International Journal of Economics, Commerce and Management
United Kingdom ISSN 2348 0386 Vol. 13, Issue 5, May 2025

https://ijecm.co.uk/

AN OVERVIEW OF ARTIFICIAL INTELLIGENCE APPROACHES TO FAKE REVIEW DETECTION IN ONLINE MARKETPLACES

Kravchuk Yelyzaveta
MSc in Enterprise Management
Sichuan University - Chengdu, China
lizakravchuk773@gmail.com

Abstract

This article presents a theoretical review of the problem of fake reviews in online marketplaces. It examines the concept of deceptive reviews, their impact on consumer trust and purchasing behavior, and the common strategies used to generate them. Given the increasing reliance on online feedback in e-commerce, the spread of fake reviews poses a significant threat to both users and businesses. This review focuses on current technological approaches to address this issue, with particular attention to the role of Artificial Intelligence (AI). The paper surveys recent advancements in the use of Machine Learning and Deep Learning techniques, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based models, for the purpose of fake review detection. These methods are capable of identifying sophisticated patterns in textual data that may signal fraudulent intention. Moreover, the review highlights the growing interest in hybrid models that combine multiple neural architectures, which have shown improved detection accuracy over single-model systems. Overall, this theoretical investigation underscores the potential of AI-driven solutions in enhancing the reliability of online review systems and fostering greater trust in digital commerce.

Keywords: Fake reviews, e-commerce, Artificial Intellegence, Machine Learning, Deep Learning

INTRODUCTION

With the expansion of the internet and the advancement of online commerce, digital marketplaces have become a crucial for international trade sector, engaging millions of consumers worldwide. According to research presented by Paraschiv et al. (2020), the number of purchases made through online platforms in 2020 increased by 40% compared to the previous years, largely due to the COVID-19 pandemic. While this trend has created new opportunities for both consumers and entrepreneurs, it has also introduced significant challenges, particularly in terms of trust and transparency. One of the most pressing issues are the widespread fake reviews, which pose a serious threat to consumer confidence (Fiedler and Kissling, 2020). As highlighted in the review conducted by Dixa in 2022, 93% of consumers pay attention to reviews before finalizing their purchasing decisions, making them an essential part of the online shopping experience (Dixa, 2022). What is more, 80% of buyers have encountered fake reviews in their shopping experience (Zhang and Chen, 2018). By and large, fake reviews can be described as inaccurate, misleading, deceptive comments which lead to wrong perception of products or services. Such reviews have an enormous effect on consumer's behaviour, decreasing consumer's trust, lowering the quality of information and undermining reliability of online marketplaces. Furthermore, they also affect the development of digital economy as they reduce the overall effectiveness of e-commerce and give an unfair advantage to less competitive sellers, distorting market competition (Sahut et al., 2024).

Under these circumstances, it is important to leverage special techniques and tools to detect fake reviews and secure a trustful online business environment. Nowadays, Artificial Intelligence is one of the most promising tools for addressing this issue, as it can learn, analyze extensive datasets, and identify recurring patterns, making it a crucial asset in combating fraudulent activities in e-commerce (Odeyemi et al., 2024). As part of using AI to detect fraudulent reviews, Deep Learning, particularly neural networks, is gaining increasing attention due to its ability to analyze text, audio, image, and video content, which are the primary formats used in reviews (Kumar et al, 2023). Another emerging topic nowadays is hybrid systems which combine different AI's subsets, such as Convolutional Neural Networks with Recurrent Neural Networks (CNN-RNN), showing a better result as while CNN focuses on identifying patterns, Recurrent Neural Networks understands the changes over time (Yin et al., 2017).

The primary aim of this research paper is to define fake reviews and investigate how various Artificial Intelligence techniques enhance the accuracy and effectiveness of detecting

them. This study explores the use of Machine Learning, Deep Learning, and neural networks to assess their ability to identify deceptive content on e-commerce platforms.

Particularly, the study attempts to:

1. Understand the concept of fake reviews, explore their motivations and creation methods, and analyze the key challenges in detecting them.
2. Assess traditional techniques for identifying fake reviews and evaluate the main limitations in their application.
3. Explore Machine Learning approaches for detecting fake reviews, including supervised, semi-supervised and unsupervised learning.
4. Examine how Deep Learning enhances the precision of fake review detection, with a focus on Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Recurrent Neural Networks (RNNs).
5. Evaluate the potential of hybrid models, such as combinations of CNNs and LTSM.

By comparing different techniques and methods, our goal is to develop an in-depth framework outlining the application of AI in detecting fake reviews and strengthening user trust in digital platforms.

Defining fake reviews and identifying the key motivation of creating them

In the modern digital economy, consumer purchasing decisions are heavily influenced by online reviews. They offer valuable insights into customer satisfaction and the quality of products. However, as online reviews gain increasing importance, the prevalence of fake reviews has also surged, leading to consumer deception and creating an unfair competitive advantage within the market (Rastogi and Mehrota, 2017). While trustworthy reviews benefit both consumers and sellers, helping consumers make informed purchasing decisions and providing retailers with valuable feedback to improve product quality, fake reviews distort this balance. Fake reviews, also known as deceptive reviews, opinion spam, or fraudulent comments, mislead potential buyers, often resulting in dissatisfaction, and preventing businesses from accurately assessing their products, ultimately harming both consumer trust and market integrity (Salminen et al., 2022).

Research by Mohawesh et al. (2021) identifies two primary motives for the creation of fake reviews: to promote your own company’s goods to gain a stronger reputation over similar products in the commercial sector or to undermine the credibility of a competitor’s offerings. Cao (2023) states that vendors can stimulate consumers to write positive comments by offering discount coupons, vouchers and cashbacks. Even though costumers may not be fully satisfied with the product, they still choose to write dishonest comments to financially benefit from the

purchase. Chowdhary and Pandit (2018) further expanded on this point, adding that another common tactic involves merchants hiring specialized writers who are experts in crafting deceptive reviews to manipulate consumer perception and boost product credibility. This strategy, combined with reward-driven comments creates a misleading digital environment in which it is difficult for consumers to distinguish real reviews from fake ones.

As fraudulent feedbacks become more widespread, their impact on e-commerce is increasingly evident. Cao (2023) highlights that they undermine key aspects of online platforms, including consumer trust, brand reputation, and fair competition. Wu et al. (2019) further emphasize that fake reviews create uncertainty among consumers, making it more challenging to assess product quality and increasing their perceived risk. When shoppers suspect that reviews may be deceptive, they become hesitant to rely on online feedback, leading to lower purchase intentions and a decline in trust in e-commerce platforms. Moreover, Shahri et al. (2023) argue that because fake reviews are often well-structured and persuasive, they can mislead consumers into purchasing low-quality products, ultimately distorting the online marketplace.

In general, fake reviews are produced through two main methods: human-written content or automated systems. With the rapid advancement of technology, automated systems have become increasingly prevalent, especially with innovations in natural language processing (NLP) and Machine Learning (ML), making the generation of dishonest reviews more efficient, sophisticated, and cost-effective (Salminen et al., 2022). Under these circumstances, artificial intelligence plays a dual role, being used not only for detecting dishonest comments but also for generating them. Furthermore, the actual amount of spam content remains unknown, making it even more challenging to detect and eliminate fraudulent reviews (Lim et al., 2010). This growing challenge has prompted researchers and e-commerce platforms to develop various methods for detecting fake reviews, ranging from user-driven approaches to advanced algorithmic techniques.

A review of traditional methods for detecting fraudulent reviews

Detecting fake reviews is a complex task because they often closely resemble genuine ones, making it challenging to distinguish between them. Creating an effective detection model is further complicated by the challenge of manually labeling reviews as real or fake, which is a time-consuming and subjective process. Since fake reviews are designed to mimic authentic ones, identifying inconsistencies can be difficult. Consequently, identifying fake reviews is typically approached as a two-class classification task, where reviews are categorized as either true or false (Wang et al., 2022). According to Liu et al. (2024), detecting fake reviews involves

analyzing multiple factors, including overall review features, reviewer behavior, and the specific review targets, to identify unusual patterns. However, basic methods that rely only on text analysis have limitations, since the interpretation of a review can vary based on its context, and certain words may have different interpretations in different situations.

In his paper, Hussain et al. (2019) outline three key steps for fake review detection. The first step involves data collection and preprocessing, where review data is gathered and cleaned to remove any noise or irrelevant information. The next step is selecting an appropriate feature engineering approach. This could include metadata-based features, which focus on external attributes like timestamps and ratings, or linguistic n-grams, which analyze the frequency and patterns of words in the reviews. Other approaches, such as behavioral features, examine reviewer habits, like posting frequency or copying content, to identify potential spam activity. These various approaches help extract meaningful patterns that can improve the accuracy of detecting fake reviews. The last step includes choosing a relevant fake review detection system (Mohavesh et al, 2021; Hussain et al, 2019).

As noted by Lim et al. (2010), one common method used by e-commerce platforms to detect fake reviews is allowing customers to vote on whether a review was helpful. However, this approach has limitations, as it depends on user engagement, which can be manipulated by spammers to artificially boost deceptive reviews. Additionally, websites like eBay, Amazon, and Walmart only allow verified buyers to leave reviews to ensure authenticity (Mayzlin et al., 2012). Another method used for detecting fake reviews is browser fingerprinting, which analyzes unique device and browser characteristics such as IP addresses, HTTP headers, and JavaScript-extracted features. This technique helps platforms identify fraudulent behavior, like multiple reviews being posted from the same device, making it harder for spammers to manipulate online ratings. However, browser fingerprinting also has its drawbacks, including privacy concerns, evasion tactics, false positives, and high computational costs (Zhang et al, 2022).

One more method to address this issue was content analysis, where experts examine the text of reviews, focusing on word choice, sentence structure, and suspicious patterns that could indicate deception. While useful, this method is time-consuming, lacks accuracy, and heavily relies on human effort (Sun et al., 2024). Another technique that also focuses on linguistic features is Linguistic Inquiry and Word Count (LIWC), which is a system for analyzing textual data that examines the frequency and proportionality of words in specific linguistic categories. Unlike other text analysis methods, LIWC is known for its user-friendly interface and affordability, making it a popular choice for businesses and researchers analyzing customer feedback (Kim, 2024).

A more advanced technique widely used today is user behavior analysis. By examining patterns such as the frequency of reviews a user posts for a particular product and the similarity between their past reviews, it becomes possible to assess the authenticity of their feedback (Al-Sultany and Hussain, 2019). Additionally, platforms can analyze various behavioral indicators, such as the timing of reviews, IP addresses, and the ratio of positive to negative ratings, to detect suspicious activity. Research shows that fake reviewers often post in short bursts, leave overly similar reviews, and deviate significantly from general rating trends, making behavior-based detection a valuable tool in combating fraudulent feedback (Mukherjee et al, 2013). This approach has been tested on Amazon, achieving an accuracy rate of 80% to 95% in identifying fake reviews (Al-Sultany and Hussain, 2019). However, nowadays all commonly used techniques mainly rely on AI features, which will be discussed in the next section.

The role of Machine Learning (ML) in identifying deceptive reviews

Machine learning is a branch of Artificial Intelligence, which is concerned with studying algorithms and deriving insights from historical data. It is a powerful tool for decision-making, commonly used in forecasting and pattern recognition, making it particularly effective for detecting fake reviews (Sarker, 2022; Janiesch et al., 2021). There are three main types of ML which are commonly applied to face this issue: supervised, semi-supervised and unsupervised learning. The main difference between these approaches lies in how they utilize labeled data. Supervised learning utilizes labeled datasets for training, making it well-suited for tasks such as classification and regression. Semi-supervised learning, on the other hand, integrates a small portion of labeled data alongside a vast quantity of unlabeled data, enhancing performance when labeled data is limited. Unsupervised learning operates exclusively on unlabeled data to identify patterns, making it effective for clustering, anomaly detection, and uncovering hidden structures in data (Arunraj et al., 2017; Prakash and Nithya, 2014).

In the context of dishonest review detection, supervised learning is the most commonly applied approach due to its high accuracy and reliability. By training on labeled datasets containing both genuine and fake reviews, supervised models can effectively learn to identify distinguishing patterns, making them highly effective for classification tasks (Mukherjee et al., 2013). Several classification algorithms have been developed for this purpose, including Naïve Bayes, Decision Trees, Logistic Regression, Random Forest, and K-Nearest Neighbors. Each algorithm has its own strengths and limitations: Naïve Bayes is effective for text classification but assumes feature independence, Decision Trees offer interpretability but are prone to overfitting, Logistic Regression works well for binary classification but struggles with complex data distributions, Random Forest enhances accuracy by combining multiple decision trees but

requires significant computational resources, and K-Nearest Neighbors, while simple, becomes inefficient with large datasets (Elmogy et al., 2021). Among these methods, the Support Vector Machine (SVM) has been identified as the most effective supervised learning algorithm for fake review detection, consistently outperforming other classifiers in terms of accuracy (Abd and Hussein, 2024).

Unsupervised learning is another technique used in e-commerce for detecting fake reviews. While it has potential in this area, its application is not as extensively explored or widely adopted as supervised learning. However, since accurately labeling datasets can be challenging, unsupervised learning provides a practical alternative in cases where supervised methods are impractical (Mohawesh et al., 2021). Principal Component Analysis (PCA), clustering methods, and anomaly detection are among the most frequently used techniques in unsupervised learning, forming the foundation for various methods in fake review detection (Cardoza and Balipa, 2023). PCA is a dimensionality reduction technique that helps reduce the complexity of high-dimensional data by transforming it into a smaller set of uncorrelated principal components, selecting those with the highest variance to improve analysis and classification accuracy (Shah and Ahmed, 2019).

Data clustering, a key process in AI, Machine Learning, and pattern recognition, involves identifying natural groupings in multidimensional data using similarity measures and is widely applied in fake review detection, data mining, compression, and some other fields (Omran et al., 2007). Fake review detection leverages clustering algorithms such as k-Means, DBSCAN, hierarchical clustering, graph-based clustering, and Gaussian Mixture Models (GMMs) to categorize suspicious reviews, identify coordinated fraudulent actions, and spot irregularities in review trends. Mothukuri et al. (2022) utilized K-means, GMM Full covariance, and GMM Diagonal covariance clustering techniques to detect fake reviews, determining that K-means achieved the highest accuracy. However, their study suggests that additional unsupervised algorithms and broader domain exploration could further enhance detection performance.

Anomaly Detection (AD) in Machine Learning is used to distinguish normal data from abnormal data, often by training models exclusively on normal instances to later identify anomalies. In recent years, AD has been increasingly combined with techniques such as Natural Language Processing (NLP) and temporal behavior analysis in order to improve detection accuracy. While this integrated approach shows considerable promise, the field remains underexplored, particularly in relation to textual data (Novoa-Paradela et al., 2024; Liu et al., 2024). One of the primary challenges in outlier detection lies in handling the diversity and complexity of data types, such as high-dimensional, spatial, or sequential data, which often require specialized algorithms. Moreover, the definition of what constitutes an **“**outlier” can vary

significantly across domains, complicating the development of universally applicable detection methods (Kannan & Park, 2017).

In response to these challenges, semi-supervised anomaly detection has gained interest, however this field of research is also still not well explored. The primary goal of semi-supervised learning (SSL) is to address the limitations of both supervised and unsupervised learning (Reddy et al., 2018). Since labeled data for fake reviews is limited and hard to obtain, combining it with a larger set of unlabeled data allows models to improve detection by making better use of available resources (Kumaran et al., 2021). Semi-supervised learning is based on three main ideas. The first is the continuity assumption, which means that if two data points are close to each other, they probably belong to the same class. The second is the cluster assumption, which says that data usually forms groups, and items in the same group likely have the same label. The third is the manifold assumption, which suggests that even though data may exist in a high-dimensional space, it actually lies on a simpler, lower-dimensional structure. (Ligthart et al., 2021).

Self-training and co-training are two common semi-supervised learning methods: self-training lets a model teach itself using confident guesses on new data, while co-training uses two models that help each other learn by sharing the predictions they make on the unlabeled data (IJAEM, 2024). Additionally, semi-supervised learning (SSL) is categorized into semi-supervised classification and semi-supervised clustering. Semi-supervised classification enhances the accuracy of the learning process, categorizing data by using both labeled and unlabeled data, where the labeled data helps guide the model to understand the unlabeled data. Semi-supervised clustering focuses on grouping similar data points together by combining labeled and unlabeled data, with the labeled data helping the model make better clusters. In fake review identification, semi-supervised classification is more commonly used as it leverages a limited amount of labeled data combined with a larger quantity of unlabeled data can enhance detection accuracy, which is particularly useful since obtaining labeled data for fake reviews is often difficult and time-consuming.

In general, machine learning (ML) offers a variety of methods for detecting fake reviews, each addressing different challenges. Supervised learning, especially with algorithms like Support Vector Machines (SVM), is the most commonly used due to its high accuracy in classifying reviews based on labeled data. In cases where labeled data is insufficient, unsupervised learning approaches like clustering and anomaly detection can reveal patterns or anomalies in review data without the necessity of labels. Semi-supervised learning (SSL), by combining labeled and unlabeled data, can boost model accuracy when labeled data is in short

supply. These approaches, each with its unique strengths, enable more robust and flexible detection systems that can adapt to the evolving tactics used in creating fake reviews.

Enhancing review credibility assessment through Deep Learning (DL)

Deep learning (DL) is a rapidly growing domain within ML and AI, known for its ability to learn from large datasets and model complex abstractions through multiple layers. Built upon artificial neural networks, DL helps computers learn patterns by starting with simple ideas and gradually understanding more complex ones, which is useful for tasks like recognizing images, speech, or text without detailed human instructions (Sarker, 2022; Tiwari et al., 2018). While DL models require significant time for training due to their many parameters, they are more efficient in testing compared to other Machine Learning algorithms (Sarker, 2021). In fake review detection, Deep Learning models are capable of processing vast volumes of review data to identify subtle signals and inconsistencies that may indicate deception. Deep learning methods like CNNs, RNNs, LSTM, autoencoders, and multilayer perceptrons have been widely used in fake review detection, showing strong performance by learning from both labeled and unlabeled data (Bathla et al., 2021).

Convolutional Neural Networks (CNNs) are advanced Deep Learning architectures that automatically learn important patterns from data without requiring manual feature design, making them more efficient and effective than older neural networks (Sarker, 2021). In contrast, Recurrent Neural Networks (RNNs) are built to manage and analyze sequences of data, using their internal memory to retain information from previous inputs. As a result, they are particularly effective in tasks involving language understanding, audio analysis, and temporal data forecasting, where the order of data matters (Mienye et al., 2024). While CNNs excel at identifying hierarchical patterns, such as key phrases, RNNs are better at capturing the flow and context of sequences, like sentence structure (Yin et al., 2017). However, basic RNNs faced issues during training, especially with remembering information over long chains of information. To address this challenge, researchers developed Long Short-Term Memory (LSTM) units and Gated Recurrent Units (GRU). With improved handling of the vanishing gradient issue, these architectures can successfully learn dependencies across much longer spans of data (Schmidt, 2019).

Nowadays, in fake review detection, the advantages of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, are often combined to improve performance. The LSTM model is a special type of Recurrent Neural Network created specifically to solve problems that occur when learning from long sequences of data, like the exploding or vanishing gradients that can happen when trying

to remember information for a long time (Okut, 2021). In simple terms, the LSTM has a unique structure made of "memory blocks" that helps it remember important information over time. These memory blocks use gates, which are like controls, decide what information to keep and what to forget, allowing the model to learn long-term patterns more effectively (Houdt et al., 2020).

CNNs outperform other models in terms of accuracy because it can extract complex, high-level features from opinions, allowing it to capture nuanced patterns in the data. This capability makes CNNs more commonly used than RNNs in this field, as they are more efficient and effective at identifying key patterns in text (Bathla et al., 2021). Still, combining both CNNs and LSTM can lead to better results. In this hybrid approach, the CNNs first look at small parts of the text to find patterns that could suggest a review is fake. Then, the LSTM reads the review in order, like how we read sentences, to understand the full meaning and writing style. This model is trained using a set of reviews that are already labeled as real or fake. It also uses word embeddings, which turn words into numbers, to help the model better understand the text and make more accurate predictions (Bhaware and Sharma, 2025).

The transformer model is another form of deep neural network that leverages self-attention to analyze dependencies across different positions in a sequence. In contrast to CNNs and advanced RNNs variants like Long Short-Term Memory (LSTM), transformer models are particularly good at managing long-range connections between elements in a sequence, while also enabling parallel processing, which makes them more efficient (Islam et al., 2023). In the field of fake review detection, transformer models such as BERT and RoBERTa are among the most commonly used due to their ability to capture contextual meaning, identify subtle linguistic patterns, and accurately differentiate between deceptive and genuine content (Alsaad and Joshi, 2024). BERT, developed by Google, is a language model that has been pre-trained and built upon the Transformer framework. It is adapted for specific tasks like fake review detection, where it extracts valuable features from both review texts and user behavior data, helping identify emotions and hidden meanings in the text (Sun et al., 2024; Azizah et al., 2023).

RoBERTa, developed by Facebook, enhances BERT's capabilities by making several key improvements. It uses a different way of hiding words during training (called dynamic masking), doesn't include the Next-Sentence Prediction task, and is trained on longer pieces of text with more data at once, helping the model understand language better. RoBERTa also uses a bigger and more detailed vocabulary and is trained on a wider variety of texts, which makes it more accurate and powerful. Despite these enhancements, RoBERTa maintains a similar base model structure to BERT, with 12 encoder layers and 768 hidden units, ensuring that it retains the core strengths of the Transformer architecture while achieving superior performance in tasks

such as fake review detection (Kolev et al., 2022). However, in the modern era of hybrid models, BERT and RoBERTa are often integrated with LSTM networks, leading to improved efficiency and performance compared to when used individually. In their study, Mohawesh et al. (2024) developed a model that leverages RoBERTa's ability to understand contextual information and LSTM's strength in capturing sequential dependencies, achieving an accuracy of 96.03% on the OpSpam dataset and 93.15% on the Deception dataset, thus outperforming existing state-of-the-art methods.

Moreover, some other models based on BERT have been developed, including ALBERT and DistilBERT. ALBERT (A Lite BERT) improves on BERT by reducing the number of parameters and sharing them across layers, making the model faster and more memory-efficient while maintaining strong language understanding. This allows ALBERT to better capture word meanings and improve text comprehension, outperforming many other deep learning models in some cases, including fake review detection (Mohawesh et al., 2021). In the meanwhile, DistilBERT is a simplified, faster version of BERT that reduces the number of layers and removes extra components to make the model more efficient while keeping strong language understanding. It is widely used when faster training and lower memory usage are needed, including in tasks like fake review detection (Mohavesh et al., 2021).

The study by Gupta et al. (2022) revealed that the RoBERTa model delivered the best results in fake review detection, with an accuracy of 69%, outperforming all other models. BERT and DistilBERT showed similar results, with DistilBERT slightly outperforming BERT (68% vs. 67% accuracy), while ALBERT recorded the weakest performance, reaching only 64% accuracy. The researchers explained that the overall lower scores compared to some benchmark studies were due to the use of a diverse, multi-domain dataset, which made the classification task more challenging but also more generalizable. In the meanwhile, other studies have shown that hybrid models outperform individual models; for example, combining RoBERTa with LSTM achieved a 96.03% accuracy rate (Mohawesh et al., 2024).

To summarize, deep learning models, especially those based on transformer architectures like BERT and RoBERTa, have proven highly effective for fake review detection. These models excel at understanding contextual information and identifying subtle patterns within the text. When combined with LSTM networks, which capture sequential dependencies, their performance improves further. For instance, the hybrid model combining RoBERTa and LSTM achieved an impressive 96.03% accuracy on the OpSpam dataset (Mohawesh et al., 2024). While RoBERTa outperforms other models like BERT and DistilBERT, the integration of these models with CNNs and LSTMs helps in achieving even better results. Furthermore, models like ALBERT and DistilBERT, which are optimized for speed and efficiency, also

contribute to advancing fake review detection by balancing performance with resource constraints. Overall, the continuous refinement of these hybrid models and their ability to process large, diverse datasets is enhancing the accuracy, reliability, and generalizability of fake review detection systems across different applications.

CONCLUSION

By and large, fake reviews have emerged as a serious issue, posing risks to both consumers and businesses worldwide. As deceptive comments become more sophisticated, the process of identifying them becomes increasingly complicated. Meanwhile, the majority of consumers rely on reviews when shopping online, which can lead to incorrect purchasing decisions and create an unfair competitive advantage on online platforms. In this context, traditional methods of detecting fake reviews have become inefficient and ineffective, highlighting the need for new technological solutions to address this problem. Artificial Intelligence (AI), specifically Machine Learning (ML) and Deep Learning (DL), offers promising approaches that have demonstrated strong performance in fake review detection. These AI-driven methods allow automated systems to analyze vast volumes of data, uncover hidden patterns, and distinguish between genuine and deceptive content with a high degree of precision.

Algorithms under supervised learning, including Random Forests and Support Vector Machines, have shown strong performance when trained on labeled datasets, while unsupervised and semi-supervised models offer solutions in scenarios where labeled data is limited or unavailable. Advanced deep learning models, such as CNNs and RNNs, provide additional improvements in detection accuracy capabilities by capturing complex linguistic tendencies and contextual relationships within review texts. Recently, new hybrid models that combine different methods, such as CNNs and LSTM, have shown significant progress in the accuracy of fake review detection. By integrating the strengths of both approaches, with CNNs identifying key phrases and LSTM capturing sequential dependencies, these models enhance the overall detection process.

Furthermore, the development of transformer-driven models, such as BERT and RoBERTa, has significantly advanced fake review detection. These models, initially trained on extensive text data and then refined for specific tasks such as identifying fake reviews, have established new benchmarks in Natural Language Processing. Their ability to understand context and semantic relationships within text enables them to identify even the most intricate fake reviews. As hybrid models and transformer-based architectures continue to develop, they

are expected to offer more precise, scalable, and flexible solutions for spotting deceptive reviews, making them invaluable in the battle against online fraud.

RECOMMENDATIONS FOR FUTURE RESEARCH

Although recent advancements in machine learning and deep learning have significantly improved fake review detection, several promising directions remain underexplored. Firstly, more attention should be given to the development of unsupervised and semi-supervised learning models. These methods are especially important due to the limited availability and high cost of obtaining accurately labeled datasets, which are often necessary for supervised learning. In particular, further exploration of hybrid semi-supervised approaches, such as combining anomaly detection with deep contextual embeddings, could lead to more flexible and effective detection systems.

Secondly, future research should focus on designing and refining hybrid model architectures that bring together the strengths of different types of models. For example, combining CNNs and LSTMs has shown promising results by using CNNs to identify local patterns and LSTMs to capture sequence information. Similarly, integrating transformer-based models like BERT or RoBERTa with recurrent networks can improve the understanding of context while preserving the ability to model temporal dependencies. These hybrid systems have the potential to offer more robust and generalizable performance across diverse and evolving types of fake review content.

Lastly, as deceptive techniques grow more advanced, it is essential that fake review detection systems continue to evolve. This includes addressing challenges posed by content generated using large language models. Techniques such as adversarial learning and reinforcement learning could support dynamic model updates and stronger resistance to manipulation. Additionally, future systems should consider practical needs like computational efficiency, cross-platform scalability, and making detection results understandable for users and platform moderators. Proactively anticipating emerging threats and incorporating adaptive mechanisms will be key to maintaining the long-term effectiveness of these systems.

REFERENCES

Paraschiv, D.M., Titan, E., loana, M.D., and Crina-Dana, I. (2022). The change in e-commerce in the context of the Coronavirus pandemic. Management & Marketing, 17, 220-233.

Fiedler, M., and Kissling, M. (2020). Fake Reviews in E-Commerce Marketing. Herald of Kyiv National University of Trade and Economics, (2), 77–86.

Dixa. (2022). 3 statistics that show how customer reviews influence consumers. https://www.dixa.com/blog/3-important-statistics-that-show-how-reviews-influence-consumers/

Sahut, J.-M., Laroche, M., and Braune, E. (2024). Antecedents and consequences of fake reviews in a marketing approach: An overview and synthesis. Journal of Business Research, 175, 114572.

Odeyemi, O., Mhlongo, N.Z., Nwankwo, E.E., and Soyombo, O.T. (2024). Reviewing the role of AI in fraud detection and prevention in financial services. International Journal of Science and Research Archive, 11, 2101-2110.

Kumar, R., Mukherjee, S., and Rana, N. P. (2024). Exploring latent characteristics of fake reviews and their intermediary role in persuading buying decisions. Information Systems Frontiers, 26, 1091–1108.

Yin, W., Kann, K., Yu, M., and Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923.

Kumar, J. (2020). Fake review detection using behavioral and contextual features. arXiv: 2003.00807.

Rastogi, A., and Mehrotra, M. (2017). Opinion spam detection in online reviews. Journal of Information & Knowledge Management, 16(4), 1750036.

Salminen, J., Kandpal, C., Kamel, A. M., Jung, S.-g., and Jansen, B. J. (2021). Creating and detecting fake reviews of online products. Computers in Human Behavior, 118, 106676.

Mohawesh, R., Xu, S., Springer, M., Al-Hawawreh, M., and Maqsood, S. (2021). Fake or genuine? Contextualised text representation for fake review detection. arXiv.

Cao, C. (2023). The Impact of Fake Reviews of Online Goods on Consumers. BCP Business & Management, 39, 420–425.

Chowdhary, N. S., and Pandit, A. A. (2018). Fake Review Detection Using Classification. International Journal of Computer Applications, 180(50), 16–21.

Wu, S., Wingate, N., Wang, Z., & Liu, Q. (2019). The Influence of Fake Reviews on Consumer Perceptions of Risks and Purchase Intentions. Journal of Marketing Development and Competitiveness, 13, 133–142.

Shahri, M. H., Haghbin, F., Raeini, Y. Q., and Monfared, N. (2023). The effects of fake reviews during stepwise topic movement on shopping attitude in social network marketing. MethodsX, 11, 102461.

Salminen, J., Kandpal, C., Kamel, A. M., Jung, S., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. Journal of Retailing and Consumer Services, 64, 102771.

Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., and Lauw, H. W. (2010). Detecting product review spammers using rating behaviors. Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM 2010), 939–948.

Wang, W., Fong, S., and Law, R. (2022). Recent state-of-the-art of fake review detection: A comprehensive review. Knowledge Engineering Review, 37, E8.

Liu, J., Quan, P., Zhang, W. (2024). A Study on Fake Review Detection Based on RoBERTa and Behavioral Features. Procedia Computer Science, 242, 1323-1330.

Hussain, N., Turab Mirza, H., Rasool, G., Hussain, I., and Kaleem, M. (2019). Spam Review Detection Techniques: A Systematic Literature Review. Applied Sciences, 9, 987.

Mayzlin, D., Dover, Y., and Chevalier, J. A. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. American Economic Review, 104(8), 2421–2455.

Zhang, D., Bu, Y., and Wang, Y. (2022). A Survey of Browser Fingerprint Research and Application. Security and Privacy, 5, e3363335.

Sun, P., Bi, W., Zhang, Y., Wang, Q., Kou, F., Lu, T., and Chen, J. (2024). Fake Review Detection Model Based on Comment Content and Review Behavior. Electronics, 13(21), 4322.

Kim, J. M., Park, K. K., Mariani, M., and Wamba, S. F. (2024). Investigating reviewers' intentions to post fake vs. authentic reviews based on behavioral linguistic features. Technological Forecasting and Social Change, 198, 122971.

Al-Sultany, G., & Hussain, S. M. (2019). Fake reviews detection through users behavior analysis. Journal of Advanced Research in Dynamical and Control Systems, 11, 737–741.

Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). Fake review detection: Classification and analysis of real and pseudo reviews (UIC-CS-2013-03). Technical Report, Department of Computer Science, University of Illinois at Chicago. https://www2.cs.uh.edu/~arjun/tr/UIC-CS-TR-yelp-spam.pdf.

Sarker, H.I. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. SN Computer Science, 3.

Janiesch, C., Zschech, P., and Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets.

Arunraj, N. S., Hable, R., Fernandes, M., Leidl, K., and Heigl, M. (2017). Comparison of supervised, semi-supervised and unsupervised learning methods in network intrusion detection system (NIDS) application. Anwendungen und Konzepte der Wirtschaftsinformatik, 6, 10–19.

Prakash, V. J., & Nithya, L. M. (2014). A survey on semi-supervised learning techniques. International Journal of Computer Trends and Technology (IJCTT), 8, 25–29.

Elmogy, A. M., Tariq, U., Mohammed, A., and Ibrahim, A. (2021). Fake reviews detection using supervised machine learning. International Journal of Advanced Computer Science and Applications, 12, 601–606.

Abd, M. J., and Hussein, M. H. (2024). Fake reviews detection in e-commerce using machine learning techniques: A comparative survey. BIO Web of Conferences, 97, 00099.

Mohawesh, R., Xu, S., Tran, S. N., Ollington, R., Springer, M., Jararweh, Y., and Maqsood, S. (2021). Fake reviews detection: A survey. IEEE Access, 9, 65771–65802.

Cardoza, A. P., & Balipa, M. (2023). A comparison of recent supervised and unsupervised learning techniques for detecting fake reviews. International Research Journal of Modernization in Engineering Technology and Science, 5, 38313.

Shah, F. M., and Ahmed, S. (2019). Fake review detection using principal component analysis and active learning. International Journal of Computer Applications, 178, 42–48.

Omran, M. G. H., Engelbrecht, A. P., and Salman, A. A. (2007). An overview of clustering methods. Intelligent Data Analysis, 11(6), 583–605.

Novoa-Paradela, D., Fontenla-Romero, O., and Guijarro-Berdiñas, B. (2024). Explained anomaly detection in text reviews: Can subjective scenarios be correctly evaluated?. Engineering Applications of Artificial Intelligence, 133, 108065.

Kannan, R., Woo, H., Aggarwal, C. C., and Park, H. (2017). Outlier detection for text data: An extended version. arXiv.

Kumaran, N., Chowdary, C. H., and Sreekavya, D. (2021). Detection of fake online reviews using semi-supervised and supervised learning. International Research Journal of Engineering and Technology (IRJET), 8(4), 650–653.

Ligthart, A., Catal, C., and Tekinerdogan, B. (2021). Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification. Applied Soft Computing, 101, 107023.

Padmanabha Reddy, Y. C. A., Viswanath, P., and Eswara Reddy, B. (2018). Semi-supervised learning: A brief review. International Journal of Engineering & Technology, 7, 81–85.

Tiwari, T., Tiwari, T., and Tiwari, S. (2018). How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different?. International Journals of Advanced Research in Computer Science and Software Engineering, 8, 1-9.

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science, 2(6), 420.

Bathla, G., Singh, R. K., Singh, P., Cambria, E., and Tiwari, R. (2022). Intelligent fake reviews detection based on aspect extraction and analysis using deep learning. Neural Computing and Applications, 34, 20213–20229.

Van Houdt, G., Mosquera, C., and Napoles, G. (2020). A review on the long short-term memory model. Neural Computing and Applications, 3, 16323–16345.

Bhaware, J., and Sharma, V. (2025). Fake review detection using hybrid model based on CNN and LSTM. International Journal of Trend in Research and Development (IJTRD), 12, 20–25.

Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., and Pedrycz, W. (2023). A comprehensive survey on applications of transformers for deep learning tasks. Neural Computing and Applications.

Alsaad, M. M. B., ans Joshi, H. (2024). Transformer-based language deep learning detection of fake reviews on online products. Journal of Electrical Systems, 20, 2368–2378.

Azizah, S. F. N., Cahyono, H. D., Sihwi, S. W., and Widiarto, W. (2023). Performance analysis of transformer-based models (BERT, ALBERT, and RoBERTa) in fake news detection. arXiv.

Kolev, V., Weiss, G., and Spanakis, G. (2022). FOREAL: RoBERTa model for fake news detection based on emotions. Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART), 2, 429–440.

Mohawesh, R., Salameh, H. B., Jararweh, Y., Alkhalaileh, M., and Maqsood, S. (2024). Fake review detection using transformer-based enhanced LSTM and RoBERTa. International Journal of Cognitive Computing in Engineering, 5, 250–258.

Gupta, P., Gandhi, S., & Chakravarthi, B. R. (2022). Leveraging transfer learning techniques - BERT, RoBERTa, ALBERT, and DistilBERT for fake review detection. In Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation (FIRE 2021), 75–82.