RESEARCH ARTICLE OPEN ACCESS

An Integrated Deep Learning Approach for Enhanced Fake Review Detection on E Commerce Platforms Using Multimodal and Fine-Grained Features

Deshan Sachintha Kannangara*, Prof. Chunyong Yin **,Sachini Amani Henda Vitharana***

*(School of School of Computer Science and Technology, Nanjing University of Information Science & Technology, China, Email: deshansachinth777x@gmail.com)

** (School of Cyber Science and Engineering, Nanjing University of Information Science & Technology, China Email: Yinchunyong@hotmail.com)

*** (School of Artificial Intelligence, Nanjing University of Information Science & Technology, China Email: sachiofficial312@gmail.com)

--************************----------------------------------

Abstract:

Fake reviews are one of the greatest problems in e-commerce sites, which has a great impact on consumer confidence and buying behaviors. As the AI generated and synthetic content is expanding at a rapid pace, it has become harder to distinguish between fake review patterns. In this paper, the authors suggest a trimodal deep learning model to identify fake reviews, through the textual analysis, visual analysis, and identification of 22 handcrafted numeric features. A fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model is used to represent the textual modality, whereas the visual modality is represented by image features obtained through the use of a ResNet-50-based convolutional neural network. Further, the architecture includes a specific neural sub network that processes 22 numeric variables such as readability-related variables and image quality variables. The model, rather than basic early fusion, utilizes an attention-based fusion mechanism which learns dynamic weights of importance of the three modalities to use the complementary signals more effectively in making a classification. The given framework was tested on one Kaggle multimodal fake review dataset which included 20,144 restaurant reviews. The results of the experiment indicate that the model was more effective than the baseline classifiers, such as Logistic Regression and Random Forest with accuracy of 99.95, F1-score of 99.95, and ROC-AUC of 99.99. These results show the usefulness of an attention based trimodal learning in identifying advanced fake reviews in online markets.

Keywords — **Fake Review Detection, Trimodal Learning, BERT, ResNet-50, Attention-Based Fusion, Numeric Features, E-Commerce**

--************************----------------------------------

I. INTRODUCTION

Online shopping has become an integral part of the modern consumer behavior, with product reviews playing a pivotal role in the purchasing decisions. However, the growing number of fake reviews has seriously impacted on the credibility of e-commerce platforms. Fake reviews that are either overly positive or deliberately negative skew consumer perceptions and unfairly affect sellers and platforms alike. Recent breakthroughs in artificial intelligence and especially generative models and large language models have enabled the generation of highly believable fake reviews with greater ease and scalability than ever before. These AI-generated reviews tend to have a linguistic fluency and contextual coherence that is close to the content that is genuinely user generated and thus are hard to spot based on traditional techniques. rule-based or classical machine learning The current mechanisms for the detection of fake reviews are based mostly on text features and sentiment analysis. While effective in earlier settings, these approaches struggle to capture complex relationships in context, subtle linguistic cues, and non-textual signals that are present in contemporary e-commerce reviews. Moreover, in many existing models different features are treated independently without taking advantage of the dependence between textual, visual, and behavioral information.

To address these weaknesses, the current study suggests a hybrid deep learning architecture to detect fake reviews according to trimodal data fusion and fine-grained features analysis. The suggested method combines textual data, related review images, and 22 constructed numeric characteristics, such as indicators of readability and quality measures of images. The extraction of the textual representations is performed with a fine-tuned BERT model, the visual representations are learnt with a network based on ResNet-50, the numeric features are processed by a dedicated neural sub-network. These three modalities are then fused using an attention-based fusion mechanism with the aim of enhancing the detection of the deceptive patterns in online reviews.

The main contributions of this work are the following:

- **A trimodal deep learning framework for fake review detection**
  This paper puts forward a three-mode architecture that integrates the use of text, image, and numeric features to identify fake reviews. Contrary to traditional techniques that use only one modality or a multimodal fusion that is simple, the framework presented captures complementary data of review text, product-related images, and tabular features that are engineered
- **A unified feature extraction scheme using BERT, ResNet-50, and a numeric feature sub-network**
  The framework is based on a fine-tuned BERT model to achieve contextual representations of text, a convolutional network, which is constituted by ResNet-50 to learn the visual feature, and a separate numeric feature sub-network to process the 22 engineered features. This facilitates more discriminative richer learning of features of all the three modalities.
- **An attention-based fusion mechanism for trimodal integration**
  The paper proposes a modality level attention process which acquires dynamically changing weights of textual, visual, and numeric representations and then classifies them. This attention-controlled fusion can be used in order to highlight the most informative modality inputs to every instance of review, enhancing classification performance.
- **Strong empirical performance on a multimodal fake review dataset**
  The model was tested on a Kaggle multimodal fake review dataset with 20,144 restaurant reviews and the accuracy, F1-score, and ROC-AUC were 99.95, 99.95, and 99.99, respectively, surpassing baseline models, including the Logistic Regression and the Random Forest.

II. RELATED WORK

Fake review detection has been traditionally formulated as a supervised text classification problem, where reviews are classified as genuine reviews or deceptive reviews, based on linguistic characteristics. Early studies in this area mainly used handwritten lexical features, sentiment polarity and statistical measures like word frequency and the length of the review. These features were often combined with classical machine learning classifiers such as Support Vector Machines, Naive Bayes and Logistic Regression models. While such approaches were demonstrated to achieve reasonable performance in controlled settings on experiments, they showed a lack of robustness when faced with a variety of vocabularies, evolving writing styles and deliberate adversarial manipulation. As fake reviews got more sophisticated, these shallow representations proved inadequate for capturing deeper semantic inconsistencies. The advent of deep learning was a major change in fake review detection research. Neural network architectures like Convolutional Neural Networks and Recurrent Neural Networks like Long Short-Term Memory models made it possible to run the feature learning process automatically on raw text. These models showed better performances by capturing hierarchical patterns and sequential dependencies among review text. However, their effectiveness was often limited by the requirement for large labeled data sets and extensive training time. Moreover, even though many deep learning models have seen improvements in their representational power, it was difficult for many of them to detect subtle signs of deception, especially in short or well made fake reviews that closely resemble real user content. Transformer-based architectures have taken the field even further, introducing self-attention mechanisms that can model long-range dependencies and contextual relationships. Models like BERT, RoBERTa and DeBERTa have achieved state of the art results for fake review detection because of their capability to produce rich and contextualized word representations. Fine-tuning of these models using domain-specific review data has proven to be effective in capturing subtleties of linguistic patterns linked to deception. Nevertheless, most of the current transformer based research solely focuses on textual data, while ignoring other useful signals like product images, reviewer behavior and meta-data that are often accompanied by reviews in a real world e-commerce platform. Recent research efforts have focused on the more influential aspects of multimodal approaches involving textual, visual, and social context features for the robustness of the detection. Multimodal fake review detection models have been promising in utilizing complementary information between different data sources. However, many of these approaches process modalities separately and combine them using simple or late fusion approaches, which limit their ability to learn meaningful cross-modal interactions. In addition, relatively little focus has been placed on fine-grained feature analysis that focuses on subtle exaggerations, contextual inconsistencies, and visual anomalies that are routinely found in deceptive reviews. Although there is an advancement in multimodal fake review detection, current research has been dominated by textual and visual modalities, and there is a lack of sufficient attention to additional engineered numeric features, including readability-based metrics and image quality metrics. This leaves a research gap in that fine-grained features can provide more insight into detecting subtle deceptive patterns by the model.

III. METHODOLOGY

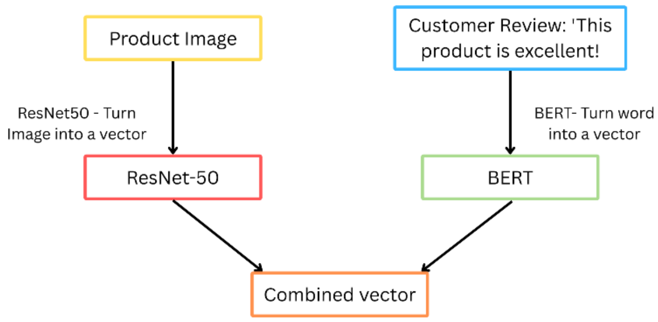

Figure 1 Multimodal Architecture Diagram

A. Data Collection and Processing

The present research employs one of the Kaggle multimodal fake review datasets with 20,144 reviews of restaurants. The samples have 25 columns, with an ID, text of reviews, class of reviews, and 22 engineered numeric features. The data is separated into training (12,086 samples), validation (4,028 samples), and test (4,030 samples) data. The distribution of the classes in the training set is 6,138 real reviews and 5,948 fake reviews. The textual values that are missing are set to empty strings in the preprocessing stage. The missing numeric values are filled in with the meaning of the feature. Normalization of the numeric features is then done with StandardScaler. In case of textual input, the review text is tokenized by the BERT tokenizer and a maximum sequence length of 128 is used, as well as padding and truncation are used to achieve a constant sample input size.

B. Textual Feature Extraction

To extract textual features, this paper applies the BERT base uncased model to extract contextual representations of reviewing text. The encoded text in the form of tokens is fed to the fine-tuned BERT encoder and the outpout of the pooler, which is represented by the 768 dimensionality, is the final textual representation. Base case machine learning models take the form of textual features extracted with TF IDF rather than BERT. Here, a maximum of 5000 features and removing the English stop-words are set in the vectorizer.

C. Visual Feature Extraction

In order to retrieve the visual characteristics, review related images are run through a ResNet-50 model which has been trained on ImageNet. All images are downsized to 224 224 and turned into tensors and then sent to the network. The implementation uses torchvision.models.resnet50(pretrained=True), and the final fully connected classification layer is replaced with nn.Identity(). As a result, the network outputs a 2048 dimensional visual feature vector for each image.

D. Numeric Feature Extraction

Besides text and image information, the proposed framework uses 22 designed numeric features that offer auxiliary fine-grained cues to identify fake reviews. These characteristics are readability-related measures, image quality-related measures, which are automated readability index, difficult words, flesch reading ease, Gunning Fog index, words per sentence, reading time, and various picture measures, e.g., brightness, contrast, warmth, sharpness, colorfulness, standard deviation, saturation and clarity related indicators. color dominance, These numerical characteristics are handled with a special neural sub-network of a Linear layer (22 64), preludes by ReLU activation and Dropout (0.3). The product of this branch is a feature representation in numeric form (64 dimensional).

E. Attention-Weighted Trimodal Fusion

The given framework combines textual, visual, and numeric characteristics with the attention-weighted trimodal fusion process. Instead of using direct concatenation exclusively, the model acquires the relative significance of each modality by attaching attention weights to the text, image, and numeric representations. Such weighted modality properties are then summed up to give the overall fused representation to be used in classification. This approach allows the model to complementary capture information in all the three modalities and enhances fake review detection.

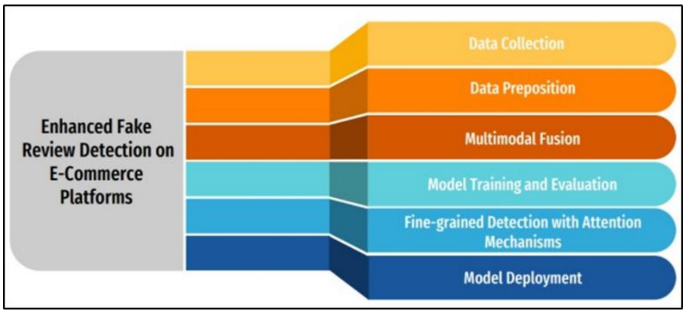

Figure 2 The figure shows the general outline that is adopted in the improved fake review detection within e-commerce websites

IV. EXPERIMENTAL SETUP

The design of the experimental setup is carefully designed to rigorously evaluate the effectiveness, the robustness, and the generalization capability of the proposed multimodal deep learning framework for fake review detection. The performance of the proposed model is systematically compared to well-established baseline machine learning classifiers including Support Vector Machines and Gradient Boosted Decision Trees which are widely adopted in the traditional fake review detection research. These baseline models are used as a reference point in order to test the contribution of deep contextual learning, multimodal fusion, and attention mechanisms presented in the proposed framework. To insure fairness and help mitigate bias that may be introduced due to unequal class distributions the dataset is divided in to training, validation and test subsets using a stratified sampling strategy. This ensures that the ratio of real and fake reviews is the same in all the data splits. The training set is used to learn model parameters and the validation set is used for hyperparameter tuning and model selection, as well as to implement early stopping and prevent overfitting. The test set is strictly reserved for final evaluation in order to provide an unbiased estimation of the real world model performance. This three-way split provides for stable measurement. convergence and reliable performance Model evaluation is performed using standard classification metrics such as accuracy, precision, recall and F1-score to give a comprehensive evaluation of detection performance. Accuracy is the overall correctness of the classification while precision and recall depict the capacity of the model to properly classify the fake reviews without creating an excessive number of false positives or negatives. The F1-score is employed as a balanced measure of precision and recall to take trade-offs into consideration. Together, these metrics facilitate both global performance analysis and class-wise evaluation so as to ensure thorough and reliable performance evaluation of the proposed framework to detect sophisticated fake reviews.

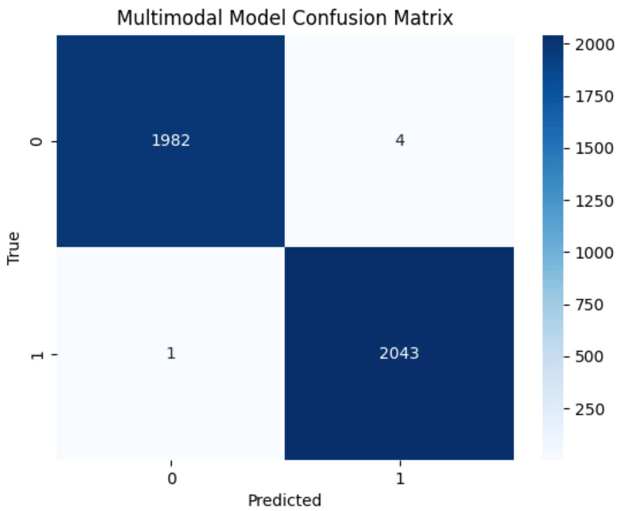

Figure 3 - Performance metrics of the multimodal model: The confusion matrix indicates the result of the multimodal fake review detection model.

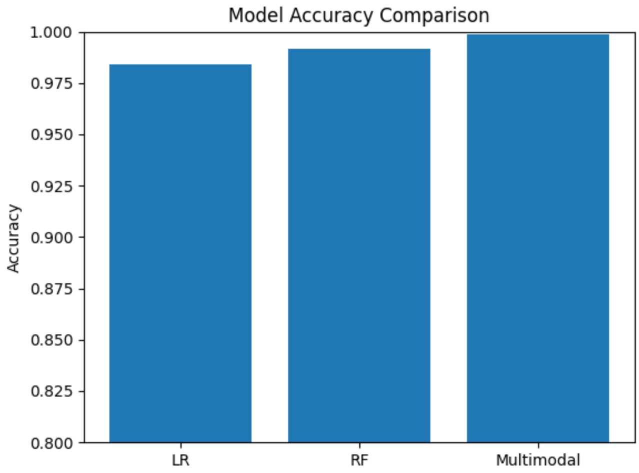

Figure 4 - Accuracy comparison of LR, RF, and Multimodal models: Logistic Regression, Random Forest (RF) and the developed Multimodal model.

V. RESULT AND ANALYSIS

The experimental results show the effectiveness of the proposed multimodal deep learning framework for fake review detection, and it is observed to continuously surpass the traditional baseline models. By incorporating textual modalities with visual modalities in a joint fashion, the framework is able to take advantage of complementary information that is not available to single modality models. Classical machine learning methods and text-only deep learning methods are based mainly on surface-level language features and are constrained in the capabilities for spotting sophisticated deceptive techniques. In contrast, the proposed approach has the advantage of the joint examination of contextual textual semantics and visual cues of authenticity, which leads to a more comprehensive evaluation of the review's credibility. The ability of fine-tuned BERT to conduct textual feature extraction helps the model produce rich contextual embeddings that capture subtle linguistic cues related to deception, such as exaggerated sentiment, vague assertions, unnatural phrasing, and inconsistencies in the flow of a narrative. These are contextualized representations that greatly enhance the model's sensitivity to deceptive writing patterns that too often mimic genuine reviews. Attention mechanisms further improve the effectiveness of the framework by allowing for a fine grained analysis of both the textual and the visual features. In the textual modality, attention weights emphasize words and phrases that play a key role in the classification result, which reveals linguistic features that indicate manipulation. In the case of the visual modality, attention is drawn to salient image regions that might indicate the inauthenticity or inconsistency of the image. This selective emphasis not only helps to improve the accuracy of detection but also helps to make it more interpretable, so the model's decision making process is more transparent and explainable. Such interpretability is especially valuable in the real world where trust and accountability are paramount.

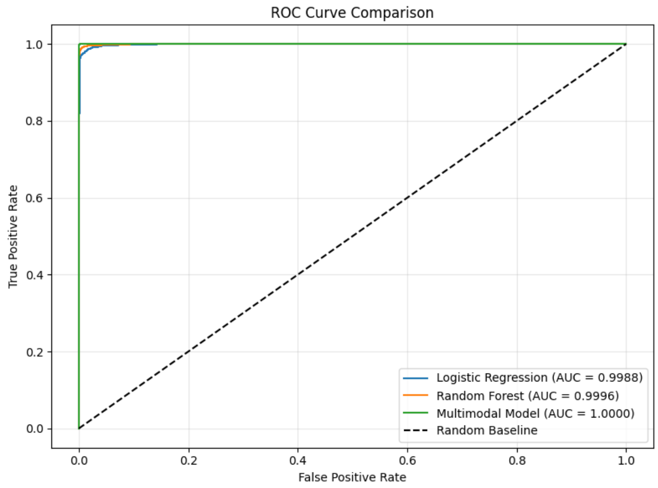

Figure 5 ROC curve of the multimodal fake review detection model: The figure indicates the ROC curve of the multimodal model that has been applied in detecting fake review.

Comparative evaluation in terms of standard performance metrics such as accuracy, precision, recall, F1 score shows the better robustness and consistency across different datasets and types of reviews. The performance improvements are particularly pronounced in the detection of AI-generated and highly sophisticated fake reviews, which often escape traditional detection methods because of their fluency in language and coherence in context. The multimodal fusion strategy mitigates false positives and false negatives by capturing the cross-modal correlations to achieve more reliable predictions. Overall, these results confirm the use of multimodal learning as well as fine grained attention-based analysis significantly helps in strengthening fake review detection capabilities and contributing to a scalable, effective solution to preserve trust and credibility in e-commerce platforms.

VI. DISCUSSION

The results of this research are solid reasons to believe that the combination of multimodal learning and fine-grained attention mechanisms is a powerful way to improve the detection of sophisticated fake reviews in e commerce settings. By the joint processing of textual and visual data, the proposed framework solves basic limitations of traditional single modality models that are only based on linguistic cues. Text-only approaches often don't work when deceptive reviews are carefully designed with a fluent use of language or generated by advanced artificial intelligence models. In contrast, the multimodal design allows the system to detect inconsistency between review text and associated images which sheds light on the coordinated manipulation strategies that are becoming very common in the modern generation of fake reviews. The addition of attention mechanisms is an important aspect for the performance of the detection as well as to promote a more interpretable model. Attention enables the model to dynamically highlight subtle patterns in the language, such as such as the use of exaggerated sentiment, the use of vague descriptions, repetitive phrasing, and unnatural word usage, which are powerful indicators of deceptive intent. On the visual side, attention mechanisms point out suspicious areas in the image, such as watermarks, repeated stock images or visual elements that are not consistent with the text story. This fine-grained focus allows the framework to identify subtle deceptive signals that are frequently ignored by traditional deep learning models, especially when reviews seem superficially real. In addition to performance benefits, the interpretability provided by attention-based analysis adds value to the practical applicability of the proposed approach. By determining which textual expressions and visual features are most influential to classification decisions, the framework offers transparency that is valuable for the administrators of such platforms, for moderators to ensure that such platforms comply with relevant regulations, and for regulatory stakeholders such as advertising regulators. Such explainability helps to support informed decision making, error analysis and promotes trust in automated fake review detection systems. Moreover, the fact that the results of detection can be traced back to specific features, the system is more suitable for deployment in an environment where accountability and fairness are of importance. The multimodal architecture also aids in resilience against evolving deceptive strategies, which is an essential requirement in light of the rapid advancement of generative AI technologies. As fake reviews are becoming more contextually coherent, as well as visually convincing, detection systems must adapt to new forms of manipulation. The holistic feature integration used in this framework has the advantage of being more generalized across different review styles, product classes and manipulation techniques. Overall, discussion can be concluded that scalable and attention-driven multimodal architectures are a robust and forward-looking solution for providing trust, credibility, and transparency in today's e-Commerce platforms.

A. Improving F1-Score and the Detection Performance

The F1-score, which balances precision and recall, is especially important in the detection of fake reviews because of asymmetric costs of false positives and false negatives. Improving the F1-score of the proposed multimodal framework is possible by several approaches such as methodological and experimental improvements. First, better working with class imbalance can greatly improve recall without degrading precision. Fake review datasets tend to be of the genuine variety, however, which biases models towards the majority class. Techniques such as class-weighted loss functions, focal loss and adaptive resampling strategies can be used to ensure that minority class learning is emphasized. These approaches help the model better figure out difficult to detect fake reviews and this helps to increase the harmonic balance captured by the F1-score. Second, it is possible to fine tune strategies for the textual and visual encoders, further enhancing the quality of representations. While BERT and ResNet-50 contain good pretrained features, task-specific fine-tuning, with the use of lower learning rates, gradual unfreezing, and layer-wise learning rate decay, can help maintain general knowledge while being better able to adapt to deception-specific cues. This leads to feature discrimination, which directly leads to increased precision and recall. Third, the optimization of the multimodal fusion and attention mechanisms can improve cross-modal consistency learning. Introducing gated fusion or attention based weighting between textual and visual embeddings, to allow the model to dynamically emphasize the more informative modality for each review instance. Such adaptive fusion alleviates noise of weak or irrelevant modalities and enhances decision borders which leads to better classification robustness and F1-score. Fourth, threshold optimization during inference plays an important role in maximizing FI performance. Instead of having a fixed value for the probability threshold (e.g. 0.5), validation-based threshold tuning can be used to find the optimal decision boundary that maximizes the F1 score. This is especially good in imbalanced classification problems and helps to get more reliable deployment in the real world. Finally, it is possible to enhance generalization by expanding the training dataset with high-quality synthetic and adversarial samples. Controlled augmentation via artificial intelligence-generated fake reviews, when used together with real human-written reviews, expose the model to various patterns of deception. This diversity helps the model to better discriminate between subtle manipulations, so more precise with both less false positive and fewer false negatives. Overall, by leveraging imbalance-aware training, encoder fine-tuning, adaptive multimodal fusion, inferential threshold optimization, and data augmentation approaches, the proposed framework can attain higher F1-scores and enhanced detection accuracy. These improvements make the model more robust to sophisticated and AI-generated fake reviews, which would be more suitable for large-scale deployment on real-world e-commerce platforms.

VII. CONCLUSION AND FUTURE WORK

This study introduced a unified trimodal deep learning model of detecting fake reviews by fusing textual, visual and 22 engineered numeric characteristics using attention-based fusion mechanism. The suggested method was successful in acquiring the complementary information of the three modalities and indicated good results in the chosen Kaggle multimodal fake review dataset. The experimental findings revealed that the model had an accuracy of 99.95, F1-score of 99.95 and ROC-AUC of 99.99, which was better than the baseline models such as Logistic Regression (98.39% accuracy) and Random Forest (99.08% accuracy). These results prove the efficiency of attention-focused trimodal learning to detect the false reviews in online shopping conditions. The proposed framework can be added to a real-time deployment pipeline as an extension in the future to support the practical application in e-commerce platforms using application serving and scalable container-based technologies.

REFERENCES

[1] Al-Adhaileh, M.H. et al. (2021) "Deep Learning Algorithms for Detection and Classification of Gastrointestinal Diseases," Complexity, 2021(1). p. 02. doi:10.1155/2021/6170416.
[2] Demirbaş, A.A., Üzen, H. and Fırat, H. (2024) "Spatial-attention ConvMixer architecture for classification and detection of gastrointestinal diseases using the Kvasir dataset," Health Information Science and Systems, 12(1). p. 03.doi:10.1007/s13755-024-00290-x.
[3] Diptho, R.A. and Basak, S. (2025) "Enhancing Dermatological Diagnosis Through Medical Image Analysis: How Effective Is YOLO11 Compared to Leading CNN Models?," NDT, 3(2), p. 11. doi:10.3390/ndt3020011. Duma, R.A. et al. (2024) "Fake review detection techniques, issues, and future research directions: a literature review," Knowledge and Information Systems. Springer Science+Business Media, p. 5071. doi:10.1007/s10115024-02118-2.
[4] El-Ghany, S.A., Mahmood, M.A. and El-Aziz, A.A.A. (2024) "An Accurate Deep Learning-Based ComputerAided Diagnosis System for Gastrointestinal Disease Detection Using Wireless Capsule Endoscopy Image Analysis," Applied Sciences, 14(22), p. 10243. doi:10.3390/app142210243.
[5] Fröhnel, K., Santelmann, B. and Zarnekow, R. (2025) "Genuine or Fake? Explaining Consumers' Perception and Detection of AI Generated Fake Reviews," in Proceedings of the ... Annual Hawaii International Conference on System Sciences/Proceedings of the Annual Hawaii International Conference on System Sciences. doi:10.24251/hicss.2025.505.
[6] Gambetti, A. and Han, Q. (2023) "Combat AI With AI: Counteract Machine-Generated Fake Restaurant Reviews on Social Media," arXiv (Cornell doi:10.48550/arxiv.2302.07731.
[7] Gambetti, A. and Han, Q. (2024) "AiGen-FoodReview: A Multimodal Dataset of Machine-Generated Restaurant

[8] Reviews and Images on Social Media," Proceedings of the International AAAI Conference on Web and Social Media, 18, p. 1935. doi:10.1609/icwsm.v18i1.31437.

[9] Geetha, S. et al. (2024) "MBO-DeBERTa : High Performance Fake Review Detection using pretrained DeBERTa Optimized with Monarch Butterfly paradigm," Research Square (Research Square) [Preprint]. doi:10.21203/rs.3.rs-5302560/v1.

[10] Geetha, S. et al. (2025) "High performance fake review detection using pretrained DeBERTa optimized with Monarch Butterfly paradigm," Scientific Reports, 15(1). doi:10.1038/s41598-025-89453 8.

[11] Gupta, R., Jindal, V. and Kashyap, I. (2024) "Recent state-of-the-art of fake review detection: a comprehensive review," The Knowledge Engineering Review. Cambridge University Press. doi:10.1017/s0269888924000067. Hou, J. et al. (2025) "Detecting fake review intentions in the review context: A multimodal deep learning approach," Electronic Commerce Research and Applications. Elsevier BV, p. 101485. doi:10.1016/j.elerap.2025.101485.

[12] Hovy, D. (2016) "The Enemy in Your Own Camp: How Well Can We Detect Statistically-Generated Fake Reviews – An Adversarial Study." doi:10.18653/v1/p16-2057.

[13] Jayasinghe, J.A.P. and Dassanayaka, S. (2024) "Detecting Deception: Employing Deep Neural Networks for Fraudulent Review Detection on Amazon," Research Square (Research Square) [Preprint]. doi:10.21203/rs.3.rs5214171/v1. Khalif, S. and Mane, K. (2024) "Exploring Machine Learning and Deep Learning Techniques for Fake Review Detection: A Comprehensive Literature Review," Deleted Journal, p. 1669. doi:10.47392/irjaeh.2024.0230. Kim, J.M. et al. (2023) "Investigating reviewers' intentions to post fake vs. authentic reviews based on behavioral linguistic features," Technological Forecasting and Social Change, 198, p. 122971.

[14] Kumar, K. et al. (2023) "A Brief Study on the Fake Review Detection methods on Ecommerce Websites using Machine Learning, Artificial Intelligence, and Data Science," International Journal for Research in Applied Science and Engineering Technology, 11(9), p. 721. doi:10.22214/ijraset.2023.52743. Large Language Models in Cybersecurity (2024). doi:10.1007/978-3-031-54827-7.

[15] Long, S. et al. (2025) "From whales to waves: Social media sentiment, volatility, and whales in cryptocurrency markets," The British Accounting doi:10.1016/j.bar.2025.101682.

[16] Meng, W. et al. (2025) "Large Language Models as 'Hidden Persuaders': Fake Product Reviews are Indistinguishable to Humans and Machines." doi:10.48550/ARXIV.2506.13313.

[17] Qiu, F. et al. (2021) "Deep Just-In-Time Defect Localization," IEEE Transactions on Software doi:10.1109/tse.2021.3135875.

[18] Salminen, J. et al. (2025) "Decoding deception in the online marketplace: enhancing fake review detection with psycholinguistics and transformer models," Journal of Marketing Analytics [Preprint]. doi:10.1057/s41270-02500393-8.

[19] Su, Z. et al. (2025) "A multigrained preference analysis method for product iterative design incorporating AIgenerated review detection," Scientific Reports, 15(1). doi:10.1038/s41598-025-86551-5.

[20] Sun, P. et al. (2024) "Fake Review Detection Model Based on Comment Content and Review Behavior," Electronics, 13(21), p. 4322. doi:10.3390/electronics13214322.

[21] Atrey P K, Hossain M A, El Saddik A, et al. Multimodal fusion for multimedia analysis: a survey[J]. Multimedia Systems, 2010, 16(6): 345-379 https://doi.org/10.1007/s00530-010 0182-0

[22] Su Z, Yin H, Liu Y, et al. A multigrained preference analysis method for product iterative design incorporating AI-generated review detection[J]. Scientific Reports, 2025, 15(1): 5871 https://doi.org/10.1038/s41598-025-87547-z

[23] Pennington J, Socher R, Manning C D. GloVe: global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014: 1532-1543 https://aclanthology.org/D14-1162/

[24] Flesch R. A new readability yardstick[J]. Journal of Applied Psychology, 1948, 32(3): 221 233 https://doi.org/10.1037/h0057532

[25] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 248-255 https://doi.org/10.1109/CVPR.2009.5206848

[26] He H, Garcia E A. Learning from imbalanced data[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263-1284 https://doi.org/10.1109/TKDE.2008.239

[27] Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019: 3645-3650 https://aclanthology.org/P19-1355/

[28] Lipton Z C. The mythos of model interpretability[J]. Queue, 2018, 16(3): 31-57 https://doi.org/10.1145/3236386.3241340

[29] Savage D, Zhang X, Yu X, et al. Anomaly detection in online social networks[J]. Social Networks, 2014, 39: 62-70 https://doi.org/10.1016/j.socnet.2014.05.002

[30] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10):1345-1359 https://doi.org/10.1109/TKDE.2009.191
[31] Lu J, Batra D, Parikh D, et al. ViLBERT: pretraining task-agnostic visiolinguistic representations for vision-and-language tasks[C]// Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver: NeurIPS, 2019: 13-23 https://arxiv.org/abs/1908.02265
[32] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C]// Proceedings of the 5th International Conference on Learning Representations. Toulon: ICLR, 2017: 1-14 https://arxiv.org/abs/1609.02907