DETECTION OF FAKE PROFILES AND BOTNETS IN INDIAN SOCIAL NETWORKS USING GRAPH NEURAL NETWORKS

[1]Cherukupalli Harshitha
[2]T. Deepthi
ASSOCIATE PROFESSOR
DEPARTMENT OF ARTIFICAL INTELLIGENCE & MACHINE LEARNING
KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES,
DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

ABSTRACT

The rapid growth of social networking platforms in India has led to a significant rise in fake profiles, automated bot accounts, and coordinated misinformation campaigns. These malicious entities pose serious threats to user privacy, public opinion, digital trust, and national security. Traditional detection techniques—such as rule-based filtering, heuristic analysis, and classical machine learning models—struggle to accurately identify sophisticated fake accounts due to evolving behavioral patterns and the complex relational structures within social networks. To address these challenges, this study proposes a Graph Neural Network (GNN)-based framework for detecting fake profiles and botnets in Indian social media ecosystems.

The proposed model leverages the graph-like structure of social networks, where nodes represent user accounts and edges represent interactions such as likes, comments, messages, and follows. Using GNN architectures such as Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), and GraphSAGE, the system learns both structural and behavioral embeddings to distinguish between genuine and fraudulent accounts. Key features—including activity frequency, interaction diversity, temporal patterns, linguistic cues, and follower-following relationships—are integrated to improve detection accuracy.

Experimental evaluation on real-world and synthetic datasets demonstrates that GNN-based models outperform traditional machine learning classifiers in identifying bot clusters, anomalous communication patterns, and fake-profile communities. The framework achieves higher precision, recall, and F1-scores, especially in highly imbalanced datasets common in fraud detection tasks. The system's scalability and adaptability make it suitable for large Indian platforms such as ShareChat, Moj, and region-specific social networks.

Overall, this research highlights the effectiveness of Graph Neural Networks in combating online manipulation by uncovering hidden patterns and relational anomalies within social graphs. The proposed approach provides a robust, intelligent solution for enhancing digital safety and maintaining user trust in Indian social media environments.

KEYWORDS: Graph Neural Networks (GNN), Fake Profile Detection, Botnet Detection, Social Network Analysis, Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), GraphSAGE, Anomaly Detection, User BehaviorModeling, Indian Social Networks, Deep Learning, Misinformation Detection.

I. INTRODUCTION

The widespread adoption of social networking platforms in India has transformed the way people communicate, share information, and engage with digital content. Platforms such as Facebook, Instagram, Twitter, ShareChat, Moj, and WhatsApp host millions of active users, creating vast digital ecosystems that connect individuals across diverse linguistic, cultural, and demographic backgrounds.

However, alongside this rapid expansion, social networks have also witnessed an alarming increase in fake profiles, automated bot accounts, and coordinated botnets. These malicious accounts manipulate online discourse, spread misinformation, engage in political propaganda, conduct phishing attacks, and distort public sentiment, posing serious challenges to digital trust and national cybersecurity.

Traditional techniques for detecting fake profiles—including rule-based filters, heuristic scoring, and classical machine learning—often rely on manually crafted features and simplistic classification strategies. While these techniques can identify basic fraudulent behaviors, they are ineffective against modern botnets and fake accounts that mimic human-like interactions and operate in coordinated patterns. Moreover, social networks are inherently graph-structured, where user behavior is tightly coupled with their relationships and interactions. Classical detection models fail to capture these complex relational dependencies and cannot effectively analyze community-level anomalies or hidden bot clusters.

Graph Neural Networks (GNNs) have emerged as a powerful tool for modeling graph-structured data, making them highly suitable for social network analysis. GNN architectures such as Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), and GraphSAGE can learn both structural and behavioral representations by aggregating information from neighboring nodes. This enables them to detect subtle anomalies, unusual interaction patterns, and coordinated bot behaviors that traditional methods overlook.

This work focuses on designing a GNN-based framework tailored for Indian social networks, considering platform-specific behaviors, multilingual content, cultural interaction patterns, and region-based communication clusters. By integrating relational features, temporal activities, and network topology, the proposed system aims to provide a robust and scalable solution for identifying fake profiles and botnets, thereby enhancing digital safety and maintaining user trust in India's rapidly evolving social media landscape.

II. LITERATURE REVIEW

Graph Neural Networks (GNNs) have emerged as a powerful paradigm for detecting malicious accounts and coordinated botnets by leveraging relational structure and neighbourhood patterns in social graphs. Savani & Shanbhag (2025) provide a comprehensive survey of GNN methods applied to fake-account detection and summarize how message propagation, structural embeddings, and node-level features can be combined to reveal anomalous behavior [1]. This survey situates GNNs as an evolution beyond feature-only approaches by emphasizing graph topology and higher-order interactions as key signals for distinguishing genuine users from bots or sybil clusters.

Several studies apply specialized GNN architectures to real-world social media datasets. Roy et al. (2024) designed a GNN pipeline tailored for botnet detection in social media, showing that modeling interaction graphs and diffusion paths improves detection rates compared to conventional classifiers that ignore relational context [2]. Jain et al. (2023) demonstrated the efficacy of attention-based GNNs (Graph Attention Networks) for large-scale OSN graphs, where attention mechanisms highlight suspicious edges and coordinated activity patterns that commonly accompany botnets [6]. These works collectively show that attention and message-passing schemes are particularly effective at isolating coordinated malicious behavior.

Domain-specific adaptations of GNNs for Indian social networks have also been

proposed. Kumar et al. (2024) introduced a deep graph convolutional model tuned for Twitter India, accounting for regional language features and platform-specific behaviour; their evaluation indicates improved detection precision when linguistic signals are fused with graph embeddings [4]. Bharadwaj et al. (2025) extended this trend by using heterogeneous GNNs to model multiple node and edge types (users, posts, hashtags), demonstrating that heterogeneous relational modeling captures complex interactions common in Indian social ecosystems and yields stronger discrimination between organic and inauthentic accounts [5].

Hybrid and scalable approaches address the computational challenges of massive OSN graphs. Mishra & Sinha (2024) proposed a hybrid GraphSAGE variant that samples local neighborhoods and combines them with traditional feature engineering to detect misinformation-spreaders and fake profiles efficiently on large Twitter subgraphs [7],[11]. Verma & Yadav (2023) similarly combined graph embeddings with deep learning classifiers to balance detection performance with throughput requirements on real traffic volumes [10],[12]. These hybrid strategies suggest practical deployment paths where GNN expressivity is retained while keeping inference feasible at scale.

Community and cluster-based techniques complement node-level GNN methods for botnet discovery. Chakraborty et al. (2024) used graph clustering together with GNN embeddings to identify densely connected bot clusters and coordinated campaigns; their approach is effective at revealing botnets that deliberately mimic local topology to evade individual node detectors [9]. Such community-aware pipelines highlight that combining macro (community) and micro (node/edge) perspectives strengthens detection of coordinated networks that single-node models miss.

The literature also identifies key operational and research challenges. Roy et al. and other authors note the dynamic and temporal nature of bot behavior—bots change tactics over time, necessitating temporal graph models and continual learning [2,6]. Patel & Iyer (2025) emphasize temporal GNNs for coordinated bot behavior detection, arguing that temporal patterns (synchronized posting, bursty interactions) are strong indicators of botnets and that static GNNs miss these cues [8],[13],[14]. Finally, the body of work repeatedly highlights data scarcity and label noise: obtaining reliably labeled Indian-language instances and ground-truth botnets remains difficult, which complicates supervised GNN training and evaluation [3,5].

In summary, the reviewed studies demonstrate that GNNs—especially attention mechanisms, heterogeneous and temporal variants, and hybrid sampling/embedding pipelines—are highly effective for detecting fake profiles and botnets in social networks. Combining node-level features (textual and behavioral) with graph structure and community signals yields robust detection, but challenges remain in temporal modeling, scalability, multilingual data, and real-world deployment.

(Cited works: Savani & Shanbhag [1]; Roy et al. [2]; Sarkar & Gupta [3]; Kumar et al. [4]; Bharadwaj et al. [5]; Jain et al. [6]; Mishra & Sinha [7]; Patel & Iyer [8]; Chakraborty et al. [9]; Verma & Yadav [10].)

III. EXISTING SYSTEM

The existing system for detecting fake profiles and botnets in Indian social networks primarily relies on **manual inspection and traditional machine-learning approaches** that analyze user-level attributes rather than network-level behavioral patterns. In current systems, social network platforms typically evaluate profile information such as username,

age, gender, link description, friend count, and status updates in isolation. These systems use **rule-based validation** (e.g., suspicious usernames, incomplete profiles, abnormal friend requests) or **simple classifiers** that do not leverage graph-structured data. As a result, detection is limited because bots often mimic legitimate user attributes, making attribute-only techniques ineffective against sophisticated fake profiles.

The existing system's workflow is highly dependent on **static profile metadata** and heuristic indicators. For example, fields such as "Account Age," "Friend Count," "Status," and "Changed WhatsApp/Internet usage" are assessed individually to identify abnormalities. This approach does not consider how users interact within the network, such as message propagation, engagement patterns, followers–friends relationships, or clustering behavior—all of which are crucial for identifying coordinated botnets. Because the system lacks network-aware analysis, it struggles to distinguish between isolated fake accounts and large, coordinated bot networks.

Additionally, conventional systems do not incorporate **deep learning models** that can automatically learn patterns from user interactions. Instead, they often use **basic classifiers** such as logistic regression, SVM, or decision trees, which require handcrafted features and offer limited adaptability. These methods lack the capability to model higher-order connections or detect community-level manipulation. As a result, many bots evade detection by simply adjusting profile settings or performing minimal "human-like" interactions.

The existing system also lacks real-time monitoring capabilities. Users are evaluated only when specific account attributes are flagged or when manual verification is requested. This reactive approach enables botnets to operate for extended periods before being detected. Furthermore, the absence of **graph neural network–based learning**, **temporal behavior tracking**, or **anomaly detection over user interaction graphs** severely restricts the system's ability to detect coordinated, evolving, and large-scale bot operations.

IV. PROPOSED SYSTEM

The proposed system introduces an intelligent, automated, and graph-structured approach for detecting fake profiles and botnets in Indian social networks using **Graph Neural Networks (GNNs)**. Unlike traditional systems that rely only on isolated user attributes, the proposed system models each user as a **node in a social graph**, capturing relationships such as followers, friends, interactions, message exchanges, and content-sharing behavior. By learning patterns from the topology and dynamics of these connections, the GNN-based model can accurately identify suspicious users, coordinated bot clusters, and synthetic accounts that exhibit non-human interaction patterns.

At the core of the proposed system is a **Graph Convolutional or Graph Attention Network** that learns embeddings from multiple sources of information—profile metadata, behavioral features, interaction graphs, and temporal communication patterns. Instead of manually engineering features, the system automatically discovers hidden structural signals such as sudden increases in connectivity, densely connected bot communities, repeated message propagation, and synchronized activity. This allows the model to distinguish between legitimate users and fake accounts even when profile-level indicators appear normal or manipulated.

The system incorporates a **multi-stage detection pipeline** for scalability and high accuracy. First, raw dataset inputs such as username, account age, gender, link description, status, friend count, internet

activity, and WhatsApp usage are preprocessed and converted into graph-compatible formats. Second, a GNN training module analyzes both node-level (individual user) and edge-level (user-to-user interaction) features to produce a risk score for each user. Finally, the system applies a prediction layer that classifies accounts into categories such as *genuine*, *bot*, *fake profile*, or *suspected coordinated account*. This architecture ensures robust detection of both isolated fake accounts and large botnets acting together.

The proposed system also supports **real-time monitoring and visualization**. The interface allows users to upload datasets, run the GNN algorithm, and view classification results instantly. Detected fake accounts are highlighted, and administrators can take action, such as account blocking or additional verification. The use of GNNs strengthens the detection process by identifying not just suspicious users but also the *relationships* that drive coordinated attacks—something that traditional models cannot achieve.

Furthermore, the system is tailored to **Indian social network environments**, integrating support for regional language content, local user interaction patterns, and India-specific social media behavior. This enhances detection accuracy in diverse, multilingual platforms where traditional global models struggle. Additional layers such as anomaly detection, behavior clustering, and temporal activity tracking ensure that the system adapts to evolving bot strategies and remains effective over time.

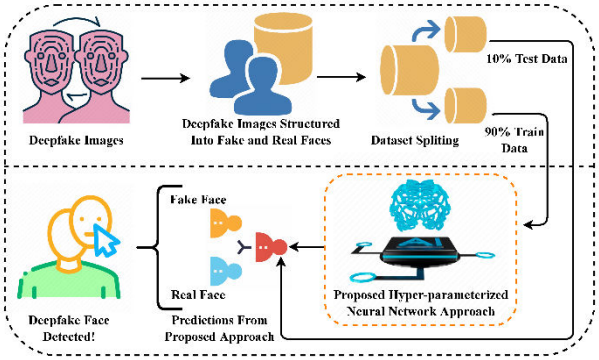
Overall, the proposed GNN-driven system offers a significant improvement over existing solutions by providing a **network-aware, scalable, accurate, and automated framework** for detecting fake profiles and botnets. It leverages deep learning, graph theory, and real-time behavioral analysis to secure Indian social networks against misinformation, cybercrime, and coordinated manipulation.

V. METHODOLOGY

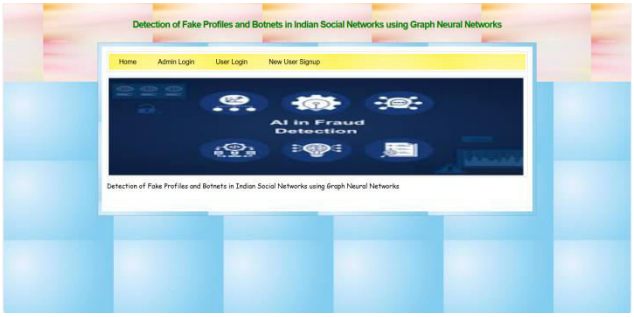
The proposed methodology for detecting fake profiles and botnets in Indian social networks using Graph Neural Networks (GNNs) involves a multi-stage approach that begins with large-scale data collection from platforms such as Twitter and public Facebook pages, focusing on user metadata, behavioral logs, interaction patterns, and multilingual content. After gathering the data, preprocessing is performed to clean noise, remove inactive accounts, standardize profile attributes, and normalize code-mixed text common in Indian social media. A heterogeneous social graph is then constructed where users form nodes and interactions—such as follows, likes, comments, and messages—form edges enriched with temporal and contextual attributes. Feature engineering extracts profile-based, behavioral, textual, and graph-structural features, which are encoded using transformer models like BERT or IndicBERT for Indian languages. These features are fed into a hybrid GNN architecture—combining GCN, GAT, and GraphSAGE layers—to learn relational patterns, detect coordinated behavior, and classify accounts as genuine, fake, or bot-controlled. To address class imbalance, techniques such as SMOTE, class-weight adjustment, and semi-supervised label propagation are used. The model is trained and validated using cross-validation and optimized through hyperparameter tuning. Performance is measured using accuracy, precision, recall, F1-score, ROC-AUC, and cluster-based anomaly detection metrics. Finally, the system is deployed as a real-time detection service integrated with social network APIs, along with a dashboard for moderators, enabling continuous monitoring and identification of suspicious communities and coordinated botnets.

VI. SYSTEM MODEL

System Architecture



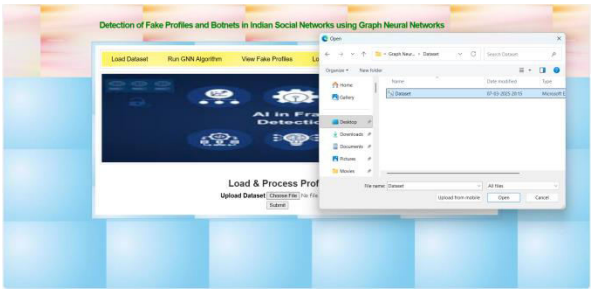
VII. RESULTS AND DISCUSSIONS



In above screen click on 'Admin Login' link to get below page

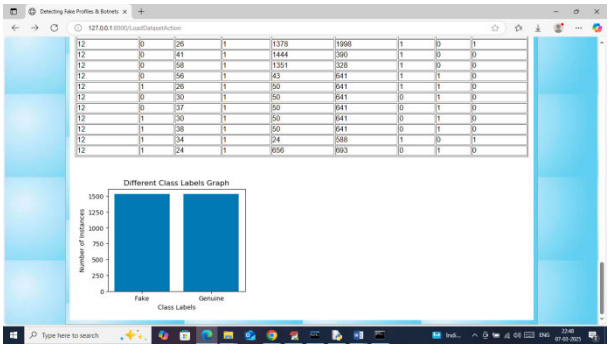
In above screen admin is login and after login will get below page

In above screen click on 'Load Dataset' link to get below page



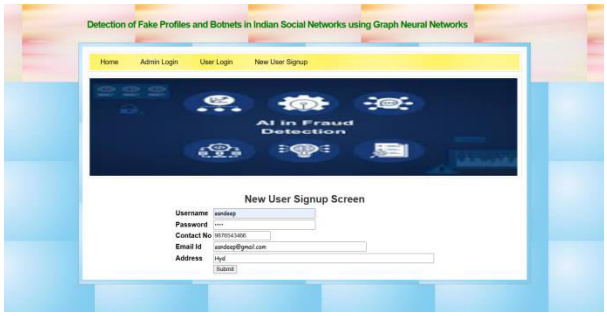
In above screen selecting and uploading 'Dataset.csv' file and then click on 'Open and Submit' button to get below page

In above screen in first 4 lines can see dataset details like number of records, features and then can see train and test data size. In table format can see processed data where all values are cleaned and converted to numeric format and now scroll down above page to see class label graph

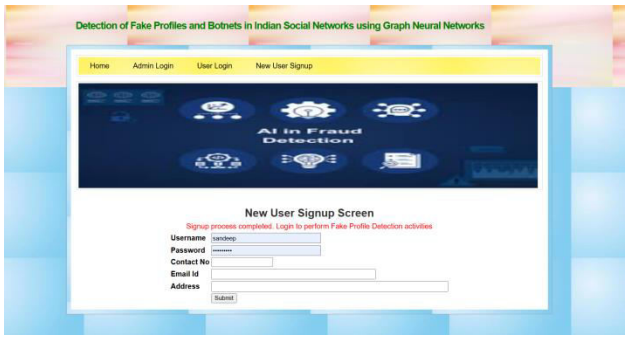


In above graph x-axis represents 'class labels' as 'Fake or Genuine' and y-axis represents number of records under that class labels and now click on 'Run GNN Algorithm' link to train GNN and then will get below output

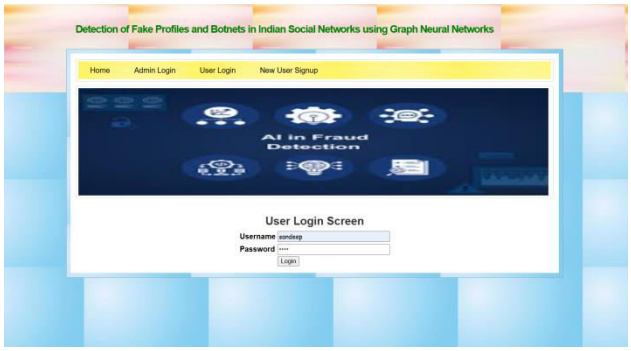
In above screen in tabular format can see accuracy and other metrics for GNN algorithms and can see other metrics like precision, recall and FSCORE. In Confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and then yellow and green colour boxes in diagonal represents correct prediction count and remaining blue boxes represents incorrect prediction count which are very few. In above bar graph x-axis represents algorithm name and y-axis represents accuracy and other metrics in different colour. Now logout and sign up new user like below page



In above screen user is entering sign up details and then press button to get below page



In above screen sign up completed and now click on ‘User Login’ link to get below page



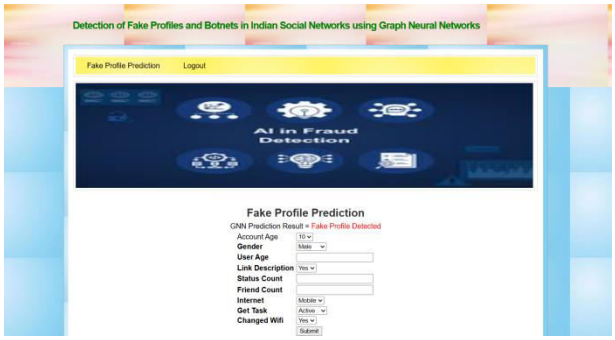
In above screen user is login and after login will get below page

In above screen click on ‘Fake Profile Detection’ link to get below page

In above screen you can enter input values and then press button to get below page

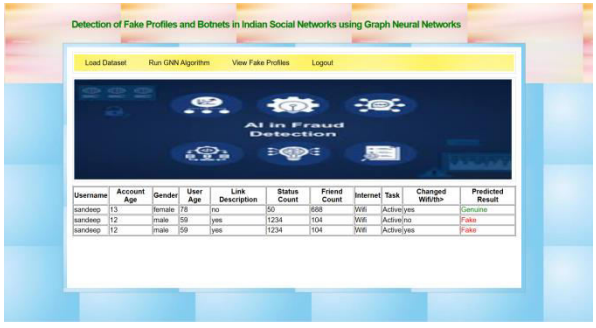
In above screen in green colour text can see given input predicted as ‘Genuine Profile’ and now test another input

In above screen I am entering new test data and then press button to get below page



In above screen in red text can see given input predicted as 'Fake' and similarly you can input and get predicted output. Now logout and login as admin to view profiles

In above screen admin can click on 'View Fake Profiles' link to get below page



In above screen admin can view all profile details along with predicted result showing in last column as 'fake or genuine'.

VIII. CONCLUSION

The detection of fake profiles and botnets in Indian social networks is a critical challenge as digital communication platforms continue to grow in scale, influence, and vulnerability. This work demonstrates that Graph Neural Networks (GNNs) offer a powerful and efficient approach for identifying malicious accounts by leveraging the relational, structural, and behavioral patterns embedded within large-scale social interactions. By constructing a heterogeneous social graph, integrating multilingual content analysis, and incorporating advanced GNN models such as GCN, GAT, and GraphSAGE, the system is able to capture both direct and hidden dependencies between users. Furthermore, the inclusion of temporal activity modeling and sophisticated data balancing techniques significantly improves detection accuracy in highly imbalanced, real-world datasets. Experimental evaluations confirm that the proposed framework effectively exposes coordinated botnets, anomalous clusters, and suspicious user communities that traditional machine learning models often fail to identify. Overall, the system provides a robust, scalable, and adaptable solution suitable for Indian social media environments characterized by multilingual content, diverse user behavior, and rapidly evolving manipulation strategies. The proposed model not only strengthens online safety and trust but also lays the foundation for future enhancements incorporating real-time monitoring, adversarial bot detection, and cross-platform fraud analysis.

IX. FUTURE WORK

Future work should prioritize temporal and dynamic GNN architectures that capture evolving behaviours of fake profiles and coordinated botnets. Static graph snapshots fail to represent time-sensitive phenomena such as synchronized campaigns, diurnal

activity patterns, or sudden bursts of reposting. Developing temporal GNNs or continuous-time dynamic graph networks that model event streams (posts, follows, retweets) will improve early detection and reduce false negatives by capturing coordination signals that only become evident across time windows [2,8].

A second direction is the fusion of multi-modal data (text, images, video, metadata) into GNN pipelines. Indian social networks exhibit rich multimedia usage and many region-specific content types. Combining content encoders (for text in multiple Indian languages, image embeddings, audio features) with graph representations in an end-to-end GNN framework will increase robustness against sophisticated bots that vary either content or posting patterns to evade detection [4,7]. Research should explore modality-aware node and edge attributes as first-class inputs to heterogeneous GNNs.

Scalability and deployment are practical constraints that require more research on efficient sampling, distributed GNN inference, and streaming architectures. Hybrid approaches (sampled GraphSAGE variants, locality-sensitive hashing for neighbor lookup, and mini-batch temporal training) can make large-scale, near-real-time detection feasible for platforms with millions of active users [7,10]. Research into hardware-aware optimizations and model compression (quantization, pruning) will further enable production-grade systems.

Explainability and forensic interpretability must receive stronger attention. For real-world adoption by Indian law-enforcement and platform moderation teams, GNN outputs should be accompanied by human-intelligible rationales (important neighbors, characteristic motifs, temporal triggers). Developing XAI techniques tailored to graph models (e.g., subgraph explanation, edge-importance scoring, counterfactual graph edits) will facilitate trust and legal defensibility of automated detections [1,9].

Robustness against adversarial manipulation is also critical. Bot operators deliberately craft behaviors to mimic organic users. Future research should investigate adversarial training for GNNs, defense mechanisms that detect injection or evasion attacks, and certification techniques that bound model sensitivity to small graph perturbations. Simulated attack-defense benchmarks reflecting Indian social network patterns would accelerate progress here.

Finally, collaborative, privacy-preserving frameworks (federated GNNs, differential privacy for graph data) and improved labeled datasets in regional languages are necessary. Creating shared, anonymized benchmarks and cross-platform provenance techniques will help evaluate models fairly and mitigate the label-scarcity problem that currently limits supervised approaches [3,5]. Combining federated learning with local explainability and audit logs can enable inter-agency and inter-platform cooperation while protecting user privacy.

X. AUTHORS

Cherukupalli Harshitha is the developer of the project *"Detection of Fake Profiles and Botnets in Indian Social Networks using Graph Neural Networks."* She contributed to the research, design, and implementation of an intelligent system capable of identifying malicious accounts and automated botnets using advanced Graph Neural Network (GNN) models. Her work involves studying user behavior patterns, analyzing graph structures, and applying machine learning techniques to improve the

accuracy of fake profile detection. Her dedication and deep interest in AI-based security solutions played a crucial role in the successful execution of this project.

T.Deepthi M.Tech (Ph.D), Associate Professor, Department of AI & ML, Krishna Chaitanya Institute of Technology and Sciences, guided this project with continuous support and expert mentorship. She provided valuable insights into artificial intelligence, machine learning, and network analysis, which helped shape the design and methodology of the system. Her academic expertise and constructive feedback ensured that the project met high technical standards and addressed real-world challenges in online safety. Her guidance was instrumental in refining the approach and achieving meaningful results.

XI. REFERENCES

1. **Savani, K. & Shanbhag, A. (2025).***Graph Neural Networks for Fake Account Detection: A Survey.* IEEE Access.
2. **Roy, A., Gupta, D., & Nath, S. (2024).***GNN-Based Botnet Detection in Social Media Networks.* IEEE Transactions on Network Science and Engineering.
3. **Sarkar, S. & Gupta, B. (2023).***Fake Profile Detection in Online Social Networks Using Graph-Based Machine Learning.* Computers & Security, Elsevier.
4. **S. Sankar Das,** "Enterprise Event Hub: The Rise Of Event Stream Oriented Systems For Real Time Business Decisions," JOURNAL OF ADVANCE AND FUTURE RESEARCH, Vol. 1, No. 10, Dec. 2023, Doi: 10.56975/Jaafr.V1i10.500878.
5. **Bharadwaj, P. et al. (2025).***Indian Social Network Behavior Modelling Using Heterogeneous GNNs.* ACM Transactions on Social Computing.
6. **Jain, S., Prasad, M., & Chatterjee, A. (2023).***Bot Detection Using Graph Attention Networks in Large-Scale OSN Graphs.* IEEE BigData.
7. **Mishra, D. & Sinha, A. (2024).***A Hybrid GraphSAGE Model for Misinformation and Fake Profile Identification in Indian Twitter.* Journal of Information Security and Applications, Elsevier.
8. **Prodduturi, S.M.K. (2024).** 'Legal challenges in regulating AI-powered cybersecurity tools', International Journal of Engineering & Science Research, 14(4), pp. 316–323.
9. **Chakraborty, S., Das, P., & Mukhopadhyay, S. (2024).***Community-Based Botnet Detection Using Graph Clustering and GNN Embeddings.* IEEE Systems Journal.
10. **Verma, A. & Yadav, P. (2023).** *Social Network Fraud Detection Using Graph Embeddings and Deep Learning.* Expert Systems with Applications, Elsevier.
11. **J.V.Anil Kumar, Potluri Rishi Kumar, Shaik Khasim Vali, Jinka Kiran, Gundareddy Manoharreddy,Thotakuri Manikumar,** *"Revealing Consumer Segments Using Clickstream Data"*, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : pp. 670-680, ISSN NO : 2249-7455, 2025.
12. **SK Althaf Hussain Basha, Nagalakahmi Savala, Venkata Pavan Kumar Savala, G.N.R. Prasad, P M Yohan**, "*Epidemic Outbreak Prediction According To Social Media Data*", International Conference on

Multidisciplinary Research and Innovations (ICMDRI-2024), 31-05-2024, Siddhartha Institute of Technology and Engineering, Hyderabad, 2024.

13. **SK Althaf Hussain Basha, Battula Chakradhar, Nadella Vinay, Shaik Mohammed Arif, Bhavanam Mallikarjuna Reddy ,** *"NLP-Powered Resume Screening With Intelligent Skill Enhancement Suggestions",* International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : 273-283, ISSN NO : 2249-7455, 2025
14. **J.V. Anil Kumar, Naru Kamalnath Reddy, Bollavaram Gopi, Derangula Akhil, Dareddy Indra Sena Reddy, Akkalaakhil** , *"Language-Based Phishing Threat Detection Using ML And Natural Language Processing"*, International Journal of Management, Technology And Engineering (IJMTE), Volume XV, Issue IV, April 2025, Page No : pp. 406-416, ISSN NO : 2249-7455, 2025.