International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Powered Smart Social Media Platform for Enhanced Content Verification and User Safety

Mittha M. S.
AI&DS Department
V.V.P.I.E.T

Kota L. V.
AI&DS Department
V.V.P.I.E.T

Adam R. R.
AI&DS Department
V.V.P.I.E.T

Chityal G. S.
AI&DS Department
V.V.P.I.E.T

Abstract - The rapid expansion of social media has improved communication but has also increased the spread of misinformation, unsafe content, and fake user activity. This paper proposes an AI-driven social media concept that focuses on detecting misleading posts, identifying harmful behavior, and ensuring user authenticity. The system incorporates machine learning, natural language processing, and behavioral analysis to evaluate content and user actions. The aim of this paper is to introduce the conceptual framework of the platform, highlight its major components, and present its potential to improve the integrity and safety of online social interactions. This work represents a basic idea submission and does not include implementation details.

Keywords - Deepfake Detection, Computer Vision, Digital Forensics, Real-Time Content Filtering

1.INTRODUCTION

This document shows the suggested format and appearance of a manuscript prepared for IJSREM journals. Accepted papers will be professionally typeset. This template is intended to be a tool to improve manuscript clarity for the reviewers. The final layout of the typeset paper will not match this template Social media platforms are now central to communication, entertainment, and information exchange. However, these platforms are often affected by false information, manipulated media, cyberbullying, and automated bot accounts. Manual moderation alone cannot handle the large volume of daily user activity.

Artificial intelligence offers promising solutions for identifying harmful content, analysing user behaviour, and flagging suspicious or misleading posts. This paper introduces the idea of a Smart AI-Based Social Media Platform that integrates intelligent content checking, automated safety mechanisms, and advanced verification features. The goal is to create a safer and more trustworthy digital environment.

2. CONCEPTUAL OVERVIEW

The proposed platform consists of three primary AI components:

A. Content Authentication Module

- Uses machine learning to analyze images, videos, and text.
- Detects manipulated or artificially generated media.
- Warns the user if shared content appears misleading.

B. Safety & Moderation Module

- Uses NLP to scan text for bullying, hate speech, or abusive language.
- Automatically restricts the visibility of harmful posts.
- Supports mental well-being by reducing exposure to aggressive content.

C. User Authenticity Module

- Flags unusual account behavior using pattern recognition.
- Detects bot-like activity based on posting frequency and interaction style.
- Adds additional security verification steps for suspicious accounts.

2.2 Tools & Technologies (Conceptual)

- **AI/ML:** TensorFlow, PyTorch, Scikit-Learn
- **NLP:** Transformers-based models
- **Backend (future implementation):** Python, Django/Flask
- **Database:** MySQL / MongoDB (optional for real development)

2.3 Future Scope

- Integration of deepfake detection for videos and images
- Multilingual harmful-content identification
- Community sentiment monitoring using AI
- Transparent AI ethics with user-friendly explanations
- Scalable cloud-based architecture for real deployment

3. CONCLUSIONS

This paper presents a conceptual model for an AI-enhanced social media system capable of authenticating content, identifying harmful interactions, and detecting fake users. By combining multiple AI techniques, the proposed idea supports a safer and more reliable user experience. Since this paper focuses only on the conceptual aspects, future work can explore

technical implementation, dataset preparation, and model evaluation.

4. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to our guide Geeta Chityal, for their valuable guidance in this project. I would also like to thank my all AI&DS faculties for sharing their experience.

5. REFERENCES

1. J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt and M. Nießner, “Face2face: Real-time face capture and reenactment of RGB videos,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Juan, USA, pp. 2387–2395, 2016.
2. A. Tewari, M. Zollhoefer, F. Bernard, P. Garrido, H. Kim et al., “High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 357–370, 2020.
3. J. Lin, “FPGAN: Face de-identification method with generative adversarial networks for social robots,” NeuralNetworks, vol. 133, no. 3, pp. 132–147, 2021.
4. R. Chesney and D. Citron, “Deepfakes and the new disinformation war: The coming age of post-truth geopolitics,” Foreign Affairs, vol. 13, no. 3, pp. 1–14, 2019.