Journal of Computational Social Science (2025) 8:72
https://doi.org/10.1007/s42001-025-00403-w

RESEARCH ARTICLE

Detecting misinformation through framing theory: the frame element-based model

Guan Wang[1] · Rebecca Frederick[1] · Jinglong Duan[1] · William B. L. Wong[1] · Verica Rupar[1] · Weihua Li[1] · Quan Bai[2]

Received: 16 January 2025 / Accepted: 18 June 2025

Abstract

In this paper, we delve into the rapidly evolving challenge of misinformation detection, specifically focusing on the nuanced manipulation of narrative frames, an under-explored area within the Artificial Intelligence (AI) community. The potential for Generative AI models to generate misleading narratives highlights the urgency of addressing this issue. Drawing from communication and framing theories, we posit that the presentation or 'framing' of accurate information can dramatically alter its interpretation, potentially leading to misinformation. In particular, the intricate user interaction in social networks plays an important role in this process, as these platforms provide an unsupervised environment for disseminating misinformation among individuals. We highlight this issue through real-world examples, demonstrating how shifts in narrative frames can transmute fact-based information into misinformation. To tackle this challenge, we propose an innovative approach that leverages the power of pre-trained large language models and deep neural networks to detect misinformation originating from accurate facts, which are portrayed under different frames. These advanced AI techniques offer unprecedented capabilities in identifying complex patterns within unstructured data, critical for examining the subtleties of narrative frames. The objective of this paper is to bridge a significant research gap in the AI domain, providing valuable insights and methodologies for tackling framing-induced misinformation, thus contributing to the advancement of responsible and trustworthy AI technologies. Several experiments are conducted, and the experimental results explicitly demonstrate the various impacts of elements of framing theory, thereby proving the rationale for applying framing theory to increase performance in misinformation detection.

Keywords Misinformation detection · Framing analysis · Framing extraction · Human-centric social good

Extended author information available on the last page of the article

Published online: 09 July 2025

1 Introduction

Misinformation in today's media landscape is growing substantively, where fake news and false or misleading information are disseminated through various media channels, e.g., news websites and online social media platforms that most people frequently use and consume information [1]. Artificial Intelligence (AI) has advanced from exclusively understanding language to Generative AI (GAI) models that can automatically generate articles, posts, and narratives with remarkable sophistication [2]. The accessibility of GAI models such as ChatGPT has expedited the creation of manipulative misinformation. In most cases, it can be difficult for readers to distinguish whether the narrative was written by a GAI model or a human author [3]. Automating fact-checking or claim validation is a well-researched task that has achieved high accuracy results with traditional misinformation detection focused on keywords [4, 5]. However, when presented with accurate facts that have been manipulated through framing to create a misleading narrative, it is difficult to identify the misinformation. The framing of accurate information by selectively highlighting certain aspects while omitting others can lead to the communication of a different message than the original, accurate narrative intended [6]. The aforementioned manipulation of accurate information by changing the perspective and frame can result in the propagation of misinformation. Thus, framing plays an important role in misinformation detection.

Framing theory illuminates the process by which communicators strategically highlight specific facets of a perceived reality within a communication text [7]. This intentional emphasis serves to advance a distinct problem definition, causal interpretation, moral evaluation, and/or treatment recommendation [7]. Framing involves the selection of certain factors about an issue or event and making them salient or emphasizing these factors over others. It is about selecting and deciding which parts of a situation or event to make salient to an audience. It also suggests how information is presented and communicated in a narrative - the story that communicates the facts in a meaningful way - can influence an individual's perception and interpretation of that information and is recognized as an important concept in the communication and social science fields [6–10]. Additionally, framing theory suggests that four frame elements contribute to how information is presented: problem definition, causal interpretation, moral evaluation, and treatment recommendation [7]. Specifically, the problem definition defines the problem by determining the actions of a causal agent along with their associated costs and benefits. It is measured by what is culturally acceptable, whereas the causal interpretation identifies the forces that cause the problem. The moral evaluation makes moral judgments about the causal agent and their effects, with the treatment recommendation offering suggestions to address the problem and the potential effects these might have. When it comes to misinformation, framing theory suggests that the manner in which information is conveyed or framed can be harnessed to persuade readers into embracing inaccurate or misleading information as truthful. By strategically highlighting specific facts or interpretations while purposefully excluding others, individuals or organizations can craft a specific narrative that aligns with their agenda to mislead the audience [6].

There are numerous research studies concerning the detection of framing, the identification of elements within a frame, and the analysis of framing itself [7, 11, 12]. However, few studies explore how frames impact the emergence of misinformation or which frame element has the greatest impact on the overall frame. It is challenging to classify misinformation stemming from factual information. Therefore, misleading information created by manipulating the frame of a truthful narrative would be undetected by traditional misinformation detection models [13]. In this study, we used three topics that have sparked public controversy to evaluate the detection model we were developing: the Three Waters Reform, COVID-19, and Nuclear Pollution. For example, an excerpt of a factual information narrative with a political frame:

"The proposed three waters reform program harks back to the Havelock North water contamination event in 2016... The government estimates that we'll need a mind-boggling $120 billion to $185 billion over the next 30 years... The government believes that four entities, aggregating all the water services across the country, offer the best and quickest opportunity to achieve the desired improvements... The review was expanded to cover all three waters, this acknowledges the inter-relationships between the three networks."

Information is presented in a straightforward and factual manner, explaining the motivation behind the reform, the expected costs, and the time frame, describing the government's belief that larger entities can achieve efficiency gains, and explaining why all three water networks were reviewed. However, an excerpt of a misleading narrative with a semantic frame that uses specific terms to associate the statement with other communication contents or features, including irony, lettering, metaphor, and so on [14]. The following example shows the satire/irony, which suggests the opposite of the original message:

"Because nothing says 'clean water' like shifting responsibility from local government to some fancy-sounding entity, right?... They even established a drinking water regulator to ensure everything meets regulatory standards, because we all know how important it is to regulate things, right?... Because who needs small, local councils when you can have these big entities making all the decisions for you? Efficiency gains are just a bonus, my friends!... Because why bother keeping it simple when you can add some unnecessary complexity?"

Satire, oversimplification, and selective framing are used to mislead as it mocks the idea of clean water as a priority, ignoring the serious health concerns that prompted the government to consider these reforms, downplays the significance of regulatory standards by sarcastically framing them as if they are unnecessary, while the actual cost estimates are not addressed seriously, dismisses the efficiency gains oversimplifying the government's rationale for proposing larger entities to handle water services and sarcastically dismisses the complexity of reviewing all three waters, suggesting that it is unnecessarily complicated.

Pre-trained Large Language Models (LLMs) and deep neural networks have been recognized as efficient and effective techniques for addressing the framing classification and misinformation detection problem, as they can learn from unstructured data and identify complex patterns that are difficult to detect using traditional methods [1].

This work builds on the well-established theoretical premise that the framing of information can significantly influence public perception, to the extent that different frames can lead to the same factual content being interpreted as misinformation [7, 15]. In this paper, we employ a narrative framework and its key elements as key considerations in the process of identifying misinformation. Our contributions of this research work include:

- We formally define misinformation that is portrayed from the facts and formulate the misinformation detection problem in the context of Generative AI.
- We propose a novel model called Framed Element-based Model (FEM), which can effectively identify misinformation stemming from portrayed facts under different framing. To the best of our knowledge, this is the first full research work, tackling the framing-based misinformation detection problem.
- We are the first to investigate how framing elements affect misinformation detection, treating each element as a separate feature for the language model to process. Our research systematically examines these elements, offering important insights into the subtle ways information can be skewed using framing. This also enhances the accuracy and effectiveness of detecting misinformation.

The rest of this work is organized as follows. In Sect. 2, we discuss related works by examining misinformation detection and framing theory. In Sect. 3, we give the formal definitions and formulate the problem. Our proposed FEM model is then explained in Sect. 4. Experimental setups and datasets are introduced in Sect. 5. In Sect. 6, four experiments are conducted to evaluate the proposed model, analyze the parameters of the four framing elements, and introduce a case study which provides tangible illustrations of the model's effectiveness. Lastly, in Sect. 7, we conclude the paper and give recommendations for future work.

2 Related work

2.1 Traditional misinformation detection

With the rise of social media, the ease with which information can be distributed and consumed has increased, allowing misinformation to also increase [1]. Traditional rule-based misinformation detection for fact-checking and fake news focused on detecting misinformation by examining who provided the information or its content. Manual fact-checking relied on the author's reputation and/or the source to determine the veracity of the information [16]. Similarly, to detect fake news on social media, the social contexts, such as explicit and implicit features of users' profiles, are evaluated to determine the credibility of the information [17]. In addition to social contexts, fake news detection focuses on the content of the text by extracting linguistic features in order to detect sensational headlines that are frequent in fake news [17]. Moreover, identifying negation keywords, such as 'no,' 'not,' or 'never,' played a significant role in enhancing the classification of rumours [18]. Traditional rule-based approaches relied on information specific to the topic to identify

misinformation correctly. Therefore, these approaches experienced limitations when detecting misinformation about a new topic [19]. These shortcomings were addressed by introducing semi-supervised and unsupervised methods [20].

2.2 Deep learning based misinformation detection

Many researchers have explored the use of deep learning techniques to automate misinformation detection, including tensor-based models, transformer architectures, as well as convolutional and recurrent neural networks [1, 21–24]. Latent patterns and spatial context were extracted from tensor-based models to construct k-nearest-neighbour graphs and belief propagation for semi-supervised misinformation detection [21]. A hybrid of convolutional neural networks (CNN) and recurrent neural networks (RNN) leverages the strengths of CNN in extracting local features and of RNN in capturing long-term dependencies to detect fake news [22]. Another RNN model found that combining sentiment, emotional, irony, and hate analysis with bagging, boosting, stacking, and voting means produced a higher accuracy than without the various analyses [24]. An evaluation of transformer-based LLMs, namely BERT variants, for use as baselines in misinformation detection can achieve comparable or better performance than more complex state-of-the-art methods [23]. More recently, a transformer-based model, MisRoBÆRTa, utilized RoBERTa and BART to outperform single-transformer misinformation detection models [25]. Finally, a hybrid deep learning model integrating feature-based models and universal sentence encoding revealed promising results on the PHEME dataset [26].

While these techniques can accurately detect misinformation without considering the narrative or frame, their challenge lies in dealing with misinformation stemming from factual events that are skewed to convey a different implication. Furthermore, they also face difficulties handling lengthy news articles that potentially contain both truthful and misleading information.

2.3 Framing theory

The frame of a piece of text can increase the salience of specific parts of information, i.e., to make information more meaningful, noticeable, or memorable [7]. An example by Entman showed that a frame can influence how a large portion of readers notice, understand, remember, evaluate, or act upon information presented to them [7]. According to Entman, the problem definition, causal interpretation, moral evaluation, and treatment recommendation are the four identifiable elements of a frame [7]. Multiple methods have been developed to detect frames using different approaches. Liu et al. detected frames from news based on the article headlines by fine-tuning a Large Language Model, i.e., BERT [27]. Alternatively, Walter and Ophir leveraged computational tools to develop a novel method, the Analysis of Topic Model Networks, for the inductive identification and categorization of frames [11]. Although both misinformation detection and frame detection are possible, the impact of frames on misinformation detection requires further research. Our proposed FEM explores this impact by incorporating the framing theory presented by

Entman to solve the earlier challenge of detecting misinformation stemming from accurate facts that are skewed to be misleading potentially [7]. Additionally, FEM discerns the respective contributions of the four framing elements to the overall accuracy of misinformation detection.

3 Preliminary

3.1 Formal definition

Definition 1: Narrative generally refers to a way of sharing stories or information, whether it is spoken, written, or shared online. In the current context, the narrative indicates the news stories and articles being disseminated online. Let $\mathcal{N} = \{n_1, n_2, \dots, n_n\}$ denote the set of narratives, a narrative can be information or misinformation in online social networks, where n_i represents a single narrative.

Definition 2: Information refers to the presentation of facts in a way that aims to convey these facts accurately. The narrative of the information is constructed to reflect its true nature and implications without distorting or omitting key elements. Let $\mathcal{I} \in \mathcal{A}$ refer to the information set:

$$\mathcal{I} = \{(fa, n) \mid fa \in \mathcal{FA} \wedge n \in \mathcal{N}\}, \tag{1}$$

where $fa \in \mathcal{FA}$ refers to a specific fact in a fact set, $n \in \mathcal{N}$ refers to a narrative of a narrative set for information, and $\mathcal{A}$ refers to a set of articles.

Definition 3: Misinformation, in contrast to information, involves presenting the same factual content within a narrative that is intentionally framed to mislead, deceive, or manipulate the audience. The key aspect of misinformation in this work is not the distorted facts themselves, but rather how they are presented in a misleading narrative. Let $\mathcal{M} \in \mathcal{A}$ represent a set of misinformation which is composed of the content and the specific narrative:

$$\mathcal{M} = \{(fa, n) \mid fa \in \mathcal{FA} \wedge n \in \mathcal{N}\}, \tag{2}$$

where $fa \in \mathcal{FA}$ refers to the same fact in a fact set as information, $n \in \mathcal{N}$ refers to the narrative of a narrative set for misinformation.

Definition 4: Frame refers to how information is structured and presented within a narrative, including the perspective from which it is told. The way facts are communicated can significantly influence an individual's perception and interpretation of the information. Framing is widely recognized as a key concept in the communication and social science fields. Mathematically, fr represents a frame, and the set of frames is $\mathcal{FR} = \{fr_1, fr_2, \dots, fr_m\}$. The relationship between a frame and a narrative of one article can be represented by $\mathcal{R} : \mathcal{N} \rightarrow \mathcal{FR}$, where $R(n_i) = fr_i$, and $\mathcal{R}$ represents the element extractor.

Definition 5: Frame Elements are the specific components used to construct a frame in articles. A frame is generally composed of four elements, and they constitute how information should be displayed in front of the readers and how the readers would perceive the content. Each article consists of four elements:

"problem definition", "causal interpretation", "moral evaluation", and "treatment recommendations". Let e represent one of the elements of a frame in an article and $\mathcal{E}_i = \{e_1, e_2, e_3, e_4\}$ represents the element set of the article a_i where e_1 represents the "problem definition", e_2 represents the "causal interpretation", e_3 represents the "moral evaluation", and e_4 represents the "treatment recommendations".

3.2 Problem formulation

The Misinformation Detection problem is defined as the process of classifying articles to identify misinformation stemming from portrayed facts under different narratives, thus misleading the audience. To achieve that, we adopt the Frame Element-based Model (FEM), incorporating the elements of framing theory extracted from the articles. The FEM is trained to understand the semantics and narratives of articles. Having a set of articles $\mathcal{A} = \{a_i, a_2, \ldots, a_n\}$, given an article $a_i \in A$, the model first extracts the frame elements $E_i = \{e_1, e_2, e_3, e_4\}$ of the article, and then encode them to get the hidden state h_i of it which is later used to calculate the probability to predict if the article is misinformation or not.

$$P(h_i) = softmax(w \cdot h_i + b), \tag{3}$$

where $P(h_i)$ is the probability that an article a_i contains misinformation, and h_i represents the last hidden state of the given article a_i or corresponding element set E_i, .

The object of the last step of predicting is defined as minimizing the loss function $\mathcal{L}$:

$$\mathbf{w}^*, b^* = \arg\min_{\mathbf{w},b} \mathcal{L}(\mathbf{w}, b) + \lambda \|\mathbf{w}\|^2, \tag{4}$$

where w^* and b^* are the target optimal weights, and the loss function $\mathcal{L}$ is defined as:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log P(h_i) + (1 - y_i) \log(1 - P(h_i)) \right], \tag{5}$$

where N is the number of samples, and y_i is the actual label of the article a_i.

4 Frame element-based misinformation detection model

In this section, the proposed Frame Element-based Model (FEM) for Misinformation detection is introduced in the context of news articles. Figure 1 demonstrates the overall framework of our proposed model, and in Algorithm 1, we showcase the steps of the whole process. In this algorithm, a refers to a news article, E refers to four elements, and e is one of the elements. $prompt_1$ is the prompt used to change the frame of an article to make it misinformation, and $prompt_2$ is the prompt used to extract frame elements from the articles.

Initially, the Frame Element Extractor is used to process the news article, extracting four framing elements: Problem Definition, Causal Interpretation, Moral

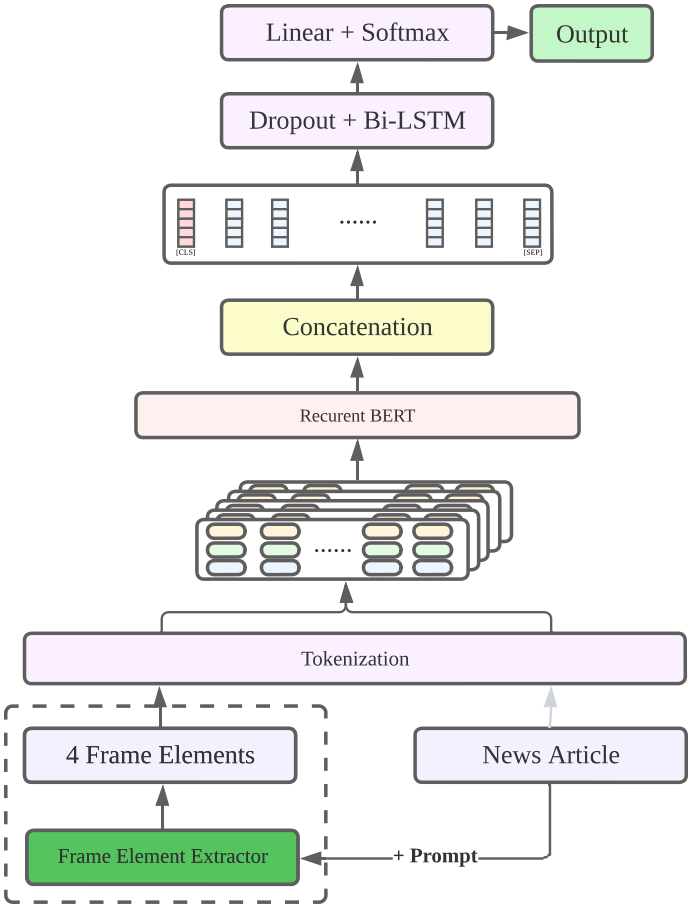


Fig. 1 The architecture of the frame element-based model

Evaluation, and Treatment Recommendation. These elements represent the core of how the information is framed. The extracted framing elements, along with the corresponding news article, are then tokenized, performing the fundamental preprocessing step in Natural Language Processing (NLP), which typically involves steps such as tokenization, encoding, and decoding.

To capture the subtle contextual nuances of each element, following the tokenization, we independently encode each element and the corresponding news article (Lines 6 to 14 in Algorithm 1). This nuanced understanding of different elements is vital as each frame carries different weights and implications for the overall narrative of this news article. The separate encoding also allows us to quantify the impact of each element, revealing the most influential aspects of how the article is framed, thereby increasing the likelihood of identifying misinformation.

Independently encoding each element and article is a strategic choice that can enhance the model's analytical precision, enabling reliable detection of

misinformation. Line 6 starts the recurrent process. An empty tensor is created in advance and used to concatenate each embedding from each iteration of the loop. In the recurrent process, we first encode the article, which serves as the main body of the input, and then encode each element.

The embeddings of each element *embE* and the news article *embA* are then concatenated to form a dense vector *emb* as the representation of the whole input, followed by a dropout layer to prevent over-fitting.

$$emb_t = concat(emb_{t-1}, h_t), \tag{6}$$

where emb_t represents the concatenated embeddings of the current time step, and h_t represents the embeddings of an element *embE* or the article *embA*.

The concatenated embeddings *emb* from the previous layer are then fed to a Bi-LSTM layer. The Bi-LSTM layer is applied to capture the holistic context after all elements and article embeddings are concatenated. This enables the model to comprehend the relationships between different elements of the article.

$$h_i = ReLU(BiLSTM(emb)), \tag{7}$$

where h_i is the representation of the input after being processed by a Bi-LSTM layer and a ReLU activation function.

A linear layer, including a Dropout, is applied to map the high-dimensional output representation h_i from the Bi-LSTM layer to the target space. The softmax function is used to obtain the probability distribution over the potential classes, which finalizes the prediction process to identify the misinformation.

$$predicts = softmax(Dropout(h_i)W^T + b), \tag{8}$$

where *predicts* is the probability distribution of the class labels, W^T is the learnable weight matrix, b refers to the bias.

Algorithm 1 Frame Element-based Misinformation detection Algorithm

Input: $D = (a,\ E)$
Output: 0 (misinformation) or 1 (information)

$Information \leftarrow Collect(sources)$
$Misinformation \leftarrow ChatGPT(information,\ prompt_1)$
$FrameElements \leftarrow ChatGPT(articles,\ prompt_2)$
Create BERT, BiLSTM, FC Layer as classifier, Dropout, ReLU
$emb := \{\}$
for $a_i, E_i \in D$ **do**
    $embA := BERT(a_i)$
    $emb := concat(emb,\ embA)$
    for $e_j \in E_i$ **do**
        $embE := BERT(e_j)$
        $emb := concat(emb,\ embE)$
    end for
    $emb := Dropout(emb)$
end for
$outputs := BiLSTM(emb)$
$h := ReLU(outputs)$
$logits := classifier(Dropout(h))$
$predicts := softmax(logits)$

5 Experiment setups

5.1 Model setup

To ensure an efficient training process, we conduct our experiments on the Paperspace[1] platform utilising the following tailored computational and training settings to the unique demands of each dataset:

- *GPU Configuration*: The model is trained over a span of 100 epochs utilising NVIDIA's A6000 48GB GPU and 45 GB 8 CPU.
- *Dropout*: To mitigate the risk of overfitting, a dropout rate of 0.3 was applied during training.
- *Language Model*: We use the bert-base-uncased version from HuggingFace, which consists of 12 transformer encoder layers, each with 12 self-attention heads and a hidden size of 768. The intermediate feedforward layer has a dimensionality of 3072.
- *Learning Rate*: The training uses an initial learning rate of 1×10^{-5}, and it is modulated following a cosine schedule with a warm-up phase. The warm-up steps vary in accordance with the specificities of each dataset.

[1] https://www.paperspace.com/.

Table 1 The statistics of the datasets after pre-processing

Dataset	Articles	Average length
The Three Waters	3262	823
Covid-19	13,386	537
Nuclear Pollution	2431	482
Mixed-topic	5915	469

- *Batch Size*: The batch size is determined based on the particular requirements and characteristics of each dataset.
- *Frame Element Extractor*: ChatGPT, as a powerful generative AI model, is used as the element extractor. Different extractors can be applied for the same purpose.

5.2 Datasets

In this section, we introduce four datasets used to evaluate our model. To assess the generalization capability of the model, we utilized four datasets, including the Three Waters Reform dataset, the COVID-19 dataset, the Nuclear Pollution dataset, and a mixed-topic dataset, i.e., the Kaggle Fake News dataset. The statistics of these datasets are displayed in Table 1.

- *Three Waters Reform* dataset is collected from The Knowledge Basket.[2] We only capture the news focus on the “Three Waters Reform" in New Zealand, a topic of substantial political discourse and interest spanning from 2017 to June 2023. This dataset accumulates a total of 1,841 articles. Following the application of our labelling process yields 3,262 articles labelled in concordance with their identified frames and frame elements.
- *Covid-19* is collected using Newsapi,[3] which is an API service that allows developers to retrieve news articles from various sources on a worldwide scale. We use “Covid-19" as keywords to retrieve news articles in the period from 01/12/2019 to 20/08/2023. These articles reflect the timely attitude to the COVID-19 pandemic. This dataset includes 13,386 articles after the pre-processing.
- *Nuclear Pollution* dataset is collected using the Newsapi as well and with the keywords “nuclear pollution" over the last 5 years. This dataset provides a comprehensive view of the discourse surrounding nuclear pollution, offering a diverse range of perspectives and information. After the data pre-processing, there are 2,431 articles with an average token length of 482.
- *Kaggle Fake News Dataset*[4] contains news articles from multiple sources, such as Reuters. For the purpose of our study, we confine our selection to the “TRUE" set and randomly select 3k articles. To augment the dataset and fulfill the research

[2] https://www.knowledge-basket.co.nz/.

[3] https://newsapi.org/.

[4] https://www.kaggle.com/datasets/stevenpeutz/misinformation-fake-news-text-dataset-79k.

objectives, we produce an additional set of data by varying the frame of existing news, ultimately resulting in 5,915 labeled samples following the implementation of our augmentation process.

5.3 Data pre-processing

Our proposed methodology begins with the collection of datasets comprising news articles from reliable sources. These articles constitute our ground truth, representing information opposite to misinformation. Accordingly, we synthesize misinformation based on the framing theory by altering the frames of our collected news articles. This process augments our datasets in a generative method at the document level and unfolds in three structured phases [28].

- *Frame Identification and Element Extraction*: Utilizing the capabilities of LLM, we first process the collected news articles to identify their frames and extract four elements of framing theory. These extracted framing elements reflect the original and unaltered state of the news articles. They are annotated with the label “1", signifying their category as information. The frames we harness in this work are selected and proven by domain experts in communication.
- *Frame Alteration*: The second stage involves the alteration of the frame, utilising ChatGPT to manipulate the article narrative while maintaining the original factual information. This step simulates the process of creating misinformation through narrative manipulation, a common way that preserves factual information but skews the frame to mislead readers. 20% of the altered narratives are verified by the domain experts.
- *Element Extraction*: In the final step, we process the narrative-manipulated articles through ChatGPT by applying a fine-grained prompt to extract the corresponding four elements of framing theory, labelled “0" along with the manipulated articles, signifying their category as misinformation. This eventually establishes the basis for comparison with the information.

This pre-processing procedure is designed to construct binary-category datasets that comprise information and misinformation, incorporating elements of framing theory, thereby enabling the nuanced training of our model. Through this process, we not only aim to create datasets that serve as the foundation for misinformation detection but also enhance our understanding of how narrative (framing theory in this work) can be utilized to generate misinformation.

5.4 Evaluation metrics and baselines

To evaluate the performance of our proposed model (FEM), we employ the Confusion Matrix as our primary evaluation measurement. The Confusion Matrix provides

a comprehensive visualization of the performance by categorizing predictions into four different classifications [29].

- *True Positives (TP)*: when predicted misinformation is actually labeled as misinformation;
- *True Negatives (TN)*: when predicted information is actually labeled as information;
- *False Positives (FP)*: when predicted information is actually labeled as misinformation;
- *False Negatives (FN)*: when predicted misinformation is actually labeled as information.

Based on the Confusion Matrix, Precision, Recall, F1-score, and Accuracy are calculated to assess the model by comparing it with other existing baseline models.

- *Accuracy* is utilized to evaluate the model's performance across all categories.

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FN| + |FP|} \tag{9}$$

- *Precision* evaluates the correctness of the positive instances (the correctness of misinformation predicted) that our model has predicted.

$$Precision = \frac{|TP||}{|TP| + |FP|} \tag{10}$$

- *Recall* presents an indication of how many of the actual positive instances our model can correctly recognize.

$$Recall = \frac{|TP|}{|TP| + |FN|} \tag{11}$$

- *F1-score* is the harmonic mean of precision, and it provides a balanced measure that takes both precision and recall into account.

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

As our baselines, we perform the fine-tuning of several highly utilised pre-trained transformer-based language models, followed by a feed-forward layer as a classifier for misinformation detection. These baselines include:

- *BERT* [30] is a groundbreaking transformer-based model in the field of NLP. It is known for its deep bidirectional training, meaning it considers the context from both the left and right sides in all layers. This leads to a more nuanced understanding of language context and semantics.
- *RoBERTa* [31] is built upon BERT by modifying key hyperparameters, training with more data, and longer training times. These changes help RoBERTa outper-

Table 2 Results on the Three Waters Dataset

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8469	0.8188	0.8127	0.8157
RoBERTa	0.8622	0.8784	0.7915	0.8327
ALBERT	0.8086	0.7651	0.8057	0.7849
XLNet	0.8545	0.8113	0.8657	0.8376
LongFormer	0.8591	0.8283	0.8516	0.8398
FEM (text+frames)	**0.9862**	**0.9695**	**0.9734**	**0.9715**
FEM (only text)	0.8652	0.8316	0.8638	0.8474
FEM (only frames)	0.9278	0.9355	0.9605	0.9478

Bold results indicate the best performance

Table 3 Results on the Covid-19 Dataset

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8372	0.8052	0.8074	0.8063
RoBERTa	0.8547	0.8539	0.7867	0.8190
ALBERT	0.8104	0.7783	0.8163	0.7968
XLNet	0.8429	0.8207	0.8629	0.8412
LongFormer	0.8546	0.8617	0.8694	0.8655
FEM (text+frames)	**0.9783**	**0.9583**	**0.9708**	**0.9645**
FEM (only text)	0.8865	0.8737	0.8826	0.8781
FEM (only frames)	0.9132	0.9195	0.9361	0.9277

Bold results indicate the best performance

form BERT on several benchmark NLP tasks. It is known for its improved robustness and efficiency.

- *ALBERT* [32] is a version of BERT optimized for lower memory consumption and increased speed. It introduces two major modifications: factorized embedding parameterization and cross-layer parameter sharing.
- *XLNet* [33] is an extension of the Transformer model. Instead of the standard transformer, XLNet uses transformer-XL [34]. It combines the best of both auto-regressive (AR) and auto-encoding (AE) models. Unlike BERT, XLNet learns to predict a word at a position in a sequence, considering all permutations of the sequence.
- *LongFormer* [35] is designed to handle longer texts. It is an extension of the standard transformer-based model, similar to BERT, but optimized for processing lengthy documents. Its key innovation is the introduction of an attention mechanism that scales linearly with sequence length.

6 Experiments and analysis

6.1 Experiment 1: model evaluation—against baselines

In this experiment, we compare the performance of our model with other baseline models on four datasets introduced in Sect. 5. The results are displayed in

Table 4 Results on the Nuclear Pollution Dataset

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8035	0.7921	0.80167	0.7969
RoBERTa	0.8167	0.8234	0.7826	0.8025
ALBERT	0.8051	0.7568	0.7864	0.7713
XLNet	0.8268	0.8035	0.8284	0.8158
LongFormer	0.8462	0.8254	0.8316	0.8285
FEM (text+frames)	**0.9538**	**0.9429**	**0.9531**	**0.9480**
FEM (only text)	0.8491	0.8365	0.8537	0.8450
FEM (only frames)	0.9035	0.9216	0.9268	0.9242

Bold results indicate the best performance

Table 5 Results on the Mixed-topic Dataset

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8354	0.8127	0.8165	0.8146
RoBERTa	0.8497	0.8503	0.7902	0.8191
ALBERT	0.8126	0.7816	0.8257	0.8030
XLNet	0.8528	0.8320	0.8783	0.8545
LongFormer	0.8736	0.8542	0.8867	0.8701
FEM (text+frames)	**0.9696**	**0.9582**	**0.9683**	**0.9632**
FEM (only text)	0.8823	0.8574	0.8929	0.8748
FEM (only frames)	0.9158	0.9207	0.9319	0.9263

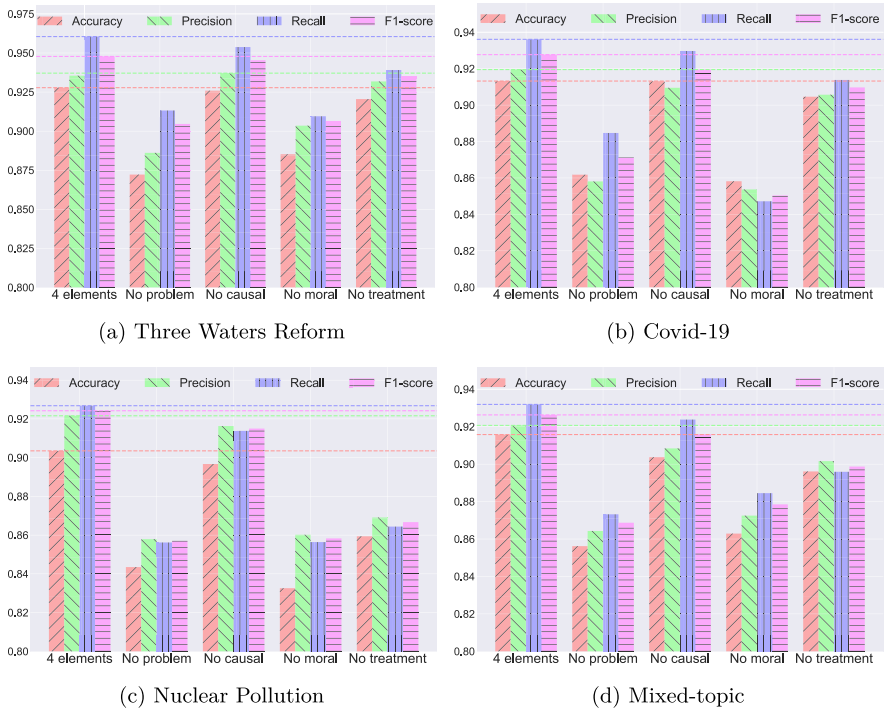
Bold results indicate the best performance

Tables 2 ,3, 4 and 5 respectively. The results on each dataset consistently show that our model (FEM) incorporating frame elements with the original news article significantly outperforms other models with only articles, presenting the importance of frame elements.

From these results, we can observe that frame elements play a crucial role in understanding and interpreting information. We also demonstrate the results of our model with only frame elements and texts as input, respectively. Compared to the baselines, the results obtained with only frame elements as input are also superior to them.

By analyzing the performance of our model with only frame elements compared to the other baselines, we can observe the significance of these elements. Frame elements contribute to a deeper semantic understanding of the content by narrowing it down to the core theme, thereby reducing distracting noise. This enables the model to grasp not just the explicit meaning but also the implicit intentions and nuances, thus increasing the probability of precisely detecting misinformation.

The experiment results in Table 5 on the Mixed-topic dataset demonstrate that frame elements can also provide a more general representation of information, making the model more adaptable and robust to variations of information.



(a) Three Waters Reform (b) Covid-19

(c) Nuclear Pollution (d) Mixed-topic

Fig. 2 Measure the performance of removing one of the elements on all four datasets

6.2 Experiment 2: parameter analysis

In this experiment, we conduct a comparative analysis to explore the contribution of each element to misinformation detection. The experimental framework analyzes the composite efficacy of the model, which is equipped with four elements: problem definition, causal interpretation, moral evaluation, and treatment recommendation. Within the area of misinformation detection with frame elements, exploring the individual contribution of distinct frame elements is vital to help us understand how frame elements influence the model's capability to grasp the veracity of information.

The model with all four elements serves as the benchmark for optimal performance, showing a high degree of accuracy, precision, recall, and F1-score. This provides a holistic, frame-element-based analysis of information, thereby enhancing the probability of identifying misinformation.

Then, we remove each frame element from all four elements, keeping the other three elements remaining. Figure 2 displays all performance metrics while Fig. 3 demonstrates the trend of F1-Scores during the training process.

From all these figures, we can observe that when the element of Problem Definition is removed from the model, a pronounced decrease in all measurements is demonstrated. This suggests that the recognition of Problem Definitions is instrumental in the precise detection of misinformation, potentially due to its role in pinpointing the core theme within the narrative that may be manipulated. Without the

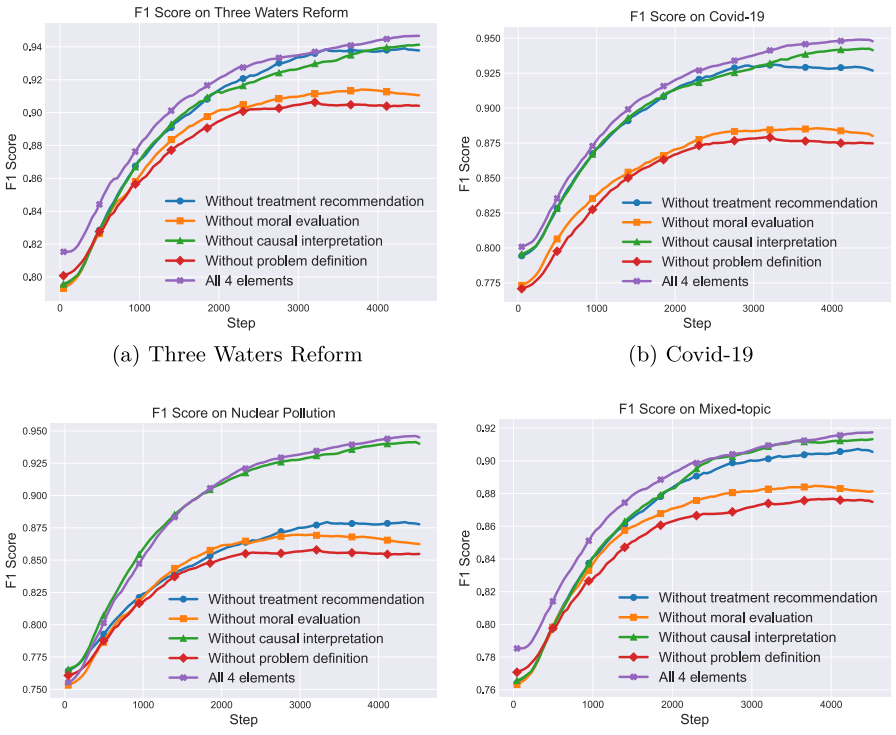
incorporation of the element of Problem Definition, the capability of the model to differentiate between true and misleading content is significantly compromised.

Meanwhile, the absence of the frame of Moral Evaluation also results in a noticeable decline in all performance metrics. It appears to be an important factor in the framing of information, indicating that it is often manipulated in the context of misinformation to elicit emotional biases or influence ethical stances.

The model, which lacks the frame of problem definition or moral evaluation, demonstrates a noticeable drop in precision and accuracy, indicating a higher rate of false positives. This implies that while the model may still identify genuine instances of misinformation, it is also more likely to incorrectly classify accurate information as misinformation.

On the contrary, a lack of the frame of Causal Interpretation or Treatment Recommendation does not indicate a substantial decline in performance metrics compared to the benchmark. This observation suggests that while they play a role in the misinformation detection process, their absence does not significantly impact the model's ability to identify misinformation.

One noticeable difference in the results demonstrated in Figs. 2c and 3c on the Nuclear Pollution dataset is the lack of a frame for Treatment Recommendation.



(a) Three Waters Reform (b) Covid-19

(c) Nuclear Pollution (d) Mixed-topic

Fig. 3 The F1-scores during the training process on all four datasets

Removing the Treatment Recommendation element also results in a lower performance across all metrics, indicating that, in the context of nuclear pollution, the treatment recommendation is likely to be a key indicator of the news articles. A lack of this element in this area could also allow misinformation, proposing ineffective or misleading responses to harness the readers. The decrement in performance of missing the frame of treatment recommendation also implies that the contribution of each element can vary depending on the subject.

6.3 Experiment 3: similarity comparison

In this experiment, we analyze the relationship between the similarity between information and misinformation under different conditions relating to the presence or absence of specific frame elements within our model. This experiment represents how closely misinformation mirrors authentic information in terms of framing. The cosine function is used to calculate their similarities:

$$sim(h_i, h_j) = \frac{h_i \cdot h_j}{\| h_i \| \| h_j \|}, \tag{13}$$

where h_i and h_j represent the final hidden states of two articles or elements from two articles.

Table 6 shows the similarities of one randomly selected article from the Three Waters Reform dataset and the overall performance (F1-score) of the model on this dataset. The similarity between the information article and the misinformation article is used as the benchmark for further analysis. The similarity of 0.86 shows that, without any specialized modification, misinformation is quite successful at resembling a genuine news article.

From Table 6, we can observe that the similarity between the information article and the misinformation article is the highest at 0.85. However, the F1 score of the model with only text on this dataset is 0.8474, which is lower than for other conditions, especially when utilising only all four elements, which hold the lowest similarity, 0.61, and the highest F1-score, 0.9478.

This experiment suggests an inverse relationship between similarity and model performance. The lower the similarity, the higher the performance is in detecting

Table 6 One single pair similarity and similarities removing one of the elements

Info vs Mis-info	Similarity	F1-score
Article Similarity	0.86	0.8474
Elements Similarity(all 4 elements)	**0.61**	**0.9478**
Elements Similarity(without problem)	0.79	0.9046
Elements Similarity(without causal)	0.62	0.9454
Elements Similarity(without moral)	0.81	0.9065
Elements Similarity(without treatment)	0.64	0.9354

Bold results indicate the best performance

misinformation. This highlights the significance of elements of framing theory in the detection process and underscores their potential to enhance detection accuracy.

We also calculate the average similarity scores between information and misinformation across four distinct datasets under different conditions. Results are displayed in Table 7. The pattern of similarity scores is relatively consistent across various topics, indicating that the manipulation of framing in misinformation follows a similar pattern. However, on the nuclear dataset, when omitting the frame of treatment recommendation, the similarity still remains at a high level, indicating the importance of treatment recommendation in this dataset.

Overall, comparing the average similarities across all datasets reveals a consistent alignment, with minimal deviations when distinct elements are omitted, except when the frame of treatment recommendation is omitted within the nuclear dataset. This exception highlights a unique pattern specific to the topic, demonstrating that different elements of framing theory exert varying levels of influence depending on the subject matter. These findings underscore the crucial importance of topic-sensitive approaches in accurately detecting misinformation.

6.4 Experiment 4: case study

In this experiment, we conduct a case study to analyze the similarities between two articles written about the same topic, each with distinct frames and frame elements. The articles focus on the government's proposed water reforms.

The first article has a political frame:

"There's a lot of change being proposed by the government... Fundamentally, they're considering shifting responsibility for our three waters: water supply, wastewater, and stormwater, from local government into four large entities... The government now believes that costs of between $120 billion and $185b will be required: between $4 and $6b per year on average... The proposed three waters reform program harks back to the Havelock North water contamination event in 2016... It's on this basis that the government has concluded that four entities, aggregating all the water services across the country, offer the best and quickest opportunity to achieve the desired improvements to the three-waters networks."

While the second article has a semantic frame to show its satire:

Table 7 Compare article average similarity with average similarity calculated using four elements

Info vs Mis-info	Three water	Covid	Nuclear	Mixed
Article Similarity	0.86	0.82	0.83	0.85
Elements Similarity(all 4 elements)	0.58	0.62	0.59	0.61
Elements Similarity(without problem)	0.83	0.81	0.82	0.83
Elements Similarity(without causal)	0.59	0.62	0.60	0.63
Elements Similarity(without moral)	0.81	0.79	0.81	0.80
Elements Similarity(without treatment)	0.60	0.64	0.78	0.63

"Oh, boy! The government is proposing some exciting changes, folks. Brace yourselves because they're considering taking control of our beloved three waters. You know, the precious water supply, wastewater, and stormwater that our local government has been responsible for?... The government estimates that we'll need a mind-boggling $120 billion to $185 billion over the next 30 years... Well, now they want to hand it over to these big entities. What a brilliant idea, right?... And get this – the government thinks it would be cheaper if larger entities took over the water services. Apparently, they can borrow more, with the government's backing, of course. I mean, who needs small, local councils when you can have these big entities making all the decisions for you?"

The similarities calculated in Table 6 of 0.86 indicate that the articles are highly similar as they both share details about the reform. However, once the frame elements are considered, the article similarity decreases to 0.61.

The political frame is informative and objective, presenting a detailed overview using formal language and statistics to support its claims. In contrast, the semantic frame is emotional and opinionated, employing colloquial language and vivid imagery to engage readers emotionally, which may risk oversimplification and bias.

To determine the classification without frame elements, our model only encodes the news article. In comparison, for classifying with the frame elements, the elements and news articles are encoded independently, with their embeddings concatenated into one vector prior to classification. The inclusion of these extra features enhances the model's performance.

For example, given the problem definition for the political frame:

"The proposed shift of responsibility for three waters from local government to four large entities known as water supply entities."

As well as the problem definition for the semantic frame:

"The proposed government takeover of three waters"

The problem definition of each article highlights its differences in framing. The politically framed article's problem definition is detailed with a neutral tone, while the semantic frame's problem definition is short with a negative perspective.

Both articles are classified as information without frame elements. However, once the frame elements are considered, the semantic frame is correctly classified as misinformation. This suggests that including the frame elements in our model contributes to the successful classification of misinformation by reinforcing the differences between the two semantically similar articles.

6.5 Discussions

We proposed a Frame Element-based Model (FEM) to distinguish misinformation from the information. Several experiments are conducted, providing crucial insights into the importance of elements of framing theory in misinformation detection. Results are evaluated by comparing them with baseline models. Furthermore, we analyzed the contribution of each element, demonstrating the different roles of the elements. By comparing the article similarities and element similarities, we gained

insights into how these elements enhance the performance of detecting misinformation. Based on the experimental results, we have the following insights:

- The results of Experiment 1 on all datasets consistently showed that incorporating elements of framing theory while detecting misinformation stemming from portrayed facts under different narratives can help improve performance.
- The parameter analysis experiment revealed that the absence of certain framing elements, particularly Problem Definition and Moral Evaluation, leads to a significant decrease in the model's accuracy and precision. This underscores the critical role these elements play in accurately detecting misinformation.
- Experiment 2 also demonstrates a finding that the specific element plays a different role on different topics, highlighting the potential impact of elements and underscoring the necessity for topic-sensitive approaches in misinformation detection, as different elements of framing theory have varying levels of impact depending on the topic.
- The similarity comparison experiment further illustrated how misinformation closely mirrors authentic information in terms of framing. The results indicated an inverse relationship between similarity and model performance: lower similarity between the information and misinformation articles led to higher performance in detecting misinformation.
- The case study shows that not applying the framing theory to the articles can lead to the result of the article with a semantic frame incorrectly classified as information, increasing the potential for misleading interpretations. It also shows that while articles may be semantically similar, the choice of framing can significantly impact the narrative, making the content appear misleading or misinterpreted.
- We leveraged LLMs in this study to identify article frames, extract frame elements, and perform frame-based content manipulation. While LLMs have shown strong performance in tasks involving information extraction and language understanding, they also present certain limitations, particularly their sensitivity to prompt phrasing, which can affect the consistency of frame detection, element extraction, and content rewriting. To mitigate this, we carefully refined and tested our prompts to ensure systematic control and reduce potential biases in the generated dataset.

7 Conclusion and future work

In this work, we introduce the Framed Element-based Model (FEM) to identify misinformation in the context of news articles incorporating the elements of framing theory. This model leverages ChatGPT and deep neural networks to detect misinformation originating from accurately portrayed facts under different frames. The efficacy of FEM is demonstrated through comprehensive performance comparisons with other methods, highlighting the effectiveness of Framed Element-based approach against traditional misinformation detection models. The contribution of each element is also evaluated and analyzed along with the similarities under

different conditions, indicating the importance of the specific element and showcasing how the narrative of an article is framed.

This work has laid a foundational understanding of how elements of framing theory influence the perception and interpretation of information. Building upon the insights obtained, there are several future directions. Future studies can delve into the impact of specific elements across various topics, such as the frame of treatment recommendations on the nuclear dataset, which shows a greater impact than on the other three datasets, raising questions that can be explored in the future. We also intend to explore how the proposed model can be integrated into broader misinformation detection systems to handle diverse real-world scenarios, including cases where the framing mechanism is not explicitly known. In addition, we plan to further strengthen the generalizability of our approach by evaluating FEM on externally sourced misinformation datasets not generated through ChatGPT. We also intend to explore comparisons with frame-aware baselines, which would provide a clearer understanding of how framing theory enhances detection capabilities in broader contexts. These extensions will complement our current findings and offer a more comprehensive evaluation of FEM's effectiveness.

Acknowledgements This project has been supported by the DCT Faculty Contestable Research Fund 2023, Auckland University of Technology (AUT), New Zealand. We acknowledge the support and resources provided by AUT and thank all the participants and collaborators who dedicated their time and effort to this work.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data availibility statement The data that support the findings of this study have been processed using Large Language Models (LLMs) to generate information that may be potentially misleading. Due to the nature of this processing, the data are not publicly available. However, the data can be made available from the authors upon reasonable request, subject to approval and with the necessary safeguards in place to ensure the responsible use of the information.

Declarations

Conflict of interest The authors declare that there is no Conflict of interest related to this work

References

1. Islam, M. R., Liu, S., Wang, X., & Xu, G. (2020). Deep learning for misinformation detection on online social networks: A survey and new perspectives. *Social Network Analysis and Mining, 10*, 1–20.

2. Longoni, C., Fradkin, A., Cian, L., Pennycook, G. (2022). News from generative artificial intelligence is believed less. In: 2022 ACM conference on fairness, accountability, and transparency (pp. 97–106).
3. Kshetri, N. (2023). Chatgpt in developing economies. *IT Professional, 25*(2), 16–19. https://doi.org/10.1109/MITP.2023.3254639
4. Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In: Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 2931–2937).
5. Tchechmedjiev, A., Fafalios, P., Boland, K., Gasquet, M., Zloch, M., Zapilko, B., Dietze, S., Todorov, K. (2019). Claimskg: A knowledge graph of fact-checked claims. In: The Semantic Web–ISWC 2019: 18th international semantic web conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18, pp. 309–324. Springer.
6. Entman, R. M. (2004). Projections of power: Framing news, public opinion, and u.s. foreign policy. University of Chicago Press
7. Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58.
8. Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
9. Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication, 49*(1), 103–122.
10. Fairhurst, G. T., & Sarr, R. A. (1996). *The art of framing*. Jossey-Bass.
11. Walter, D., & Ophir, Y. (2019). News frame analysis: An inductive mixed-method computational approach. *Communication Methods and Measures, 13*(4), 248–266.
12. Touri, M., & Koteyko, N. (2015). Using corpus linguistic software in the extraction of news frames: Towards a dynamic process of frame analysis in journalistic texts. *International Journal of Social Research Methodology, 18*(6), 601–616. https://doi.org/10.1080/13645579.2014.929878
13. Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR), 53*(5), 1–40.
14. Scheibenzuber, C., Neagu, L.-M., Ruseti, S., Artmann, B., Bartsch, C., Kubik, M., Dascalu, M., Trausan-Matu, S., & Nistor, N. (2023). Dialog in the echo chamber: Fake news framing predicts emotion, argumentation and dialogic social knowledge building in subsequent online discussions. *Computers in Human Behavior, 140*, Article 107587. https://doi.org/10.1016/j.chb.2022.107587
15. Chong, D., & Druckman, J. N. (2007). Framing theory. *Annual Review of Political Science, 10*(1), 103–126.
16. Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics, 10*, 178–206. https://doi.org/10.1162/tacl_a_00454
17. Shu, K., Wang, S., Liu, H. (2018). Understanding user profiles on social media for fake news detection. In: 2018 IEEE conference on multimedia information processing and retrieval (MIPR) (pp. 430–435). IEEE
18. Kwon, S., Cha, M., Jung, K., Chen, W., Wang, Y. (2013) Prominent features of rumor propagation in online social media. In: 2013 IEEE 13th international conference on data mining (pp. 1103–1108). https://doi.org/10.1109/ICDM.2013.61
19. Vlachos, A., Riedel, S. (2014) Fact checking: Task definition and dataset construction. In: Danescu-Niculescu-Mizil, C., Eisenstein, J., McKeown, K., Smith, N.A. (eds.) Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science, pp. 18–22. Association for Computational Linguistics, Baltimore, MD, USA . https://doi.org/10.3115/v1/W14-2508 . https://aclanthology.org/W14-2508
20. Oshikawa, R., Qian, J., Wang, W.Y. (2018) A survey on natural language processing for fake news detection. CoRR **abs/1811.00770**
21. Abdali, S., Bastidas, G.G., Shah, N., Papalexakis, E.E. (2020) Tensor embeddings for content-based misinformation detection with limited supervision. Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities, 117–140
22. Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid cnn-rnn based deep learning approach. *International Journal of Information Management Data Insights, 1*(1), Article 100007.

23. Pelrine, K., Danovitch, J., & Rabbany, R. (2021). The surprising performance of simple baselines for misinformation detection. *Proceedings of the Web Conference, 2021*, 3432–3441.
24. Pillai, S. E. V. S., Hu, W.-C. (2023). Misinformation detection using an ensemble method with emphasis on sentiment and emotional analyses. In: 2023 IEEE/ACIS 21st international conference on software engineering research, management and applications (SERA) (pp. 295–300). https://doi.org/10.1109/SERA57763.2023.10197706
25. Truică, C.-O., & Apostol, E.-S. (2022). MisrobÆrta: Transformers versus misinformation. *Mathematics*. https://doi.org/10.3390/math10040569
26. Alzahrani, A., Baabdullah, T., Almotairi, A., Rawat, D. B. (2023). A hybrid deep learning architecture for misinformation detection on social media. In: 2023 IEEE 24th international conference on information reuse and integration for data science (IRI) (pp. 199–204). https://doi.org/10.1109/IRI58017.2023.00040
27. Liu, S., Guo, L., Mays, K., Betke, M., Wijaya, D. T. (2019). Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence. In: Proceedings of the 23rd conference on computational natural language learning (CoNLL) (pp. 504–514).
28. Bayer, M., Kaufhold, M.-A., & Reuter, C. (2022). A survey on data augmentation for text classification. *ACM Computing Surveys, 55*(7), 1–39.
29. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22–36.
30. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
31. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692
32. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942
33. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, **32**.
34. Dai, Z., Yang, Z., Yang, Y., Carbonell, J. G., Le, Q., Salakhutdinov, R. (2019) Transformer-xl: Attentive language models beyond a fixed-length context. In: Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 2978–2988).
35. Beltagy, I., Peters, M. E., Cohan, A. (2020) Longformer: The long-document transformer. arXiv:2004.05150

Authors and Affiliations

Guan Wang[1] · Rebecca Frederick[1] · Jinglong Duan[1] · William B. L. Wong[1] · Verica Rupar[1] · Weihua Li[1] · Quan Bai[2]

✉ Weihua Li
weihua.li@aut.ac.nz
https://scholar.google.co.uk/citations?user=-Ecc4U8AAAAJ

Guan Wang
guan.wang@autuni.ac.nz

Rebecca Frederick
ngt8261@autuni.ac.nz

Jinglong Duan
jinglong.duan@autuni.ac.nz
https://scholar.google.com/citations?user=UNAu0WcAAAAJ

William B. L. Wong
william.wong@aut.ac.nz
https://scholar.google.co.uk/citations?user=iRIYZy4AAAAJ

Verica Rupar
verica.rupar@aut.ac.nz
https://scholar.google.co.uk/citations?user=d5U1FNkAAAAJ

Quan Bai
quan.bai@utas.edu.au
https://scholar.google.co.uk/citations?user=V6p5hfkAAAAJ

[1] School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand

[2] School of Information and Communication Technology, University of Tasmania, Hobart, TAS 7001, Australia