Brand Integrity Nexus is built to help organizations protect their online identity and reduce risks linked to fake products and fraudulent digital activity

Monika Jaglan[1], Gobind Kumar Gautam[2], Sweta Sah[3], Vaishnavi Asthana[4], Shaiba Parveen Ansari[5]
[1,2,3,4,5] *Department of Computer Engineering, MGM COET, Noida, Uttar Pradesh, India – 201309*

Abstract- Instagram has seen a noticeable increase in fake and misleading posts, which has made it harder for brands to preserve a genuine online presence. In this work, we introduce Brand Integrity Nexus, a system that uses multiple AI techniques to study images, captions, and other post details to identify content that may be suspicious or misleading.

The system integrates automated data collection (Instaloader, Apify), text–image preprocessing, embedding generation (BERT, DistilBERT, CLIP, ResNet), and blockchain-based hashing for evidence integrity. Initial results confirm successful development of a clean dataset, consistent embeddings, and early similarity analysis, establishing a foundation for a future multimodal classifier.

I.INTRODUCTION

Instagram's fast growth has increased the spread of counterfeit promotions and impersonation pages. Detecting such posts is challenging due to:

- multimodal content (image + caption + hashtags),
- high post volume,
- misleading visuals and ambiguous hashtags,
- lack of automated brand-protection tools.

This project builds an offline multimodal research pipeline capable of collecting real Instagram posts, preprocessing content, generating embeddings, and storing suspicious-post hashes on blockchain for tamper-proof verification.

Objectives :

1. Automate Instagram data collection
2. Clean and organize text, pictures, and related post information so they can be used effectively.
3. Create meaningful text and image representations using modern natural language and computer-vision models.
4. Detect caption–image inconsistencies.
5. Store hashes of flagged posts via blockchain.
6. Create a structured dataset for future ML training.

Research Questions:

1. Can embeddings detect mismatched or misleading brand content?
2. Does preprocessing improve similarity accuracy?
3. How well do BERT and CLIP capture counterfeit cues?
4. Can blockchain ensure secure evidence logging?

II.LITERATURE REVIEW

Earlier research on detecting fake products shows that combining different types of information gives better accuracy. Language models such as BERT, DistilBERT, and RoBERTa have proven useful when working with captions, short descriptions, and hashtags. Similarly, image-based models like ResNet and ViT are able to capture visual cues, patterns, and details that are often linked to brand authenticity.

Multimodal models like CLIP show strong performance in aligning images with text, making them ideal for detecting mismatched captions. Prior research also supports blockchain for enhancing evidence integrity through content hashing.

These works collectively validate the use of multimodal embeddings + blockchain for counterfeit detection.

III.METHODOLOGY

3.1 Data Collection

Data is scraped from Instagram using:

- Instaloader – captions, comments, hashtags, image URLs, metadata.
- Apify Scraper – structured JSON and engagement data.

Duplicates and corrupted posts are removed.

3.2 Preprocessing

Text Preprocessing

- Remove emojis, URLs, and special characters
- Tokenization and lowercasing
- Hashtag and keyword extraction
- Normalizing metadata

Image Preprocessing

- Resize + normalization
- Duplicate & low-quality detection (pHash)
- Standardized storage mapping

3.3 Dataset Construction

SQLite database stores:

- Clean captions
- Image paths
- Metadata
- Embeddings
- Blockchain-hash references

All linked using unique post IDs.

3.4 Feature Extraction

Text Embeddings:

- BERT
- DistilBERT
- RoBERTa

These models capture meaning, intent, and authenticity cues in captions.

Image Embeddings:

- ResNet-50
- CLIP Vision Encoder
- ViT

Identify logos, products, visual anomalies.

Multimodal Similarity

CLIP compares caption–image alignment.

Low similarity indicates potentially misleading or counterfeit content.

3.5 Experimental Evaluation

The objective of this phase is validation—not final classification.

Key checks include:

- embedding quality and dimensionality consistency,
- caption–image similarity distribution,
- basic clustering behaviour,
- database correctness.

This confirms the system's readiness for future training.

3.6 Blockchain Integration

Any post that seems doubtful is converted into a SHA-256 hash and saved in a private Ethereum network (using Ganache) so that its record remains secure and unchanged.

The main advantages are:

- immutability
- evidence integrity
- tamper-proof logging

3.7 Future Expansion

The pipeline is scalable to platforms like YouTube and Facebook, enabling embeddings from thumbnails, titles, and descriptions.

Expected Outcomes

1. Reliable data scraping and cleaning
2. Rich multimodal dataset in SQLite
3. Working embedding extraction pipeline
4. Early counterfeit-indicator identification
5. Tamper-proof storage of suspicious posts

Architecture Overview

1. Data Ingestion – Raw Instagram data
2. Preprocessing – Clean text & images
3. Storage Layer – SQLite
4. Embedding Layer – Text/vision encoders
5. Analysis Layer – Similarity scoring, clustering
6. Blockchain Layer – Hash storage
7. Dashboard Layer (future) – Visualization

Implementation Tool

1. Collection of data from Instaloader, Apify
2. Preprocessing of the data using Python (regex, NLTK, spaCy, OpenCV, PIL, imagehash)
3. Database to store data such as SQLite, Pandas
4. Text and image models such as BERT, DistilBERT, ResNet, CLIP, and ViT.

5. Machine-learning frameworks like PyTorch and scikit-learn.
6. Blockchain tools including Ganache, Solidity, and Web3.py.
7. Visualization libraries such as Matplotlib and Seaborn.

Coun

terfeit Detection Survey — Summary

Figure 1 below compares three types of counterfeit-detection approaches — NLP-based caption analysis, HOG-based logo detection, and shape-based visual detection — using two benchmarking datasets (Series 1 and Series 2). Shape-based detection consistently performs the best across both datasets, reaching close to 90% accuracy, while NLP-based detection shows moderate performance between 78%–80%, also captions can be misleading or incomplete. HOG-based logo detection performs well on Dataset 2 but shows reduced effectiveness on Dataset 1, likely due to logo distortion or low-quality images.

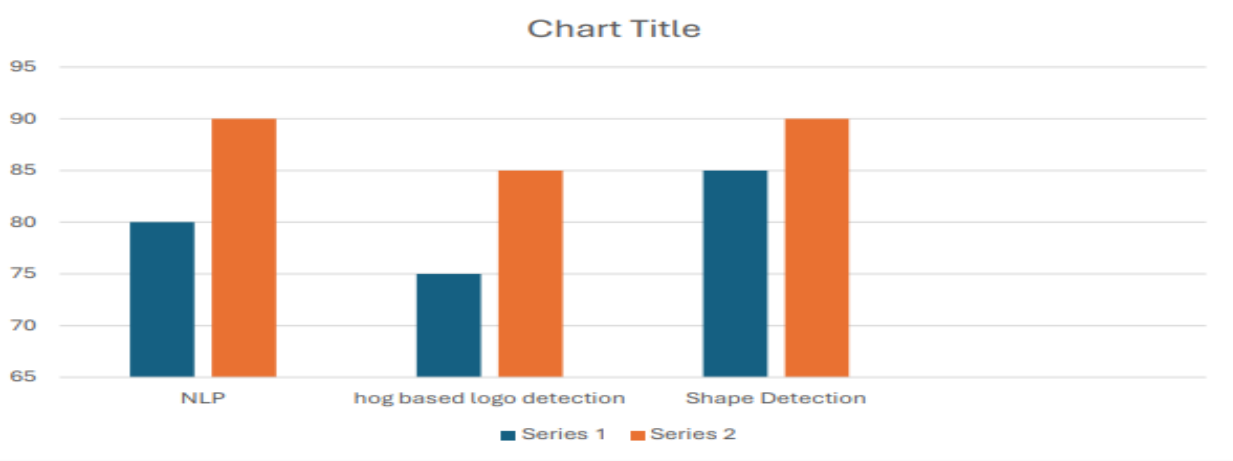

Figure 1: Comparison of Detection Accuracy Across Different Methods

This bar chart evaluates how three individual features—image-based product analysis, text description analysis, and colour detection—perform in identifying counterfeit items. Image-based analysis shows the highest reliability at 90%, demonstrating that visual cues (such as product shape, labeling, and packaging) play a key role in spotting counterfeit products. Text descriptions rank second, as they help catch inconsistencies between the posted caption and the actual product. Colour detection shows the lowest accuracy, indicating colour alone is not a strong indicator of authenticity.

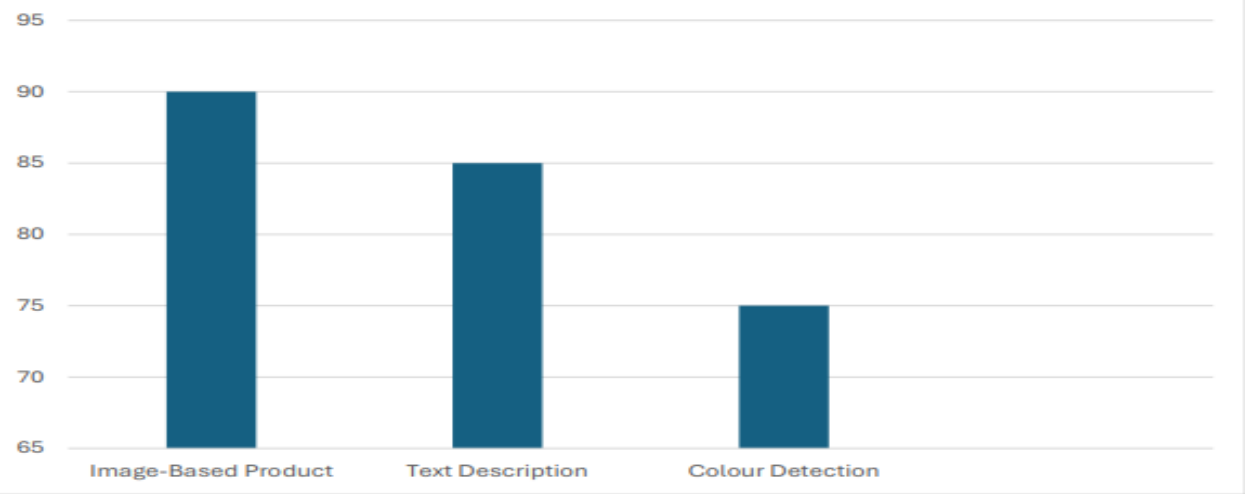

Figure Y: Performance comparison of image-based, text-based, and colour-based cues in counterfeit product detection

IV.BRAND AUTHENTICITY-SUMMARY

The survey conducted by Dina Younis (2025) examines how user-generated content on social media influences the perceived authenticity of brands and how this perception affects consumer behavior. Respondents reported that genuine, natural, and experience-based posts increase their trust in a brand, whereas overly edited or promotional content reduces authenticity. The study highlights that authenticity strongly impacts users' willingness to engage with a brand, recommend it to others, and make purchase decisions. Overall, the findings show that users depend heavily on authentic social media content when forming opinions about brands, making authenticity a key factor in digital brand management.

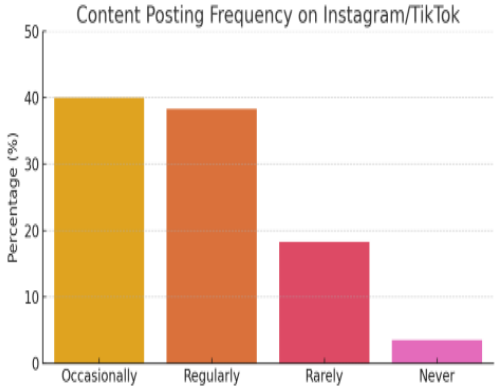

Figure 3. Content Posting Frequency on Instagram/tiktok

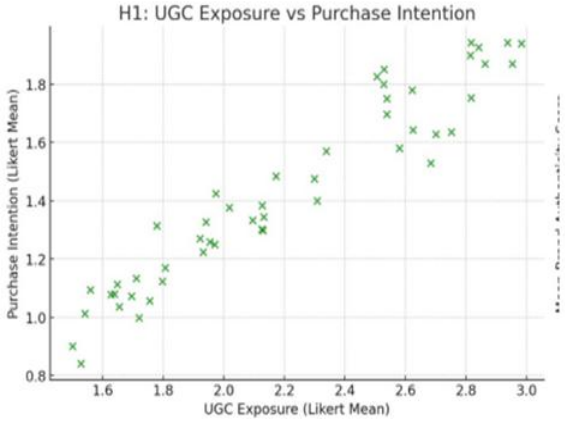

Figure 4. A Scatter Plot Illustrating the H1

V.CONCLUSION

The initial stage of this project makes a solid base for building a system that can identify potentially fake or misleading brand-related content. By bringing data collection, cleaning steps, embedding generation, and secure blockchain storage, together the system forms a strong starting point for future development.

Moving forward, the plan is to train a full detection model and eventually enable real-time monitoring features.

REFERENCE

[1] Devlin et al. (2019) describe BERT, a widely used transformer model for language tasks.
[2] Sanh and colleagues (2019) present DistilBERT, a smaller variant of BERT.
[3] He et al. (2016) introduced the ResNet architecture for image recognition.
[4] Dosovitskiy's work (2021) explains the concept of Vision Transformers.
[5] Radford et al. (2021) developed CLIP, which learns jointly from images and text.
[6] Zarei and team (2020) explored detecting impersonation on social platforms.
[7] Zhang et al. (2019) discuss blockchain privacy and safety aspects.
[8] Wolf and collaborators (2020) give an overview of the Transformers library.