The Informed Fake News Advisor (IFNA): Toward Sociotechnical Solutions for Fake News Detection

Namosha Veerasamy and Danielle Badenhorst
CSIR, Pretoria, South Africa

nveerasamy@csir.co.za
dbadenhorst@csir.co.za

Abstract The detection of fake news is a multidimensional challenge that demands solutions extending beyond purely computational approaches. Although advances in natural language processing, machine learning, deep learning, and multimodal analysis have strengthened technical capabilities, misinformation continues to proliferate. Fake news thrives within environments shaped by complex social interactions, platform-specific advantages, and human judgement. Social factors (such as user profiles, sharing behaviours, engagement metrics, network structures, and crowd-sourced credibility signals) play a critical role in how misinformation is created, propagated, and perceived, yet these contextual nuances are often overlooked in algorithmic models. These social dimensions operate alongside technical elements, including linguistic cues, visual content, temporal dissemination patterns, and hybrid feature integration. Drawing on a review of the recent literature, this work synthesises sociotechnical elements to inform the development of an integrated approach to the detection of fake news. We introduce the SHAPE conceptual framework to guide the development of the Informed Fake News Advisor (IFNA). This conceptual framework will guide the creation of IFNA, which will consist of detection tools that combine technical precision with sensitivity to social context. By framing fake news detection as a sociotechnical problem, IFNA shifts the focus from isolated technical optimisation towards a holistic design philosophy, supporting the development of solutions that are both effective in detection and responsible in deployment within complex information ecosystems.

Keywords: Fake news, Fake news detection, Sociotechnical systems, Framework

1. Introduction

Misinformation can include a range of information reliability that moves from verified facts to distortion (e.g., exaggeration and selective use of evidence), or rumour (maybe true and maybe false), and finally to outright falsehoods (lies and fake news) (Hendricks and Vestergaard, 2018). Relative concepts related to fake news include satire, yellow journalism, junk news, pseudo-news, hoax news, propaganda news, advertorial, false information, fake information, misinformation, disinformation, mal-information, alternative fact, and post-truth (Wang, 2020).

Figure 1 shows the relationship between many terms relevant to fake news. Aïmeur et al. (2023) provide the following definitions:

- Misinformation is false information that is shared without the intention of misleading or causing harm,
- Disinformation is false information that is shared to intentionally mislead,
- Malinformation is genuine information that is shared with the intention of causing harm.

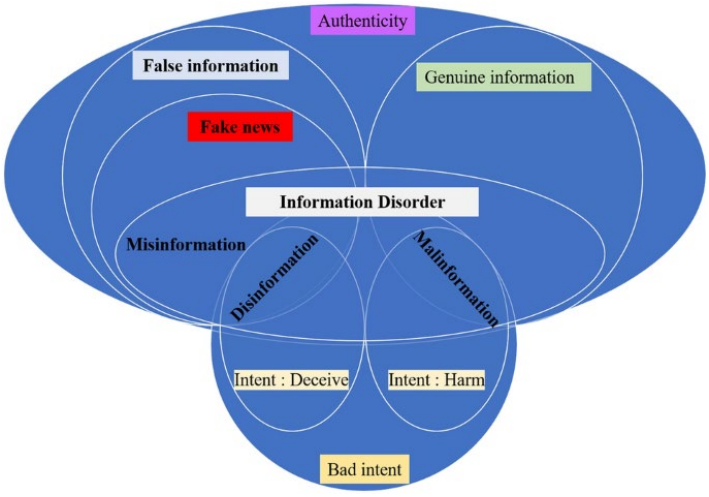

Figure 1: Relationship of Terms Related to Fake News (Aïmeur et al., 2023)

Stanford University defines fake news as news articles that are intentionally and verifiably false and could mislead readers (Allcott and Gentzkow, 2017).

Fake news entails the use of false stories and news that may mislead or lure readers for financial, political, or other purposes. Fake news can be harmful for many reasons (Shu et al., 2019):

- People are misled and accept false beliefs
- Changes the way people respond to true news
- Breaks the trustworthiness of the entire news ecosystem

Fake news has become a significant problem, as the intentional spread can lead to misleading information being mistaken for genuine news. Due to the widespread use of digital technologies, such as the Internet and social media, fake news can proliferate at a rapid rate. Fake news has become a significant problem, as the intentional spread of misinformation can lead to misleading information being mistaken for genuine news. Additionally, the use of digital technologies, such as the Internet and social media, can lead to the rapid proliferation of fake news.

Addictive news platforms lack editorial controls, allowing wide dissemination of misinformation via algorithmic news feeds, and without the necessary digital media literacy skills, people frequently fall victim to dubious claims they encounter in this context (llaire-Duquette et al., 2025). Fake news can mislead the public, create distorted perceptions, and negatively influence decision-making. Meanwhile, humans have been shown to be not proficient in differentiating between truth and falsehood when overloaded with deceptive information (Zhou et al., 2020). The sheer volume of information that users are overwhelmed with makes it difficult to detect fake news.

An additional significant challenge in these contemporary information ecosystems is the increasing use of social media to access news. This phenomenon is particularly acute in regions such as South Africa, where 73% of respondents report concern about their ability to discern true from false information online: a figure comparable to Nigeria and the United States, and substantially above the global average of 58% (SANEF, 2025). In the United States, 38% of adults regularly obtain news from Facebook and 35% from YouTube, underscoring the reliance on platforms that often lack robust editorial oversight (Pew Research Center, 2025). Meanwhile, in South Africa, 33% of users access news via TikTok and 42% via YouTube, among the highest rates on the continent (SANEF, 2025).

Fake news can also take many different forms, such as fake reviews, false political statements, and misleading advertisements. This affects the ways that people make decisions, can impact brand reputation and can also damage consumer trust. Fake news can sometimes be challenging to detect due to ambiguities in language and the deliberate crafting of content that mimics the tone and structure of genuine news. Detection algorithms can sometimes fail to detect these subtle nuances, including the use of satire, humour, and opinion writing. Such pieces do not adhere to journalistic practices, such as objective reporting. The manipulation of images and video clips can make fabricated stories appear more convincing. Clickbaiting techniques also draw the readers in with misleading and sensational headlines.

Contemporary detection systems face substantial constraints; technical challenges may impact the effectiveness of fake news detection systems. If algorithms are trained on datasets that are not large and diverse enough, bias can be introduced. The limited availability of high-quality training data for fake news detection is a big issue for training supervised learning models (Zhang and Ghorbani, 2020). This will impact the ability of these tools to detect fake news. Automated detection of fake news is simply not enough to flag content. Machine models can help identify patterns, but they can also overlook clever uses of language that aim to evade detection. Confirmation biases, echo chambers, and cognitive dissonance remain challenges, as fact-checking can still be rejected. New strategies for creating fake news will also emerge. Bots can be used to create fake news, which can then be disseminated. Online users may also not take accountability for content that is posted, shared and commented on. Inadvertently, they also contribute to the proliferation of the problem.

The acceptance and proliferation of fake news reflect complex interactions between social, psychological, and technical factors. Existing resources for protecting against misinformation demonstrate limited efficacy, depending on individuals' ability to discriminate, corroborate, and verify the authenticity of information. Technocentric approaches alone prove insufficient to address this multidimensional challenge; a more comprehensive framework that integrates human cognition, social dynamics, and technical capacity is necessary.

From a cybersecurity perspective, Caramancion et al. (2022) argue that disinformation, including fake news, can be framed as a cybersecurity threat as it exploits human cognitive vulnerabilities to compromise the integrity, availability, and trustworthiness of information, analogous to traditional cyberattacks on systems or networks. Recognising this as a cyber threat prompts the need for socio-technical defences addressing both human- and system-level vulnerabilities.

The contribution of this work is therefore not the introduction of a novel detection algorithm, but the development of a design-oriented framework that synthesises fragmented research across cognitive psychology, social network analysis, and computational fake news detection. By framing fake news detection as a sociotechnical design problem, this paper provides a theoretical foundation and research agenda for the development of more effective and responsible detection systems.

The rest of the paper is structured as follows: Section 2 reviews current approaches to fake news detection and proliferation, discussing cognitive theories, social-networks-based approaches, and technical/computational approaches. Section 3 proposes the SHAPE model, which helps guide the user in techniques for dealing with information processing. Section 4 provides a discussion, limitations, and future work. Section 5 concludes the paper.

2. Current Approaches to Fake News Detection and Proliferation

Fake news proliferation detection methods have been studied in various disciplines. Techniques range from cognitive theories and approaches (Section 2.1), social network-based approaches (Section 2.2), and technical or computational approaches (Section 2.3). Fake news detection methods can rely on a technocentric paradigm only, which has its limitations (Section 2.4).

2.1 Cognitive Theories and Approaches

Psychology plays a fundamental role in shaping how we process information based on the reinforcement of certain beliefs and our current emotional state. Cognitive biases, emotional responses, and motivational factors are deeply embedded in our psychological functioning. Forms of confirmation bias include (University of California Irvine, 2025)

- Familiarity Bias: The repetition of misinformation makes it feel more believable simply because people are exposed to it frequently, a phenomenon enhanced by social media's high-volume information flow.
- Availability Bias: People tend to believe information that is more easily recalled or accessible in memory, which often favours sensational or emotionally charged fake news stories.
- Confirmation bias: People may tend to strongly believe and share fake news when it supports their pre-existing beliefs

Biases are mental shortcuts that reduce the cognitive effort required to process information. However, it also makes us vulnerable to fake news. Biases, when combined with emotional responses, social pressures, and digital algorithms, can reinforce prior beliefs and predispose users to believe fake news due to the way content is presented. Our interpretation of fake news is influenced by our existing beliefs, emotional state, and social priorities, as our brains tend to opt for the fastest route to processing information without engaging in proper critical thinking and fact-checking processes.

When humans experience emotions such as fear, joy, anxiety, or anger, it can affect their ability to evaluate the credibility of information and share content without proper verification. Deeply rooted in our psyche is the need for gratification, autonomy, social acceptance, and altruism; Sharing content satisfies these needs, even if people are being deceived or sharing misleading information. This form of fast sharing for approval, inclusion, fulfilment, and entertainment has created a society where emotionally charged and sometimes biased or sensationalised content is spreading rapidly on various digital platforms.

Fake news can be used as a form of emotional manipulation to amplify feelings of anger, joy, anticipation, or excitement. The use of sensationalist or emotionally charged expressions/imagery can trigger emotional reactions that override rational and critical thinking. When users are in emotional states such as anger, happiness, excitement, or agitation, they may be more susceptible to believing fake news, as their judgement can be clouded by the intense emotional responses they are experiencing. A lack of education and scientific knowledge can influence some individuals to be more susceptible to fake news. Lower education and knowledge

levels are associated with increased engagement with fake news on social media, such as clicking links or sharing posts, thus helping to spread misinformation (Mihai-Ionuț and Irina, 2019).

Some of the cognitive factors that have influenced the spread of fake news include (Beauvais, 2022):

Confirmation bias, where individuals seek information that aligns with their existing beliefs,

Political partisanship affecting political beliefs,

Prior exposure and the illusory truth effect, whereby people believe a statement to be true as it is repeated frequently,

Type of reasoning: intuitive thinking vs. analytic thinking,

Lack of critical thinking,

Little time spent on news,

To be "in the know",

Low educational level,

Low science knowledge, and

Trust in "elite" messages.

Beauvais (2022) also mentions the following psychological factors that have led to the spread of fake news:

- Attraction to novelty and emotionally evocative content, often leading to impulsivity in sharing,
- High emotional states and low emotional intelligence can increase susceptibility,
- Anti-science attitudes also play a role.

Shane (2020) also discusses some of the key psychological reasons that influence our susceptibility to misinformation (and fake news).

- Cognitive miserliness: Humans prefer to use simpler and easier ways of solving problems that require the least mental effort. This is part of what makes our brains so efficient: we do not want to think very hard about everything. However, this also leads us not to put in enough effort into things that may require more effort, especially when we encounter news stories and fail to use critical thinking to first evaluate the content.
- Dual process theory: There are 2 basic ways of thinking. System 1 is an automatic process that requires little effort, and System 2 is an analytical process that requires more effort. Since we are cognitive misers, we tend to default to System 1 (easy form) thinking. Automatic processing is a simpler and faster approach; we tend to make easy judgements, but it can be inefficient, as we can miss crucial details.
- Heuristics are used as indicators for quick judgments instead of complex analysis. For example, relying on a social endorsement heuristic, such as a post on a social media site that appears to come from someone who can be trusted. However, the post could have been falsely created to impersonate the person and may not be a completely reliable indicator that the news story is true.
- Cognitive dissonance occurs when a person encounters information that contradicts beliefs, which may result in the rejection of credible information to alleviate the dissonance.
- Motivated reasoning: It is when people use reasoning skills to believe what they want to believe, instead of determining the truth. People's rational faculties can lead to misinformed beliefs, rather than lazy thinking.
- Pluralistic ignorance: a lack of understanding about what others in society think and believe. This can lead people to incorrectly believe that others share a particular political view when, in fact, it is held by a very small number of people. A related phenomenon is the false consensus effect, in which people overestimate the number of others who share their views.

Kahnemann (2011), in his renowned book "Thinking, Fast, and Slow", explains that the way to block errors that originate in System 1 is simple in principle: recognise the signs that you are in a cognitive minefield, slow down, and ask for reinforcement from System 2 (Kahnemann, 2011). Saltor et al. (2023) showed that reflective thinking styles and actively open-minded thinking predicted better fake news discrimination, with actively open-minded thinking being the strongest predictor. Saltor et al. (2023) explain that reflectiveness, that is, conscious and

deliberate thinking processes, can intervene in the fast default responses provided by intuitive thinking if prompted. The human brain is naturally inclined toward rapid, low-effort cognitive processing, which increases reliance on intuitive rather than analytical thinking. However, the key to preventing the propagation of fake news will be to slow down automatic thinking processes and apply more deliberate, critical thinking. This will be further discussed in the conceptual IFNA framework to help guide the user on principles for the prevention of fake news.

2.2 Social Network Approaches

Fake news is predominantly distributed using social media platforms or Online Social Networks (OSNs); this mode of news diffuses faster and more broadly than the truth (Vosoughi et al., 2018). Additionally, OSNs maximise engagement; this skews content visibility, favouring posts similar to previously engaged content while suppressing alternative perspectives. Critically, this approach has pointed to negative societal consequences, potentially reducing content diversity and amplifying divisive material (Bouchaud, 2024). The' unique characteristics of the platforms contribute to this issue: unlike traditional media, social media users do not choose their news sources, instead relying on proprietary algorithms that provide targeted, often non-transparent information (Shu et al., 2017).

Within these OSNs, connections on social media form rich, interconnected graph structures. These graphs can represent users or content as nodes, connected by edges (links) that represent various relationships, such as content similarity, reposts, friendships, etc. (Lakzaei et al., 2024). Utilising these graphs, Social Network Analysis (SNA) approaches can utilise the structural dimensions of these sharing networks, behaviour patterns, and interactions, to provide additional context to solely content-based methods (as discussed in Section 3.3). Furthermore, SNA can be applied at varying network levels (node-level, ego-level, triad-level, community, or whole network), allowing rich information and analysis opportunities (Zhou and Zafarani, 2019).

2.2.1 Graph-based detection methods

Graph-based detection uses relationship modelling between news articles, users, content, and external sources as graphs. This includes techniques such as Propagation-Based Methods (Section 2.2.1.1), Knowledge-Driven Methods (Section 2.2.1.2), Social Context and Heterogeneous Graphs (Section 2.2.1.3) and Multi-Modal and Evidence-Aware Graphs (Section 2.2.1.4).

- Propagation-Based Methods

Broadly, these deep-learning methods utilise Graph Neural Networks (GNNs) to consider how fake and real news propagate differently. Fake news spreads through social networks differently in comparison to other forms of content. Using graph structures, such as network centrality (the degree to which a node is central to a network), it is possible to identify indicators such as: (a) rapid dissemination, (b) clustering, and (c) influential nodes that are associated with fake news (Sivasankari and Vadivu, 2022).

GNNs have proven particularly effective in capturing propagation patterns, frequently outperforming text-only baselines, especially in early detection scenarios and in multilingual contexts where language-agnostic approaches are required (Han et al., 2021; Harby and Zulkernine, 2023). The temporal and dynamic nature of information diffusion has motivated the development of temporal-delay GCNs and dynamic graph snapshots, which capture the evolving topology of network structures over time (Song et al., 2022, 2021). These temporal extensions address limitations of static graph approaches: these systems can model how information flows evolve as fake and real news traverse networks at different velocities and scales. Various GNNs have been developed, including Graph Convolution Neural Network (BiGCN-A, BiGCN-B), Graph Attention Neural Network (BiGAT), GraphSage (BiSAGE), Graph Convolution with ARMA filters (BiARMA), and Simplified Graph Convolution Neural Network (BiSGCN) (Harby and Zulkernine, 2023), with varying levels of accuracy.

- Knowledge-driven Methods

Knowledge-driven Methods construct knowledge graphs by extracting entities and relationships from news content and cross-referencing them against external databases such as Wikidata. This method frames fake news detection as a fact-checking problem, wherein claims extracted from news articles are compared against verified factual assertions using graph matching or embedding-based techniques (KEHGNN-FD, DEAP-FAKED, MiLk-FD) (Mayank et al., 2022; Phan et al., 2023; Xie et al., 2024). Heterogeneous knowledge graphs (integrating diverse entities, such as news articles, semantic entities, topics, and/or factual claims) can provide rich representations in comparison to homogeneous graphs. Heterogeneity can introduce nuanced reasoning about the relationships

between claims and external knowledge and has resulted in improved detection precision in the literature (Xie et al., 2024).

- Social Context and Heterogeneous Graphs

Beyond propagation patterns and factual verification, heterogeneous graph methods model complex interactions among users, news content, comments, and topics as integrated graph structures. Representative approaches include DAGA-NN (Domain-Adversarial Graph Attention Neural Networks) and Graph Attention Networks (GATs) (Chang et al., 2024; Yuan et al., 2021). These approaches capture multiple dimensions of credibility assessment simultaneously: user reputation and historical reliability, community structure and clustering patterns, and cross-domain relationships that may indicate coordinated inauthentic behaviour (Phan et al., 2023). Graph Attention mechanisms and domain-adversarial networks enable these methods to weight the significance of different node and edge types, allowing the model to learn which social signals are most indicative of fake news (Yuan et al., 2021).

- Multi-Modal Graphs

Contemporary detection methodologies integrate multiple modalities: text, images, and external evidence into unified graph structures. Multi-modal graph approaches, such as TEMGNNs (Temporal Enhanced Multi-modal Graph Neural Networks), MEFaND, and SAFE (Social Attention Fusion Engine), use GNNs to integrate features across modalities and explicitly model semantic relationships between textual claims and visual content (Chang et al., 2024; Sormeily et al., 2024; Yuan et al., 2021). These methods address the credibility that manipulated images and videos lend to false narratives: these systems can better detect incongruities between claims and corroborating evidence that might be missed by unimodal approaches.

2.3 Technical Approaches

Fake news detection via computational means has garnered much attention in the literature. Considering the influx of information disseminated online, detecting fake news at scale computationally could address the limitations posed by manual fact-checking: a practice that is seemingly becoming impossible given the volume of digital content (Shah et al., 2024). Computational approaches that use machine learning and natural language processing (NLP) (Section 3.3.1), deep learning (Section 3.3.2), transformer models (Section 3.3.3), or ensemble systems (Section 3.3.4) to detect fake news with promising accuracy.

2.3.1 Machine learning and natural language processing

Machine learning techniques aim to automatically detect and classify fake news by analysing textual patterns and linguistic features. Traditional machine learning models such as logistic regression, decision trees, support vector machines (SVM), random forests, and Naive Bayes are commonly used for the detection of fake news. These models typically rely on features extracted from news content using NLP techniques such as tokenisation, lemmatisation, part-of-speech tagging, and term frequency-inverse document frequency (TF-IDF) (Prachi et al., 2022).

2.3.2 Transformer-based models

Transformer-based architectures have achieved state-of-the-art performance in fake news detection tasks, largely displacing earlier deep learning approaches. The foundational Bidirectional Encoder Representations from Transformers (BERT) model and its variants, including Robustly Optimised BERT Pre-training Approach (RoBERTa) and XLNet, have the capacity to capture semantic and contextual information via bidirectional attention mechanisms. These models benefit substantially from transfer learning, whereby pre-trained representations developed on large text corpora are fine-tuned on fake news detection tasks, reducing the requirement for task-specific training data (Aljawarneh and Swedat, 2024; Kaliyar et al., 2021; Raza and Ding, 2022).

- Large Language Models (LLMs)

LLMs have shown promising results in fake news detection through three primary mechanisms: (i) fake text classification, (ii) fact-checking and verification, and (iii) contextual analysis (Papageorgiou et al., 2024). Of particular interest is the ability to decompose claims, retrieve relevant evidence, and integrate sources, to minimise extensive human annotation. Techniques such as stepwise prompting and weak supervision have been shown to enhance accuracy and reduce hallucinations, previously constraining these methods (Kuntur et al., 2025). However, limitations are still present, as various fine-tuned BERTs still outperform LLMs in detection

accuracy, though LLMs are still used as "advisors" to provide multi-perspective instructive rationales (Hu et al., 2024).

2.3.3 Deep learning

Deep learning architectures have demonstrated substantially superior performance compared to traditional machine learning methods, particularly when applied to complex and large text corpora. Convolutional Neural Networks (CNNs) excel at capturing local patterns and n-gram features in textual data through hierarchical feature extraction. By applying convolutional filters to text sequences, CNNs identify frequently occurring lexical patterns characteristic of fake news (Kumar et al., 2020). Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks address a fundamental limitation of CNNs by explicitly modelling sequential dependencies and contextual information within news articles. Hybrid models that combine CNN and RNN, such as CNN-LSTM, RNN-CNN, and Bidirectional Gated Recurrent Unit-Bidirectional LSTM architectures (Bi-GRU-Bi-LSTM), thus incorporate both local and sequential information, often achieving substantially higher accuracy than unimodal approaches (Kumar et al., 2020; Nasir et al., 2021).

2.3.4 Ensemble systems

Ensemble learning methodologies combine predictions from multiple models to achieve improved robustness and accuracy, mitigating the individual weaknesses inherent in single-model approaches. Multimodal Ensembles represent a particularly sophisticated development, combining textual, visual, and social context features through integrated transformer-based and deep learning models (Ahmad et al., 2020; Ali et al., 2022; Hakak et al., 2021).

2.4 Current Limitations of Technocentric Approaches

Multiple elements limit the sole reliance on purely algorithmic or technocentric methods for the detection of fake news. The changing nature of misinformation tactics creates an adversarial challenge that static models cannot resolve, necessitating continuous retraining and deployment cycles to maintain performance as bad actors adapt to detection methods (Vishnupriya et al., 2024).

Furthermore, the lack of high-quality, diverse and community-driven datasets is a major barrier. Further compounding this is the dependence of the supervised learning technique on labelled data, which restricts scalability and adaptability. In low-resource settings or when considering data sets in different languages, this poses significant challenges (Klein and Wueller, 2017). Fundamental barriers can still constrain widespread applicability: this includes data bias in datasets, small or imbalanced datasets leading to under- or over-fitting concerns and limited empirically validated use-cases (Hamed et al., 2023).

3. SHAPE: The Conceptual Framework for IFNA

A more hybrid model is needed that combines technical methods of source verification and fact-checking with human oversight. AI tools can be used to identify potential fake news, but humans themselves will have influence and judgement. This section provides a more detailed discussion of combined cognitive, social network-based, and technical approaches for detecting fake news.

Previous work includes the IFAKT Framework, which proposes a fake news detection framework (Veerasamy et al., 2024). The recommended guidelines in the IFAKT framework include considering the following issues: cultivating awareness around the source by checking credibility and investigating unknown sources, and deconstructing information by analysing headlines, verifying facts, and identifying logical fallacies. Increase media literacy through cross-referencing and fact-checking, and practice critical thinking skills to minimise confirmation bias and foster lateral thinking. Lastly, building healthy habits through open discussion and being cautious, while remaining continually aware.

The BBC has also adopted the SIFT method to spot misinformation (Ruggeri, 2024). SIFT stands for

- STOP: do not automatically share the post, comment
- Investigate the source: Find out where the post came from. Conduct a web search, as the platform search results can be misleading and may appear to be from a reputable website. Investigate whether the source is a reputable media outlet. For sources involving individuals, the user must verify whether the individual is an expert in the subject. Determine the purpose of content stemming from businesses. Check if there are underlying objectives, such as sales, advertising, substantiating a donor, or political endorsement.

- Find Better Coverage: If, after checking the source, there are still some doubts, it may be necessary to do further analysis by using fact-checking services.
- Trace back to its original context: Even if a claim has been reported on a credit media outlet, it may not necessarily be original or may be taken out of context. Many people may make claims on social media, and posts can go viral. It is important to determine the original source of the claim.

To synthesise the various cognitive, social network, and technical elements of misinformation detection, a conceptual framework called SHAPE is proposed. The framework will lead to the future development of an Information Fake News Advisor (IFNA), which will consist of technical tools to assist users in identifying fake news.

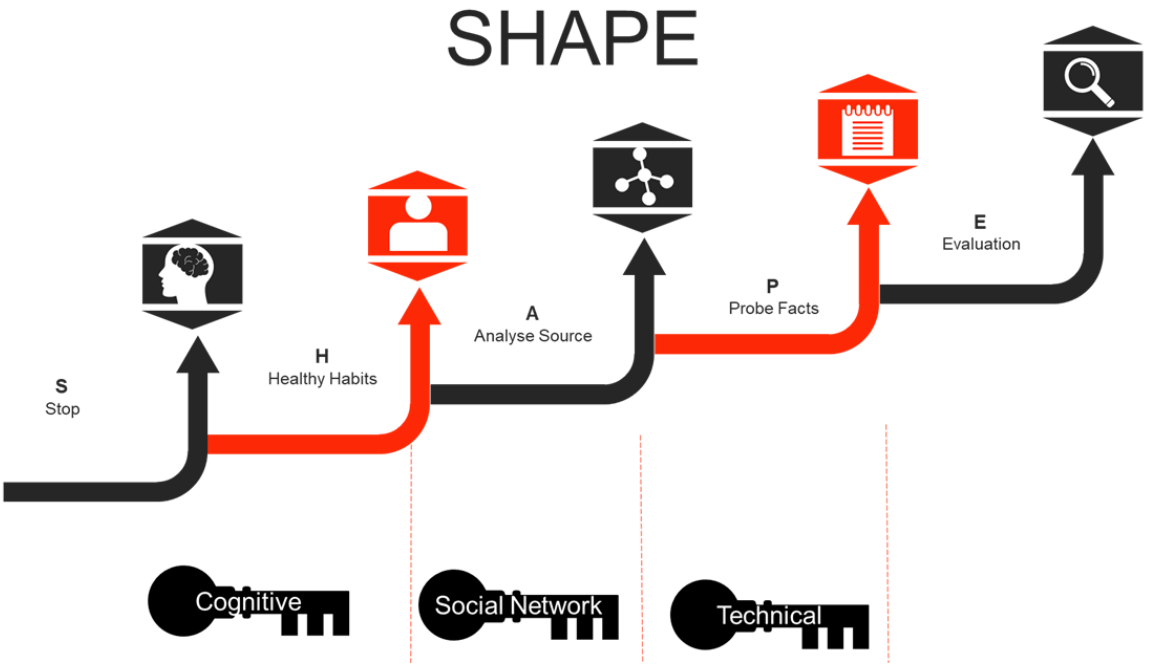

Figure 2: Conceptual Framework for IFNA

SHAPE (see Figure 2) is defined as:

S: Stop. Stop automatic processing but instead apply more lateral thinking and cognitive effort. Avoid the mental shortcuts of bias (familiarity, availability and confirmation bias). Users need to apply more analytical and critical thinking by using various socio-technical approaches. The aim is to think slower and not faster to actually analyse the content with various other cognitive processes, as well as social networks and technical techniques and tools.

H: Healthy Habits. Develop healthier habits to assess content and emotional responses. It is vital that users take a step back before sharing or reacting to determine whether the information is just trying to trigger a strong emotional response. As users utilise more socio-technical methods and tools to assess content, over time more healthy habits for information sharing can be developed. Users need to keep updated on media literacy and cybersecurity skills to learn about the spreading of misinformation and the latest cyber threats that may pose as real information.

A: Analyse source: Various platforms can be used to share and transfer information. These can include mainstream media, such as television, radio, and print, as well as digital social media platforms like Instagram, TikTok, Facebook, and Twitter/X. Users should verify the source of the content – is it a reputable site or author that can be trusted for fact-based information? Established news outlets, academic institutions and government agencies can be more trustworthy than random posts, headlines, adverts, reviews or claims. Cross-checking against other independent sources can also help verify the accuracy of stories.

Social-network detection approaches analyse sharing and behaviour patterns to identify certain indicators. Studying social-network detection approaches can help provide users with additional context, such as identifying influential nodes that spread the content, as well as insights into clustering and rapid dissemination. By

employing social-network detection techniques such as propagation-pattern analysis, network-based user credibility assessment, hybrid network content integration, and platform considerations, users can gain insight into how certain posts are favoured and others are suppressed. This can help users better assess the nature of the content.

P: Probe facts: The use of fact-checking websites can assist in verifying stories and debunking false claims. Perform a deeper headline analysis. Consider whether sensationalistic tactics have been applied with excessive emotion and punctuation. Users should reflect on whether the content has logical fallacies-is the story logical or feasible in the relevant context. Technical and computational methods can be used for the detection of fake news through the analysis of textual patterns and linguistic analysis.

E: Evaluate. Fundamentally, humans have at their disposal a set of socio-technical approaches as well as their human intuition to evaluate and make a judgement call. The key aspects will be to ask logical questions, examine the source, and determine whether the information is exaggerated and unlikely. Users can utilise technical techniques to assess the content, consult fact-checking platforms, and leverage tools and apps to help flag potential fake news. Overall, users need to develop stronger practices to cultivate a more discerning eye when faced with the flood of online content. There is no silver bullet for fake news detection, but with practice, discerning practices, and more critical thinking, users can develop better responses to the inundation of fake news.

4. Discussion

This study proposes sociotechnical integration as a necessity for scalable fake news detection. Technocentric approaches, while advancing detection accuracy, inevitably encounter limitations rooted in their methodologies: machine learning systems cannot anticipate novel adversarial strategies, graph-based methods depend on rich network data that are often unavailable on emerging platforms, and transformer models remain insensitive to the emotional and social contexts that motivate sharing behaviour. In contrast, psychological or cognitive approaches lack the scalability and precision necessary to detect false content at the current pace of dissemination.

The proposed SHAPE framework thus argues for the integration of technical tools and human judgement as complementary components within a sociotechnical system. When considering limitations, the framework remains conceptual and requires empirical validation through user studies and iterative refinement. Moreover, this work adopts a largely individualistic framing, emphasising user-level cognition; deeper systemic interventions addressing platform architectural choices, which are deliberately optimised for engagement rather than veracity, remain underexplored. Future work could involve empirical validation within the broader IFNA toolkit through controlled user studies. As a conceptual framework, SHAPE requires empirical validation. Future work will involve qualitative user studies and quantitative assessment of IFNA-assisted decision-making, measuring improvements in fake news discrimination, confidence, and response behaviour relative to unaided judgement.

5. Conclusion

Fake news detection remains a multidimensional challenge. It is not an issue that can be addressed solely by technical or social approaches. This paper argues that current algorithmic methods, whilst increasingly sophisticated, cannot yet adequately address misinformation in isolation from human cognition. The SHAPE framework bridges this gap by integrating cognitive theories, social network analysis, and computational detection into a coherent sociotechnical approach, guiding users to modulate their information processing through deliberative thinking, awareness of emotional manipulation, source evaluation, evidence verification, and critical judgement. The empirical development and iterative refinement of IFNA will require collaboration between disciplines and communities. Ultimately, addressing fake news effectively demands not a single solution, but a collaboration of disciplines and techniques.

Ethics Declaration: No ethical approval was required for this literature study.

AI Usage Declaration: Generative artificial intelligence (AI), specifically large language models (LLMs), was used in a limited capacity to help articulate some of the authors' ideas, concepts and explanations. No tools were used to write original content or generate *de novo* text. AI tools used to aid the writing process include ChatGPT and Perplexity.

References

Ahmad, I., Yousaf, M., Yousaf, S., Ahmad, M.O., 2020. Fake News Detection Using Machine Learning Ensemble Methods. Complexity 2020.

Aïmeur, E., Amri, S., Brassard, G., 2023. Fake news, disinformation and misinformation in social media: a review. Soc. Netw. Anal. Min. 13.

Ali, A.M., Ghaleb, F.A., Al-Rimy, B.A.S., Alsolami, F.J., Khan, A.I., 2022. Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique. Sensors 22.

Aljawarneh, S.A., Swedat, S.A., 2024. Fake News Detection Using Enhanced BERT. IEEE Trans. Comput. Soc. Syst. 11.

Allcott, H., Gentzkow, M., 2017. Social media and fake news in the 2016 election. Journal of Economic Perspectives.

Beauvais, C., 2022. Fake news: Why do we believe it? Joint Bone Spine.

Bouchaud, P., 2024. Skewed perspectives: examining the influence of engagement maximization on content diversity in social media feeds. J. Comput. Soc. Sci. 7.

Caramancion, K.M., Li, Y., Dubois, E., Jung, E.S., 2022. The Missing Case of Disinformation from the Cybersecurity Risk Continuum: A Comparative Assessment of Disinformation with Other Cyber Threats. Data (Basel). 7.

Chang, Q., Li, X., Duan, Z., 2024. Graph global attention network with memory: A deep learning approach for fake news detection. Neural Networks 172.

Hakak, S., Alazab, M., Khan, S., Gadekallu, T.R., Maddikunta, P.K.R., Khan, W.Z., 2021. An ensemble machine learning approach through effective feature extraction to classify fake news. Future Generation Computer Systems 117.

Hamed, S.K., Ab Aziz, M.J., Yaakub, M.R., 2023. A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion. Heliyon.

Han, Y., Karunasekera, S., Leckie, C., 2021. Graph Neural Networks with Continual Learning for Fake News Detection from Social Media. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12892 LNCS.

Harby, A.A., Zulkernine, F., 2023. A Comparative Analysis of Graph Neural Networks for Fake News Detection. In: Proceedings - International Computer Software and Applications Conference.

Hendricks, V.F., Vestergaard, M., 2018. Reality lost: Markets of attention, misinformation and manipulation, Reality Lost: Markets of Attention, Misinformation and Manipulation.

Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., Qi, P., 2024. Bad Actor, Good Advisor: Exploring the Role of Large Language Models in Fake News Detection. In: Proceedings of the AAAI Conference on Artificial Intelligence.

Kahnemann, Daniel., 2011. Thinking, fast and slow. Farrar, Straus and Giroux.

Kaliyar, R.K., Goswami, A., Narang, P., 2021. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimed. Tools Appl. 80.

Klein, D.O., Wueller, J.R., 2017. Fake news: A Legal Perspective. Journal of Internet Law 20.

Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., Akbar, M., 2020. Fake news detection using deep learning models: A novel approach. Transactions on Emerging Telecommunications Technologies 31.

Kuntur, S., Wróblewska, A., Paprzycki, M., Ganzha, M., 2025. Under the Influence: A Survey of Large Language Models in Fake News Detection. IEEE Transactions on Artificial Intelligence 6.

Lakzaei, B., Haghir Chehreghani, M., Bagheri, A., 2024. Disinformation detection using graph neural networks: a survey. Artif. Intell. Rev. 57.

llaire-Duquette, G., Hasni, A., Drouin, J.N., Groleau, A., Mahhou, A., Legault, A., Khayat, A., Carignan, M.E., Ayotte-Beaudet, J.P., 2025. Primary school pupils' ability to detect fake science news following a news media literacy intervention: Exploration of their success rate, evaluation strategies, self-efficacy beliefs, and views of science news. Journal of Digital Educational Technology 5, 2506.

Mayank, M., Sharma, S., Sharma, R., 2022. DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection. In: Proceedings of the 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2022.

Mihai-Ionuț, P., Irina, E.N.E., 2019. Influence of the educational level on the spreading of Fake News regarding the energy field in the online environment. In: Proceedings of the International Conference on Business Excellence. Sciendo, pp. 1108–1117.

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.

Papageorgiou, E., Chronis, C., Varlamis, I., Himeur, Y., 2024. A Survey on the Use of Large Language Models (LLMs) in Fake News. Future Internet.

Pew Research Center, 2025. Social Media and News Fact Sheet [WWW Document]. URL https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/ (accessed 11.6.25).

Phan, H.T., Nguyen, N.T., Hwang, D., 2023. Fake news detection: A survey of graph neural network methods. Appl. Soft Comput.

Prachi, N.N., Habibullah, M., Rafi, M.E.H., Alam, E., Khan, R., 2022. Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms. Journal of Advances in Information Technology 13.

Raza, S., Ding, C., 2022. Fake news detection based on news content and social contexts: a transformer-based approach. Int. J. Data Sci. Anal. 13.

Ruggeri, A., 2024. The "Sift" strategy: A four-step method for spotting misinformation [WWW Document]. BBC. URL https://www.bbc.com/future/article/20240509-the-sift-strategy-a-four-step-method-for-spotting-misinformation (accessed 10.28.25).

Saltor, J., Barberia, I., Rodríguez-Ferreiro, J., 2023. Thinking disposition, thinking style, and susceptibility to causal illusion predict fake news discriminability. Appl. Cogn. Psychol. 37.

SANEF, 2025. South Africa's Digital News Landscape in 2025 [WWW Document]. South African National Editors' Forum. URL https://sanef.org.za/south-africas-digital-news-landscape-in-2025/ (accessed 11.6.25).

Shah, N., Mitalia, N., Dhanare, R., 2024. Enhancing the Precision and Efficiency of Fake News Prediction: A Comprehensive Review. In: 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL). pp. 335–341.

Shane, Tommy., 2020. The psychology of misinformation: Why we're vulnerable [WWW Document]. First Draft News.

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, 22–36.

Shu, K., Wang, S., Liu, H., 2019. Beyond news contents: The role of social context for fake news detection. In: WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining.

Sivasankari, S., Vadivu, G., 2022. Tracing the fake news propagation path using social network analysis. Soft comput. 26.

Song, C., Shu, K., Wu, B., 2021. Temporally evolving graph neural network for fake news detection. Inf. Process. Manag. 58.

Song, C., Teng, Y., Zhu, Y., Wei, S., Wu, B., 2022. Dynamic graph neural network for fake news detection. Neurocomputing 505.

Sormeily, A., Dadkhah, S., Zhang, X., Ghorbani, A.A., 2024. MEFaND: A Multimodel Framework for Early Fake News Detection. IEEE Trans. Comput. Soc. Syst. 11.

University of California Irvine, 2025. Misinformation - Get the Facts [WWW Document]. https://guides.lib.uci.edu/Misinfo/bias.

Veerasamy, Namosha., Khan, Zubeida., Badenhorst, D., 2024. Examining Misinformation and Deep Fakes . In: Proceedings of the 25th European Conference on Knowledge Management. Academic Conferences International Limited, pp. 870–877.

Vishnupriya, G., Jeriel, A., Anish, R., Ajay, G., Giftson, A.J., 2024. Combating Fake News in the Digital Age: A Review of AI-Based Approaches. In: 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). IEEE, Pune, India.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science (1979). 359.

Wang, C.C., 2020. Fake news and related concepts: Definitions and recent research development. Contemporary Management Research 16.

Xie, B., Ma, X., Wu, J., Yang, J., Fan, H., 2024. Knowledge Graph Enhanced Heterogeneous Graph Neural Network for Fake News Detection. IEEE Transactions on Consumer Electronics 70.

Yuan, H., Zheng, J., Ye, Q., Qian, Y., Zhang, Y., 2021. Improving fake news detection with domain-adversarial and graph-attention neural network. Decis. Support Syst. 151.

Zhang, X., Ghorbani, A.A., 2020. An overview of online fake news: Characterization, detection, and discussion. Inf. Process. Manag. 57.

Zhou, X., Jain, A., Phoha, V. V., Zafarani, R., 2020. Fake News Early Detection. Digital Threats: Research and Practice 1.

Zhou, X., Zafarani, R., 2019. Network-based Fake News Detection. ACM SIGKDD Explorations Newsletter 21.