International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Fake Social Media Profile Detection and Reporting

G Sarayu
Presidency University
Bengaluru,India
khushikuhu12@gmail.com

Fatima Noori
Presidency University
Bengaluru,India
fatimanoori9801@gmail.co m

Sony Priya
Presidency University
Bengaluru,India
sonypriya42022@gmail.c om

Ms. Bhavya B
Assistant Professor,
School of Computer Science and Engineering,
Presidency University
Bengaluru,India
bhavya.b@presidencyuniversity.in

ABSTRACT :

The rapid growth of social media platforms has led to an increase in the creation and misuse of fake profiles, posing significant threats to user security, online trust, and information integrity. Fake profiles—ranging from bots and impersonators to sockpuppets—are often used for malicious activities such as spreading misinformation, executing scams, and manipulating public opinion. This article explores the various methods employed to detect fake accounts, including machine learning algorithms, natural language processing (NLP), image analysis, and social network graph theory. It also addresses the major challenges involved in fake profile detection, such as evolving attacker strategies, data privacy limitations, and scalability issues. Furthermore, the article outlines best practices for users and highlights emerging technologies that promise more accurate and real-time detection. The study emphasizes the need for a multi-layered, adaptive approach to combat the growing sophistication of fake profiles on social media.

Introduction

In the digital age, social media platforms like Facebook, Twitter, Instagram, and LinkedIn have become integral to how people connect, communicate, and consume information. While these platforms offer immense benefits, they are also vulnerable to exploitation—particularly through the creation and use of fake profiles. These fraudulent accounts, often designed to mimic real users, are used for a wide range of malicious purposes, including spreading misinformation, launching phishing scams, manipulating public opinion, and engaging in cyberbullying.

Fake profiles not only pose a risk to individual users but also threaten the credibility of the platforms themselves. With billions of active users and massive volumes of content generated daily, detecting and mitigating fake profiles has become a critical challenge for social media companies. Traditional manual moderation methods are insufficient at this scale, which has led to the development of automated, intelligent systems powered by artificial intelligence (AI), machine learning (ML), and natural language processing (NLP).

This article delves into the nature of fake profiles, the technologies used to detect them, the challenges involved in doing so, and the evolving strategies needed to stay ahead of increasingly sophisticated threats. By understanding the scope and impact of fake profile activity, we can better appreciate the importance of robust detection mechanisms in securing the future of online interactions.

What is a Fake Profile?

A fake profile on social media is a user account that is not tied to a real individual or organization. These profiles may be:

Bots: It automates the accounts content posting and interacts with users. Sockpuppets: Fake identities used to manipulate discussions or spread propaganda.

Catfish accounts: Fake personas created to deceive others, often for emotional manipulation or financial gain.

Why Fake Profiles are Dangerous

Spreading Misinformation

Fake profiles can amplify false narratives, especially during elections or crises.

Scamming Users

Impersonation and phishing scams are common among fake accounts.

Undermining Trust

A large number of fake users can degrade the quality and trustworthiness of a platform.

Data Breaches & Harvesting

Fake accounts often attempt to gather personal information from real users.

How Fake Profile Detection Works

1. Machine Learning Models

Algorithms analyze account behaviors, content, and metadata to distinguish between real and fake profiles.

Features considered:

Posting frequency and patterns Account age
Number of followers vs following Content originality and sentiment
Engagement metrics (likes, comments, shares)

2. Natural Language Processing (NLP)

NLP is used to analyze language patterns in posts and messages. Fake profiles often use repetitive, unnatural, or overly generic language.

3. Image Analysis

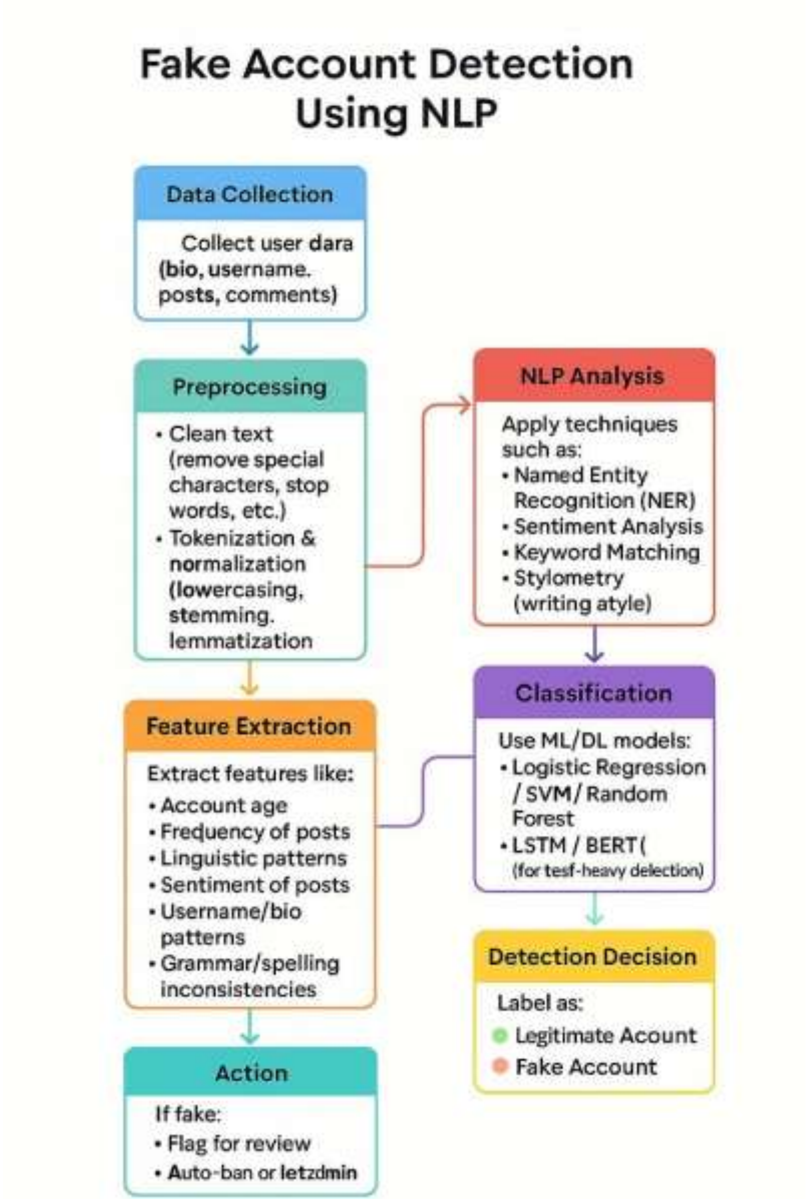
Reverse image search to detect stolen profile pictures

Analyzing image metadata and face consistency in multiple posts

4. Graph-Based Methods

Social network graphs are studied to see how an account is connected. Fake profiles often have weak or unusual network patterns.

Flow Chat



Models used

Natural Language Processing (NLP) NLP is a field of Artificial Intelligence that enables machines to understand, interpret, and respond to human language. NLP combines linguistics and machine learning techniques to process large amounts of text or speech data.

Linear Regression Linear Regression is a supervised machine learning algorithm used for predicting a continuous output. It models the relationship between a dependent variable (target) and one or more independent variables (features) by fitting a linear equation to the observed data:

$y=mx+cy = mx + cy=mx+c$

Artificial Neural Networks (ANN) ANNs are computing systems inspired by the human brain.They consist of layers of interconnected nodes (neurons), where each node performs computations. ANNs are powerful tools for solving complex problems like image recognition, language processing, and classification tasks. They learn patterns through a process called backpropagation and adjust weights to minimize error during training.

Challenges in Detecting Fake Profiles

Evolving Techniques Fake account creators constantly adapt to avoid detection.

Privacy Constraints Limited access to user data due to privacy laws can make detection harder.

False Positives Real users may be incorrectly flagged, damaging user experience.

Scale Platforms like Facebook or Twitter have billions of users, making manual detection impractical.

Real-World Tools & Techniques

Botometer: Analyzes Twitter accounts and scores them based on their likelihood of being bots.

Facebook's AI-based detection: Combines behavioral analysis with deep learning.

CAPTCHAs: Used to filter out automated bots.

Two-Factor Authentication (2FA): Helps prevent fake account creation.

Best Practices for Users

Be cautious when accepting friend/follow requests from unfamiliar profiles. Verify suspicious profiles through reverse image search and profile scrutiny. Report fake or abusive accounts to platform moderators.
Avoid sharing sensitive information publicly.

Future Directions

More Advanced AI Models: Including generative AI to predict and counter new fake profile strategies.

Decentralized Identity Systems: Blockchain-based identities may help verify users securely.

Conclusion

The prevalence of fake profiles on social media has evolved into a critical cybersecurity concern with far-reaching implications for individuals, organizations, and society at large. These deceptive accounts serve as powerful tools for malicious actors—used to spread disinformation, manipulate political discourse, execute fraud, and undermine digital trust. Their increasing sophistication, driven by advances in AI and automation, demands an equally advanced and adaptive approach to detection and prevention.

This article has explored the various techniques employed to identify fake profiles, including machine learning models, natural language processing, image and metadata analysis, and graph-based social network analysis. While each of these methods contributes uniquely to the identification process, their true power lies in their integration—working together as part of a layered defense mechanism that can effectively filter out inauthentic activity across platforms.

However, the fight against fake profiles is far from over. Detection systems face significant challenges, including the dynamic tactics used by attackers, the vast scale of user data, privacy restrictions, and the constant risk of false positives. These obstacles underscore the importance of continuous innovation and collaboration between technology providers, platform developers, policymakers, and end users.

Referece

- Kaur, H., Singh, M., & Singh, K. (2020). Detection of fake profiles on social media using NLP and machine learning techniques. Procedia Computer Science, 173, 104-112.
- Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing (3rd ed. draft). Stanford University.·
- Ng, A. Y. (2011). Machine Learning Specialization. Stanford University.
- Grover, S., & Mark, G. (2019). Detecting fake accounts in social networks using regression-based approaches. ACM Conference on Web Science.
- Al-Qurishi, M., et al. (2019). Fake profile detection on social media using deep learning techniques. IEEE Access, 7, 128990-129003.
- Detecting Fake Accounts on Social Media Subrahmanian, V. S., Azaria, A., Durst, S., et al. (2016)Science, 354(6308), 1090-1094.
- A Machine Learning Approach for Fake Profile Detection in Online Social Networks Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010)
IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.