iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65676

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

Detecting Fake Comments Using AI (Artificial Intelligence)

Rucha A. Desai[1], Prof. Mukul Jagtap[2]
[1]*Student, Keystone School of Engineering, Pune*
[2]*Professor, Keystone School of Engineering, Pune*

Abstract: ***In today's digital landscape, online comments can make or break reputations. But what if those comments are fake? Artificially inflated or deflated, fake comments can distort public opinion, undermine trust, and wreak havoc on businesses and individuals alike. To combat this growing menace, we can take the help of Artificial Intelligence (AI). By leveraging machine learning, natural language processing, and deep learning, we can develop sophisticated systems to detect and weed out fake comments. Let's delve into the world of AI-powered fake comment detection by exploring the latest techniques, from sentiment analysis to behavioral modeling, and examine real-world case studies of successful implementations.***
Keywords: Fake Comment Detection, AI-Powered Moderation, Online Deception Analysis, Social Media Authentication, Machine Learning for Comment Filtering.

I. INTRODUCTION

Growth in the Global Internet market has led to rapid online traffic. Growth of online social media platforms has enabled users to share their opinions and engage in discussions with others. However, this increased online activity has also led to the proliferation of fake comments, which can spread misinformation, manipulate public opinion, and undermine trust in online communities. Fake comments can take many forms, including spam, phishing, and propaganda, and can be generated by humans or automated bots.

Artificial Intelligence (AI) has emerged as a promising solution for detecting fake comments. Various AI-powered techniques such as machine learning, deep learning and natural language processing, can analyze comment text, user behavior and network patterns to identify fake comments. Previous studies have shown significant promise in detecting fake comments with some of them reporting accuracy rates of over 90%.

In the early 2000's Fake comment detection began with simple rule-based systems, using keyword filtering and IP blocking. By 2010's (early era of ML) Machine learning algorithms, such as logistic regression and decision trees, were applied to detect fake comments. In recent years (2015-Present) Deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have significantly improved fake comment detection accuracy.

Despite the growing interest in AI-powered fake comment detection, there is a need for a comprehensive review of the existing literature. This survey aims to provide an overview of the current state of AI-powered fake comment detection, including the techniques, datasets, and evaluation metrics used. We also identify the challenges and limitations of existing approaches and highlight future research directions.

II. PRELIMINARY

The detection of fake comments is a challenging task, as they often mimic the characteristics of genuine comments. Traditional methods for detecting fake comments, such as keyword filtering and IP blocking, have proven to be ineffective against sophisticated fake comment generators.

Here is an overview of how fake comments can be detected using an AI-powered model

1) Step 1: Data Collection

- Collect a large dataset of comments from various sources, including social media platforms, online forums, and review websites.
- Ensure the dataset includes both genuine and fake comments.

2) Step 2: Data Preprocessing

- Preprocess the collected data by:
- Tokenizing the comments into individual words or phrases.

©IJRASET: All Rights are Reserved

- Removing stop words, punctuation, and special characters.
- Converting all text to lowercase.
- Removing duplicate comments.

3) *Step 3: Feature Extraction*
- Extract relevant features from the preprocessed data, such as:
- Sentiment analysis: Determine the sentiment of the comment (positive, negative, or neutral).
- Syntax and semantics: Analyze the comment's syntax and semantics to identify patterns and anomalies.
- User behavior: Analyze the user's behavior, such as their posting history and engagement patterns.
- Network analysis: Analyze the comment's position within the network, including its relationships to other comments and users.

4) *Step 4: Model Selection*
- Select a suitable AI algorithm for fake comment detection, such as:
- Machine learning algorithms: Support Vector Machines (SVM), Random Forest, or Gradient Boosting.
- Deep learning algorithms: Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or Long Short-Term Memory (LSTM) networks.

5) *Step 5: Model Training*
- Train the selected AI algorithm using the preprocessed data and extracted features.
- Split the dataset into training, validation, and testing sets.
- Use techniques like cross-validation and hyperparameter tuning to optimize the model's performance.

6) *Step 6: Model Evaluation*
- Evaluate the trained model using metrics such as:
- Accuracy
- Precision
- Recall
- F1-score
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
- Compare the model's performance with baseline models or other state-of-the-art models.

7) *Step 7: Model Deployment*
- Deploy the trained model in a real-world setting, such as:
- Integrating the model with social media platforms or online forums.
- Using the model as a plugin for content management systems.
- Creating a standalone application for fake comment detection.

8) *Step 8: Continuous Monitoring and Improvement*
- Continuously monitor the model's performance in real-world settings.
- Collect feedback from users and update the model accordingly.
- Refine the model by incorporating new features, updating the algorithm, or using transfer learning.

III. ADVANTAGES

1) *Improved Accuracy:* AI-powered fake comment detection can achieve higher accuracy rates compared to traditional methods, which rely on manual moderation or simple keyword filtering.
2) *Increased Efficiency:* AI can analyze large volumes of comments quickly and efficiently, reducing the need for manual moderation and saving time and resources.
3) *Scalability:* AI-powered fake comment detection can be easily scaled up or down to accommodate changing volumes of comments, making it an ideal solution for large online communities.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

4) *Real-time Detection:* AI can detect fake comments in real-time, preventing them from spreading misinformation or causing harm to individuals or organizations.
5) *Reduced False Positives:* AI-powered fake comment detection can reduce the number of false positives, which can help to minimize the risk of mistakenly removing genuine comments.
6) *Improved User Experience:* By removing fake comments, AI-powered fake comment detection can improve the overall user experience, making online communities safer and more trustworthy.
7) *Enhanced Reputation Management:* AI-powered fake comment detection can help organizations to protect their reputation by removing fake comments that may be damaging to their brand.
8) *Compliance with Regulations:* AI-powered fake comment detection can help organizations to comply with regulations, such as the General Data Protection Regulation (GDPR), by removing fake comments that may be in violation of these regulations.
9) *Cost Savings:* AI-powered fake comment detection can help organizations to save costs by reducing the need for manual moderation and minimizing the risk of reputational damage.
10) *Data-Driven Insights:* AI-powered fake comment detection can provide organizations with data-driven insights into the nature and scope of fake comments, which can inform their content moderation strategies and improve their overall online presence

IV. DISADVANTAGES

1) *Bias in AI Models:* AI models can be biased if they are trained on biased data, which can lead to inaccurate results and unfair treatment of certain groups.
2) *False Positives:* AI-powered fake comment detection can result in false positives, where genuine comments are mistakenly identified as fake.
3) *False Negatives:* AI-powered fake comment detection can also result in false negatives, where fake comments are mistakenly identified as genuine.
4) *Over-Reliance on Technology:* Over-reliance on AI-powered fake comment detection can lead to a lack of human oversight and judgment, which can result in inaccurate results.
5) *Evasion Techniques:* Sophisticated fake comment generators can use evasion techniques, such as code obfuscation or using multiple accounts, to evade detection by AI-powered fake comment detection systems.
6) *High Computational Costs:* Training and deploying AI models for fake comment detection can require significant computational resources and can be costly.
7) *Data Quality Issues:*AI-powered fake comment detection requires high-quality data to train and validate the models. Poor data quality can lead to inaccurate results.
8) *Lack of Transparency:* AI-powered fake comment detection systems can be opaque, making it difficult to understand how they arrive at their decisions.
9) *Dependence on Data:* AI-powered fake comment detection systems are only as good as the data they are trained on. If the data is biased or incomplete, the system's performance will suffer.
10) *Need for Continuous Updates:* AI-powered fake comment detection systems require continuous updates to stay effective, as new types of fake comments and evasion techniques emerge.

V. CONCLUSION

Detecting fake comments with the help of AI can be game changer in maintaining authenticity of online interactions. The key to effective fake comment detection lies in the use of advanced machine learning and deep learning techniques such as Generative Adversarial Networks (GANs) and Transformers. These techniques help AI models to learn anomalies in language that are indicative of fake comments.

Sentimental Analysis also plays a crucial role in fake comment detection. By analyzing the sentiment of the comments, AI models can identify comments that are attempting to manipulate public opinion or spread misinformation.

With advantages, the AI-powered model also has limitations. Accuracy of the AI-powered model can be compromised by biased training data and the use of evasion techniques by sophisticated fake comment generators.

Despite these challenges, by leveraging advanced machine learning and deep learning techniques, we can create more authentic and trustworthy online communities.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] H. Aghakhani, Aravind Machiry, Shirin Nilizadeh, Christopher Krügel and Giovanni Vigna, “Detecting Deceptive Reviews Using Generative Adversarial Networks”, Computer Science 2018 IEEE Security and Privacy Workshops (SPW) DOI:10.1109/SPW.2018.00022, Corpus ID: 44115116.

[2] R. Mohawesh et al., "Fake Reviews Detection: A Survey," in IEEE Access, vol. 9, pp. 65771-65802, 2021, doi: 10.1109/ACCESS.2021.3075573.

[3] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar and M. S. Rahman, "A Comprehensive Review on Fake News Detection With Deep Learning," in IEEE Access, vol. 9, pp. 156151-156170, 2021, doi: 10.1109/ACCESS.2021.3129329.

[4] M. Abdulqader, A. Namoun and Y. Alsaawy, "Fake Online Reviews: A Unified Detection Model Using Deception Theories," in IEEE Access, vol. 10, pp. 128622-128655, 2022, doi: 10.1109/ACCESS.2022.3227631.

[5] Abhijeet A Rathore , Gayatri L Bhadane , Ankita D Jadhav , Kishor H Dhale, Jayshree D Muley, 2023, Fake Reviews Detection Using NLP Model and Neural Network Model, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 12, Issue 05 (May 2023).

[6] Pengfei Sun, Weihong Bi, Yifan Zhang and Qiuyu Wang, "Fake Review Detection Model Based on Comment Content and Review Behavior" November 2024 Electronics 13(21):4322, DOI:10.3390/electronics13214322

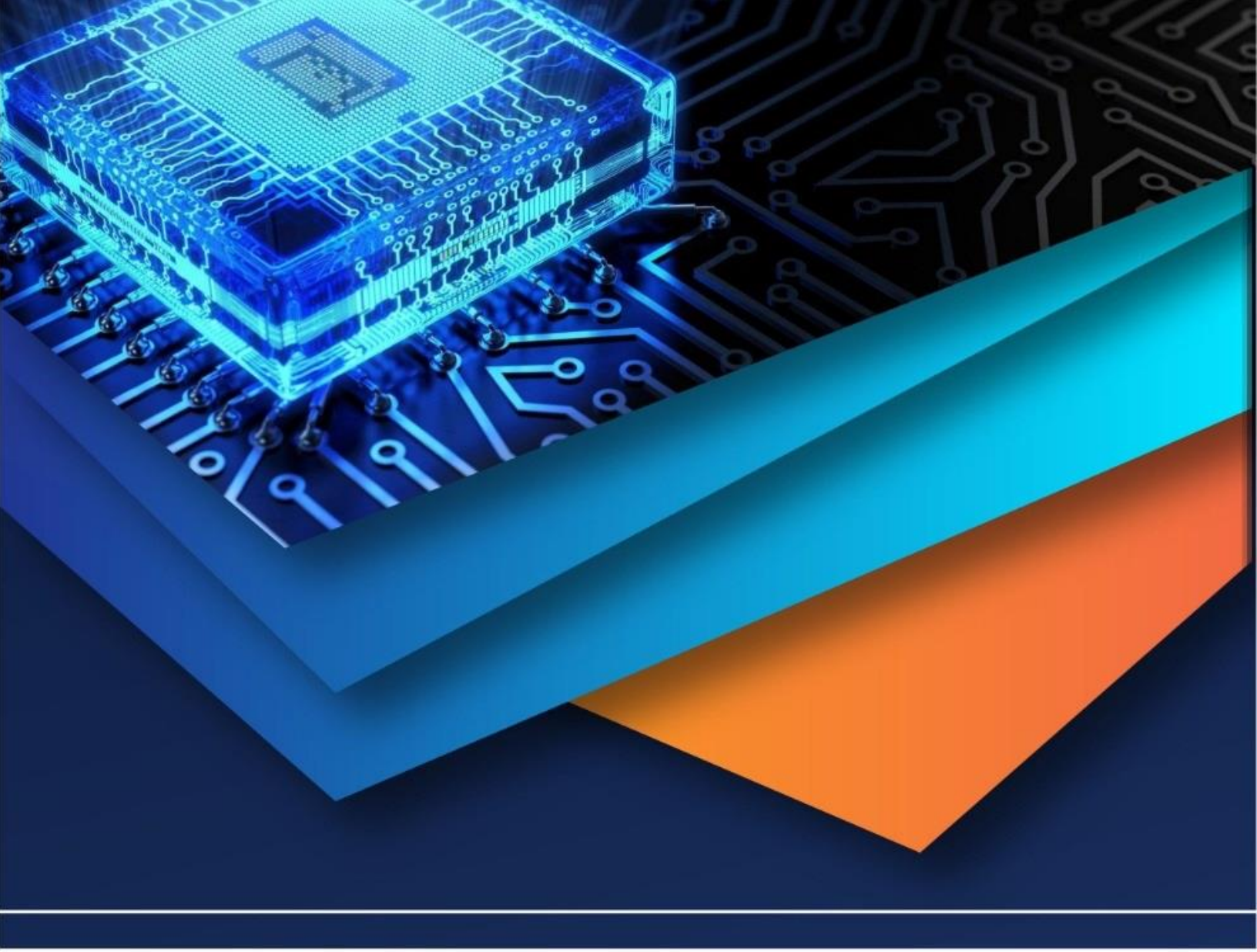
[7] Gupta, R., Kashyap, I., Jindal, V. (2024). A Comprehensive Survey on Fake Review Detection System with Future Directions. In: Santosh, K., Nandal, P., Sood, S.K., Pandey, H.M. (eds) Advances in Artificial-Business Analytics and Quantum Machine Learning. COMITCON 2023. Lecture Notes in Networks and Systems, vol 1053. Springer, Singapore. https://doi.org/10.1007/978-981-97-4860-0_1

[8] Jinhao Liu, Pei Quan, Wen Zhang, “A Study on Fake Review Detection Based on RoBERTa and Behavioral Features”, Procedia Computer Science, Volume 242, 2024, Pages 1323-1330, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2024.08.131. (https://www.sciencedirect.com/science/article/pii/S1877050924018507)

[9] Rami Mohawesh, Haythem Bany Salameh, Yaser Jararweh, Mohannad Alkhalaileh, Sumbal Maqsood, Fake review detection using transformer-based enhanced LSTM and RoBERTa, International Journal of Cognitive Computing in Engineering, Volume 5, 2024, Pages 250-258, ISSN 2666-3074, https://doi.org/10.1016/j.ijcce.2024.06.001.

[10] Maneerat Puttarattanamanee, L. Boongasame, Karanrat Thammarak, “A Comparative Study of Sentiment Analysis Methods for Detecting Fake Reviews in E-Commerce”, 1 June 2023, Computer Science, Business DOI:10.28991/hij-2023-04-02-08 Corpus ID: 264069226.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)