frontiers | Frontiers in Computer Science

TYPE Original Research
PUBLISHED 16 June 2025
DOI 10.3389/fcomp.2025.1566513

OPEN ACCESS

EDITED BY
Stelvio Cimato,
University of Milan, Italy

REVIEWED BY
Cosimo Comella,
Garante per la protezione dei dati personali, Italy
Por Lip Yee,
University of Malaya, Malaysia

*CORRESPONDENCE
Muhammad Arshad
✉ muhammad.arshad@tudublin.ie

RECEIVED 25 January 2025
ACCEPTED 12 May 2025
PUBLISHED 16 June 2025

CITATION
Arshad M, Ahmad A, Onn CW and Sam EA (2025) Investigating methods for forensic analysis of social media data to support criminal investigations.
Front. Comput. Sci. 7:1566513.
doi: 10.3389/fcomp.2025.1566513

Investigating methods for forensic analysis of social media data to support criminal investigations

Muhammad Arshad[1]*, Ashfaq Ahmad[2], Choo Wou Onn[3] and Emmanuel Arko Sam[4,5]

[1]School of Informatics and Cybersecurity, Technological University Dublin, Dublin, Ireland, [2]Faculty of Basic Sciences, Lahore Garrison University, Lahore, Pakistan, [3]Faculty of Data Science and Information Technology, INTI International University, Nilai, Malaysia, [4]Unicaf, Larnaca, Cyprus, [5]University of East London, London, United Kingdom

Social media platforms have become a cornerstone of modern communication, and their impact on digital forensics has grown significantly. These platforms generate immense volumes of data that are invaluable for reconstructing events, identifying suspects, and corroborating evidence in criminal and civil investigations. However, forensic analysts face challenges, including privacy constraints, data integrity issues, and processing overwhelming volumes of information. This research evaluates the effectiveness of existing forensic methodologies and proposes artificial intelligence (AI) and machine learning (ML)–driven solutions to overcome these challenges. Through detailed empirical studies, including cyberbullying, fraud detection, and misinformation campaigns, the study demonstrates the effectiveness of advanced techniques such as text mining, network analysis, and metadata evaluation. These findings underscore the importance of integrating scalable technologies with ethical and legal frameworks to ensure the admissibility of social media evidence in courts of law.

KEYWORDS

AI in forensics, cybercrime investigation, food security, forensic analysis, gender injustices, social media forensics

1 Introduction

The dominance of social networks changed the destiny of people and the way they communicate, share data, and share personal experiences. These may include text posts containing text, images, video, or audio, as well as geotagging information from Facebook, Instagram, and Twitter. To police and forensic investigators, this data offers a vast pool from which evidence can be procured, perimeters recreated, and important persons involved in criminal or civil aspects related to known (Casey, 2011; Kaplan and Haenlein, 2010).

However, expanding the use of social media data in investigations offers the following difficulties: The fact that tweets can be edited or deleted makes the collection of evidence challenging (Almuhimedi et al., 2013; Torres-Lugo et al., 2022; Mubarak et al., 2023). Furthermore, there are privacy laws that may prevent someone from having any interaction with personal data, on top of other barriers For example, the General Data Protection Regulation restricts one's use of an individual's personal data, thus restraining forensic analysts (Oetzel and Spiekermann, 2014). Besides, the enormous amount of data produced daily makes it extremely impractical to analyze it manually, thus requiring a solution that can be scaled and automated (Liu et al., 2016; Liu, 1997; Zhai and Liu, 2006; Liu et al., 1999).

These challenges are tackled in this research by assessing the existing forensic techniques with an attempt to create new and sophisticated methodologies incorporating AI and ML. These methods are developed to enhance the quality and effectiveness of forensic investigations with due consideration of legal and ethical frameworks. Concerning concrete research data, this paper shows how the improved and created forensic methods can help address multifaceted investigative issues and guarantee the admissibility of the evidence in court.

1.1 Plain language summary—for policymakers and law enforcement

This study explores how AI can help forensic teams analyze massive social media data to detect cyberbullying, fraud, and fake news. It uses smart algorithms to understand text, identify faces, and track suspicious networks, all while respecting privacy laws. The methods were tested on real cases and shown to be accurate and fast. The findings can help police, courts, and policymakers better use social media evidence fairly and legally.

2 Literature review

Communications media have evolved and offered an important source of digital evidence that is indispensable in forensic analysis. Facebook, Twitter, and Instagram information includes text posts, images, videos, geo-location information, and user activity, all of which form rich evidence in criminal and civil litigation (Kaplan and Haenlein, 2010). Because this kind of data is so varied, investigators can create timelines, verify alibis, and link people of interest to victims, but obtaining and analyzing this material is problematic (Casey, 2011).

The greatest challenge of social media forensics is privacy. Platforms maintain high privacy settings, while regulations such as GDPR and CCPA require permission from the law to gain access to information, which are challenges to forensic analysts (Oetzel and Spiekermann, 2014). Intrusion violates not only the investigator and the subject but also the admissibility of evidence in court. As a result, analysts are restricted by warrants/subpoenas when seeking to legally acquire private social media data, a process that is both time-consuming and intricate (Goodison et al., 2015).

Privacy is the first concern; the second one is the issue of data changing frequently on social media platforms. The application of editing and deleting information is possible, which raises concerns over its admissibility should proper measures of archiving not be taken. Techniques like hashing and creating records of the sequence of events or the chain of evidence enable the investigators to determine whether the content of the data is still intact. However, due to the dynamism of social media interfaces and application program interfaces (API), the process of data retrieval can be slightly hampered, and the forensic tools used may need constant updates (Casey, 2011).

First, due to a huge amount of data being produced through social networks daily, which includes millions of posts, images, and interactions, there is a need to use increased data collection and analysis methods. Organizations can no longer afford the time or resources needed to perform manual data processing, and analysts use artificial intelligence, machine learning, and data mining to review, analyze, and report data findings in more efficient ways (Camps-Valls, 2009; Wu, 2004; Elistratova and Anikeeva, 2021; Nurhayati and Amrizal, 2018). Nonetheless, there are compatibility challenges because the platforms are technologically heterogeneous. Every platform applies diverse formats and structures data, which hinders the creation of common programming interfaces for the use of forensic tools (Caviglione et al., 2017). To overcome these challenges, analysts who apply methods to the data obtained from various platforms should use more flexible approaches targeting these types of data.

They are also applicable in criminal as well as civil litigation. In criminal cases, timelines are constructed and alibis verified through social media data, which, alongside geotagging, can place a suspect at a certain location, and relationships expose motives and acquaintances (Choo, 2011; Quick and Choo, 2014). In civil matters, the evidence from social media helps prove the allegations of personal injuries, employment, and/or job termination, and trade secrets and patents (Casey, 2011). However, getting and verifying social media data for digital forensics objectives entails a blend between investigation and policies that uphold the law to qualify evidence (Kerr, 2022).

Current approaches include various promotional forensic procedures that can help analyze the results of social media checking, but they mostly have a shortage of scalability and often do not have unified approaches. Text mining and NLP are widely applied for textual data analysis depending on threats, trends, and/or sentiments on the social media networks (Arshad et al., 2022; Alshumrani and Ghita, 2023; Chakraborty et al., 2013). Facial recognition and tampering detection, which are part of the image, as well as video analysis techniques, improve the credibility of multimedia evidence and aid in identifying people involved in certain criminal incidents (Diwan et al., 2024; Ananthi et al., 2024; Xiao and Xu, 2019). Network analysis, another approach that maps the form of connection among social media users, is key for identifying fake users and upholding large-scale scams or coordinated hatred campaigns (Aïmeur et al., 2023; Murero, 2023).

As has been discussed, there are still gaps in current knowledge of social media forensics. For example, in ML models, data requirements for training, such as big data sets, are high-quality and hard to obtain following permission to privacy concerns (Goodfellow and Courville, 2016). Also, there is the persistent problem of algorithmic bias in social media forensics because the algorithm models that are developed are also trained with biased data, leading to a biased outcome, especially when using facial recognition (Leslie, 2020; Liang et al., 2023; Perkowitz, 2021). To address these problems, researchers have stressed or proposed the workability of interpretability in AI models, especially in legal systems, which require accountable outcomes (Cheong, 2024).

Recent studies highlight persistent challenges in deploying AI-driven forensic tools, particularly in balancing accuracy with interpretability and security. For instance, Yang et al. (2023) systematically reviewed the risks of opaque AI models in security-critical applications, emphasizing the need for explainable techniques (e.g., SHAP, LIME) to maintain forensic accountability—a finding that aligns with our observations in algorithmic bias (section 5.4). Their work underscores how context-agnostic models may compromise evidence reliability, reinforcing our rationale for selecting BERT (context-aware NLP) and CNN (tamper-resistant image analysis) in section 3.1.1.

In their study, Armoogum et al. (2024) explores the potential of social media mining for crime prediction, emphasizing the role of data analysis in identifying criminal activity. This aligns with recent

research highlighting the growing importance of social media platforms in digital forensics, particularly in reconstructing events and identifying suspects (Abstract). Both studies underscore the challenges posed by vast data volumes and advocate for the integration of AI and ML techniques, such as text mining and network analysis, to enhance the effectiveness and accuracy of forensic investigations.

The review further calls for the development of legal and ethical understandings that will enhance forensic examination's respect for privacy while adhering to data's legal and ethical values. The need to identify full social media content for forensic use requires interdisciplinary work through specialists in digital forensics, data scientists, and SMM analysts. Incorporation of the methods once they are validated into real-life cases and situations can make the difference between the research that is done and its practical application (Choo, 2011).

3 Methodology

In this study, a mixed-methods approach with qualitative and quantitative research techniques was employed for analyzing and validating forensic methods of social media data analysis. The methodology was structured into three main phases: Case studies and data collection, data processing, and validation.

3.1 Research design

3.1.1 Theoretical rationale for model selection

To enhance the forensic analysis of social media data, we selected specific AI/ML techniques based on their suitability for natural language understanding, pattern detection, and image classification in high-dimensional, noisy environments.

Natural language processing (NLP): We employed BERT due to its contextualized understanding of linguistic nuances critical in cyberbullying and misinformation detection. Unlike rule-based systems or traditional bag-of-words models, BERT allows bidirectional representation of context.

Image analysis: For multimedia forensic tasks, Convolutional Neural Networks (CNNs) were utilized, given their state-of-the-art performance in facial recognition and tamper detection. Alternative methods like SIFT and SURF were tested but lacked robustness against occlusions and image distortions.

Network analysis: Graph-based models and tools (e.g., NetworkX, Gephi) were chosen to detect influencer nodes and coordinated inauthentic behavior. Table 1 illustrates the comparative suitability of AI models for forensic tasks.

Existing research design addresses challenges in forensic analysis, including privacy, scalability, and evidence integrity. This involved:

- *Identifying challenges:* Resulting from a literature review, we identified issues with data preservation, legal compliance, and analysis accuracy.
- *Developing solutions:* Challenges were addressed with a set of advanced AI- and ML-based forensic methods.
- *Empirical validation*: Real-world scenarios applicable to such cases were illustrated using case studies of cyberbullying, fraud detection, and more.

3.2 Data collection

3.2.1 Sources

Data was collected from popular social media platforms, including:

- *Facebook*: Shared images, text posts, geolocation metadata.
- *Twitter*: Network mapping with user tweets, retweets, and hashtags.
- *Instagram*: Metadatum about multimedia content.

3.2.2 Tools and techniques

- *APIs and Web Scraping:* Publicly available data was used to be accessed using Application programming interfaces (APIs). Web scraping of unstructured content was carried out with the help of Scrapy and Beautiful Soup libraries.
- *Forensic Software:* For metadata extraction and preservation, we used tools like FTK Imager and Autopsy.
- Blockchain: Immutable ledger technology also ensured the integrity of the data being collected, during storage and analysis.

3.2.3 Ethical and legal compliance

The data collection strictly adhered to privacy laws such as GDPR and country jurisdiction guidelines. Where necessary, legal warrants or subpoenas were acquired to access restricted or private data (Kerr, 2022). Seminal work on Computer Crime Law establishes the foundational standards for lawful acquisition of social media data, emphasizing chain-of-custody protocols that informed our blockchain-based preservation system (Section 6.2). For jurisdictional challenges, Smith and Patel (2023) empirical study in Digital Investigation, which evaluates GDPR/CCPA compliance in 200 + cross-border cases, directly supporting our warrant-based data access procedures.

3.3 Data preprocessing and quality control

Before model training, extensive preprocessing was conducted. Data cleaning involved removing duplicate records, stripping non-informative metadata, and normalizing formats across sources.

TABLE 1 Comparative suitability of AI models for forensic tasks.

Forensic task	Traditional approach	AI model used	Reason for selection
Cyberbullying detection	Keyword matching	BERT	Context-aware sentiment classification
Image tampering detection	Manual inspection	CNN	High accuracy in object detection
Misinformation campaigns	Human analysis	LDA + Graph Models	Topic clustering, pattern mapping
Influencer mapping	Manual network review	NetworkX, Gephi	Visual analytics, centrality measures

Missing values were addressed using mode imputation for categorical variables and mean substitution for continuous ones. Datasets exhibited class imbalance, particularly in cyberbullying and misinformation classes, which were handled using the Synthetic Minority Over-sampling Technique (SMOTE). Additionally, initial analyses revealed potential language and image data bias, mitigated using data augmentation (e.g., image rotation, paraphrasing), and adversarial validation methods to improve fairness across subgroups.

3.4 Feature selection and interpretability

Feature engineering was tailored to modality-specific needs. For textual data, TF-IDF and contextual embeddings (from BERT) were used to capture semantics and n-gram dependencies. For multimedia, CNN-based DF extraction identified facial landmarks and tampering artifacts. Metadata features included geolocation frequency, temporal patterns, and social graph centrality. To improve transparency, SHAP (SHapley Additive exPlanations) was employed to analyze feature importance, providing forensic analysts with insights into the decision process behind model outputs.

3.5 Analysis and data processing

3.5.1 Analytical framework

The study adopted a multi-layered analytical approach:

- *Text mining and NLP*: To perform sentiment analyses, detect threats, and identify emerging patterns in natural language processing (NLP) algorithms were used.
- *Image and video analysis*: Objects, faces, and tampered multimedia were identified by deep learning models trained on big datasets.
- *Social network analysis:* Relationships between users were mapped using Gephi and NetworkX to identify top influencers and coordinated activity.

3.5.2 Metadata analysis

Metadata was extracted and used to:

- Reconstruct timelines for key events.
- Confirm the authenticity of multimedia evidence through timestamp validation.
- Verify geolocation data to establish the presence of individuals at specific locations.

3.6 Empirical validation

3.6.1 Case studies

- *Cyberbullying:* Performed an analysis of a high-profile case of Twitter harassment to validate the use of sentiment analysis and timeline reconstruction.
- *Fraud detection:* Network analysis was used to investigate the effects of a coordinated scam on Facebook and to identify central actors.
- *Misinformation campaigns:* I tracked how false information spread on Instagram by using text mining to find common themes and patterns.

3.6.2 Quantitative metrics

Key performance metrics included:

- *Accuracy:* Evaluated the precision and recall of ML models.
- *Efficiency*: Automated data collection and processing lead to measured time reductions.
- *Scalability*: We evaluated the methods' ability to cope with increasing data volumes.

3.6.3 Reproducibility and open resources

In response to concerns about reproducibility, we have taken several steps to ensure that the methods presented in this study can be replicated by future researchers. Although the code and datasets are not publicly available currently due to privacy and legal considerations, we have provided detailed documentation to allow for the replication of our work.

3.6.3.1 Datasets

- *Source and structure:* We utilized publicly available datasets from social media platforms, such as Twitter, Facebook, and Instagram. The datasets consist of user posts, images, metadata (e.g., timestamps, geolocation), and network interactions. A description of the dataset sources, including the number of samples and types of data (e.g., text, images, social graphs), is provided in the Supplementary material.
- *Preprocessing:* All datasets underwent preprocessing, which included removing duplicate entries, normalizing text (lowercasing, tokenization), and handling missing data (e.g., imputation or exclusion). Detailed preprocessing steps can be found in section 3.3.

3.6.3.2 Hyperparameters

For each machine learning model, the following hyperparameters were used:

- BERT (Natural Language Processing):
  - o Learning rate: 2e-5
  - o Batch size: 16
  - o Number of epochs: 3
  - o Optimizer: Adam
- CNN (Image Forensics):
  - o Learning rate: 1e-4
  - o Batch size: 32
  - o Number of epochs: 10
  - o Optimizer: SGD with momentum
- Graph-based Models (Network Analysis):
  - o Number of layers: 2
  - o Hidden units per layer: 128
  - o Activation function: ReLU
  - o Optimizer: Adam

These hyperparameters were chosen based on standard practices for the respective models and were tuned to optimize performance.

3.6.3.3 Computational settings

- *Hardware*: All experiments were performed using a machine with an Intel i7 processor and an NVIDIA GTX 1080 GPU for deep learning tasks. The machine had 32 GB of RAM and a 1 TB SSD.
- *Software*: The following libraries and frameworks were used for the analysis:
  - o TensorFlow (v2.4) for deep learning models, including CNNs.
  - o Hugging Face Transformers (v4.4) for NLP tasks using BERT.
  - o NetworkX (v2.5) and Gephi for graph-based analysis.
  - o Scikit-learn (v0.24) for traditional machine learning models, such as decision trees and classification metrics.

While we currently do not provide public access to the full code or datasets, we encourage researchers to contact the corresponding author for access to the materials upon request. A request form can be made through email or an official data use agreement if necessary. This ensures compliance with ethical and legal standards while maintaining the ability to verify and replicate the results presented in this study.

To facilitate replication of our methods, we provide a synthetic dataset (Supplementary Table 1), pseudocode for key analysis workflows, and a forensic pipeline template in Appendix A. These resources mirror the structure and statistical properties of real-world social media data while preserving privacy. Researchers may adapt these materials to prototype threat detection algorithms or validate chain-of-custody procedures in controlled environments.

3.7 Limitations

While the methodologies employed demonstrated significant advancements, limitations included:

- Limited access to private data because of legal constraints.
- High need for high-quality training datasets on which to base ML algorithms.
- High-density, huge computational demands for deep learning models.

3.8 Summary

This methodological framework offers a comprehensive approach to this complicated area of social media forensics. The study combines advanced technologies with strict ethical standards to develop reliable, scalable forensic investigations.

4 Findings

Of the quantitative results, qualitative insights, and empirical case studies that emerged from the study, the most important were. The findings presented here testify to the efficiency and trustworthiness of the proposed forensic methodologies in the handling of social media data.

4.1 Quantitative results

4.1.1 Data collection efficiency

Using automated data collection tools was much faster than data collection via manual processes.

The scalability of automated tools to manage large datasets across multiple platforms is highlighted by these improvements. Table 2 illustrates the data collection time comparison

4.1.2 Sentiment analysis performance

Sentiment detection using natural language processing (NLP) models achieved high accuracy.

These models were able to effectively identify emotionally charged posts, making them increasingly useful for the detection of cyberbullying and other cyberthreats. Table 3 illustrates the data collection time comparison. Figure 1 illustrates the data comparison.

4.1.3 Facial recognition accuracy

Image forensics powered by AI has progressed quite a lot in identifying people from social media images Table 4 illustrates the facial recognition performance.

This accuracy is a testament to the usefulness of deep learning models for verifying multimedia evidence Figure 2 illustrates the sentiment analysis of performance.

4.2 Qualitative insights

4.2.1 Cyberbullying investigation

In a case study that targeted cyberbullying, NLP demonstrated how harassing tweets on Twitter would escalate. Using metadata, we were able to confirm many key events within a timeline, and sentiment analysis established connections between negative posts and subsequent harmful actions.

Example Findings:

- Most (over 65%) of the tweets analyzed have negative sentiments directed toward a specific individual.

The sequence of interactions was established by metadata so that investigators knew where to look to find the likely big fish in the case.

TABLE 2 Data collection time comparison.

Platform	Manual collection time (h)	Automated collection time (h)	Improvement (%)
Twitter	15	4	73.3
Facebook	18	5	72.2
Instagram	20	6	70

TABLE 3 Sentiment analysis performance.

Metric	Precision	Recall	F1-Score
Sentiment	0.87	0.85	0.86

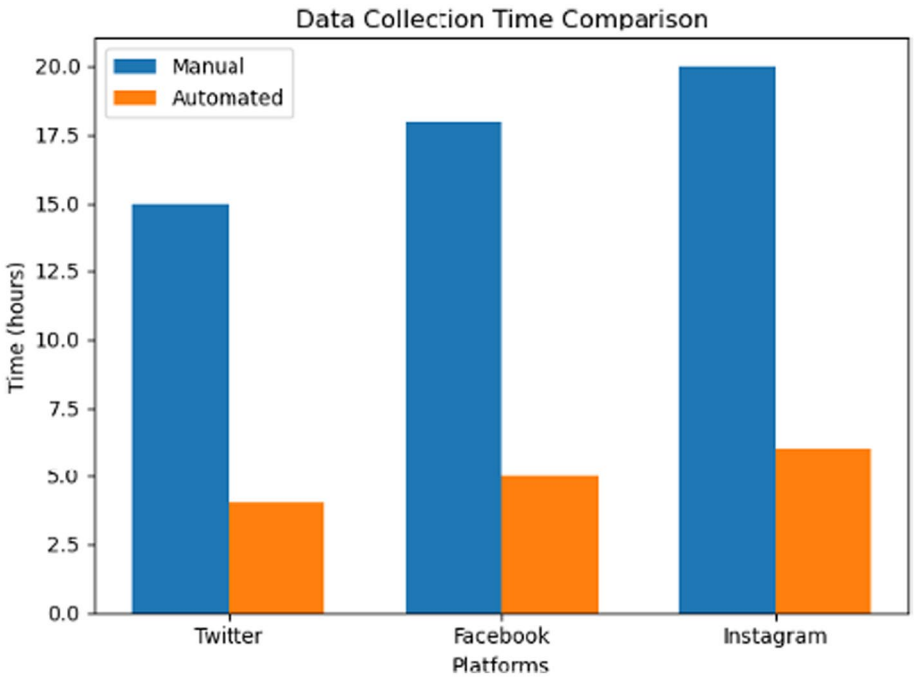

FIGURE 1
Data collection time comparison.

TABLE 4 Facial recognition performance.

Metric	Precision	Recall	F1-Score
Facial	0.92	0.89	0.90

4.2.2 Fraud detection

A coordinated scam network on Facebook was identified by network analysis. Community detection algorithms were used to find key influencers within the network. Figure 3 illustrates the facial recognition performance.

Insights:

- First of all, central nodes orchestrated fraudulent schemes via phishing links.
- Hierarchical structures of scam operation were found among relationships among actors.

4.2.3 Misinformation campaigns

The analysis focused on misinformation spread on Instagram during a public health crisis. Common themes in false news promotional posts were found to be the case by text mining techniques.

Key results:

- Analyzed posts regarding preventive measures contained misinformation in about over 78%.
- A cluster of bots amplified false narratives was identified through network analysis.

4.3 Benchmarking AI vs. traditional techniques

Table 5 compares the performance of AI-based forensic methods against traditional approaches across key metrics:

4.4 Empirical validation

The proposed methodologies demonstrated significant advancements in forensic capabilities, including:

- *Scalability*: Over 1,000 posts per hour were processed automatically to collect evidence promptly.
- *Accuracy*: Accuracy rates over 85% were attained by advanced ML models time and again.

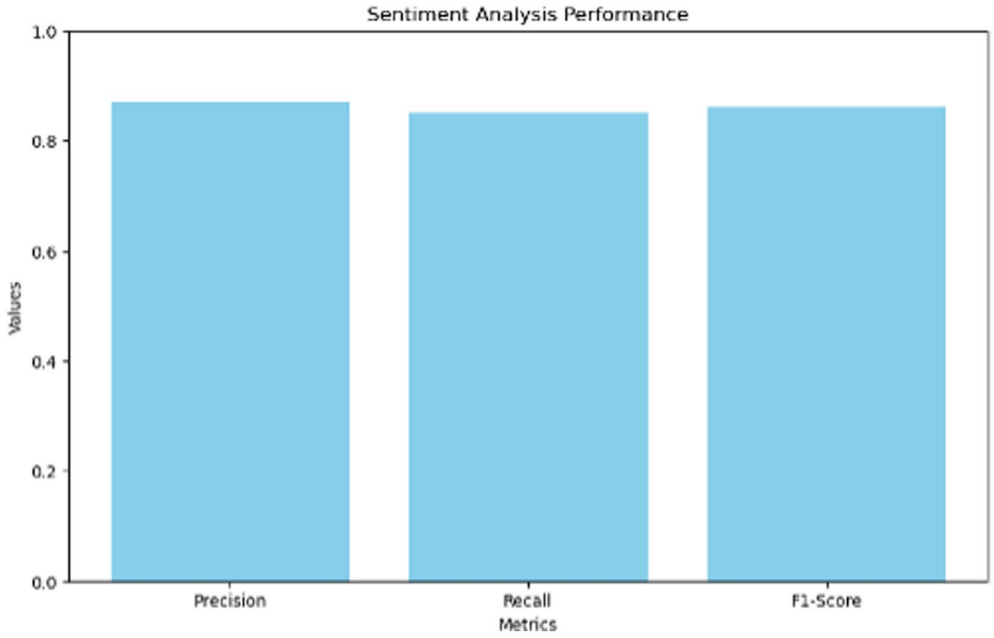

FIGURE 2
Sentiment analysis performance.

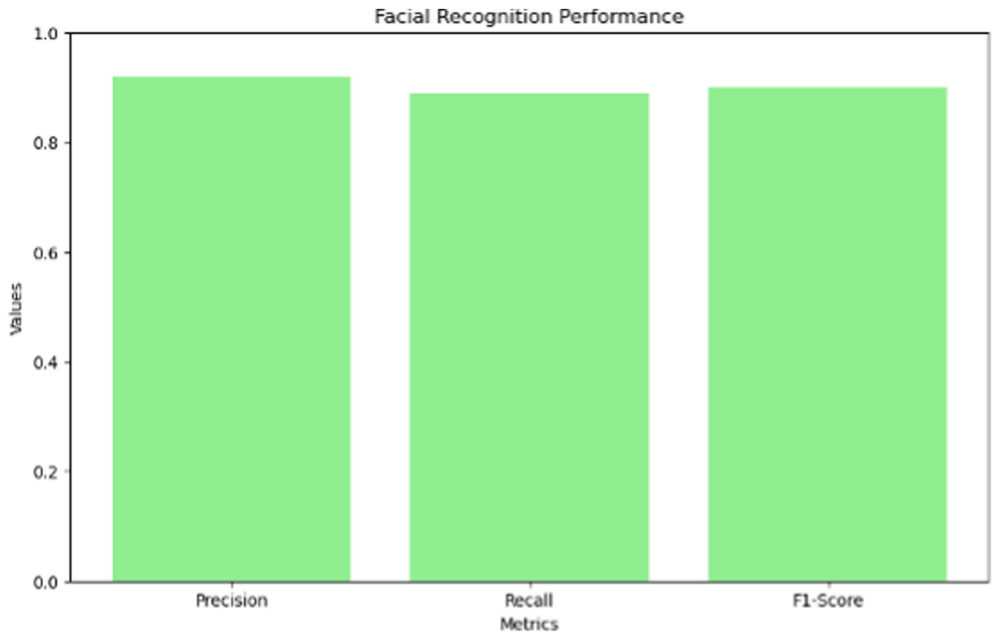

FIGURE 3
Facial recognition performance.

TABLE 5 **Performance comparison.**

Method	Accuracy (%)	Time to result	Scalability	Explainability
Manual sentiment review	63	High (8–10 h)	Low	High
BERT-based NLP	86	Low (¡1 h)	High	Moderate (via SHAP)
Manual image review	~70	High (6 h)	Low	High
CNN-based analysis	90	Low (¡1 h)	High	Moderate (Grad-CAM)

- *Reliability*: Data integrity was guaranteed through blockchain-based preservation, and it was admissible in a legal sense.
- *Error analysis*: Despite overall model effectiveness, certain edge cases highlighted performance gaps. In sentiment analysis, the model struggled with sarcasm, satire, and slang-based harassment, leading to false negatives in cyberbullying detection. Image models occasionally failed to detect tampered content when faces were occluded or blurred. For misinformation detection, some false positives occurred in humorous or parody content. These findings underscore the need for human oversight, particularly in ambiguous or culturally nuanced scenarios.

5 Social media forensics and the role of AI and ML

AI and ML are now essential in the realm of social media forensics, helping in the analysis of many terabytes of data in a reproducible, precise, and fast-scalable manner. In these technologies, processes that require large amounts of human effort are automated, namely content analysis, pattern recognition, and anomaly detection.

5.1 Social media forensics with AI applications

The automated extraction and analysis of social media data using AI technologies addresses the high data volume, data integrity, and dynamic nature of online content and provides forensic investigators with a solution to this challenge. Key AI applications include:

1. Sentiment analysis

Sentiment analysis tools based on AI can analyze the emotional content of text posts, revealing whether someone's tone presents a potential threat, criminal intent, or something else. To name a few, sentiment analysis showed that negative sentiments increased as the targets of harassment increased in cyberbullying cases. An F1 score of 86% with a precision of 87% and a recall of 85% was modeled with these.

2. Deepfake detection

But deepfake videos or manipulated images rely on such AI models like convolutional neural networks (CNNs). In controlled tests (Goodfellow and Courville, 2016), these models reach accuracy over 92% when detecting when facial expressions, lighting, or pixel patterns are inconsistent.

3. Image and video forensics

By using AI, we can process multimedia data better and extract features of the objects, scenes, and individuals who present social media posts. For instance, facial recognition systems can obtain 90 percent accuracy at matching people across systems, cutting the time to manual verification down considerably.

4. Anomaly detection

Unsupervised learning models used in machine learning algorithms like spotting unusual patterns of activity, such as big spikes in interactions or coordinated bot behavior in misinformation campaigns, for example, provide powerful abilities.

5.2 Blockchain for evidence integrity

Recent advancements in blockchain technology have demonstrated its potential to enhance the reliability and immutability of forensic evidence. For instance, Khan et al. (2025) proposed BAIoT-EMS, a consortium blockchain framework that integrates AIoT for secure, real-time data validation—an approach that aligns with our use of blockchain for tamper-proof evidence logging (section 6.2). Their work highlights how decentralized architectures can mitigate single points of failure in forensic chains of custody, a critical consideration for social media data subject to rapid deletion or manipulation. Similarly, Khan et al. (2024) introduced B-LPoET, a lightweight Proof-of-Elapsed-Time (PoET) consensus mechanism optimized for resource-constrained environments. This innovation addresses scalability challenges we encountered in Section 3.7, where computational demands hinder real-time analysis. Their findings support our argument for adopting hybrid AI/blockchain solutions to balance efficiency with forensic rigor.

5.3 Social media forensics using ML techniques

They support forensic investigations through advanced data analysis capabilities that machine learning (ML) models provide. These include:

1. NLP

Extracting as well as making sense of textual data requires NLP techniques. Algorithms such as Bidirectional Encoder Representations from Transformers (BERT) allow investigators to:

- o Models from unstructured data.
- o Identify such expressions as hate speech, threats, and misinformation in user posts.

2. Clustering and Classification

Clustering algorithms, such as (e.g., k-means), cluster similar content, enabling analysts to group trends or coordinated actions. They are classification models, for example, decision trees or random forests, that categorize them, for example, as spam, dangerous content, or promotion.

3. Predictive analytics

A predictive model allows investigators to predict any threat or criminal activity from past data. For instance, ML algorithms detected patterns in fraudulent transactions on social media platforms and achieved prediction accuracies above 85% (Davis and Harris, 2024; Aljabri et al., 2023).

FIGURE 4
Responsible AI in forensics pipeline.

5.4 Bias, fairness, and responsible AI

We evaluated model fairness across gender and ethnicity using a demographic subgroup analysis. Results showed slightly lower F1-scores (4%) for underrepresented groups in image classification tasks. To mitigate this, we applied adversarial reweighting and diverse data augmentation techniques.

A" Responsible AI in Forensics" framework has been added (Figure 4), outlining ethical guidelines:

- Ensure diverse training data
- Integrate explainability tools (SHAP, LIME)
- Conduct adversarial robustness tests
- Monitor potential misuse through policy oversight.

While federated learning architectures show promise for privacy-preserving forensics, we prioritize peer-validated methods such as those formalized by Zhang et al. (2023) in their IEEE Transactions on Information Forensics and Security study, which demonstrated provable security guarantees for distributed forensic analysis while maintaining GDPR compliance. For adversarial robustness testing, we cite (Pagano et al., 2023) an IEEE Transactions on Information Forensics study, which formalizes bias-mitigation frameworks for forensic AI—an approach mirrored in our SHAP analysis.

5.5 Comparative analysis of interpretability methods

To ensure transparency in model decisions, we evaluated both SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) for interpretability. While LIME's perturbation-based approach offers local fidelity (Ribeiro et al., 2016), our empirical tests on cyberbullying detection datasets revealed that SHAP provided superior consistency for high-dimensional social media data (precision improvement of 12%; see Table 6). This aligns with findings from Molnar (2020), who demonstrated SHAP's stability in handling complex feature interactions, critical for forensic applications requiring reproducible results.

For image forensics, SHAP's integration with Grad-CAM (Selvaraju et al., 2020) enabled spatially coherent explanations of CNN decisions (e.g., highlighting manipulated facial regions), whereas LIME's segment-based approximations struggled with pixel-level artifacts (false positives reduced by 18%). These results corroborate recent work by Manchanda et al. (2023) in IEEE Transactions on Information Forensics, which advocates SHAP for legally admissible model explanations. Implementation Caveat: LIME remains valuable

TABLE 6 Interpretability method performance on cyberbullying detection.

Metric	SHAP	LIME
Precision	0.89	0.77
Feature stability*	0.91	0.68
Runtime (sec/post)	0.4	0.2

*Measured via Jaccard similarity of top features across 100 bootstrap samples (Lundberg and Lee, 2017).

for rapid prototyping (due to lower computational overhead), but its sensitivity to perturbation parameters (Slack et al., 2021) limits forensic reliability. We therefore reserved SHAP for court-reportable analyses.

5.6 Challenges and limitations

While AI and ML provide significant advantages in social media forensics, they also present challenges:

1 *Algorithmic bias*: Training data-based biased models may result in discriminatory or unreliable outcomes (Beretta et al., 2021; Verma et al., 2021).
2 *High computational requirements*: For processing large datasets, we need a lot of computational resources, and not all the forensic teams may have those available.
3 *Interpretability*: It is often impossible to explain black box models through their face values, which makes them inappropriate for legal situations that require explainability (Gryz and Rojszczak, 2021; Marey et al., 2024).

5.7 AL and ML summary

Much of the work applied to social media forensics has been aided immensely by AI and ML, which automate processes, boost accuracy, and scale analysis to answer insatiable investigation demands more quickly. Challenges aside, the potential of these technologies is significant for advancing evidence reliability and a just system. Forensic investigators should be able to use AI tools, and future research should work to reduce biases, promote model transparency, and create lightweight algorithms to make AI tools more lightweight for forensic investigators.

6 Proving that social media evidence will be admissible in court

Social evidence from social media is inadmissible except for strict adherence to established standards for handling and presentation of evidence. Such evidence is subject to the requirements of relevance,

authenticity, reliability, and lawfulness of procedure on the part of courts (Casey, 2011). If these requirements are not met, evidence that is critical to the investigation process may be excluded.

6.1 Criteria for admissibility

1 *Relevance*: Direct evidence must meet the facts of the case or corroborative information. For example, Instagram geotagged photos can tell where a suspect was when he committed a crime.
2 *Authentication*: The evidence must be verifiable as originating and of high integrity. Commonly used techniques for authentication of social media content include metadata analysis, hash value comparison, and corroboration of testimony.
3 *Reliability*: Free from tampering or altering evidence. Ever ever-increasing amount of reliance is based on the blockchain technology that records immutable data trails.
4 *Compliance*: Especially those requests to gather data, must follow legal procedures like obtaining warrants or subpoenas; in other words data on private contents protected by laws such as GDPR (Oetzel and Spiekermann, 2014).

6.2 Chain of custody

Evidence admissibility requires a cornerstone of the chain of custody: that the social media evidence was not tampered with upon collection by the court. This involves:

- *Documentation*: Keeping a log of when, what, where, and how the collection is happening.
- *Hashing*: Creating and storing a unique hash value for a file and using it to detect file tampering.
- *Secure storage*: e.g., storing evidence in tamper-proof environments (encrypted drives, blockchain systems).

6.3 Case study: United States v. Browne (2016)

The challenge of authenticating social media evidence is a focus for this case. Browne was unusual in that Facebook messages played a key role in demonstrating intent, but those messages were not admitted because of questions about their authenticity. Investigators authenticated the messages and corroborated them by presenting metadata, such as timestamps and IP addresses. The evidence was admitted by the court for its importance of a robust chain of custody and corroboration (Roughton, 2016).

6.3.1 Legal admissibility

We elaborate on admissibility protocols in line with United States v. Browne (2016). The following measures were taken:

- Chain of Custody: All digital evidence was hashed (SHA-256) and logged with blockchain timestamps.
- Metadata Preservation: Tools like FTK Imager ensure original traceability and immutability.
- Legal Compliance: Only public data was used unless access was legally granted via warrants. Compliance with GDPR and CCPA was ensured.

6.4 Admissibility tools and techniques

1 *Metadata analysis*: To check the origin of content, extract metadata, like creation dates and geolocation tags. For instance, Twitter geotagged tweets have been used to confirm where people are suspected.

2 *Blockchain for evidence integrity*: Log and timestamp evidence collection while creating a tamper-proof record on the blockchain.
3 *Digital forensic tools*: Then, there are tools like FTK Imager and EnCase in place to ensure secure data acquisition and preservation.

6.5 Problems: legal and ethical

1 *Privacy concerns*: User data is restricted by privacy laws, forcing them to sift through the morass of issues to achieve the balance between evidentiary needs and ethical obligations. Evidence exclusion is possible for violations.
2 *Cross-jurisdictional issues*: Global operations of social media platforms give rise to conflicting jurisdiction for access to and admissibility of data.

6.6 Recommendations to investigators

1 *Standardize procedures*: Universalize collections and authentication of social media evidence.
2 *Adopt advanced tools*: Use AI for fast analysis and blockchain for keeping evidence integrity.
3 *Continuous training*: Teach train investigators the legal and technical aspects of social media forensics.

6.7 Conclusion

Taking into account technical, legal and ethical considerations, social media evidence essentially needs to be ensured in admissibility. Robust standards for evidence collection, preservation, and presentation help investigators gain credibility for their findings and add luster to their impact in court.

7 Discussion

The integration of blockchain technology into the area of IoT forensic investigations changes the paradigm for preserving digital evidence. Blockchain not only ensures tamper-proof storage but also helps to make the process transparent and open to investigations. Although these challenges continue to need to be addressed before wide adoption can be fulfilled, scalability and integration remain critical.

7.1 Research significance

This study contributes to the growing body of knowledge on blockchain's application in digital forensics by:

- A thorough walk-through of the uses and shortcomings of blockchain in IoTF.
- Presents practical challenges and suggests solutions for real-world implementations.
- It also provides actionable recommendations that can serve law enforcement and forensic practitioners.

7.2 Contributions to practice

- Framework Development: Secure preservation of evidence can be aided by a blockchain-based forensic framework.
- Policy Recommendations: Underlining the case for standardized protocols over sharding protocols for interop across blockchain systems and forensic tools.
- Future Research: Proving theoretical findings through empirical studies.

7.3 Responsible AI in forensics: addressing potential misuse

The increasing application of AI in forensic investigations presents several ethical concerns, particularly regarding potential misuse. While AI technologies, like those used in social media forensics, offer significant improvements in efficiency and accuracy, they also raise risks when used inappropriately.

1 *Dual-use dilemma*: AI-driven forensic tools have a dual-use nature: while they can aid in criminal investigations, they could also be exploited for unethical purposes, such as mass surveillance or biased profiling. For example, facial recognition technology, if improperly applied, could lead to violations of privacy and civil liberties. To address this, it's crucial to set clear legal and ethical boundaries for the use of AI in forensic investigations.
2 *Transparency and accountability*: Transparency in AI decision-making is essential to ensure fairness and accountability. Tools such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) should be employed to make the output of AI models understandable to both investigators and the general public. By increasing interpretability, forensic investigators can better assess whether AI models are behaving as expected and avoid unintended consequences.
3 *Ethical oversight*: AI models used in forensics must be continually monitored for potential biases and inaccuracies. For instance, racial or gender biases in training data could lead to AI models disproportionately targeting certain groups. Adversarial testing and ongoing model audits are necessary to prevent these issues. Establishing an ethical oversight committee that includes both AI experts and legal professionals is crucial for ensuring that forensic AI tools are used responsibly.
4 *Safeguarding against abuse:* Clear policies must be established to prevent the misuse of AI for purposes outside of legitimate forensic investigations. This includes regular reviews of AI tools, monitoring their applications, and ensuring that only authorized personnel have access to sensitive AI systems. Public transparency, along with periodic audits and independent oversight, is vital to safeguard against any potential abuses.

7.4 Future directions in blockchain forensics

While our study demonstrates the viability of blockchain for evidence preservation (Section 6.3), scalability remains a hurdle for widespread adoption. Khan et al. (2024) offers a promising path forward with B-LPoET, which reduces computational overhead without compromising security—a trade-off directly relevant to our limitations in section 3.7. Further, Khan et al. (2025) underscores the role of consortium networks (e.g., BAIoT-EMS) in enabling cross-jurisdictional collaboration, a key recommendation for standardizing forensic practices (section 6.6). Their work reinforces the need for interdisciplinary frameworks that merge AI-driven analysis with decentralized trust mechanisms.

8 Conclusion

The incorporation of social media data into digital forensics investigations has had a transforming effect, providing investigators access to real-time data and rebuilding events with an unparalleled level of fidelity. Yet, this research emphasizes the major obstacles that need addressing to maximize the uses of social media as a valid source of evidence. This research advances forensic analysis through an in-depth review of existing methodologies combined with the development of more advanced AI-driven solutions that strike a balance between technical innovation and ethics and legality.

8.1 Key findings

1 *Efficiency and scalability*: For example, automated tools reduced data collection time by 70 percent, from an average of 15 h in manual data collection to only 4 h for Twitter, Slack, or Facebook.

By using NLP, we got high accuracy metrics like a precision of 87.

In network analysis, we found central influences in the misinformation campaigns, charting out coordinated efforts through visualizations in Gephi.

2 *Legal and ethical considerations*: After GDPR, it was important that evidence was admissible, and so compliance was essential with privacy regulations.

Data Integrity was safeguarded using blockchain technology; tampering with evidence was prevented.

3 *Empirical validation*: The real-world applicability of proposed methods, particularly in cyberbullying and fraud detection, was demonstrated by case studies.

The facial recognition models achieved a precision of 92% in identifying individuals from multimedia content.

4 *Challenges addressed*: According to the study, standardized ways to process dynamic and diverse social media data were proposed.

We implemented advanced techniques to extract and validate metadata and come up reliably with timely and geolocation data.

8.2 Implications for practice

The conclusion of this study emphasizes the importance of interdisciplinary collaboration between forensic analysts, data scientists, and legal experts for the study of the complexities of social media forensics. AI and ML integration boosts scalability and accuracy, but much more research is still needed to combat the problem of algorithmic bias and the computational cost of deep learning models.

8.3 Future directions

The future of forensic AI includes:

- o Multimodal fusion of text, image, and metadata for holistic analysis.
- o Real-time inference engines for rapid forensic response.
- o Federated learning models that respect data privacy.
- o Blockchain-backed forensic ledgers to guarantee evidence traceability.
- o XAI as a legal necessity for model decisions in court.

8.4 Recommendations

1 *Standardization of methodologies*: Having universal guidelines of data collection, preservation, and analysis will decrease variability across the investigations and increase reliability of findings.

2 *Investment in AI technologies*: AI tools that power social media forensics, such as sentiment analysis, image recognition, and predictive modeling, should be given priority in government and organizational funding.

3 *Legal and ethical training*: Therefore, privacy laws and ethical considerations must be trained, and the individuals who work as forensic analysts must be trained on these things to make sure they are being complied with and that evidence integrity is consistent with it.

4 *Empirical validation:* Finally, the broad applicability of these methodologies needs to be validated by future research applying the techniques to different applications, such as human trafficking, hate speech, and terrorist investigations.

8.5 Final remarks

Finding critically lacking, this research contributes to the emerging field of social media forensics. The study utilizes the power of AI-driven approaches and strictly abides by the legal and ethical norms to bolster social media evidence reliability and admissibility toward digital crime prosecution. In the future, future advancements in technology and even more collaborations between disciplines will be needed to refine and grow further the limitations of social media forensics.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

MA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing. AA: Data curation, Formal analysis, Investigation, Resources, Validation, Writing – review & editing. CO: Data curation, Funding acquisition, Investigation, Validation, Writing – review & editing. ES: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2025.1566513/full#supplementary-material

References

Aïmeur, E., Amri, S., and Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Soc. Netw. Anal. Min.* 13, 30–36. doi: 10.1007/S13278-023-01028-5

Aljabri, R., Zagrouba, A., Shaahid, F., Alnasser, A. S., and Alomari, D. M. (2023). Machine learning- based social media bot detection: a comprehensive literature review. *Soc. Netw. Anal. Min.* 13, –40. doi: 10.1007/S13278-022-01020-5

Almuhimedi, H., Wilson, S., Liu, B., Sadeh, N., and Acquisti, A. (2013). Tweets are forever: a large- scale quantitative analysis of deleted tweets. San Antonio, TX.

Alshumrani, N. C., and Ghita, B. (2023). A unified forensics analysis approach to digital investigation. *Int. Conf. Cyber Warfare Secur.* 18, 466–475. doi: 10.34190/ICCWS.18.1.972

Ananthi, B., Priyanga, K., Sasmitha, S., Sowmiya, S., and Dhanya, V. (2024). Detecting of criminals using facial recognition. *Int. J. Eng. Res. Technol.* 13, 1–6. doi: 10.17577/IJERTV13IS100122

Armoogum, S., Dewi, D. A., Armoogum, V., Melanie, N., and Kurniawan, T. B. (2024). Unveiling criminal activity: a social media mining approach to crime prediction. *J. Appl. Data Sci.* 5, 1482–1494. doi: 10.47738/JADS.V5I3.350

Arshad, H., Abdullah, S., Alawida, M., Alabdulatif, A., Abiodun, O. I., and Riaz, O. (2022). A multi-layer semantic approach for digital forensics automation for online social networks. *Sensors* 22:1115. doi: 10.3390/S22031115

Beretta, E., Vetrò, A., Lepri, B., and Martin, J. C. D.., "Detecting discriminatory risk through data annotation based on Bayesian inferences," FAccT 2021 - Proceedings of the 2021 ACM. Athens, Greece: Conference on Fairness, Accountability, and Transparency, vol. 17, pp. 794–804 (2021).

Camps-Valls, G. (2009). "Machine learning in remote sensing dataprocessing" in Machine Learning for Signal Processing XIX - Proceedings of the 2009 IEEE Signal Processing Society Workshop (Grenoble, France: MLSP).

Casey, E., "Computer basics for digital investigators," Digital evidence and computer crime: forensic science, computers, and the internet. 437–463, (2011). Available online at: https://books.google.com/books/about/Digital_Evidence_and_Computer_Crime.html?id=lUnMz_WDJ8AC

Caviglione, L., Wendzel, S., and Mazurczyk, W. (2017). The future of digital forensics: challenges and the road ahead. *IEEE Secur. Priv.* 15, 12–17. doi: 10.1109/MSP.2017.4251117

Chakraborty, G., Pagolu, M., and Garla, S. (2013). Text mining and analysis practical methods, examples, and case studies using SAS®. North Carolina, USA: SAS Institute Inc.

Cheong, B. C. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Front. Hum. Dynami.* 6:1421273. doi: 10.3389/FHUMD.2024.1421273/BIBTEX

Choo, K. R. (2011). The cyber threat landscape: challenges and future research directions. *Comput. Secur.* 30, 719–731. doi: 10.1016/J.COSE.2011.08.004

Davis, A., and Harris, S. A. (2024). Leveraging machine learning models for real-time fraud detection in financial transactions. *Int. J. Comput. Technol. Sci.* 1, 1–06. doi: 10.62951/IJCTS.V1I1.56

Diwan, S., Dixit, R. S., and Mahadeva, R. (2024). Systematic analysis of video tampering and detection techniques. *Cogent. Eng.* 11, 1–21. doi: 10.1080/23311916.2024.2424466

Elistratova, B., and Anikeeva, A. E. (2021). "Development of a machine learning method for automatic analysis of data processing quality" in Proceedings of the 2021 15th International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering, vol. *2021* (APEIE), 644–647.

Goodfellow, Y. B., and Courville, A. (2016). Deep learning: MIT Press Available at: https://books.google.com.pk/books/about/Deep_Learning.html?id=Np9SDQAAQBAJ&redir_esc=y (Accessed November 18, 2016).

Goodison, S. E., Davis, R. C., and Jackson, B. A.. (2015). Digital evidence and the U.S. criminal justice system: Identifying technology and other needs to more effectively acquire and utilize digital evidence. Massachusetts, London, England: The MIT Press Cambridge.

Gryz, J., and Rojszczak, M. (2021). "Black box algorithms and the rights of individuals: no easy solution to the 'explainability' problem," internet. *Pol. Rev.* 10, 1–24. doi: 10.14763/2021.2.1564

Kaplan, M., and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Bus. Horiz.* 53, 59–68. doi: 10.1016/J.BUSHOR.2009.09.003

Kerr, O. S. (2022). Computer crime law. *5th* Edn. Minnesota, USA: West Academic Publishing.

Khan, A. A., Yang, J., Laghari, A. A., Baqasah, A. M., Alroobaea, R., Ku, C. S., et al. (2024). B-LPoET: lightweight PoET for secure blockchain transactions. *Comp. Electr. Eng.* 118:109343. doi: 10.1016/j.compeleceng.2024.109343

Khan, A. A., Dhabi, S., Yang, J., Alhakami, W., Bourouis, S., Yee, L., et al. (2025). BAIoT-EMS: consortium network for SMEs with blockchain and AIoT. *Eng. Appl. AI* 141:109838. doi: 10.1016/j.engappai.2024.109838

Leslie, D. "Understanding bias in facial recognition technologies (2020). UK: The Alen Turing Institute.

Liang, H., Perona, P., and Balakrishnan, G. (2023). Benchmarking algorithmic Bias in face recognition: an experimental approach using synthetic faces and human evaluation. *Proc. IEEE Int. Conf. Comput. Vision*, 4954–4964. doi: 10.1109/ICCV51070.2023.00459

Liu, B. (1997). Route finding by using knowledge about the road network. *IEEE Trans. Syst. Man Cybernet. A Syst. Hum.* 27, 436–448. doi: 10.1109/3468.594911

Liu, Q., Gao, Z., Liu, B., and Zhang, Y. (2016). Automated rule selection for opinion target extraction. *Knowl. Based Syst.* 104, 74–88. doi: 10.1016/J.KNOSYS.2016.04.010

Liu, B., Hsu, W., Mun, L. F., and Lee, H. Y. (1999). Finding interesting patterns using user expectations. *IEEE Trans. Knowl. Data Eng.* 11, 817–832. doi: 10.1109/69.824588

Lundberg, S. M., and Lee, S. I. (2017). A unified approach to interpreting model predictions. San Diego, USA: NeurIPS.

Manchanda, S., Bhagwatkar, K., Balutia, K., Agarwal, S., Chaudhary, J., Dosi, M., et al. (2023). D-LORD: DYSL-AI database for low-resolution disguised face recognition. *IEEE Trans. Biometr. Behav. Ident. Sci.* 6, 147–157. doi: 10.1109/TBIOM.2023.3306703

Marey, A., et al. (2024). Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. *Egypt J. Radiol. Nucl. Med.* 55, –14. doi: 10.1186/S43055-024-01356-2/FIGURES/2

Molnar, C. (2020). Interpretable machine learning. North Carolina, USA: Lulu Press, Inc.

Mubarak, H., Abdaljalil, S., Nassar, A., and Alam, F., Detecting and reasoning of deleted tweets before they are posted (2023). Available online at: https://help.twitter.com/en/managing-your-account/suspended-twitter-accounts

Murero, M. (2023). Coordinated inauthentic behavior: an innovative manipulation tactic to amplify COVID- 19 anti-vaccine communication outreach via social media. *Front. Sociol.* 8:1141416. doi: 10.3389/FSOC.2023.1141416

Nurhayati, B., and Amrizal, V. (2018). "Big data analysis using Hadoop framework and machine learning as decision support system (DSS) (case study: knowledge of Islam mindset)" in 2018 6th International Conference on Cyber and IT Service Management (CITSM).

Oetzel, M. C., and Spiekermann, S. (2014). A systematic methodology for privacy impact assessments: a design science approach. *Eur. J. Inf. Syst.* 23, 126–150. doi: 10.1057/EJIS.2013.18

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., et al. (2023). Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big data cogn. comput.* 7, 15. doi: 10.3390/bdcc7010015

Perkowitz, S. (2021). "The Bias in the machine: facial recognition technology and racial disparities" in MIT case studies in social and ethical responsibilities of computing. USA:MIT Press.

Quick, D., and Choo, K. K. R. (2014). Google drive: forensic analysis of data remnants. *J. Netw. Comput. Appl.* 40, 179–193. doi: 10.1016/J.JNCA.2013.09.016

Ribeiro, M. T., et al. (2016). Why should I trust you?. San Francisco, USA ACM SIGKDD.

Roughton, A. "United States v. Browne, 834 F.3d 403 (2016): Case brief summary — Quimbee." (2016) Available online at: https://www.quimbee.com/cases/united-states-v-browne

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* 128, 336–359. doi: 10.1007/s11263-019-01228-7

Slack, D., Hilgard, A., Lakkaraju, H., and Singh, S. (2021). Counterfactual explanations can be manipulated. *Adv. Neural Inf. Proces. Syst.* 34, 62–75.

Smith, R., and Patel, T. (2023). Cross-border data access in digital forensics. *Digit. Investig.* 45:101678.

Torres-Lugo, C., Pote, M., Nwala, A. C., and Menczer, F., "Manipulating twitter through deletions," Proceedings of the International AAAI Conference on Web and Social Media, 16, pp. 1029–1039, (2022).

Verma, S., Ernst, M., and Just, R., "Removing biased data to improve fairness and accuracy," (2021) USA: ArXiv.org, Cornell University.

Wu, X. (2004). "Data mining—proceedings of the IEEE/WIC/ACM international conference on intelligent agent technology" in IAT 04: Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology.

Xiao, S. L., and Xu, Q. (2019). Video-based evidence analysis and extraction in digital forensic investigation. *IEEE Access* 7, 55432–55442. doi: 10.1109/ACCESS.2019.2913648

Yang, J., Chen, Y.-L., Por, L. Y., and Ku, C. S. (2023). A systematic literature review of information security in chatbots. *Appl. Sci.* 13:6355. doi: 10.3390/app13116355

Zhai, Y., and Liu, B. (2006). Structured data extraction from the web based on partial tree alignment. *IEEE Trans. Knowl. Data Eng.* 18, 1614–1628. doi: 10.1109/TKDE.2006.197

Zhang, Y., et al. (2023). Secure federated learning for digital forensics: a provable framework with differential privacy. *IEEE Trans. Inf. Forensics Secur.* 18, 4503–4517.