International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

A Review on AI-Generated Fake News Detection using Advanced Machine and Deep Learning Techniques

Sunaina Bairagi
M. Tech Scholar
Department of Information Technology
Mahakal Institute of Technology
Ujjain, India

Shivani Sharma
Assistant Professor
Department of Information Technology
Mahakal Institute of Technology
Ujjain, India

Abstract* - This review paper seeks to deliver an exhaustive examination of the identification of AI-generated misinformation, emphasising the aims, methodologies, and results pertinent to contemporary machine learning (ML) and deep learning (DL) techniques. The main goal is to find out how advanced computational models can find fake news stories made by advanced AI systems, like huge language models and multimodal generators, which often copy human writing and multimedia quite well. The methodology entails a comprehensive analysis of both conventional machine learning techniques, including Logistic Regression, Support Vector Machines, Random Forests, and ensemble methods, as well as deep learning architectures such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and transformer-based models like BERT, RoBERTa, and GPT. We also look at hybrid methods that mix feature engineering with deep contextual embeddings, as well as adversarial training strategies, multimodal analysis, and ways to link these with knowledge graphs and fact-checking systems. The results show that DL models, especially transformer-based and hybrid frameworks, are more accurate, robust, and adaptable than traditional ML methods. This makes it possible to find small linguistic, semantic, and multimodal discrepancies in AI-generated content. New trends like real-time detection, explainable AI, and policy-driven interventions make detection even better and help build confidence and transparency in society. This review underscores the essential function of machine learning (ML) and deep learning (DL) in countering AI-generated misinformation, stressing the necessity for ongoing research, dataset curation, and the amalgamation of technological, regulatory, and educational initiatives to maintain information integrity in the digital age.

Keywords- AI-generated news, Fake news detection, Machine learning, Deep learning, Natural language processing (NLP), Transformer models, Recurrent neural networks (RNNs), Long short-term memory (LSTM).

I. INTRODUCTION

Artificial intelligence (AI) has come a long way in a short amount of time, and it has changed the way content is made. Machines can now write text that looks very much like human writing, which has led to new uses and big problems, especially with the spread of fake news made by AI. These made-up stories can change people's minds, affect elections, and cause societal turmoil, so it's really important to find them quickly. Conventional methods like manual fact-checking, rule-based systems, or keyword matching are not enough to deal with these complex AI-generated texts. This is because they often have coherent language, contextually relevant arguments, and stylistic fluency that make them look like real journalism. To find them effectively, you need to use advanced machine learning (ML) and deep learning (DL) models. Machine learning techniques offer a fundamental structure for categorising news articles by discerning patterns from annotated datasets comprising genuine and fabricated news

Some examples of supervised algorithms are Logistic Regression, Random Forests, Gradient Boosting, and Support Vector Machines (SVM). These algorithms look at linguistic, syntactic, and semantic features like n-grams, term frequency-inverse document frequency (TF-IDF), part-of-speech tags, sentiment polarity, readability scores, and metadata characteristics like source credibility and publishing patterns. Ensemble approaches make things even more stable by combining predictions from several classifiers. This lowers the number of false positives and makes it easier to generalise across different news domains [1]. Deep learning models provide an enhanced analytical layer, as neural networks may autonomously identify intricate contextual and semantic patterns frequently overlooked by conventional machine learning methods. Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Gated Recurrent Units (GRUs) are good at modelling the flow of context and sequential dependencies in text. This lets them find small inconsistencies, logical contradictions, or strange language patterns that might mean the content was made up. BERT, RoBERTa, XLNet, and GPT-based detectors are all transformer-based models that use attention mechanisms to figure out how important each token is in a sentence. This helps them find writing style, coherence, and semantic relationships that are out of the ordinary and may point to AI-generated misinformation. Fine-tuning these models on news datasets that are specialised to a certain field helps them keep up with new AI writing styles and themes, which improves their ability to find things. Hybrid methods

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

that combine machine learning (ML) and deep learning (DL) techniques have shown a lot of promise. They use both human-engineered features like syntactic irregularities, semantic deviations, and sentiment anomalies, as well as deep contextual embeddings from neural networks. This makes for a complete detection system that benefits from both human insights and the ability of deep learning to learn features automatically. Adversarial training techniques, where models are trained on a variety of AI-generated fake news samples, make them more resistant to new ways of making content and less likely to be attacked in new ways. Even with these improvements, there are still problems, such as the lack of high-quality labelled datasets, problems with model interpretability, and the fact that AI text generators are changing quickly to get around detection systems. Researchers are still looking into ways to make systems more adaptable, resilient, and efficient, such as contrastive learning, few-shot learning, multimodal analysis that combines text with images or metadata, and reinforcement learning-based detection. In conclusion, using advanced machine learning and deep learning models to find fake news made by AI is a key and scalable way to keep information honest, lower the risks of false information in society, and build trust in digital communication. This shows how important it is to keep coming up with new ideas for model development, dataset curation, and evaluation methods to keep up with AI-generated content that is getting more and more complex.

II. LITERATURE REVIEW

Iasulaitis 2025 et al. The propagation of fake news has become easier due to widespread use of social media and messaging. It also happened at this election. Since data is continuously produced, manual verification is labor-intensive. One way to tackle this problem is creating a model that can automatically detect misinformation. This work proposes a first fact-checked dataset of Brazilian fake news. FactPolCheckBr is fully in Brazilian Portuguese and comprises 1,873 fake news of the 2022 presidential elections. News articles were scraped from the web, clustered around topics and tagged with candidate bias through a political content analysis. The dataset will investigate patterns of misinformation and machine learning-based fake news detection [2].

Alghamdi 2024 et al. Social media's cross-linguistic dissemination of fake news constitutes a serious societal risk in many low-resource languages such as Swahili and Indonesian. Detection techniques now suffer from shortcomings, particularly as the pre-trained language model (PLM) mBERT cannot extend sequence lengths and their training was noisy. The proposed multilingual fake news detection (MFND) framework uses extractive and abstractive summarization to shorten the news articles while preserving relevant information for their respective classifications. The content is classified through mBERT. Testing on a multilingual dataset showed superior performance over the state of the art. This demonstrates that eliminating redundant content improves accuracy and efficiency, allowing for faster and sturdier fake news detection [3].

Akhtar 2023 et al. Distribution of fake news and misinformation is a serious concern in social media and other online platforms. Fake News has disrupted lives largely and influenced perception and decision-making.

While it is important, especially in political and economic contexts, FNaD which stands for Fake News and Disinformation (FNaD) detection model videos which are based on AI and ML but not much research has been done on this area. Also, it is important for limiting supply chain disruptions (SCDs). By using various AI and ML techniques, this study builds a robust FNaD detection model. The model relies on case studies from Indonesia, Malaysia, and Pakistan and more. The model showed practical utility in endorsing managerial decisions, serving as a valuable addition to both supply chain and AI-ML research, suggesting implementable decisions and future research directions [4].

Verma 2023 et al. Online social media platforms are cheap, easily accessible and fast way to communicating which also helps in the easy spread of low-quality fake news. This misinformation presents a sizable threat to society's various stakeholders by undermining trust in media and confusing audiences. This paper addresses the challenge by proposing a novel framework, Message Credibility MCred to detect fake news using local and global text semantics and languages. The framework integrates BERT to capture global semantic relationships among words in sentences and CNNs to fetch local N-gram features. Results on a widely-used Kaggle benchmark reveal that MCred improves detection accuracy by 1.10% over the state-of-the-art models. This illustrates the power of utilizing local and global textual information [5].

Sastrawan 2022 et al. Fake news refers to false information that is intentionally shared for a specific purpose. Its unchecked spread can have grave political and social consequences. To lessen this, many studies focus on detecting fake news by using smart computation. In this study, the authors employed deep learning methods using various architectures such as CNN, Bidirectional LSTM, and ResNet in conjunction with pre-trained word embedding patterns. Four different datasets were used, each undergoing data augmentation via back-translation to collect balanced classes. The results of the experiments indicate that the Bidirectional LSTM works better than CNN and ResNet for all datasets. Hence, the Bidirectional LSTM is capable to model sequentially and contextually more effectively than CNN and ResNet [6].

TABLE 1: LITERATURE REVIEW

Authors/Year	Method	Research gap	Findings
Salamanos/2022[7]	Meta-graph with relational, semantic, topical features; graph neural networks	Traditional cascade-level classification ignores user relationships and content semantics	Meta-graph improves detection accuracy 3–4%, additional 1% with semantics

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Xu/2022[8]	Graph-based semantic structure mining; neighborhood propagation; graph structure learning	Sequential models fail long-distance claim-evidence integration; include redundant information	GET outperforms state-of-the-art models in comprehensive fake news detection
Deepak/2021[9]	Comprehensive review of deep learning models, datasets, evaluation metrics, literature	Need deeper understanding of spread, detection, and information ecosystem impact	Identifies current progress, datasets, and promising research directions for detection
Oliva/2021[10]	Experiments with simple and complex datasets using NLP and ML models	Dataset construction significantly impacts performance; model complexity often misaligned	Appropriate dataset-model alignment critical; simpler models suffice for simple datasets
Juan/2020[11]	Review of visual features, detection methods, multimedia fake news challenges	Limited understanding of visual content's role in fake news detection	Visual content enhances detection; understanding features aids multimedia fake news detection

III. EMERGENCE OF AI-GENERATED CONTENT

Due to improvements in natural language generation, the rise of generative models and the widespread availability of AI tools, the emergence of AI-generated content has become so popular. Generative frameworks mimic the style, tone, and target audience of established journalists. This approach enables the rapid production of news content through large language models like GPT and T5. Cloud platforms and APIs offer wide accessibility which enables both legitimate and illegitimate uses [12]. Content created by artificial intelligence also involves audio, video, images and deepfakes to boost credibility. This explosion affects the way we consume news, shape opinions and amplifies misinformation and highlights the need for detection systems, media education and regulations.

A. Advancements in Natural Language Generation (NLG)

The fast-paced advancement of natural language processing (NLP) technologies have been a significant contributor to the emergence of AI-generated content. Text generated by modern large (or transformer-based) language models (LLMs) such as (GPT, T5, etc.) resembles human-like text which closely mirrors natural syntax, grammar and semantic flow. Essentially, the type of tools that can write a coherent article, a good summary, or a catchy social media post with a little human prompt, have made it increasingly difficult to spot. These models can maintain their context across sentences; this explains their ability to write convincing stories or highly readable news pieces. It can even be leveraged to produce misinformation at scale [13]. As a result, advancements in NLG have led to the efficiency and volume of content creation undertaken at scale. At the same time, it is triggering unprecedented challenges as regards upholding information integrity in general and fake news in digital media ecosystems.

B. Rise of Generative Models

Generative AI models such as GPT, BERT, T5 that rely on transformer-based frameworks have been able to generate any complete content like an article, report or social media post. Models are trained on huge datasets to mimic language patterns, contextual relationships or domain-specific knowledge so that realistic content can be produced. In the context of news, generative models can imitate the writing style of a journalist, replicate the format of a fact-based news event, and adapt to a target audience thereby optimizing the persuasive compliance of the AI-based news [14], [15]. The emergence of such models has sped up content generation, which allows for the rapid sharing of information at levels that are hard to check physically. In spite of the opportunities these technologies provide for automated summarization, translation and personalization, they also involve risks. For example, adversaries may use generative models for disinformation, opinion manipulation and synthetic news campaigns, complicating traditional fact-checking and verification.

C. Accessibility of AI Tools

AI-generated content is on the rise due to the availability of AI tools for everyone. Thanks to cloud-based APIs, open-source platforms, and ease of use, almost everyone can create realistic text, images, and multimedia without much effort. This equality of access to AI tech lowers the entry barriers making it easier for several users to create content. Even though an easy approach helps legitimate applications like automated journalism, summarization of content, and educational tools, it also helps create deception. These tools could be misused by malicious actors to create fake news and manipulate narratives or impersonate credible sources easily [16], [17]. Due to the availability of AI-powered content generation, there is an increase in synthetic news, making its detection difficult and thereby underscoring the need for robust verification to ensure the integrity of digital media.

D. Integration with Multimedia Content

Increasingly technologies of Artificial Intelligence are producing believable images, videos, audios along with text that will lend credence to misinformation. Generative adversarial networks, or GANs, deepfake technologies and multimodal AI models can produce realistic images and sounds that may empower AI text and thus, misinformation. This union gives the fake news the appearance of authenticity when spurious visuals or audio can back up the misleading narrative, making the audience more likely to trust it [18]. For instance, by using AI, you can get a video of a personality who is making false statements or get some images for fake news. The challenge for detection systems is that they are required to work on multiple content modalities simultaneously to capture alterations involved in a digital image-video and text generation. This would help avoid the spread of sophisticated AI-enabled misinformation.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

E. Impact on News Consumption

A sum of the overall increased amount of news consumer through the AI-generated content, the speed at which they consume news increased as it gave one version and also the credibility of news reports. One of the simplest uses of Ai and machine learning are content generation. Ai models can not only create content at scale, but they can also Create fake news and decimate any credible report as the audience is exposed to unlimited news. The increased use of social media changes the public's opinion and also shape narratives across platforms. In addition, news produced by AI gets used for audiences in mind that employ persuasive language and style to capture and enhance credibility. This causes consumers to do so unknowingly depending on fake or biased information which would make them make incorrect decisions [19]. The impact of scam AI content must be clearly defined to take necessary action. It's essential to find the right balance while ensuring that the technology is not impacted by any over-regulation.

IV. IMPORTANCE OF DETECTING AI-GENERATED FAKE NEWS

The fake news generated by the AI tools threatens the society. It may harm the integrity of information, manipulate public opinion and harm the economy. Machine learning and deep learning techniques can detect fake news and rumors by analyzing the textual patterns, semantics, sentiment, and dissemination behaviors of such information. These systems can stop misinformation from spreading which preserves credible information. It can also prevent social and political machinations, financial disruption, effective crisis response, and media reliability. Ensuring quick detection allows accurate information to reach the users, builds confidence in digital channels, and empowers informed choices, shielding persons, communities, and institutions from the repercussions of AI false news.

A. Preserving Information Integrity

Essential to a well-informed society is information integrity. The integrity of news dissemination is collateral damage due to AI-generated fake news, which creates content that closely resembles legitimate journalism, and misleads people in realising the truth. When people end up believing fake news, they do not get the chance to know the real story. When you detect AI-generated fake news, it ensures that the facts remain accessible to all and the manipulation can only go so far [20]. Machine learning and deep learning techniques detect inconsistencies in texts by analyzing linguistic patterns and semantics in context (to establish if they are fabrications). Detection systems maintain information integrity on online platforms. This upholds the credibility of the platforms while strengthening trust in digital content. Furthermore, it mitigates the risk of common misinformation becoming widespread. Such integrity is necessary in an age of misinformation.

B. Preventing Social Manipulation

Fake news, above all that produced using artificial intelligence, can be used strategically to influence social behaviour, change public opinion or interfere in the electoral and other political processes. When community leaders present fabricated narratives or events as real, they create discontent, discord and destabilized peace in communities. It is critical to curbing the impact of AI-generated fake news on society. Models of machine learning and deep learning can analyze large amounts of online content. This process helps to identify patterns that are indicative of misinformation campaigns, misinformation content or AI-generated content. Detection systems minimize the risk of harmful narratives gaining popularity, thereby limiting the scope of public opinion manipulation. Preventing social manipulation protects the individual autonomous decision-making but also preserves the integrity of public reasoning and conversation during critical moments. These become possible when the information is accurate, truthful or verifiable and not manufactured.

C. Mitigating Economic Impact

Fake news generated through artificial intelligence can have many economic consequences when false information is circulated against companies, markets or financial policies. Fake news about expenses or another company could cause fluctuations in share prices, investor panic, and reputational damage that creates real economic harm. For instance, false news regarding a company's financial condition can cause sudden stock sell-offs or affect confidence. By detecting false information using learning algorithms and deep learning models, we can avert this economic collapse due to news and happenings [21]. These models examine the circumstances surrounding abnormal dissemination patterns of textual sentiment and other metadata to pinpoint likely fabrications. Detection systems help in minimizing the costs related to the economic impact. Consequently, investors, businesses, and the economy as a whole are protected from clickbait. This will help ensure greater accuracy of value-based market decisions not driven by AI-generated or fraudulent storylines.

D. Ensuring Public Safety

AI-generated fake news can directly threaten public safety and exacerbate the consequences of natural disasters, pandemics, and social crises. If someone shares incorrect or misleading information about safety measures, medical advice, or emergency responses it can create mass panic. Furthermore, it may also cause unsafe behavior or delayed access to accurate information and resources. Detection of such content is essential to reduce damage and help reliable information reach the audience. Advanced detection models can quickly identify false or misleading information circulating online, giving authorities, health agencies and the media a head start. During a crisis like pandemic the unverified treatments fairly declared as false news can cause death if believed [22]. Using a machine learning and deep learning system for AI-generated misinformation detection will help to prevent any public harm, better crisis management and informed decision making among citizens when information is time-sensitive and accurate.

E. Maintaining Trust in Media

The credibility of media depends on the trustworthy nature of the content. Fake news generated by AI is ruining trust by making it harder for people to distinguish fake news from true

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

news. Trust is lost once scepticism sets in, people begin to disengage, and misinformation spreads further.

Fig. 1 Maintaining Trust in Media

In order to restore the credibility of media organisations, one must be able to detect AI fake news. Machine learning and deep learning algorithms scan language patterns, validate facts, and examine content consistency to identify false stories. These systems keep people from being over-exposed to made-up information that makes them trust digital and traditional news sources [23]. As important as individual knowledge is for society, it is citizen's trust in the credibility of media that has the power to create stable societies. Trustworthy news helps citizens actively participate in society.

V. CHALLENGES IN FAKE NEWS DETECTION

Currently available models of artificial intelligence do generate text similar to that of human beings. Therefore it is very difficult for both human beings and tweaked algorithms to identify fake news. A simple modification of language, context and sentiment can easily evade detection. Moreover, due to your social media messaging apps news sites many reach millions of users within minutes, it is therefore challenging to contain in real time fake news. As more and more multimodal content is created, it will be difficult to detect fakes. Everyday malicious actors adjust their behaviours to avoid detection by algorithms while noticeable but as all this is important, there is still the challenge of ensuring effective detection of the malicious actions with respect to privacy, data security, and ethically deploying AI [24].

A. Sophistication of AI-Generated Content

Recent artificial intelligence models, especially advanced natural language generation systems and large language models, are so sophisticated that the text it generates is almost indistinguishable from that written by a human [25]. Thanks to these models, coherent narratives, relevancy across sentences, and writing styles can all be generated. Hence, it becomes extremely hard for humans or traditional rule-based algorithms to identify fake news. The use of manipulative wording and tone plays an important role in how facts are presented. As a result, these sophisticated AI-generated fake news are often not detected by simple detection systems that look for keywords, shallow patterns, or simple heuristics.

B. Rapid Spread Across Multiple Platforms

The amazing velocity with which fake news is disseminated over social media, chat applications, online news platforms is phenomenal and may get to millions in minutes. Due to the rapid proliferation of a fact or false claim, it becomes difficult to detect and respond to it. Misinformation becomes widespread before traditional verification methods or manual fact-checking processes can contain it, as these are usually too slow [26]. Thus, detection systems have to cope with the challenge of processing a huge volume of content in real time and with accuracy. This clearly points to the need for automated, scalable and adaptive methods to tackle the infectious spread of disinformation.

C. Multimodal Content Complexity

Fake news is no longer limited to mere text, but now also images, videos, audio including deepfake. This is done to make these ads more convincing. The fact that the content is multimodal makes detection difficult as the system needs to look at different content all at once. Spotting inconsistencies or manipulations in videos, images, audio, and their corresponding textual cues relies on powerful algorithms with understanding of semantic, visual, and auditory patterns. Accomplishing such sophisticated multimodal analysis is computationally intensive and technologically challenging. It needs large processing power, appropriate annotated dataset and integrated AI models capable of efficiently and scalably identifying falsehoods in various formats [27].

D. Evasion Tactics by Malicious Actors

Creators of fake news constantly evolve their strategies to Misinformation is becoming harder to detect as false information tools evolve. They may change phraseology, use synonym replacements, reorganize the sentence structure, or modify the metadata and data type to get past the algorithm. There are those who use AI content-generating tools to pass off their posts as legit, and others manipulate the posting schedule to slip under the radar [28]. This ongoing adaptation spurs a never-ending arms race wherein contemporary detection systems are invariably learning and updating to counter new evasion tactics. This enables automated detection tools to function against fresh disinformation campaigns.

E. Ethical and Privacy Concerns

Detecting fake news usually demands investigation of user behavior and content and interaction patterns in the social network. The involvement of sensitive information may create serious user privacy and data security issues as well, in this process. To discourage the use of negative or undesirable behavior researchers and developers must ensure basic principles of ethical AI-is the AI systems are transparent, fair, and accountable, and that they minimize the risk of misuse or unintended bias. Finding an equilibrium between precise, real-time detection and safeguarding individual rights poses a challenge. It requires the implementation of privacy-preserving techniques, secure data protocols, and ethical AI governance frameworks.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

IV. Machine Learning Approaches for Fake News Detection

Machine learning techniques have been used to detect fake news using textual and metadata features. Models that are supervised such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, and Random Forest classify the news articles using labelled datasets of real and fake. Misinformation detection is improved through the extraction of several lexical, syntactic, semantic, sentiment and metadata features [29]. Ensemble techniques use multiple classifiers to improve accuracy and robustness, while hybrid approaches marry handcrafted features with neural network embeddings for deeper context analysis. Fake news generation by AI is continuously evolving. So models are trained on this type of adversarial training to make them robust.

A. Supervised Models

Fake news detection relies on supervised machine learning models that learn through labeled datasets of real and fake news articles. These models can then classify news articles as authentic or fabricated. Algorithms commonly used include Logistic Regression, Support vector machines (SVM), Decision Trees and Random Forests. Logistic Regression uses the features to evaluate the probability of a news article being a fake. On the other hand, SVM separates the data points in a feature space in such a way that the classes can be distinguished effectively [30]. Decision Trees and Random Forests recursively partition the data features, building a tree-like model to make predictions. These models allow for a structured, interpretable framework to automate news classification.

B. Feature Engineering

Feature engineering is essential for transforming fake news data into useful and analyzable information in machine learning settings. The discovery of patterns that are unusual or repetitious can be aided by these lexical features. The sentence structure and part-of-speech tags reveal inconsistencies in writing [31]. Semantic features include word embeddings and topic distributions, which capture meaning and context. Meanwhile, sentiment analysis detects unusually exaggerated emotions or tone anomalies common in fake news. Various metadata details which might include source reliability, publication time, and author patterns offer more insights. Components developed with great caution permit ML models to analyse true and false stories.

C. Ensemble Methods

Fake News detection ensemble approaches combine the predictions of multiple classifiers to improve performance and overcome the limitation of individual model. Methods like bagging, boosting, and stacking combine the strengths of multiple algorithms to improve the overall performance of a model [32]. Random Forest employs bagging with multiple decision trees while Gradient Boosting focuses on correcting the error of weak learners. The variations in writing styles, domains and topics in news articles can be dealt with ensembles. When different models are integrated together, the misclassification gets reduced. There is an improved robustness. Thus, your misclassification rate is less with ensemble methods. These methods are also good for fake news detection.

D. Hybrid Approaches

To improve the performance in detecting fake news hybrid approaches combine traditional feature engineering with deep learning embeddings. The implicit understanding afforded by deep neural networks through embeddings permits a more refined and high-level understanding of a text's context than the explicit signals that one may attain through the use of handcrafted grammatical, lexical, syntactic and sentiment features. For instance, the word or sentence embedding of Word2Vec, GloVe, BERT, etc., captures semantic relations that are difficult to encode manually [33]. The blending of these two methods results in hybrid models that derive their insights from humans and automatically through deep learning. Combining these two approaches ameliorates detection of AI-generated and other highly sophisticated fake news, improving accuracy and generalizability.

E. Adversarial Training

Adversarial training is a technique that strives to improve the robustness of fake news detection models to new content. Models are trained not only on regular real and fake news datasets but also on adversarially generated samples that look like human-generated AI news and other cartels. When the model faces difficulty during training, it learns to detect small inconsistencies and deceitful behaviour - things that normal classifiers can't pick up on. Adversarial training of ML models allows the models to adapt to novel styles, emerging AI-generated content, and deliberately misleading techniques. Also, enhancing its long-term effectiveness.

VII. Deep Learning Models for Fake News Detection

Deep learning techniques involving complex textual and multimodal features are widely used for fake news detection. RNNs can process sequential data. They have the ability to capture dependencies across sentences that allow them to detect subtle linguistic cues. Long Short-Term Memory Networks (LSTMs), a type of RNN, can deal with long-term dependencies of longer texts. Thus, this ultimately allows for improved contextual understanding. CNNs distinguish the local pattern of words or images that imply misleading information. Transformer-based models such as BERT and GPT use attention techniques to recognise the global context and semantic inconsistencies [34]. Incorporating CNNs with RNNs/LSTMs and transformers, hybrid models improve detection accuracy and robustness across content type.

A. Recurrent Neural Networks (RNNs)

RNNs are deep learning models that enable the analysis of any kind of sequential data. As a result, RNNs are able to look at the flow and structure of words in news articles and social media posts. Unlike the traditional neural networks, the RNN architectures maintain a kind of memory that allows them to use the information in the previous inputs. RNNs can identify patterns, context cues, and subtle language glitches that may indicate a text is fake or misleading. Fake news detection

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

tasks can leverage RNNs (recurrent neural networks) because they can identify inconsistencies, repeats or unnatural phrases.

B. Long Short-Term Memory Networks (LSTMs)

Long Short-Term Memory networks (LSTMs) are a special kind of Recurrent Neural Network (RNN) designed to overcome the limitations of RNNs, particularly the difficulty to learn long-term dependencies. In regular RNNs, they lose focus on meaningful context from past inputs in the sequence. This happens as they go deeper inside the text resulting in limited effectiveness on longer texts like articles and social media posts. LSTMs overcome this limitation through their unique memory cell structure which is able to remember useful information and forget unhelpful ones to maintain context over longer sequences [35]. LSTMs manage distant words conveniently. Therefore advantageous in fake news detection on long texts, as relevant cues or inconsistencies may lie far apart in the content. LSTMs improve the accuracy and robustness of fake news detection systems, enabling them to preserve essential linguistic and contextual relationships while eliminating noisy input, making them indispensable in modern natural language processing applications.

C. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of deep learning model originally created for image recognition but has been successfully adapted for text and other types of data. CNNs can strongly detect patterns, phrases, sentence structure, and other elements in input data such as news posts, and visual elements, on social media, etc. In the fake news detection area, CNNs analyze text, detecting repetitions, strange word associations, or unique styles to determine if something is being fabricated or not. Furthermore, CNNs can be used to analyze the images or multimedia associated with news articles in order to flag manipulations or inconsistencies that contribute to misinformation [36]. By automatically identifying these localized features, CNNs lessen the importance of manual creation of features and make it easy to scale in large data sets. With their prowess in detecting text and image anomalies, CNNs have become an essential part of contemporary fake news detectors.

D. Transformer-Based Models (BERT, RoBERTa, GPT)

Models built on transformer architectures, including BERT, RoBERTa, and GPT, have transformed natural language processing by enabling the effective use of global context through attention mechanisms. Unlike traditional sequential models, transformers analyze the relationships that each word in a text has with all other words simultaneously. This gives it the ability to look for contradictions and disagreements, as well as subtle manipulation that may be embedded across sentences and paragraphs. BERT, GPT models can procured large-scale corpora and fine-tuned on datasets with domains which enables high precision to identify fake or misleading news [37]. Transformers do a good job of managing long documents as they retain context over long sequences. Due to the incorporation of semantic understanding and sophisticated pattern identification, the transformer-based framework presents a robust and scalable option for the accurate and reliable automated detection of fake news. It also proves to outperform many conventional deep learning set-ups in terms of metrics.

E. Hybrid Deep Learning Models

The incorporation of hybrid deep learning models involves combining different architectures like Convolution Neural Network (CNN) and Recurrent Neural Network (RNN) or Long Short-Term Memory (LSTM) based and transformer-based models to create robust systems to detect fake news. Through the blend of the abilities of CNNs to capture local patterns, RNNs/LSTMs to do contextual modeling (sequential dependencies), and transformers to do global attention, hybrid models can analyze both fine-grained and high-level features across text and multimodal [38]. This enables them to perceive slight nuances in the language, inconsistencies that appear later on and also tampering of the images, videos, audio provided with the text. Hybrid techniques take care of the limitations of a single model architecture which enhances the detection performance of the model along with adaptability to different types of data. In particular, they can counter complex fake news generated by artificial intelligence which may involve text and other media. Thus, hybrid deep learning models can offer a solution for fake news detection that is comprehensive, scalable and highly accurate.

VIII. EMERGING TRENDS AND FUTURE DIRECTIONS

A new artificial intelligence (AI) system that can detect anything from text to images and videos comprising deep fakes is being developed as part of a multimodal AI research project in Australia. The real-time detection and automation can allow misinformation to be identified quickly, while adversarial training makes the model robust against evasive content. Merging detection with fact-checking and knowledge graphs enhances cross-lingual verification accuracy and platform diversity [39]. Also, regulations, policy, and media literacy initiatives help ethical frameworks and inform the public to spot false information. As a whole, these approaches represent a comprehensive, adaptive, and reliable strategy to fight fakes news.

A. Multimodal Detection Systems

As it becomes tougher to identify whether components of content were created with artificial intelligence, the need for multimodal detectors has arisen. A multimodal detector can analyze text, image, video, and audio simultaneously. Multimodal detection systems utilize various architectures from deep learning. The deep learning architectures CNN, RNN/LSTM, transformer are applied on images/text sequences/context of text, respectively, to detect fraudulent content. This method enables detection of complex AI-generated disinformation, like deepfake videos or misleading photos in news reporting [40]. As these systems can build associations between text and media, they can recognize instances of discord when one would not. Future study will aim for improved computational efficiency, multimodal large-scale dataset handling, and feature extraction across

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

modalities for galleries. The multimodal detection is a big trend in the battle against the AI-generated news. It is going to provide a reliable, accurate, and scalable solution reflecting the nature of misinformation.

B. Explainable AI (XAI) in Detection

Explainable AI (XAI) contributes to AI-generated news identification. It is important to ensure that the model explains itself for trust and accountability, in case a model identifies something as fake. Deep learning models, such as transformers and hybrid networks, operate as "black boxes". They provide high accuracy but little interpretability. XAI approaches highlight the important features, phrases and visual cues impacting a classification in order to enhance decision-making transparency [41]. The reliability of model findings can also boost the confidence of users to engage these tools for automated detection. Techniques like attention visualization, SHAP, and LIME provide interpretable insights into complex AI models. XAI will become a focus in the near future. Researchers plan to enhance the transparency of these models. They will also reduce bias and help modulate their explanations.

C. Real-Time Detection and Automation

With misinformation travelling at terrific speeds through social media, messaging apps and online news portals, real-time detection and automation are becoming vital against the blight of AI-created misinformation. In modern detection systems, it has become increasingly obvious that processing streaming data is more efficient than storing it, as alerts have to be raised and suspicious items identified before they reach the masses [42]. Through NLP and deep learning combined with network analysis, fake news propagation is detected promptly using automated pipelines. Real-time detection also uses adaptive models to track AI generation techniques, language changes and visual manipulations as they develop. The large scale monitoring across different platforms now made possible due to cloud computing, edge devices and scalable architectures. Future plans call for measures including reducing latency, improving computational efficiency, and building self-learning models that automatically adapt to sophisticated misinformation strategies, allowing timely intervention to mitigate the societal negative effects of AI-driven fake news.

D. Adversarial Training and Robust Models

As content generated with the help of AI becomes more advanced, detection systems face adversarial traffic that is purposely created to avoid detection by automated systems. Adversarial training is a prevailing method to improve robustness. It entails training the model with adversarially crafted or manipulated data to enhance resilience to AI-generated deceptive news. Methods for improving misinformation detection systems involve generating synthetic adversaries which can either be adversarial examples or generative models that simulate possible misinformation patterns [43]. The system is better able to detect subtle manipulation of text, images, and multimodal content. Powerful models also use hybrid structures and ensemble learning to lower false negatives and increase overall accuracy. In future research endeavors, the emphasis will be on merging adversarial training with real-time detection. This will guarantee that our models will continue to work well against AI-generated fake news campaigns. These campaigns will be continuously evolving. We will also create frameworks that can enable security against new attack strategies. At the same time, we will ensure interpretability and scalability.

E. Integration with Fact-Checking and Knowledge Graphs

Merging AI detection techniques with automated fact-checking and knowledge graph offers a hopeful technique for verifying the authenticity of news. Knowledge graphs organize information into a semantic network of entities, facts and relationships, enabling detection systems to spot logical mistakes, contradictions and unverifiable claims. These integrated systems refer to trusted databases to assess news content, helping to flag false information and misleading claims and prove it to the user. Fact checking eliminates the sole reliance on linguistic and stylistic properties. This enhances the accuracy of detecting the output of an artificial intelligence model (AIM) as well as its misinformation [44]. Future directions include building dynamic knowledge graphs that update automatically in real-time, enhancing automated reasoning, and developing multilingual capabilities to detect misinformation on a wide range of platforms, ultimately enabling a more comprehensive and effective approach to combatting AI-generated fake news.

F. Policy, Regulation, and Media Literacy

To properly tackle AI generated content technology, regulation and society must be engaged. Policies and regulations of content generation can formulate the standards for transparency, ethical deployment of AI, and accountability for the same, as per the culture. Governments, media organizations and technology companies are collaborating to implement frameworks to ensure AI technology-use is responsible and control the spread of fake news. Programs that develop media literacy seek to make people better able to determine what is true and false. The future directions advocates for international cooperation, regulation enforcement and synergy between educational initiatives and detection technologies to create an ecosystem that enables innovation without harm and inspires confidence in digital information.

IX. CONCLUSION

To wrap it up, utilizing advanced machine learning (ML) and deep learning (DL) models for detection of AI generated fake news is essential for maintaining credibility of information and preventing any societal, political and economic risks. The advancements in AI content creation, such as large language models and multimodal models, has made fake news very realistic and hard to identify. To complement traditional fact checking and rule-based methods, ML algorithms such as Logistic Regression, SVM, Random Forests and ensemble methods can be deployed that investigate linguistic, syntactic, and semantic and metadata features to detect inconsistencies. Figures like RNN, LSTM, CNN, and transformer-based BERT and GPT are deep learning architectures that can successfully enhance detection by picking up long-term dependencies, local patterns, and global contextual relationships in text and multi-modal contents. By mixing handmade features with neural embeddings, hybrid models enhance robustness. Further, adversarial training allows systems to get ready for the growing development of AI-generated misinformation. The latest trends which include multimodal detection, real-time automation, explainable AI, knowledge graph integration, fact-checking frameworks, and policies, regulations, and media literacy initiatives offer a comprehensive, adaptive, and scalable approach against disinformation. A continuous innovation of ML and DL processes is essential despite context-specific limitations like data scarcity, interpretability and rapid dynamic content. These strategies represent a robust approach for identifying, mitigating and preventing AI-generated fake news, ensuring credibility, trustworthiness and alignment to community needs in the information process.

REFERENCES

[1] S. Kuntur, A. Wróblewska, M. Paprzycki, and M. Ganzha, "Fake News Detection: It's All in the Data!," pp. 1–12, 2025, [Online]. Available: http://arxiv.org/abs/2407.02122

[2] S. Iasulaitis *et al.*, "FactPolCheckBr: a dataset of fake news fact-checked during the 2022 Brazilian presidential elections," *Lat. Am. Data Sci.*, vol. 5, no. 1, pp. 28–37, 2025, doi: 10.53805/lads.v5i1.76.

[3] J. Alghamdi, Y. Lin, and S. Luo, "Fake news detection in low-resource languages: A novel hybrid summarization approach," *Knowledge-Based Syst.*, vol. 296, no. March, p. 111884, 2024, doi: 10.1016/j.knosys.2024.111884.

[4] P. Akhtar, A. Mujahid, G. Haseeb, and U. Rehman, "Detecting fake news and disinformation using artificial.pdf," pp. 633–657, 2023.

[5] P. K. Verma, P. Agrawal, V. Madaan, and R. Prodan, "MCred: multi-modal message credibility for fake news detection using BERT and CNN," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 8, pp. 10617–10629, 2023, doi: 10.1007/s12652-022-04338-2.

[6] I. K. Sastrawan, I. P. A. Bayupati, and D. M. S. Arsa, "Detection of fake news using deep learning CNN–RNN based methods," *ICT Express*, vol. 8, no. 3, pp. 396–408, 2022, doi: 10.1016/j.icte.2021.10.003.

[7] M. Paraschiv, N. Salamanos, C. Iordanou, N. Laoutaris, and M. Sirivianos, "A Unified Graph-Based Approach to Disinformation Detection Using Contextual and Semantic Relations," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 16, no. Icwsm, pp. 747–758, 2022, doi: 10.1609/icwsm.v16i1.19331.

[8] W. Xu, J. Wu, Q. Liu, S. Wu, and L. Wang, "Evidence-aware Fake News Detection with Graph Neural Networks," *WWW 2022 - Proc. ACM Web Conf. 2022*, vol. 2022, no. April 2022, pp. 2501–2510, 2022, doi: 10.1145/3485447.3512122.

[9] D. P, T. Chakraborty, C. Long, and S. K. G, "Deep Learning for Fake News Detection," *Inf. Retr. Ser.*, vol. 42, no. 2, pp. 71–100, 2021, doi: 10.1007/978-3-030-62696-9_4.

[10] C. Oliva *et al.*, "Fake news detection: When complex problems demand complex solutions," 2021.

[11] J. Cao, P. Qi, Q. Sheng, T. Yang, J. Guo, and J. Li, "Exploring the Role of Visual Content in Fake News Detection," pp. 141–161, 2020, doi: 10.1007/978-3-030-42699-6_8.

[12] V. Nair, D. J. Pareek, and S. Bhatt, "A Knowledge-Based Deep Learning Approach for Automatic Fake News Detection using BERT on Twitter," *Procedia Comput. Sci.*, vol. 235, no. 2023, pp. 1870–1882, 2024, doi: 10.1016/j.procs.2024.04.178.

[13] L. Xiao, Q. Zhang, C. Shi, S. Wang, U. Naseem, and L. Hu, "MSynFD: Multi-hop Syntax Aware Fake News Detection," *WWW 2024 - Proc. ACM Web Conf.*, no. February, pp. 4128–4137, 2024, doi: 10.1145/3589334.3645468.

[14] X. Liu *et al.*, "FKA-Owl: Advancing Multimodal Fake News Detection through Knowledge-Augmented LVLMs," *MM 2024 - Proc. 32nd ACM Int. Conf. Multimed.*, no. October 2024, pp. 10154–10163, 2024, doi: 10.1145/3664647.3681089.

[15] S. Abdali, S. Shaham, and B. Krishnamachari, "Multi-modal Misinformation Detection: Approaches, Challenges and Opportunities," *ACM Comput. Surv.*, vol. 57, no. 3, 2024, doi: 10.1145/3697349.

[16] M. Z. Nawaz, M. S. Nawaz, P. Fournier-Viger, and Y. He, "Analysis and Classification of Fake News Using Sequential Pattern Mining," *Big Data Min. Anal.*, vol. 7, no. 3, pp. 942–963, 2024, doi: 10.26599/BDMA.2024.9020015.

[17] M. A. B. Al-Tarawneh, O. Al-irr, K. S. Al-Maaitah, H. Kanj, and W. H. F. Aly, "Enhancing Fake News Detection with Word Embedding: A Machine Learning and Deep Learning Approach," *Computers*, vol. 13, no. 9, pp. 1–25, 2024, doi: 10.3390/computers13090239.

[18] S. Kuntur, M. Krzywda, A. Wróblewska, M. Paprzycki, and M. Ganzha, "Comparative Analysis of Graph Neural Networks and Transformers for Robust Fake News Detection: A Verification and Reimplementation Study," *Electron.*, vol. 13, no. 23, pp. 1–17, 2024, doi: 10.3390/electronics13234784.

[19] P. Dhiman, A. Kaur, D. Gupta, S. Juneja, A. Nauman, and G. Muhammad, "GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection," *Heliyon*, vol. 10, no. 16, p. e35865, 2024, doi: 10.1016/j.heliyon.2024.e35865.

[20] J. Alghamdi, S. Luo, and Y. Lin, "A comprehensive survey on machine learning approaches for fake news detection," *Multimed. Tools Appl.*, vol. 83, no. 17, pp. 51009–51067, 2024, doi: 10.1007/s11042-023-17470-8.

[21] X. Yu, S. N. Luo, Y. Wu, Z. Cai, T. W. Kuan, and S. P. Tseng, "Research on a Capsule Network Text Classification Method with a Self-Attention Mechanism," *Symmetry (Basel).*, vol. 16, no. 5, 2024, doi: 10.3390/sym16050517.

[22] A. Onan, "Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 7, 2023, doi: 10.1016/j.jksuci.2023.101610.

[23] S. Alyoubi, M. Kalkatawi, and F. Abukhodair, "The Detection of Fake News in Arabic Tweets Using Deep Learning," *Appl. Sci.*, vol. 13, no. 14, 2023, doi: 10.3390/app13148209.

[24] H. Yang *et al.*, "A Multi-Layer Feature Fusion Model Based on Convolution and Attention Mechanisms for Text Classification," *Appl. Sci.*, vol. 13, no. 14, 2023, doi: 10.3390/app13148550.

[25] P. Dhiman, A. Kaur, C. Iwendi, and S. K. Mohan, "A Scientometric Analysis of Deep Learning Approaches for Detecting Fake News," *Electron.*, vol. 12, no. 4, pp. 1–31, 2023, doi: 10.3390/electronics12040948.

[26] H. Padalko, V. Chomko, and D. Chumachenko, "A novel approach to fake news classification using LSTM-based deep learning models," *Front. Big Data*, vol. 6, 2023, doi: 10.3389/fdata.2023.1320800.

[27] N. Põldvere, Z. Uddin, and A. Thomas, "The PolitiFact-Oslo Corpus: A New Dataset for Fake News Analysis and Detection," *Inf.*, vol. 14, no. 12, 2023, doi: 10.3390/info14120627.

[28] C. O. Truică and E. S. Apostol, "It's All in the Embedding! Fake News Detection Using Document Embeddings," *Mathematics*, vol. 11, no. 3, 2023, doi: 10.3390/math11030508.

[29] S. K. Hamed, M. J. Ab Aziz, and M. R. Yaakub, "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*, vol. 9, no. 10, p. e20382, 2023, doi: 10.1016/j.heliyon.2023.e20382.

[30] D. K. Sharma and S. Garg, "IFND: a benchmark dataset for fake news detection," *Complex Intell. Syst.*, vol. 9, no. 3, pp. 2843–2863, 2023, doi: 10.1007/s40747-021-00552-1.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

[31] E. Essa, K. Omar, and A. Alqahtani, “Fake news detection based on a hybrid BERT and LightGBM models,” *Complex Intell. Syst.*, vol. 9, no. 6, pp. 6581–6592, 2023, doi: 10.1007/s40747-023-01098-0.

[32] E. Rahmanian, “Fake news: a classification proposal and a future research agenda,” *Spanish J. Mark. - ESIC*, vol. 27, no. 1, pp. 60–78, 2023, doi: 10.1108/SJME-09-2021-0170.

[33] A. K. R. Nadikattu, “Machine Learning vs Deep Learning Models for Detecting Fake News: A Comparative Analysis on Fake-NewsNet Dataset,” *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4382241.

[34] B. Koloski, T. Stepišnik Perdih, M. Robnik-Šikonja, S. Pollak, and B. Škrlj, “Knowledge graph informed fake news classification via heterogeneous representation ensembles,” *Neurocomputing*, vol. 496, pp. 208–226, 2022, doi: 10.1016/j.neucom.2022.01.096.

[35] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, “Fake News Classification using transformer based enhanced LSTM and BERT,” *Int. J. Cogn. Comput. Eng.*, vol. 3, no. March, pp. 98–105, 2022, doi: 10.1016/j.ijcce.2022.03.003.

[36] F. Sakketou, J. Plepi, R. Cervero, H. J. Geiss, P. Rosso, and L. Flek, “FACTOID: A New Dataset for Identifying Misinformation Spreaders and Political Bias,” *2022 Lang. Resour. Eval. Conf. Lr. 2022*, pp. 3231–3241, 2022.

[37] D. Rohera *et al.*, “A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects,” *IEEE Access*, vol. 10, pp. 30367–30394, 2022, doi: 10.1109/ACCESS.2022.3159651.

[38] C. M. Lai, M. H. Chen, E. Kristiani, V. K. Verma, and C. T. Yang, “Fake News Classification Based on Content Level Features,” *Appl. Sci.*, vol. 12, no. 3, pp. 1–21, 2022, doi: 10.3390/app12031116.

[39] K. M. Fouad, S. F. Sabbeh, and W. Medhat, “Arabic fake news detection using deep learning,” *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 3647–3665, 2022, doi: 10.32604/cmc.2022.021449.

[40] E. Amer, K. S. Kwak, and S. El-Sappagh, “Context-Based Fake News Detection Model Relying on Deep Learning Models,” *Electron.*, vol. 11, no. 8, pp. 1–13, 2022, doi: 10.3390/electronics11081255.

[41] J. Alghamdi, Y. Lin, and S. Luo, “A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection,” *Inf.*, vol. 13, no. 12, pp. 1–28, 2022, doi: 10.3390/info13120576.

[42] S. Pandey, S. Prabhakaran, N. V. S. Reddy, and D. Acharya, “Fake News Detection from Online media using Machine learning Classifiers,” *J. Phys. Conf. Ser.*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012027.

[43] B. Palani, S. Elango, and K. Vignesh Viswanathan, *CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT*, vol. 81, no. 4. Springer US, 2022. doi: 10.1007/s11042-021-11782-3.

[44] A. M. Ali, F. A. Ghaleb, B. A. S. Al-Rimy, F. J. Alsolami, and A. I. Khan, “Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique,” *Sensors*, vol. 22, no. 18, 2022, doi: 10.3390/s22186970.