Computer Fraud and Security
ISSN (online): 1873-7056

Generative AI for Claims Evidence Interpretation and Fraud Analysis

Sulabh Jain

Berkshire Hathaway Homestate Companies, USA

Abstract

The insurance claims process generates vast volumes of unstructured evidence that present substantial challenges for human adjudicators to analyze comprehensively and consistently. Generative artificial intelligence has emerged as a transformative technology for automating evidence interpretation and fraud detection across the insurance industry. Large language models process narrative evidence from claims descriptions, witness statements, and medical records to extract key facts and identify inconsistencies. Vision transformer architectures analyze claims imagery, including property damage photographs and accident scene documentation, to detect manipulation and assess damage severity. Multimodal transformer architectures integrate textual and visual information simultaneously, enabling correlation between written descriptions and photographic evidence. Fraud detection employs supervised machine learning models trained on historical claims data, unsupervised anomaly detection systems, and behavioral pattern analysis. Generative AI systems reduce document review time substantially while improving fraud detection accuracy when augmenting traditional rule-based indicators. Synthetic data generation addresses data scarcity challenges by creating realistic fraudulent claim examples for training purposes. However, significant technical challenges persist, including hallucinations where models generate factually incorrect information, reduced generalizability in fine-tuned models, adversarial attacks, and bias risks. Explainability requirements demand transparent reasoning for fraud flagging decisions through attention mechanism visualization and feature importance measures. Insurance regulators are putting more pressure on transparency and auditability in automated claims decisions, which will require full documentation of the decision and testing for bias across demographic categories. Privacy issues also require the safeguarding of sensitive policyholder data. Success will require balancing the recent and transformative capabilities of ethics and governance, oversight by humans, and regulatory compliance, which will be necessary to ensure fairness and accuracy in claims processing.

Keywords: Generative Artificial Intelligence, Insurance Fraud Detection, Claims Evidence Interpretation, Multimodal Transformers, Explainable AI

I. Introduction

The insurance claims process generates massive volumes of unstructured evidence. This evidence can consist of photographs, medical records, police reports, witness statements, expert assessments, and claims narratives. Human adjudicators face a substantial challenge throughout data comprehensively and consistently. The vast amount of information in modern claims will create bottlenecks in processing workflows. Traditional manual review methods struggle to keep pace with increasing claim volumes. Adjusters must examine multiple evidence sources simultaneously while maintaining accuracy and consistency across decisions [1].

Generative artificial intelligence has emerged as a transformative technology for automating evidence interpretation and fraud detection. Large language models, vision transformers, and multimodal architectures now enable systematic extraction and synthesis of complex claims evidence. The implications of these digitized operations will represent a profound shift from rule-based automation to intelligent systems that can have some contextual and nuanced understanding. The systems can process natural language, analyze images, and integrate information across multiple formats. This capability addresses longstanding challenges in claims processing that resisted conventional automation attempts [1].

These systems can recognize fraud indicators, contradictions, and anomalies that may not meet the threshold or standards for most human reviewers. These systems can also identify patterns that would take considerable time to manually review. The industry loses billions to fraudulent claims while legitimate claimants have to wait for claims verification. Traditional fraud detection relied on rigid rule sets and some statistical anomalies. These methods often generated high

false positive rates while missing sophisticated fraud schemes. Generative AI introduces adaptive detection capabilities that improve continuously through exposure to new fraud patterns [2].

Nevertheless, there are critical technical, operational, and governance challenges that are still underaddressed in the literature. The technology raises novel risks: model hallucination, amplification bias, and even adversarial manipulation. High reliability and regulatory compliance are required in insurance operations, demanding a thoughtful application of that technology. Organizations will have to decide when and where their need for automation and the benefits it provides should take precedence over human oversight and explainability. Regulatory frameworks remain dynamic in response to AI decision-making in insurance contexts. Industry practitioners need guidance on responsible deployment that maximizes benefits while managing risks [2].

This article investigates how generative AI systems systematically process claims evidence while maintaining accuracy and fairness. The discussion examines technical architectures employed in evidence interpretation. It explores fraud detection methodologies and quantifies operational impacts. The content addresses technical challenges, including hallucinations, domain generalizability, and adversarial attacks. Governance requirements, including explainability, regulatory compliance, and privacy protection, receive detailed attention. The technology promises substantial operational benefits but requires careful implementation to address inherent constraints and regulatory requirements. Insurance organizations need frameworks for evaluating when and how to deploy generative AI in claims operations [1].

II. Generative AI Architectures for Claims Evidence Analysis

Contemporary generative AI systems employ multiple technical architectures working in concert. Large language models process unstructured narrative evidence from claims descriptions, witness statements, and medical records. These models extract key facts and identify inconsistencies. They assess claim coherence against policy terms through natural language understanding. Modern language models can comprehend complex insurance terminology and policy language. They identify relevant information from lengthy documents that might span hundreds of pages. The systems recognize entities such as locations, dates, medical conditions, and parties involved in incidents [3].

The models can compare claim narratives against supporting documentation automatically. This includes medical records, police reports, and prior statements. The systems identify logical inconsistencies that may indicate fraudulent embellishment or deliberate deception. Natural language processing techniques specifically designed for contradiction detection compare statements across multiple documents. The models identify temporal inconsistencies where timelines conflict between different evidence sources. They detect contradictions in causal relationships described in claim narratives versus supporting documentation. Semantic similarity measures identify when different terms describe inconsistent events or conditions [3].

Language models also extract structured information from unstructured text. They populate claim data fields automatically from narrative descriptions. This capability reduces manual data entry requirements while improving data quality. The models identify key claim characteristics, including incident type, severity indicators, and relevant policy provisions. They flag claims requiring specialized expertise or additional investigation based on content analysis. This triage capability ensures complex claims receive appropriate routing while straightforward claims move through expedited processing [3].

Vision transformer architectures analyze claims, evidence imagery. This includes property damage photographs, accident scene documentation, and medical imaging. The systems detect image manipulation and assess damage severity. They identify evidence replication patterns suggesting fraud. Vision transformers can recognize when the same damage appears in multiple unrelated claims. This capability helps detect organized fraud rings using recycled evidence. The models compare images across large claim databases to identify duplicate or highly similar photographs. They detect common image manipulation techniques, including splicing, copy-move forgery, and digital enhancement [4].

Vision systems also estimate damage severity and repair costs from photographs. They identify vehicle damage patterns consistent with the described accident mechanisms. The models recognize inconsistencies between claimed damage and photographic evidence. For property claims, vision transformers assess structural damage, water damage extent, and fire damage patterns. They compare damage characteristics against typical patterns for claimed incident types. This capability

helps identify staged accidents or exaggerated damage claims. The systems can also verify the authenticity of submitted imagery by detecting signs of digital manipulation or stock photography usage [4].

Multimodal transformer architectures integrate textual and visual information simultaneously. These models correlate written descriptions against photographic evidence. They flag discrepancies that might indicate falsified or misleading claims. For example, a claim narrative describing severe vehicle damage can be verified against accident scene photographs. The models check whether visible damage aligns with the described impact points and collision dynamics. They identify cases where textual descriptions significantly exaggerate photographic evidence [4].

Multimodal systems excel at detecting contradictions between written descriptions and visual evidence. This integration provides more comprehensive fraud detection than text-only or image-only systems. The models can process medical claim narratives alongside diagnostic imaging to verify consistency. They compare property damage descriptions against multiple photographs to assess overall claim validity. Multimodal architectures also correlate police report descriptions with accident scene imagery. This comprehensive evidence correlation identifies subtle inconsistencies that single-modality systems might miss. The technology enables simultaneous processing of all available evidence types for holistic claim assessment [4].

AI Architecture Type	Processing Capabilities	Application in Claims
Large Language Models	Extract key facts and identify inconsistencies from narrative evidence	Process claims descriptions, witness statements, and medical records
Vision Transformers	Detect image manipulation and assess damage severity	Analyze property damage photos and accident scene documentation
Multimodal Transformers	Integrate textual and visual information simultaneously	Correlate written descriptions against photographic evidence
NLP Systems	Compare statements across multiple documents for contradictions	Identify temporal inconsistencies and causal relationship conflicts
Entity Recognition	Identify locations, dates, medical conditions, and involved parties	Populate claim data fields automatically from unstructured text

Table 1: Generative AI Architectures for Claims Evidence Processing [3, 4]

III. Fraud Detection Methodologies and Operational Impact

Generative AI fraud detection employs multiple complementary techniques. Supervised machine learning models that learn from historical claims data. Within these models, non-fraudulent claims are distinguishable from fraudulent claims. Prediction accuracy improves over time as more verified instances of fraud are integrated into training datasets. These models may predict fraudulent behaviors that human reviewers miss or take more time to identify as harmful. Supervised models learn complex feature interactions that indicate fraud risk. They recognize combinations of claim characteristics that correlate with fraudulent intent. The systems adapt to emerging fraud tactics as training data reflects new fraud schemes [5].

Supervised models employ various machine learning algorithms, including gradient boosting, random forests, and neural networks. These algorithms process hundreds of claim features simultaneously. Features include claim amount, incident details, claimant history, provider networks, and temporal patterns. The models generate fraud probability scores indicating the likelihood that specific claims warrant investigative attention. Score thresholds determine which claims receive automated approval, manual review, or detailed investigation. Organizations continuously refine these thresholds based on fraud detection performance and operational efficiency metrics [5].

Unsupervised anomaly detection systems use generative adversarial networks and variational autoencoders. These systems identify claims exhibiting statistically unusual patterns. They work even without explicit fraud labels. The models compare new claims against normal claim distributions. Claims that deviate significantly receive higher fraud

risk scores. Unsupervised methods prove particularly valuable for detecting novel fraud schemes not represented in historical training data. They identify outliers based on statistical properties rather than learned fraud patterns [6].

Generative adversarial networks learn normal claim distributions through adversarial training processes. The generator network creates synthetic claims resembling legitimate claims, while the discriminator network distinguishes real claims from synthetic ones. This process produces models that understand normal claim characteristics deeply. New claims substantially different from learned distributions trigger anomaly alerts. Variational autoencoders compress claim information into latent representations and reconstruct claims from these representations. Large reconstruction errors indicate anomalous claims that differ significantly from typical patterns [6].

Behavioral pattern analysis examines individual claimant histories and networks of related claimants. The systems identify suspicious circumstances such as repeated high-value claims. They detect coordination with specific service providers or participation in organized fraud rings. Network analysis techniques map relationships between claimants, service providers, witnesses, and legal representatives. The models identify clusters of related parties exhibiting coordinated claim patterns. They detect when multiple claimants use identical witnesses, frequent the same medical providers, or employ the same legal representation. These network patterns often indicate organized fraud operations [6].

Temporal analysis examines claim timing patterns. The systems identify claimants filing claims immediately before or after policy changes. They detect patterns of claims filed shortly after policy inception or just before policy expiration. Geographic analysis identifies concentrations of claims in specific areas that may indicate local fraud rings. The models also examine claim complexity progression, where claimants file increasingly sophisticated or higher-value claims over time. This progression sometimes indicates learning behavior as fraudsters refine their tactics [6].

Quantified impact assessments demonstrate substantial operational benefits. GenAI-enabled claims processing reduces document review time significantly. This enables faster claims resolution while improving accuracy through systematic analysis. Traditional manual review required adjusters to read through extensive documentation, cross-reference multiple sources, and manually identify inconsistencies. Generative AI systems perform these tasks in minutes rather than hours. The technology allows adjusters to focus on judgment and decision-making rather than information gathering and basic analysis [5].

Fraud detection accuracy improvements occur when generative models augment traditional rule-based indicators. The models show particularly strong performance in identifying emerging fraud patterns not present in historical training data. Traditional rule-based systems required manual updates as new fraud tactics emerged. Generative models adapt more dynamically to evolving fraud landscapes. They identify subtle pattern variations that rigid rules cannot detect. The technology proves especially effective for complex fraud schemes involving multiple coordinated parties or sophisticated deception tactics [6].

False positive reduction through refined scoring algorithms ensures high-risk claims receive appropriate attention. This minimizes disruption to legitimate claimants. High false positive rates in traditional fraud detection systems created friction for honest claimants. Legitimate claims flagged incorrectly experienced processing delays and additional scrutiny. Generative AI systems with more sophisticated pattern recognition reduce these unnecessary interventions. The technology distinguishes more accurately between unusual but legitimate claims and genuinely fraudulent submissions. This discrimination improves customer satisfaction while maintaining fraud detection effectiveness [5].

A critical innovation addresses data scarcity challenges in fraud model development. Insurance organizations often possess imbalanced datasets. Confirmed fraud cases represent only a small fraction of total claims volume. This creates fundamental statistical challenges for supervised machine learning. Traditional algorithms struggle with severely imbalanced class distributions. They tend to optimize for the majority class accuracy at the expense of minority class detection. This bias undermines fraud detection effectiveness where the minority class represents the critical target [6].

Generative approaches synthesize realistic fraudulent claim examples from confirmed fraud patterns. The synthetic data maintains statistical relationships with legitimate claims. This enables more balanced training datasets that improve model generalizability. Synthetic data generation uses confirmed fraud cases as templates. The models introduce controlled variations that maintain fraud characteristics while creating diverse examples. This augmentation prevents overfitting to limited real fraud cases. The synthetic claims provide sufficient training examples for algorithms to learn robust fraud detection patterns [6].

Organizations implementing synthetic data augmentation achieved substantial fraud detection precision improvements. False positive flagging rates decreased significantly. The balanced training data enables models to learn fraud indicators more effectively without sacrificing legitimate claim processing. Synthetic data generation also addresses privacy concerns in model training. Real fraud cases often contain sensitive personal information. Synthetic data removes direct connections to real individuals while preserving statistical properties necessary for effective model training. This approach enables broader model development and testing without compromising data privacy [5].

Detection Method	Operational Mechanism	Primary Use Case
Supervised Learning Models	Learn patterns from historical verified fraud cases	Distinguish legitimate claims from fraudulent submissions
Generative Adversarial Networks	Learn normal claim distributions through adversarial training	Identify claims with statistically unusual patterns
Variational Autoencoders	Compress and reconstruct claim information	Detect anomalous claims through reconstruction errors
Behavioral Pattern Analysis	Examine claimant histories and related networks	Identify coordination with service providers and fraud rings
Synthetic Data Generation	Create realistic fraudulent claim examples from patterns	Address imbalanced datasets and improve model training

Table 2: Fraud Detection Methodologies and Techniques [5, 6]

IV. Technical Challenges and Limitations

Despite substantial promise, generative AI fraud detection faces significant constraints. Hallucinations represent critical risks in claims processing. Large language models sometimes generate plausible but factually incorrect information. These fabrications can include invented policy provisions, nonexistent medical conditions, or false evidence summaries. False conclusions can result in wrongful claim denials or fraudulent claim approvals. This risk requires human oversight and verification of AI-generated conclusions [7].

Hallucinations occur because language models predict plausible text continuations rather than retrieving verified facts. The models lack inherent mechanisms to distinguish between actual knowledge and likely-sounding fabrications. Insurance applications require factual accuracy where hallucinations create unacceptable risks. Organizations must implement validation mechanisms that verify AI-generated information against source documents. Human reviewers need training to recognize potential hallucinations and verify critical claims before acting on AI recommendations. System designs should present source evidence alongside AI-generated summaries to facilitate verification [7].

Fine-tuned language models often suffer from reduced generalizability. Performance degrades when applied to claims outside their specific training domain. Models developed mostly on auto insurance claims may be ineffective in processing homeowners' claims. Models with training predominately from one geographic area may struggle with claims from a region with different regulations, terms, and fraud trends. This constraint requires continuous model updating and domain expansion. Organizations must maintain multiple specialized models for different claim types. They need processes for identifying when claims fall outside model training domains [8].

Domain shift represents a fundamental challenge in machine learning deployment. Models learn statistical patterns specific to their training data. When deployment data differs from training data, performance deteriorates. Insurance claims exhibit substantial variation across product lines, geographies, and time periods. Models must either train on sufficiently diverse data or recognize when domain-specific expertise applies. Organizations need frameworks for evaluating model confidence and detecting out-of-domain inputs that require human expertise [8].

Another emerging area of concern is adversarial attacks. Fraudsters intentionally create claims for the sole purpose of eluding AI detection. Fraud tactics evolve alongside detection capabilities. This creates an ongoing arms race between

fraud detection systems and fraudulent actors. Sophisticated fraudsters may probe detection systems to identify decision boundaries. They craft claims that fall just below fraud score thresholds or exploit known model weaknesses. The systems require continuous updating to address new evasion techniques [7].

Adversarial robustness testing evaluates model vulnerability to intentional manipulation. Red team exercises, where internal teams attempt to craft undetectable fraudulent claims, identify system weaknesses. Organizations must monitor for systematic evasion patterns that indicate adversarial exploitation. Model updates should address identified vulnerabilities while maintaining detection of traditional fraud patterns. The adversarial challenge requires ongoing investment in model improvement and fraud tactic monitoring [7].

Bias risks persist where generative AI systems perpetuate historical discrimination. Models can amplify biases against protected demographic groups. Historical claims data may reflect past discrimination in claims handling. Models trained on this data learn discriminatory patterns as legitimate decision factors. This creates fairness and legal compliance risks. Biased models may flag claims from certain demographic groups at higher rates regardless of actual fraud risk. Such discrimination violates fair lending laws and insurance regulations prohibiting unfair discrimination [8].

Bias mitigation requires proactive testing and correction. Organizations must evaluate model performance across demographic groups defined by protected characteristics. Statistical parity metrics assess whether fraud detection rates differ significantly across groups. Equalizing false positive rates across groups ensures that legitimate claimants face similar scrutiny regardless of demographics. Bias correction techniques include reweighting training data, adjusting decision thresholds by group, or constraining model optimization to maintain fairness criteria. Ongoing monitoring ensures that bias does not emerge as models update or deployment patterns change [8].

Challenge Type	Description	Mitigation Approach
Hallucinations	Models generate plausible but factually incorrect information	Implement validation mechanisms and human oversight
Reduced Generalizability	Performance degrades on claims outside training domain	Maintain multiple specialized models for different claim types
Adversarial Attacks	Fraudsters craft claims to evade AI detection	Conduct red team exercises and continuous model updating
Bias Risks	Systems perpetuate historical discrimination patterns	Perform bias testing and equalize false positive rates across groups
Domain Shift	Models struggle with varied product lines and geographies	Evaluate model confidence and detect out-of-domain inputs

Table 3: Technical Challenges in Generative AI Fraud Detection [7, 8]

V. Explainability, Governance, and Regulatory Compliance

Explainability and interpretability remain critical requirements. Modern insurance fraud management requires explainable AI approaches. Systems must provide transparent reasoning for fraud flagging decisions. They must cite specific evidence and identify risk factors. Black box models that generate fraud scores without explanation prove insufficient for insurance applications. Adjusters need to understand why specific claims received high fraud scores. Investigators require specific fraud indicators to guide their inquiries. Claimants have the right to understand why their claims received additional scrutiny [9].

Effective implementations employ attention mechanism visualization. These visualizations show which evidence elements most influenced fraud scoring. Attention weights indicate which portions of claim narratives, which photographs, or which historical patterns contributed most significantly to fraud assessments. Heat maps overlay attention weights on input documents, highlighting suspicious elements. This guidance helps human reviewers quickly

identify concerning evidence for further evaluation. Attention mechanisms provide model-intrinsic explanations based on actual computational processes [9].

Example-based explanations compare flagged claims to similar historical fraud cases. These explanations leverage case-based reasoning familiar to insurance professionals. The system retrieves confirmed fraud cases exhibiting similar characteristics to the flagged claim. Side-by-side comparisons illustrate why the current claim resembles known fraud patterns. This explanation style helps adjusters recognize fraud indicators they might have missed. It also builds adjuster trust in AI recommendations by grounding them in concrete precedents. Example-based explanations work particularly well for complex fraud schemes where multiple interacting factors indicate fraud [9].

Feature importance analysis identifies which claim characteristics contributed most substantially to fraud assessment. Global feature importance shows which factors generally drive fraud scores across all claims. Local feature importance explains why specific factors mattered for individual claims. Feature importance helps organizations understand model decision logic and identify potential problems. If protected demographic characteristics show high importance, this indicates problematic bias requiring correction. Feature importance also guides fraud investigation by highlighting which aspects of claims deserve detailed scrutiny [9].

Human-in-the-loop frameworks maintain final decision authority with qualified adjusters. GenAI systems serve as evidence synthesis and pattern recognition tools. They do not function as autonomous fraud determination systems. This ensures human judgment remains central to fraud decisions while leveraging AI capabilities for efficiency. The human-in-the-loop design addresses both technical and governance concerns. Humans provide common sense reasoning that models lack. They override incorrect AI recommendations and recognize novel situations requiring specialized expertise [10].

Effective human-AI collaboration requires appropriate interface design. Systems should present AI recommendations clearly while preserving adjuster agency. Displays should show confidence levels so adjusters know when to exercise additional skepticism. Explanations must be brief but provide sufficient information to facilitate an informed decision. Workflow integration should enhance rather than replace adjuster expertise. Training programs must prepare adjusters to work effectively with AI tools, understanding both capabilities and constraints [10].

Insurance regulators increasingly require transparency and auditability in automated claims decisions. GenAI implementations must maintain comprehensive decision documentation. This includes evidence inputs, model processing steps, fraud indicators identified, and confidence scores. Documentation supports human reviewer determinations and regulatory audits. Regulators need to verify that automated systems comply with insurance laws and regulations. Comprehensive audit trails enable after-the-fact review of decisions. This proves essential when claimants dispute claim denials or handling procedures [10].

Regulatory compliance requires demonstrating that AI systems do not discriminate in claims handling. Organizations must conduct bias testing across demographic groups. Regular model performance audits ensure fairness and accuracy. Testing protocols should examine fraud detection rates, false positive rates, claim denial rates, and processing times across protected groups. Statistical tests determine whether observed differences exceed acceptable variation. When bias appears, organizations must implement corrections before continued deployment. Regulatory examinations increasingly include AI system reviews where regulators assess model governance, testing procedures, and fairness metrics [10].

Privacy considerations require appropriate data protection. Generative AI systems process sensitive policyholder information. They will have to follow privacy requirements related to laws and regulations, such as GDPR and state insurance privacy laws. Safeguarding personal data means encryption, access control, and data retention. Principles of data minimization constrain informational collection and retention to that which is required. Organizations must document data flows showing how personal information moves through AI systems. Privacy impact assessments identify risks before system deployment. Breach notification procedures address potential security incidents [10].

Model training creates particular privacy challenges. Training data often includes detailed personal and medical information. Organizations must prevent model memorization of specific training examples. Differential privacy techniques can be applied to the training process by injecting noise before data is compiled, making the models unable to identify whether individual training examples were used or not. Another technique, federated learning, allows AI models to be trained on distributed data without the migration of sensitive data. In addition, synthetic data is usable in place of

actual training data for applications where it is required or useful, and involves lower privacy risk. These privacy-preserving techniques enable effective model development while protecting sensitive information [10].

By engaging with both transformative capabilities and undying constraints, the use of these models ethically, socially, and responsibly becomes more visible. These rapidly changing ways of accelerating and improving insurance can contribute to better outcomes, all the while protecting fairness, transparency as well and regulatory compliance. Delivery of success will depend on the balance of innovating responsibly and managing the risk. To do this, the organization must have a clear governance framework to assess to determine acceptable use-case scenarios, levels of human oversight, and when to start and stop pre-approved testing. The executive level understands the benefits of using the models versus the risks, and they need to be connected as they weigh the best decision for future deployment. Technical teams need resources and authority to implement systems properly. Claims professionals require training to work effectively with AI tools while maintaining professional judgment and ethical standards [10].

Requirement Type	Implementation Method	Stakeholder Benefit
Attention Visualization	Display evidence elements influencing fraud scoring	Help reviewers identify concerning evidence quickly
Example-Based Explanations	Compare flagged claims to historical fraud cases	Build adjuster trust through concrete precedents
Feature Importance Analysis	Identify claim characteristics driving assessments	Guide fraud investigation and detect bias
Human-in-the-Loop Frameworks	Maintain adjuster authority with AI as support tool	Ensure human judgment remains central to decisions
Comprehensive Documentation	Record evidence inputs, processing steps, and confidence scores	Support regulatory audits and claimant dispute resolution

Table 4: Explainability and Governance Requirements [9, 10]

Conclusion

Generative artificial intelligence represents a transformative technology for insurance claims, evidence interpretation, and fraud detection. Large language models, vision transformers, and multimodal architectures enable comprehensive analysis of complex claims evidence that historically required extensive manual review. These systems extract key information from unstructured narratives, analyze photographic evidence for manipulation or inconsistencies, and correlate information across multiple evidence sources. Fraud detection capabilities employ supervised learning from historical fraud cases, unsupervised anomaly detection identifying unusual patterns, and behavioral analysis recognizing coordinated fraud schemes. The technology delivers substantial operational benefits by reducing claims processing time while improving fraud detection accuracy. Synthetic data generation addresses fundamental challenges in fraud model training caused by severely imbalanced datasets where confirmed fraud cases remain rare. However, significant technical challenges persist, including model hallucinations that generate factually incorrect information, reduced generalizability when models encounter claims outside their training domains, adversarial manipulation by sophisticated fraudsters, and bias risks that perpetuate historical discrimination. Considerations for explainability compel transparent reasoning through visualizing attention, explanations based on examples provided, and using feature importance analysis. Human-in-the-loop frameworks permit humans (certified adjusters) to maintain the final authority in decision-making, while using the power of the AI models for speed and efficiency. For compliance considerations with an insurance regulator, extensive documentation of all decisions made, testing for bias in data by demographic categories, and privacy measures related to sensitive information previously disclosed to policyholders must be adhered to. Successful implementation demands balancing innovation with conscious governance, recognition that generative A.I. is a powerful tool to augment human expertise and is not a replacement for professional judgment. Organizations must develop a clear governance framework, invest in rigorous testing, and maintain ongoing monitoring for fairness and accuracy. This technology has

the potential to fundamentally reshape insurance claims operations, and by embracing generative A.I., insurance providers will realize the stated promises and gain a competitive advantage, all while carefully balancing capabilities and constraints to thoughtfully deploy A.I. systemically.

References

1. Sanjay Menon, et al., "Transform Insurance Claims with Generative AI," ValueLabs, 2025. [Online]. Available: https://www.valuelabs.com/resources/blog/ai-ml/transforming-insurance-claims-with-generative-ai/
2. Emanuele Costa and Nadine Moore, "GenAI Will Write the Future of Insurance Claims," Boston Consulting Group, 2025. [Online]. Available: https://www.bcg.com/publications/2023/the-future-of-insurance-claims
3. Petros Boulieris, et al., "Fraud detection with natural language processing," 2023. [Online]. Available: https://link.springer.com/article/10.1007/s10994-023-06354-5
4. Santhosh Raminedi, et al., "Multi-modal transformer architecture for medical image analysis and automated report generation," Scientific Reports, 2024. [Online]. Available: https://www.nature.com/articles/s41598-024-69981-5
5. Chris Raimondo and Vidhya Sekhar, "How insurers can leverage the power of generative AI," EY, 2023. [Online]. Available: https://www.ey.com/en_us/insights/insurance/how-insurers-can-leverage-the-power-of-generative-ai
6. Paul Kovalenko, "Insurance Fraud Detection Using Machine Learning," Langate, 2025. [Online]. Available: https://langate.com/news-and-blog/insurance-fraud-detection-using-machine-learning/
7. AutomationEdge, "The Role of Generative AI in Anomaly Detection for Claims and Fraud Analytics." [Online]. Available: https://automationedge.com/blogs/anomaly-detection-for-fraud-with-generative-ai/
8. E Igba, et al., "Synthetic Data Generation Using Generative AI to Combat Identity Fraud and Enhance Global Financial Cybersecurity Frameworks," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/389509666_Synthetic_Data_Generation_Using_Generative_AI_to_Combat_Identity_Fraud_and_Enhance_Global_Financial_Cybersecurity_Frameworks
9. Alphaeus Dmonte, et al., "Claim Verification in the Age of Large Language Models: A Survey," arXiv, 2024. [Online]. Available: https://arxiv.org/abs/2408.14317
10. Kieran Norton, et al., "How can tech leaders manage emerging generative AI risks today while keeping the future in mind?" Deloitte, 2025. [Online]. Available: https://www.deloitte.com/us/en/insights/topics/digital-transformation/four-emerging-categories-of-gen-ai-risks.html