.0%
Round Control (STEEL)	71.0%	79.0%	75.0%
SReLLM	**74.5%**	**82.0%**	**78.2%**

DISCUSSION

In this study, we introduced SReLLM, a retrieval-augmented large language model (LLM) framework for fake news detection. By employing a multi-round evidence retrieval mechanism, SReLLM effectively addresses challenges such as outdated sources, misinformation spread, and the long-tail phenomenon. Experimental results show that SReLLM outperforms traditional machine learning, deep learning, and existing retrieval-augmented models, achieving the highest F1-Ma scores across multiple datasets: 74.5% on LIAR, 82.0% on CHEF, and 78.2% on PolitiFact. These results demonstrate SReLLM's superiority over models like STEEL and other baseline approaches. Its dynamic web-based retrieval and adaptive multi-round search enhance evidence collection, improving transparency through human-readable justifications. Despite its effectiveness, SReLLM has some limitations. The reliance on a static blacklist for filtering misinformation may be insufficient in a rapidly evolving digital landscape. Additionally, input context length constraints could hinder the model's ability to process complete information, and computational demands pose challenges for large-scale deployment.

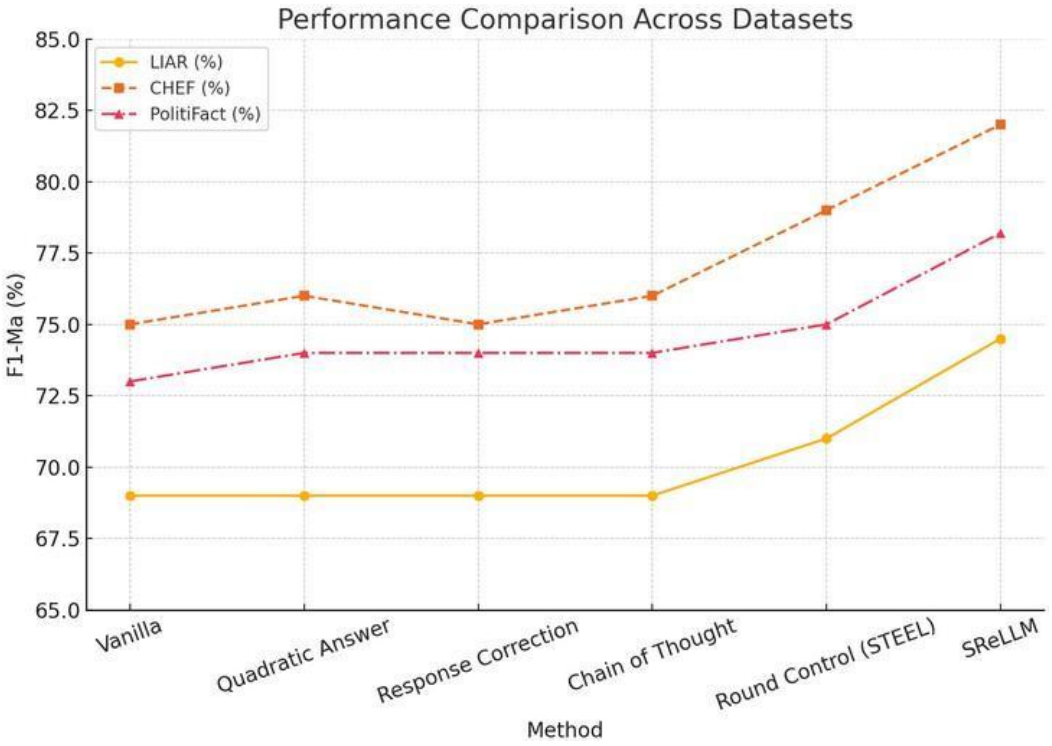

Figure 6: Category of selected research paper

Journal of Information Systems Engineering and Management
2025, 10(39s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

ACKNOWLEDGMENT

We would like to thank the university's research departments for providing the Manuscript number IU/R&D/2025-MCN0003541 according to their guidelines. This identification makes it easier to track and communicate about our research as it moves through the publication process. We would also like to express our appreciation to everyone who helped to build this work.

REFRENCES

[1] Haoran Wang and Kai Shu. “Explainable claim verification via knowledge-grounded reasoning with large language models”. In: Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore, 2023, pp. 6288–6304.

[2] D. M. J. Lazer, M. A. Baum, and Y. et al. Benkler. “The Science of Fake News”. In: Science 359.6380 (2018), pp. 1094–1096.

[3] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu. “Fake News Detection on Social Media: A Data Mining Perspective”. In: ACM SIGKDD Explorations Newsletter 19.1 (2017), pp. 22–36.

[4] S. Vosoughi, D. Roy, and S. Aral. “The Spread of True and False News Online”. In: Science 359.6380 (2018), pp. 1146–1151.

[5] Xinyi Zhou and Reza Zafarani. “A survey of fake news: Fundamental theories, detection meth- ods, and opportunities”. In: ACM Computing Surveys (CSUR) 53.5 (2020), pp. 1–40.

[6] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal. “FEVER: A Large-scale Dataset for Fact Extraction and Verification”. In: Proceedings of NAACL (2018), pp. 809–819.

[7] N. J. Conroy, V. L. Rubin, and Y. Chen. “Automatic Deception Detection: Methods for Finding Fake News”. In: Proceedings of the Association for Information Science and Technology 52.1 (2015), pp. 1–4.

[8] K. Shu, S. Wang, and H. Liu. “Beyond News Contents: The Role of Social Context for Fake News Detection”. In: Proceedings of WSDM (2019), pp. 312–320.

[9] H. Zhang, J. Li, and X. Wang. “A Deep Learning Approach for Fake News Detection Based on Content and Context Features”. In: Neurocomputing 448 (2021), pp. 150–162.

[10] T. Nakamura, K. Levy, and G. Mark. “Fact-checking Misinformation: User Motivations and Intentions”. In: Proceedings of CHI (2020), pp. 1–14.

[11] A. Hanselowski, A. PVS, F. Caspelherr, D. Chaudhuri, T. M¨oller, and I. Gurevych. “A Ret- rospective Analysis of the Fake News Challenge Stance-Detection Task”. In: Proceedings of COLING (2018), pp. 1859–1874.

[12] Botambu Collins, Dinh Tuyen Hoang, Ngoc Thanh Nguyen, and Dosam Hwang. “Trends in combating fake news on social media - a survey”. In: J. Inf. Telecommun. 5.2 (2021), pp. 247– 266.

[13] Karish Grover, S. M. Phaneendra Angara, Md. Shad Akhtar, and Tanmoy Chakraborty. “Pub- lic wisdom matters! Discourse-aware hyperbolic fourier co-attention for social text classifica- tion”. In: NeurIPS. 2022.

[14] Neema Kotonya and Francesca Toni. “Explainable automated fact-checking: A survey”. In: Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020. Barcelona, Spain (Online), 2020, pp. 5430–5443.

[15] Singh, U., Shukla, S., & Gore, M. M. (2024). Functional Connectivity and Graph Embedding-Based Domain Adaptation for Autism Classification from Multi-site Data. Arabian Journal for Science and Engineering, 1-20.

[16] Nicola Capuano, Giuseppe Fenza, Vincenzo Loia, and Francesco David Nota. “Content-based fake news detection with machine and deep learning: A systematic review”. In: Neurocomputing 530 (2023), pp. 91–103.

[17] Xichen Zhang and Ali A. Ghorbani. “An overview of online fake news: Characterization, de- tection, and discussion”. In: vol. 57. 2. 2020, p. 102025.

[18] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. “Emergent abilities of large language models”. In: 2022.
[19] Canyu Chen and Kai Shu. “Combating misinformation in the age of LLMs: Opportunities and challenges”. In: CoRR abs/2311.05656 (2023).
[20] Gautier Izacard, Patrick S. H. Lewis, Maria Lomeli, Lucas Hosseini, Fabio Petroni, Timo Schick, Jane Dwivedi-Yu, Armand Joulin, Sebastian Riedel, and Edouard Grave. “Few-shot learning with retrieval-augmented language models”. In: Journal of Machine Learning Research 24 (2023), 251:1–251:43.
[21] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. “REALM: Retrieval-augmented language model pre-training”. In: Proceedings of the 37th International Conference on Machine Learning, ICML 2020. 2020.
[22] Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christo- pher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, Xu Jiang, Karl Cobbe, Tyna Eloundou, Gretchen Krueger, Kevin Button, Matthew Knight, Benjamin Chess, and John Schulman. “WebGPT: Browser-assisted question-answering with human feedback”. In: vol. abs/2112.09332. 2021.
[23] Wenhao Yu, Dan Iter, Shuohang Wang, Yichong Xu, Mingxuan Ju, Soumya Sanyal, Chen- guang Zhu, Michael Zeng, and Meng Jiang. “Generate rather than retrieve: Large language models are strong context generators”. In: The Eleventh International Conference on Learning Representations (ICLR 2023). 2023.
[24] Akari Asai, Sewon Min, Zexuan Zhong, and Danqi Chen. “Retrieval-based language models and applications”. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts, ACL 2023. Toronto, Canada, July 2023, pp. 41–46.
[25] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R. Narasimhan, and Yuan Cao. “ReAct: Synergizing reasoning and acting in language models”. In: The Eleventh International Conference on Learning Representations (ICLR 2023). 2023.
[26] Liangming Pan, Xiaobao Wu, Xinyuan Lu, Anh Tuan Luu, William Yang Wang, Min-Yen Kan, and Preslav Nakov. “Fact-checking complex claims with program-guided reasoning”. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023). 2023, pp. 6981–7004.
[27] Weijia Shi, Sewon Min, Michihiro Yasunaga, Minjoon Seo, Rich James, Mike Lewis, Luke Zettlemoyer, and Wen-tau Yih. “REPLUG: Retrieval-augmented black-box language models”. In: CoRR abs/2301.12652 (2023).
[28] Zhengbao Jiang, Frank F. Xu, Luyu Gao, Zhiqing Sun, Qian Liu, Jane Dwivedi-Yu, Yim- ing Yang, Jamie Callan, and Graham Neubig. “Active retrieval-augmented generation”. In:
Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023. Singapore: Association for Computational Linguistics, 2023, pp. 7969–7992.
[29] Singh, U., Shukla, S., & Gore, M. M. (2024). Detection of autism spectrum disorder using multi-scale enhanced graph convolutional network. Cognitive Computation and Systems, 6(1-3), 12-25.
[30] Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton- Brown, and Yoav Shoham. “In-context retrieval-augmented language models”. In: vol. 11. 2023, pp. 1316–1331.
[31] Singh, U., Shukla, S., & Gore, M. M. (2022, December). An improved feature selection algorithm for autism detection. In 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1-8). IEEE.
[32] Jing Ma, Wei Gao, and Kam-Fai Wong. “Detect Rumors on Twitter via Deep Dynamic Multi- Task Learning”. In: Proceedings of ACL. 2019.
[33] Lianwei Wu, Jingyuan Yang, Xianpei Huang, and Shengping Zhang. “Evidence-Aware Hierar- chical Interactive Attention Networks for Explainable Claim Verification”. In: Proceedings of IJCAI. 2020.
[34] Ngoc Phuoc An Vo and Kyumin Lee. “Hierarchical User Representation with Multiple Self- Supervised Tasks for Fake News Detection”. In: Proceedings of ACL. 2021.
[35] Huimin Xu, Lianwei Wu, Shengping Zhang, and Xianpei Huang. “Graph Enhanced Truth Discovery for Fake News Detection”. In: Proceedings of WWW. 2022.

[36] Yuxuan Liao, Shu Wang, Yifan Wang, and Qing Li. “MUSER: Multi-Source Evidence Rea- soning for Fact Verification”. In: Proceedings of KDD. 2023.
[37] Yuning Hu, Jianlong Zhang, Zhiwei Zhang, and Pengfei Wang. “ReRead: Revisiting Readabil- ity for Fake News Detection”. In: Proceedings of SIGIR. 2023.
[38] OpenAI. GPT-3.5-turbo. Available at: https://openai.com/. 2022.
[39] Wei-Lin Chiang, Ziwei Xiao, Rui Zhong, Shang-Wen Xu, and J. Zico Kolter. “Vicuna: An Open-Source Chatbot Optimized for Dialogue”. In: 2023.
[40] Zhenghao Liu, Jianhao Ma, Shengding Tang, Lei Li, Zhiyuan Liu, and Maosong Sun. “WE- BGLM: A Retrieval-Augmented Generation Model for Online Question Answering”. In: Pro- ceedings of KDD. 2023.
[41] Li Pan, Zhen Xu, Yiming Liu, Kaiyu Fang, and Mingze Ma. “ProgramFC: Enhancing Fact- Checking with Programmatic Reasoning”. In: Proceedings of ACL. 2023.
[42] Sadia, H., Abbas, S. Q., & Faisal, M. (2022). A Bayesian network-based software requirement complexity prediction model. In Computational Methods and Data Engineering: Proceedings of ICCMDE 2021 (pp. 197-213). Singapore: Springer Nature Singapore.
[43] Ankita Srivastava. (2024). Improving Blockchain Security: Integrating Encryption and Hashing Techniques. International Journal of Intelligent Systems and Applications in Engineering, 12(22s), 740-747
[44] Farooq Ahmad, Mohammad Faisal, "Assessing Similarity between Software Requirements: A Semantic Approach", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.15, No.2, pp. 38-53, 2023. DOI:10.5815/ijieeb.2023.02.05
[45] Ahmad, Shamim & Tripathi, Dr. (2023). A Review Article on Detection of Fake Profile on Social-Media. International Journal of Innovative Research in Computer Science and Technology. 11. 44-49. 10.55524/ijircst.2023.11.2.9.
[46] Mishra, Alok, and Halima Sadia. 2023. "A Comprehensive Analysis of Fake News Detection Models: A Systematic Literature Review and Current Challenges" Engineering Proceedings 59, no. 1: 28. https://doi.org/10.3390/engproc2023059028
[47] Kumar, D., & Ahamad, F. (2023, December). Opinion extraction from big social data using machine learning techniques: A survey. In *AIP Conference Proceedings* (Vol. 2916, No. 1). AIP Publishing.
[48] Kumar, D., & Ahamad, F. (2024). Opinion Extraction using Hybrid Learning Algorithm with Feature Set Optimization Approach. Journal of Electrical Systems. 20. 1266-1276. 10.52783/jes.3694.