ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

Fusing Generative AI and Agentic Workflows for Fraud Mitigation in InsurTech

Carlos Mendoza
Asst Professor

Abstract

In a hybrid generative and agentic artificial intelligence architecture, generative AI generates fraud-detection signals based on historical events while agentic AI provides decision-suggestion commands to the human stakeholder in-the-loop. This synergistic approach overcomes the limitations of pure generative and agentic frameworks in detecting fraud patterns in modern InsurTech platforms. Rapidly growing digitalization and increased dependence on digital channels increase the susceptibility of insurance platforms to behavioral fraud. Advanced technologies such as AI and Machine Learning (ML) have the potential to mitigate the risk of billions of dollars lost through these frauds. To harness the power of both generative and agentic AIs, a novel hybrid system is proposed. Key stakeholders of the system include the InsurTech venture for which the system is being built, customers of the insurance platform, the development team, and the operational team responsible for maintaining and managing the AI pipelines.

Data aspects of the architecture encompass novel feature-engineering techniques, a sophisticated data-pipeline approach, and an incremental learning mechanism to prevent performance deterioration due to concept drift. An unsupervised-ML-based segmentation model has been instantiated for the generation of training datasets and enables scalable continuous-signal generation in a real-time production environment. The quality of signals in the fraud-detection domain is not always suitable for direct operationalization. Hence, the operational design leverages the hybrid nature of the architecture by enabling human evaluation of the suggestion signals in a real-time mode. The primary fraud-detection application of the system is claims related, although it has the capability to generate fraud signals for policy underwriting, channel fraud, and behavioral fraud use cases as well.

Keywords : Generative Artificial Intelligence,Agentic AI Systems,Fraud Detection,InsurTech,Machine Learning,Deep Learning,Explainable AI (XAI),Anomaly Detection,Predictive Analytics,Natural Language Processing (NLP),Multi-Agent Systems,Risk Assessment,Synthetic Data Generation,Real-Time Fraud Detection,Intelligent Decision Systems.

1. Introduction

Advancing AI-enabled intelligence in Fraud Detection is vital for InsurTech companies and a challenging AI engineering task, especially the detection of insurance fraud. Despite its popularity and cost efficiency, probability-based AI has natural weaknesses making it incapable of tackling any problem. Fraud detection, a core operation for insurance companies using AI, requires the ability to identify unseen patterns and new classes in highly imbalanced data, and relate various concepts, none being in the historical data, instantiated with heterogeneous signals that need to be interpreted and understood for right decisions. Generative AI, a recently popular area of AI, can solve these issues normally using natural textual descriptions or large datasets as prompts. As it is less explored in fraud detection, Generative AI is used to design a probabilistic Feature Engineering Pipeline and Formulation. Deception and fraud detection studies reveal that behavioral and contextual factors play an important role in fraud detection. Therefore, besides generative AI a second, agentic AI using proxy or surrogate models with metaphorical story-

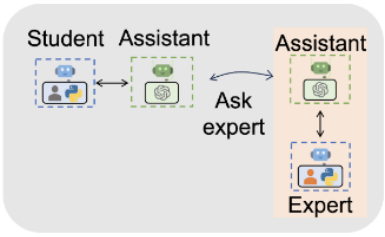
ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

tellers reasoning and decision-making using behavioral signals provide engines for explaining unseen patterns as (non-)probabilistic agents. However, Agentic AI faces its own challenges. With the right design, Generative and Agentic AI can complement each other. A hybrid framework using both Generative and Agentic AI is proposed to provide an industry standard – Intelligent Fraud Detection – for fast, accurate, and cost-efficient fraud detection on Modern InsurTech Platforms.

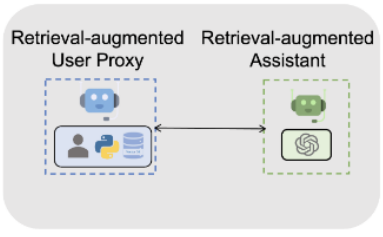
1.1. Background and Significance

Detection of fraudulent behavior is a central concern for the InsurTech industry and one of the most pressing use-case for leveraging modern machine learning algorithms. Open-ended data availability on policyholders and their individual interactions with the insurance provider presents multiple opportunities where detection of abnormal or deviant behavior could either mitigate Insurance Fraud in real-time (Claims Fraud) or improve dynamic risk assessment (Underwriting Fraud). Modern companies have started capturing behavioral patterns on interactions with their websites and mobile applications to identify among good and deviant customers (Channel Fraud). Data being the new Oil, an organization might not have a shortage of data but building relevant features that can identify these frauds in real time or batch analytics are essential to put these future derived models into production.

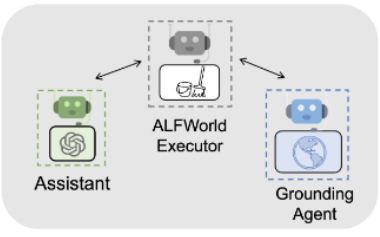
Fraud Signal refers to the congregation detection signal for Fraud related events that will be communicated to Counter Fraud functions on Fraud Crime types like Claims Fraud, Underwriting Fraud, Policy Fraud, Channel Fraud, Behavioral Fraud etc. This would serve as threshold levels for further investigation and follow-up action. The Fraud Signals would play a pivotal role in future model deployment and validation for Fraud detection related Analytics. To make the Fraud Business Use-Case more reliable and accurate, Fraud Signals will undergo periodic Dashboard Validation based on a historical view with a specific time horizon.



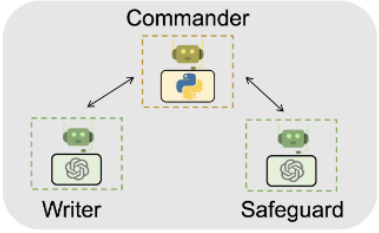
A1. Math Problem Solving



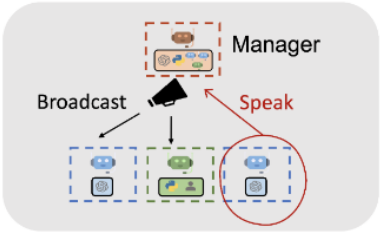
A2. Retrieval-augmented Chat



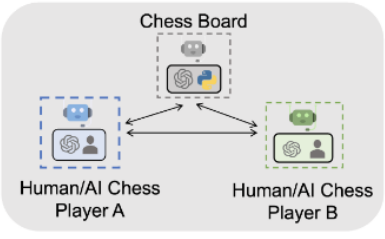
A3. Decision Making



A4. Multi-agent Coding



A5. Dynamic Group Chat



A6. Conversational Chess

Fig 1: Agentic Frameworks for Generative AI

1.2. Research design

A contemporary intelligent fraud detection framework designed with a typical InsurTech architecture is addressed though each of its components. The construction is frequently underpinned by practical use-cases and supported by clear evidence. Ultimately, it employs a hybrid Generative-AI and Agentic-AI approach that encompasses key phases of the InsurTech business process, exploiting different facets of the pol-system data landscape. Fraud detection represents one of the most important focal points as InsTech companies process a huge amount of data in real-time. Detection of fraud is very challenging because fraudsters behave like others, but often with different patterns.

Detecting bias-based, intent-based, or decision-based fraud is therefore crucial. Passengers can manipulate the underwriting. They may also abuse other policies. In a travel-pol-world, travel-pol-fraud-signals include fraud beyond compartments. Hybrid signal-detection in policy and claim phase, for agentic, non-agentic channels; decision-and-eternity gap; multi-modal signals; agent-intent. Model-driven governs online-learning approach to detecting 'concept-drifts' & 'pattern-shifts'. All fraud detections observe 'Plato's fairness' as-sharp-a-bias as a-predict-or-decision.

Equation 1: Multimodal feature fusion

Step 1: Encode each modality

Let:

$$h_i^{(s)} = f_s\left(x_i^{(s)}\right)$$
$$h_i^{(t)} = f_t\left(x_i^{(t)}\right)$$
$$h_i^{(g)} = f_g\left(x_i^{(g)}\right)$$

where:

- $f_s(\cdot)$= encoder for structured fraud features,
- $f_t(\cdot)$= text encoder / LLM embedding,
- $f_g(\cdot)$= graph or relation encoder.

Step 2: Concatenate all encoded features

$$h_i = \begin{bmatrix} h_i^{(s)} \\ h_i^{(t)} \\ h_i^{(g)} \end{bmatrix}$$

Step 3: Project to a shared fraud representation

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

Apply a linear map and nonlinearity:

$$z_i = \phi(Wh_i + b)$$

This is the first complete equation:

$$\boxed{z_i = \phi\left(W\begin{bmatrix} f_s(x_i^{(s)}) \\ f_t(x_i^{(t)}) \\ f_g(x_i^{(g)}) \end{bmatrix} + b\right)}$$

2. Theoretical Foundations

The proposed solution supports intelligent fraud detection in modern InsurTech platforms. A hybrid framework, based on generative and agentic AI, transforms transaction data into relevant signal features for the fraud detection domain. A redefined problem space categorizes fraud signals into claims, underwriting, behavioral, and channel fraud detection, enabling use-case specification, data curation, model training, and threat mitigation. The current manuscript focuses on theory and architecture.

Modern InsurTech platforms manage a variety of transactions among numerous participants and offer high data velocity and volume. Risk is present in all transactions. Fraud in insurance represents the combination of three types of risk: instances of bad faith committed by the insured party during an engagement with the insurer that violate the agreed terms while being undetected; instances of bad faith committed by the insurer during an engagement with the insured party that violate the agreed terms while being undetected; and instances of influence or pressure from non-participant third parties seeking to shift detected losses to another party. Despite this complexity, most corporates rely on rule-based systems to detect fraudulent activities or reduce fraud risk. Hybrid solutions based on generative and agentic AI can improve human fraud-prevention capabilities.

2.1. Generative AI in Fraud Detection

The application of Generative Pretrained Transformers (GPTs) and similar Large Language Models (LLMs) has surged in recent years across numerous domains and industries. Several research works advocate its suitability for decision-making in domains characterized by sparse, context-specific knowledge. In the insurance domain, a significant factor contributing to the success of organizations is their ability to efficiently manage loss-making or post-loss situations. In addition to fraud execution, a major driver for loss scenarios is the false negatives that slip past the fraud detection mechanisms. These facts suggest that knowledge-centric fraud detection would benefit from utilizing the knowledge-rich Generative AI technology. Nevertheless, the fraud detection domain has seen little progress in adopting these technologies.

Research on the Large Language Models highlighted their potential to revolutionize both the enterprise world and daily life through their unique operational strength: self-generated natural sentence structures capable of answering user queries. Yet, their impact has not extended beyond text generation and conversational agents. Within the Fraud Detection domain, work around ChatGPT shows that the knowledge-rich LLM technology can also effectively act as a knowledge worker, providing responses to complex questions rooted in particularized knowledge. The Fraud Signal domain thus emerges as a significant area of

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

implementation. Leveraging its ability to effectively and accurately answer queries rooted in defined knowledge, Generative AI can assist agents in conducting high-stakes conversations, thereby enhancing the enterprise's risk management capabilities and reducing fraud losses.

Aspect	Derived from article	Concise interpretation
Research problem	Insurance fraud on modern InsurTech platforms is difficult to detect because fraud is adaptive, data is imbalanced, signals are heterogeneous, and new fraud patterns emerge	Classical rule-based or purely predictive systems are not enough
Main aim	Propose an intelligent fraud detection framework combining generative AI and agentic AI	The paper's central contribution is a hybrid architecture
Why generative AI	Generates fraud-detection signals from historical, contextual, and multimodal data	Useful for pattern discovery and knowledge-rich signal formation
Why agentic AI	Provides decision suggestions, reasoning support, and interpretable action guidance	Useful for human-in-the-loop judgment and operational decisions
Why hybridize both	Each approach compensates for the other's weaknesses	Better balance of detection, explanation, and actionability

Table 1: Research Problem, Aim, and Core Contribution

2.2. Agentic AI and Decision-Making

Agentic AI pertains to automated systems that decide or act independently, directed by objectives but competent in strategy selection, akin to RL agents. For detection-oriented operations, agentic AI can autonomously devise a fraud intervention strategy for an involved scenario, albeit a simulated mode holding a non-deterministic layer may resonate better with stakeholders. Decision Trees (DTs) are considered a foundational representation in this realm. DT-generative PGMs create accessible and interpretable states, while GNN-empowered and deep models unveil interrelations in intricate scenarios. Alternative PGMs explore supervised safety concerns, yet agentic AIs necessitate endorsement from a clientele wary of actions lacking credible justification.

Agentic AI caters to situations where autonomous decision-making is requisite, melding Generative AI and Decision Support systems—independent models empowering independent users, albeit not directly substituting core operations. Recent Agentic AI deployments exhibit DTs for situational prudence, incorporated into risk-aware dialogue frameworks; DNNs predicting user behavior steering other agentic models. Frictionless interactions foster user acceptance. Within financial domains, explainable strategies attract users who forfeit agency amid opaque actions. Agentic AI is designed for contexts that demand autonomous decision-making, combining the creative capabilities of generative AI with the structured guidance of decision support systems. Rather than replacing core operations, these systems function as independent models that empower users to make informed

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

choices. Recent implementations highlight the integration of digital twins (DTs) to simulate and evaluate scenarios for improved situational awareness, alongside deep neural networks (DNNs) that anticipate user behavior and guide other agentic components. This layered intelligence enables more adaptive and risk-aware interactions. Crucially, seamless and frictionless user experiences play a central role in building trust and acceptance. In domains like finance, where decisions carry significant consequences, users are more inclined to engage with agentic systems that prioritize explainability and transparency, especially when opaque processes might otherwise diminish their sense of control.

Fig 2: Rise of Agentic AI in Financial Decision-Making

2.3. Synergies Between Generative and Agentic Approaches

Generative AI is typically used to simulate Properties-of-Interest to assist other downstream predictive models. Such Generative-AI enabled systems can be made able to reason about the potential actions of clients or fraudsters by supplementing Generative AI with a hybrid Agentic-AI, combining goal-based planning and reasoning about actions with behavioural time-series modelling. The resulting system can be used to model expected behaviours often ignored by downstream predictive friction-detection models tailored to fraud and fraud-related breaches such as cyber-risks, privacy risks, money-laundering behaviours, and non-ethical behaviours such as slavery and child-exploitation.

Generative AI assist in Fraud Detection in Online Behaviour with an AI-System Architecture and Simulation Workbench enabling a Hybrid Approach involving Development of Use-Cases where the Primary Detection-Focus is Generated and Enabling Generative Behavioural-Based Modelling using Trigger-based Agentic AI. Agentic AI Systems bring about improvisation during peak traffic and using Paradigmatic Agent Prototypes enable Abduction, more like the Broad Use-of-Response-Surfaces-and-Meta-Agents Approach in Serious Gaming AI. Failure-Condensed-Knowledge AI-Paladins facilitate Error-Proofing using Meta-Agents along-Development-Cycle.

Equation 2: Unsupervised segmentation for fraud-signal generation

Step 1: Assign each sample to the nearest cluster

For event i, define cluster assignment c_i:

$$c_i = \arg\min_{k \in \{1,\dots,K\}} \| z_i - \mu_k \|^2$$

Step 2: Define the clustering objective

The total within-cluster variance is:

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \| z_i - \mu_k \|^2$$

where

$$r_{ik} = \begin{cases} 1, & \text{if } c_i = k \\ 0, & \text{otherwise} \end{cases}$$

Step 3: Final segmentation equation

$$\boxed{J_{\text{seg}} = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \| z_i - \mu_k \|^2}$$

with

$$\boxed{c_i = \arg\min_{k} \| z_i - \mu_k \|^2}$$

3. Architectural Overview

The proposed fraud detection solution can be viewed as an intelligent agent acting on behalf of insurance organizations, preferably at the InsurTech end of the spectrum. A more detailed description of the context, constituents, data pipeline, feature engineering, models, and modeling components is provided here.

3.1. Context and Constituents

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

Fraud detection is just one of many possible application areas for data-driven online learning. The main constituents of the data pipeline at the intelligent-agent level can be viewed as the following five-dimensional Cartesian coordinate system:

Application context. This provides temporal, geographical, and domain-sensitive information related to the applicant, policyholder, policy, insurer, claim, channel, payment, and external data from sources or institutions providing background checks, driver's licenses, medical records, transactions, and social activities, among other information.

Fraud type. All possible fraudulent actions committed by all involved fraudsters, including claimants, policyholders, agents, brokers, and employees.

Intended audience. The different stakeholders of the model's predictions, ranging from claim handlers to the underwriting governance body, external fraud laboratories, and the policyholders themselves.

Model type. All possible models developed in response to chronic, epidemic, or emerging fraud types mapped onto the previously identified application context. In a longitudinal analysis, a subdivision into sufficient data for supervised learning and insufficient data for semi-supervised learning would usually be appropriate.

Function type. This represents the cyclical nature of online learning at work and is included primarily to underline the fact that every fraud type in every context for every intended audience should be monitored continuously and that an online-learning component should always be integrated into the modeling ecosystem.

3.2. Data Pipeline and Feature Engineering

The role of data is paramount to the robustness—and the eventual success—of the fraud system. Apart from obvious security, quality, and access aspects, three key dimensions require special consideration. First, temporal considerations are critical to the data-gathering frequency and latency allowed for real-time predictions. Second, factors such as ownership or channel, monitoring systems, and external data service applications, along with their elements and digital links, should be explored at the Schmitt Test level. Third, business-consumer privacy regulations govern the gathering and storage of data elements linked to embedded guarantees.

Initial asset monitoring should focus on automated ongoing monitoring of all available data sources, detection of emergent-use-drifting source links and components, signature-tracking model candidates, and qualification of the individuals to be warned in real time when such drifts are detected. Only a small fraction of the available data sources are actively scraped by the enterprise or controlled cost-efficiently. Data sources such as unstructured OTC channels, negative social behaviors and backgrounds, highly negative fingerprints of money laundering, KYC, KYB, AML, and Chek etc. checks are not cheap but they are fee-constrained real-time external data sources highly sensitive to fiduciary aspects.

3.1. System Context and Stakeholders

The hybrid architecture is implemented in a real-world context, responding to the requirements of a modern InsurTech platform engaged in health-related insurance activity. Country-specific laws, company regulations, and internationally acknowledged standards, such as ISO 27001 and PCI DSS, impose the need to handle sensitive, extra-special categories of personal data, such

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

as that associated with healthcare. Consequently, official personal data protection and processing guidance must be followed. In addition, the business relies on the open-source software practice, developing in five different product lines whose backends are connected to a shared technology ecosystem.

Domain experts and fraud analysts propose fraudulent behaviors, which are then translated by data scientists into supervised learning tasks. The proposed and implemented framework detects two types of fraud in insurance policies—namely underwriting fraud and policy fraud—and a wide range of fraud in claims, including both claim fraud and channel fraud. On the business side, the adaptive model management and monitoring framework aims to keep the model capabilities up to date through online learning, continuous validation, and automatic retraining.

Equation 3: Generative anomaly / fraud evidence score

Step 1: Model normal behavior distribution

Suppose the historical distribution within a segment $c_i$is Gaussian:

$$z_i \mid c_i = k \sim \mathcal{N}(\mu_k, \Sigma_k)$$

Then the density is:

$$p(z_i \mid c_i = k) = \frac{1}{(2\pi)^{d/2} \mid \Sigma_k \mid^{1/2}} \exp\left(-\frac{1}{2}(z_i - \mu_k)^T \Sigma_k^{-1}(z_i - \mu_k)\right)$$

Step 2: Convert likelihood into anomaly score

Rare events should have high anomaly score, so define:

$$A_i^{(\text{gen})} = -\log p(z_i \mid c_i)$$

Substitute the Gaussian density:

$$A_i^{(\text{gen})} = -\log\left[\frac{1}{(2\pi)^{d/2} \mid \Sigma_{c_i} \mid^{1/2}} \exp\left(-\frac{1}{2}(z_i - \mu_{c_i})^T \Sigma_{c_i}^{-1}(z_i - \mu_{c_i})\right)\right]$$

Step 3: Simplify

Using $-\log(ab) = -\log a - \log b$and $-\log e^{-u} = u$,

$$A_i^{(\text{gen})} = \frac{d}{2}\log(2\pi) + \frac{1}{2}\log \mid \Sigma_{c_i} \mid + \frac{1}{2}(z_i - \mu_{c_i})^T \Sigma_{c_i}^{-1}(z_i - \mu_{c_i})$$

So the complete generative fraud-evidence equation is:

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

$$A_i^{(\text{gen})} = \frac{d}{2}\log(2\pi) + \frac{1}{2}\log \mid \Sigma_{c_i} \mid + \frac{1}{2}(z_i - \mu_{c_i})^T \Sigma_{c_i}^{-1}(z_i - \mu_{c_i})$$

3.2. Data Pipeline and Feature Engineering

Continuous machine learning systems assume the need for ongoing training and adaptation. Yet efficiently managing data drift is often more important than the underlying feature extraction and learning processes. Outdated models that make poor predictions while remaining within operational tolerances still allow the perpetration of fraud, which is often planned at the level of months rather than days.

Training and inference phases rely on the same data pipeline but follow different paths. At inference time, the pipeline must operate with low latency—sub-second for claims fraud detection—while keeping complexity under control. Labelling input data for training poses another challenge that cannot be avoided in an online learning setting: adaptations rely on additional training batches, usually generated manually to a manageable, but require constant monitoring to ensure that incidents anticipated are indeed detected at the time of writing. Failure to do so may lead to significant reputational damage and loss of expertise by the organization if incident detection is no longer achieved.

Model performance can also be increased by augmenting product-specific visual data with social network information. Fraudsters are not working in silos, and successful visual inspection of insurance fraud in social media requires additional filtering based on social network attributes. Enhanced data sources can be temporally assigned as available. Data drift should trigger scrutiny of the detected fraud signals, and insufficient support at these lower levels might together explain a lack of detections at the higher levels of aggregation.

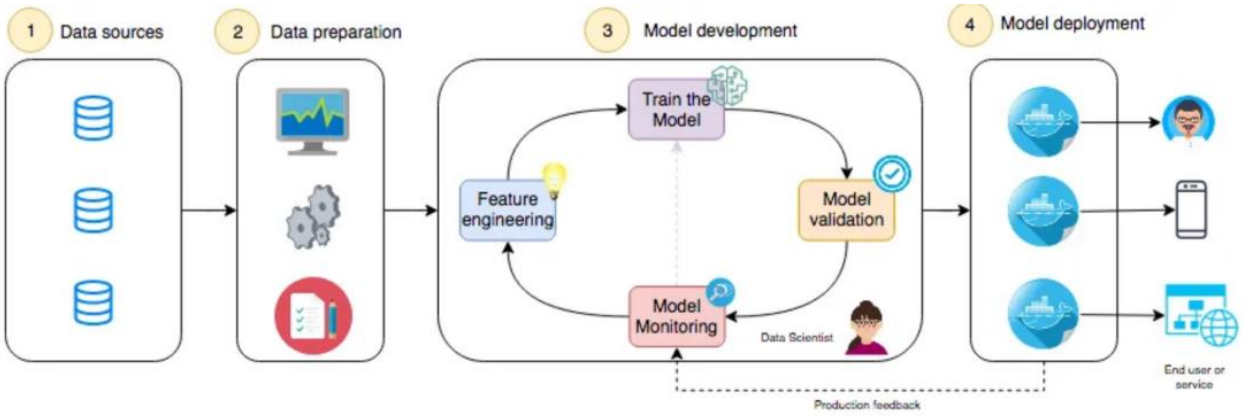


Fig 3: Data Pipeline and Feature Engineering

3.3. Model Architecture and Components

The AI model architecture comprises three primary components: an applications-bounded Large Language Model (LLM)

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

discretizing textual data, a multi-modal multi-domain Fraud Detection Model (FDM) learning from structured data, and a Knowledge Graph (KG) underlying datasets for both models in textual and tabular format.

The applications-bounded LLM dynamically expands its knowledge base by querying a Custom-Knowledge-Base (CKB). Configured as a multimodal querying and answering system, it generates answers with a description text (explanatory sentence), formal description (Structured Query Language), and resources in JSON format. An example use is generating Synthetic Data for an FDM input feature to capture the relationships hidden in the training data.

4. Fraud Signals and Use-Cases

Fraud detection is a challenging task due to the multitude of ways in which service users can abuse a system, both opportunistically and strategically. The agentic Generative AI-Hybrid Detection System presented here identifies various types of abuse signals in insurance platforms, such as claims fraud, underwriting and policy fraud, and channel and behavioral fraud. Beyond detection, two clear applications further demonstrate the versatility of the hybrid architecture.

Fraudulent claims pose a complex challenge to insurance providers, with the potential to exploit system weaknesses arising from the digitalization of business processes. Traditional fraud detection approaches employ machine learning classifiers to identify suspicious claims based on diverse data sources. These classifiers generate a probability score indicating the extent of potential fraud for each claim. A Generative AI-based solution follows a similar methodology. Recent developments, including foundation models leveraging generative pre-training and multimodal training on vast multi-domain datasets, have demonstrated strong performance across numerous vision and language benchmarks and tasks.

The advantage of the Generative AI-Hybrid Detection System is its ability to model abuse signals in a multi-layered approach, encoding feature similarities and differences in a semantically meaningful way. The model combines a decisioning layer with a Generative AI runtime application that automatically and dynamically resolves incoming decision points from the decisioning model. This model architecture is consequently applicable for claims fraud detection.

Architectural layer	**Main elements in the article**	**Function in the framework**	**Output**
Data source layer	Policy data, claims data, channel data, customer interactions, payments, external checks, social/activity signals	Collect raw enterprise and external inputs relevant to fraud	Multi-source raw data
Data engineering layer	Feature engineering, pipeline design, temporal handling, privacy-aware processing, drift monitoring	Clean, transform, and prepare fraud-relevant features	Structured and enriched features
Knowledge layer	Knowledge graph, custom knowledge base, textual and tabular datasets	Supports reasoning, query expansion, and contextual grounding	Contextualized domain knowledge

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

Architectural layer	Main elements in the article	Function in the framework	Output
Generative AI layer	Application-bounded LLM, synthetic data generation, textual reasoning	Creates fraud signals, explanatory text, query responses, and semantically enriched features	Fraud signals, descriptions, SQL/JSON-like outputs
Agentic AI layer	Decision support logic, interpretable decision models, surrogate/proxy reasoning, intervention strategies	Produces actionable recommendations for investigators or business users	Decision suggestions and intervention guidance

Table 2: Derived Architecture of the Proposed Hybrid Fraud Detection Framework

4.1. Claims Fraud Detection

Fraudulent claims are among the most serious threats to the insurance industry. The resulting monetary losses may endanger the providers' viability and harm honest customers who have to pay higher premiums. Staged theft or damage to one's own insured goods are common fraud patterns, as are false claims for medical expenses or injuries. In jurisdictions that treat injury mitigation as a legal duty, the absence of mitigation may itself be evidence of fraud. Other common signs of claims fraud are exaggerated claims, the timing of a loss or claim, or abrupt changes in a customer's claims pattern. Special enclaves of claims fraud, such as income-abuse schemes, may rely on complex scenarios and the active support of external accomplices. Such schemes may involve the submission of fictitious documents or the fabrication of the existence or loss of an insured asset. Other frauds can appear simple but may be committed only by insiders authorized to grant insurance cover for the risk assumed.

Prevention may require careful selection of intermediaries to create a "culture" of non-fraudulent business and adequate training and control of employees in charge of claims processing. Data-driven solutions assist risk managers in designing proper prevention strategies and may warn about unusual risks in a specific location or segment. Yet, a fraud-facilitating culture can be more difficult to change, and the main treatment is to detect fraud patterns on claims data. The patterns have a complicated nature that requires advanced machine-learning and natural-language-processing methods based on supervised models. Such models require labeled data for training and testing, but creating a labeled dataset is one of the hardest parts of developing a claims-fraud-detection solution.

4.2. Underwriting and Policy Fraud

Modern InsurTech platforms must be prepared for fraud not only in insurance claims but also in the underwriting and policy life cycles. Such fraud includes the creation of fictitious policies and clients, which damages the insurance channel, and the proposition of unsustainable insurance products with fraudulent intent, undermining clients' trust. To combat these frauds, models can be trained to produce a definitive answer when confronted with a suspected policy or client in policy life cycle operations or that is to be sold in an operations life cycle.

The high level of transparency and security in the blockchain, associated with the general movement of the insurance market, enables conversations of models that can reduce the detection of underwritting fraud and policy fraud in sales through Machine

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

Learning Operations with Model Governance. These models aim to enable real-time decision making or alerts to previous business principles of the pockets where flows are being directed. Based on the concept of technology purpose in the insurance channel, it all reduces risk and enables a focus on client satisfaction lawfully.

Fig 4: Trust AI Agents in Underwriting Decisions

4.3. Channel and Behavioral Fraud

Fraud throughout sales and distribution processes is often perpetrated by agents, intermediaries, business partners, or customers connected to these processes. A classic example is an insurance salesman using their own social security number to buy a life insurance policy. In these cases, fraud may occur in multiple dimensions: the insured may not be the true party at risk; a third party, usually a relative, may be intended to benefit but not disclosed; and the choice of the insurance coverage may not be aligned with the principal's real needs. Reasoning about relational fraud in insurance may leverage any combination of three essential types of fraud signals: anomalies relating to historical claims behavior of the sales and distribution processes; unnatural values associated with geographical or demographical locations for the salesforce; and unfeasible behaviors detected on the conversations of assigned agents.

Another important source of fraud consists of channel partners operating outside of the legal agreements establish with the insurance company. This may result in the use of official product documentation for lines of business that are not technically supported or in the offering of unauthorized benefits to the clients. Behavioral fraud signals aim to detect possible behaviors in the channel that depart from the company's ethical standards and channel rules. Salespersons may fabric fake profiles in the

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

social networks to place connections with customers not usually accessible to them, like high net worth customers. Examining the connections established by agents, such signals may capture situations where agents have motives for vindictiveness or malice.

Equation 4: Agentic decision utility function

Let the action space be:

$$\mathcal{A} = \{\text{approve,flag,investigate,reject,escalate}\}$$

Step 1: Define decision features

Suppose the decision depends on:

- fraud risk,
- explanation confidence,
- compliance cost,
- customer-friction penalty,
- expected financial loss avoided.

Represent them by $\psi_m(z_i, a)$, $m = 1, \dots, M$.

Step 2: Weighted utility

$$U(a \mid z_i) = \sum_{m=1}^{M} w_m \ \psi_m(z_i, a)$$

This says the agent scores every action by a weighted combination of objectives.

Step 3: Optimal agentic recommendation

Choose the action with maximum utility:

$$a_i^* = \arg\max_{a \in \mathcal{A}} U(a \mid z_i)$$

Hence the complete equation set is:

$$\boxed{U(a \mid z_i) = \sum_{m=1}^{M} w_m \ \psi_m(z_i, a)}$$

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

$$\boxed{a_i^* = \arg\max_{a \in \mathcal{A}} U(a \mid z_i)}$$

5. Training, Evaluation, and Adaptation

Training the various pipelines within a fraud detection framework requires substantial quantities of labeled data when supervised learning approaches are used. In the context of training these components, there are several issues to consider, including data curation in terms of volume and privacy, definition of suitable evaluation metrics, and a protocol for validating the models. An online learning approach can reduce the performance risk associated with a changing data distribution during operation.

Maintaining gratuitous privacy and security of sensitive data within the fraud detection pipeline is critical. Consequently, production-ready implementations of the pipelines use pseudonymized, racially-annotated attributes and curated information from public databases rather than production data when training fraud detection models. Pseudonymization involves replacing sensitive attributes with a token that can be used to query a separate base for reconstruction. All fraud detection models are implemented outside the core data processing and control plane of the firm. Inference is executed internally against the fraud detection models using de-identified attributes, and inbound junk is filtered into a separate records base at the border of the data privacy and processing domain.

5.1. Evaluation Metrics and Validation Protocols

Determining the overall effect of fraud detection efforts is inherently multifaceted; measuring detection performance of individual models is complicated by the need for active management to discourage fraudulent behavior. Nevertheless, several metrics can be employed to track success within the respective areas, such as:

- **Claims fraud detection**: Composite signal strength (potentially normalized by claim count), claim originator reliability score, crown jewel risk alert and untracked high-risk property coverage count.

- **Underwriting and policy fraud**: Composite signal strength, ratio of accepted flagged quotes/policies, policy originator reliability score, flagged accounts on policy reissue.

- **Channel and behavioral fraud**: Composite channel fraud score, e-commerce transaction anomaly count, anomalous bot detection count.

Real-time inference speeds and resource scaling for training and online learning can be monitored independently of fraud detection performance. The adaptive learning procedure allows continuous evaluation of prediction fidelity as new gold standard labels are added, enabling drift detection and consideration for retraining. Online evaluation should also include continual examination of transfer learning to ensure sensitivity to coverage change, not just catastrophic transition in the opposite direction.

5.2. Data Curation and Privacy Considerations

Well-configured online learning systems can greatly improve the model performance by incorporating new data and insights. However, privacy-related concerns regarding the training data need to be carefully considered. The utilization of personal data in

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

training machine-learning models poses risks such as sensitivity disclosure and attribute inference. Mitigating these risks usually involves de-identification of the data by removing identifying information. Nevertheless, current research shows that de-identification does not completely prevent disclosure and inference attacks. The training data should therefore be revisited to ensure proper safeguards are in place.

For the models to remain updated with the changing dynamics, data history is crucial. Automatic logging of events allows for wealth of past events being recorded in an inexpensive manner. Recording these events for online learning is paramount to model improvement while avoiding potential biases in the data and maintaining fairness. Once sufficient past data accumulated, segmentations can be performed by different attributes for detection and prediction. Origins of disparities in distribution (concept drift, feature drift, instance drift, semantic drift) have to be monitored and recorded, as impact on the model varies on case basis and its mitigation requires different approaches.

Maintaining up-to-date models in dynamic environments depends heavily on the availability and quality of historical data. Automatic event logging plays a critical role by enabling the continuous and cost-effective collection of large volumes of past interactions and outcomes. This accumulated data supports online learning, allowing models to adapt over time while also helping identify and mitigate potential biases, thereby promoting fairness. As sufficient data is gathered, it can be segmented across various attributes to enhance detection and predictive capabilities. However, it is equally important to monitor and record different forms of distributional shifts—such as concept drift, feature drift, instance drift, and semantic drift—since each type affects model performance differently. Understanding the origin and nature of these disparities enables the application of appropriate mitigation strategies, ensuring sustained model reliability and effectiveness.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

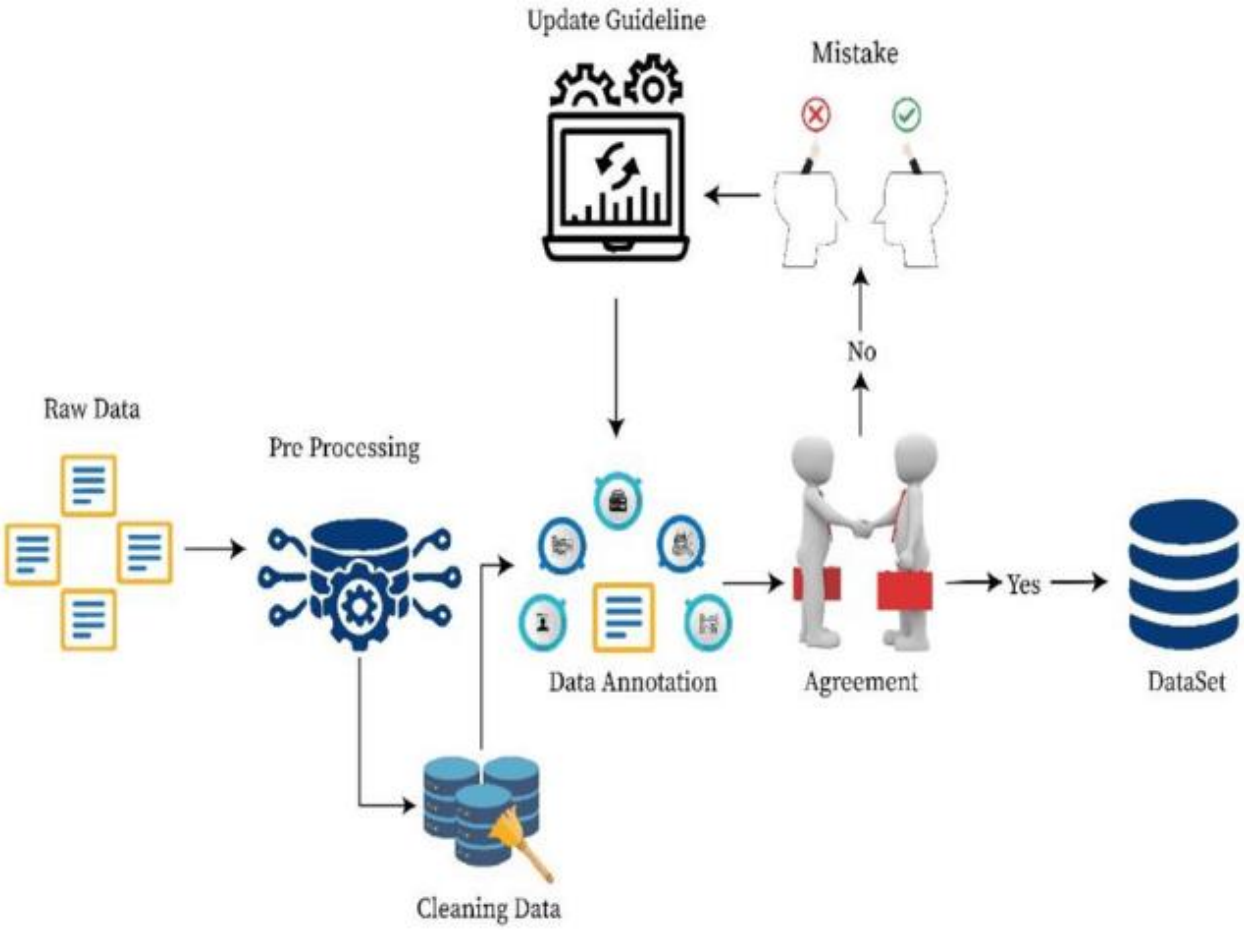


Fig 5: Guide to Data Curation in Machine Learning

5.3. Online Learning and Drift Mitigation

Detecting fraud, particularly in financial services, demands a high level of accuracy, and traditional machine learning techniques fall short of achieving this ideal. Consequently, many companies in production environments prefer not to utilize models that could classify fraud at both extremes. Instead, they create models that are optimized for capturing a greater share of the positive class at the expense of precision. Because the trade-off decisions are typically based on business volumes, a certain quality level is set for the fraud signal at the model development level, and business units accept the rest. Notably, this is not a criticism; the models have been built exactly to suit this kind of quality decision.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

With the introduction of online learning, however, the production setup becomes crucial for maintaining the performance of the models. In most financial services applications, data labeling is expensive and, in some cases, not even close to real time (e.g., labeling for approval of claims). Thus, the conventional online learning approach of retraining models frequently using the latest data is prone to fail. For example, in claims fraud detection, it may not make sense to retrain a claims fraud model every hour since the labeling of claims takes much longer than that. However, if the production latency for that model is not huge, inferences and the inferred signals related to claims of fraud and non-fraud offer valuable signals for online learning. The quality of these labels tend to be accepted in a larger window. Model lifecycle management encompasses monitoring, retraining, explainability, incident response, and business-continuity planning. Regular real-time alerts on increasing detection recurrence, use-case periodicity, or associated bank-grade FTA failures must be generated. Distributed fraud-control incident-response teams perform offline, knowledge-mining analysis to identify new use-case signals, suggesting dual-action mandates for the model development function. Management of online strategy includes active model-verification, -logging, and -testing protocols, in keeping with risk-driven principles. Available explainability frameworks for collaborative design and operations provide further thrice-validated assurance of post-load-balance identification directly incorporating spatiotemporal dynamics.

Equation 5: Final hybrid fraud signal

So we now combine:

- the generative anomaly evidence $A_i^{(\text{gen})}$,
- and the agentic decision utility $U(a_i^* \mid z_i)$.

Step 1: Normalize both components

Because the anomaly score and utility may have different scales, convert them to $[0, 1]$with a sigmoid:

$$\tilde{A}_i = \sigma(A_i^{(\text{gen})}) = \frac{1}{1 + e^{-A_i^{(\text{gen})}}}$$
$$\tilde{U}_i = \sigma(U(a_i^* \mid z_i)) = \frac{1}{1 + e^{-U(a_i^* | z_i)}}$$

Step 2: Weighted fusion

Let $\alpha \in [0,1]$be the blending coefficient. Then:

$$S_i = \alpha \tilde{A}_i + (1 - \alpha)\tilde{U}_i$$

Substituting the sigmoid terms gives:

$$\boxed{S_i = \alpha \frac{1}{1 + e^{-A_i^{(\text{gen})}}} + (1 - \alpha) \frac{1}{1 + e^{-U(a_i^* | z_i)}}}$$

Step 3: Operational fraud flag

The article repeatedly refers to fraud signals as thresholds for further investigation.

So define a threshold τ:

$$\hat{y}_i = \begin{cases} 1, & S_i \geq \tau \\ 0, & S_i < \tau \end{cases}$$

6. Operationalization and Deployment

Real-time inference is a key requirement, as latency failures are likely to undermine stakeholder trust in the solution. Ingested fraud signals therefore need to be processed before their first use in the simplest possible manner, ideally with a single forward pass through a single model. Beyond fraud signal detection, however, use-case-specific post-processing logic must apply, potentially for every fraud signal category. With the current bank-grade latency target of 500 ms post-metric collection factored in, a separate model, additional decision vegetation layer, or judicious framing of the original problem – for example, reducing deep learning to lower-latency classical ML – are necessary.

Access to data from multiple contact and capacity channels enables further model pooling or resource-sharing opportunities. Supply chain architecture can identify regionally related substances, for example, vehicles and buildings, and relevant weather-related conditions, and cross-reference location and identity behaviour with further tabular features for pricing and terms, enabling safe channel profiling through network analysis. Current design must nevertheless remain realistic, avoiding dependence on yet immature technologies, for example, real-time penetration testing, deepfake detection, or large-scale LLM-drill-down. Testing risk is further heightened since considerable use-case-adjusted adverse-expense-profile margins are already included in current operational budgets.

6.1. Real-Time Inference and Latency Targets

Serving a large number of clients in real time can easily become a bottleneck in the operationalization of a machine-learning fraud signal. A data-driven backend that supports extensive real-time inference capabilities can be built using online processor services such as Azure Functions or AWS Lambda. These platforms allow the deployment of model query endpoints as URL stubs. API services then manage the online requests to the models by distributing the traffic among active instances and automatically growing or shrinking the number of instances in accordance with the demand. This guarantees real-time processing with latency measured in milliseconds for detection. Input for the models can be pre-processed before the model response latency even begins, while the models remain updated with the latest available data.

Accidentally processing million-dollar file upload requests and similar situations are avoided by utilizing the incoming traffic control capabilities of these platforms. Establishing all-stakeholder demand, sufficient traffic management, and a jointly establish incident-response service allows such fraud signals to remain focused on their intended fraud-detection goals.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

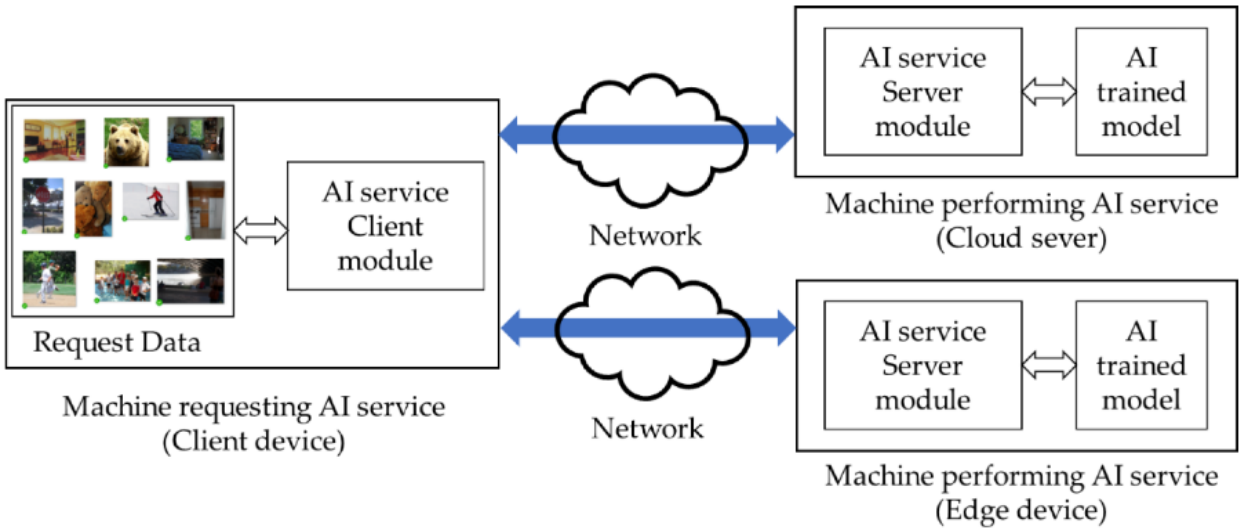


Fig 6: Real-Time Inference and Latency Targets

6.2. Model Management and Versioning

Recognizing the dynamics of fraud risk, the ML models must evolve accordingly. Model management functions address the spectrum of activities related to the training, validation, and operational use of ML models, including updating models triggered by confirmation bias. Automatic detection of fraud through ML pipelines can inadvertently escalate minor fraudulent activities into material risk, which in turn may prompt the training and deployment of dedicated models, distinct from those aimed at increased cost prediction accuracy.

Central to model management are versioned artifacts—models and their features, matrices for inference, performance, and testing metrics. Artifact versioning enables ML observability. Assumptions around data drift, model stability, or temporal change of context are discussed and, if necessary, new inference lookups and test matrices defined. Supporting cloud infrastructure enables versioned artifact curation and management.

6.3. Monitoring, Explainability, and Incident Response

Operational fraud-detection systems must be monitored to ensure that I/O and performance characteristics are within expected bounds. A data-monitoring solution capable of detecting quality drift in data streams supports the overall goal of deploying a solution that is explainable and capable of online learning. Support for Spanish-language capabilities via automated translation of textual features demonstrates consideration of fairness, accountability, and transparency (FAT) principles. Integration of monitoring capabilities also facilitates compliance with regulatory guidelines requiring incident-response mechanisms for AI-based systems.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

Regulatory guidance calls for the establishment of incident-management processes, including escalation points and associated responsibilities, for the investigation of any incident involving the fraud-detection solution that raises an alert flag or is otherwise generated during normal monitoring activities. Legal and data-privacy concerns related to the use of AI-based solutions in high-stakes decision-making processes require deployment of effective monitoring, review, and testing practices that are essential for ensuring system performance and mitigating risk.

Fraud category	Examples mentioned or implied in article	Likely signals/features	Intended operational use
Claims fraud	Exaggerated claims, false medical/injury claims, staged loss, fictitious documents, fabricated asset loss	Claim timing anomalies, claim history change, document inconsistencies, claimant behavior, external corroboration gaps	Flag suspicious claims for investigation before payout
Underwriting fraud	Misrepresentation at onboarding, anti-selection, false declarations	Input-rule violations, inconsistent applicant attributes, undeclared risk indicators, suspicious quote patterns	Detect risky or deceptive applications during underwriting
Policy fraud	Fictitious policies, fraudulent policy modification, policy lifecycle abuse	Reissue anomalies, originator risk score, abnormal policy changes, linked-entity irregularities	Prevent policy abuse and fraudulent policy management
Channel fraud	Misuse by agents, intermediaries, channel partners, unsupported offerings, unauthorized benefits	Distribution anomalies, geographic irregularities, product/channel mismatch, partner behavior deviations	Monitor sales/distribution abuse and third-party misconduct

Table 3: Fraud Categories, Signals, and Use-Cases Derived from the Paper

7. Risk, Ethics, and Legal Compliance

The deployment of AI systems generates a wide range of risks, especially when humans are not in the loop, or the decision can cause damage or even loss of life. For example, consumers lose trust when a fraud detection system generates false negatives or falsely accuses innocent people. Therefore, an intelligent fraud detection system must satisfy common security requirements, such as availability, integrity, and confidentiality; fairness, accountability, and transparency (FAT); and regulatory compliance. Addressing common security requirements and FAT is vital in any AI system, including fraud prevention. The choice of technology integration must consider privacy and data protection, legal implications in the event of mistakes, and the stability of internal and external relations so that functional requirements are met according to expectations.

Privacy and security of personal information are nowadays important topics in most democratic countries. Multifactor authentication, data encryption during transit and at rest, data leakage protection, and other common standards offered by cloud providers should be used. The customer is an external stakeholder. There must be a justification for collecting personal

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

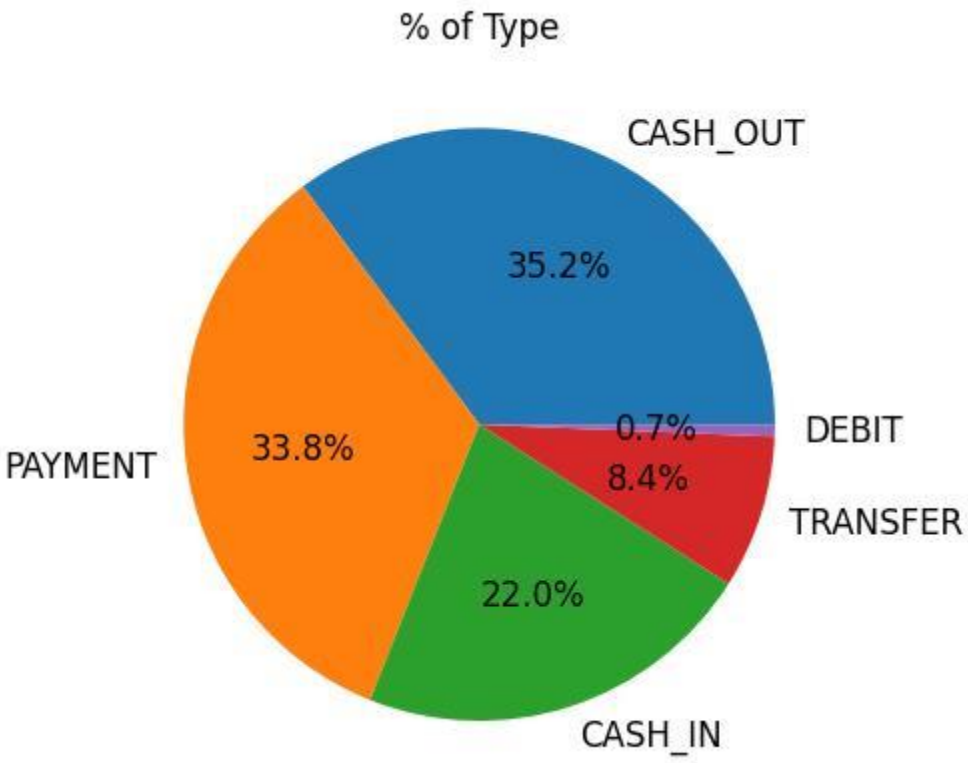
information about customers. A valid justification is that with this information, an intelligent fraud detection system can learn to prevent fraud, generally for a more personalized experience and protection of the customer. Even so, privacy regulations such as GDPR, HIPAA, and LGPD and the ePrivacy Directive must be followed. Data must be anonymized whenever possible, customers must have the right to limit how their data is used, and users must be notified in case of data breaches.

7.1. Privacy, Security, and Data Governance

Modern insurance policies require organizations to safeguard sensitive customer information from data breaches and to guarantee effective cybersecurity. Digital privacy regulations (e.g., the European General Data Protection Regulation and the California Consumer Privacy Act) establish minimum thresholds for the collection, retention, and usage of personal information with which organizations must comply. Fraud detection models inevitably require sensitive customer data for effective operations, and machine learning systems inherently involve heightened cybersecurity threats.

The application of privacy-preserving, model-aware, secure and trustworthy federated learning techniques is particularly important in the context of model training. This entails multi-party collusion-protected secure multiparty computation and secret sharing protocols designed to jointly build a model while encrypting intermediary results, thus preventing the exposure of any of the training clients' data. Such methods also decrease the risk of leakage of sensitive information. Moreover, user data are retained only for as long as required to build and maintain the fraud detection models. Privacy- and security-preserving data augmentation and anonymization techniques (e.g., data synthesis, data masking, query-based black-box approaches, closed-box model inversion, and clustering) reduce the risk of data leakage further.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07



7.2. Fairness, Accountability, and Transparency

Use of ML models and related technologies in supporting decisions in such sensitive areas places requirements of fairness, accountability, and transparency (FAT) on model designers and deployers often extending beyond the regulations imposed by supervisory authorities. Examples of these requirements include:

- Documenting all model development and operational phases, even if innovative techniques and/or guidelines such as those from the Institute of Electrical and Electronics Engineers (IEEE) are followed. This is crucial, especially when the selection of the black-box-style ML model appears to contradict the explanation needs for that specific use case.

- Assessing performance at different segments of the data without relying on age, gender, or ethnicity as segmenting attributes. Effort is invested to evaluate the risk of bias and fairness of the deployed models by exploring different measures of optimal fairness and engaging in fairness-aware reweighting of dataset instances.

- Explaining the ML system decisions along with supporting evidence. A wide array of model-explanation methods are investigated, and the selected explanations are communicated a posteriori.

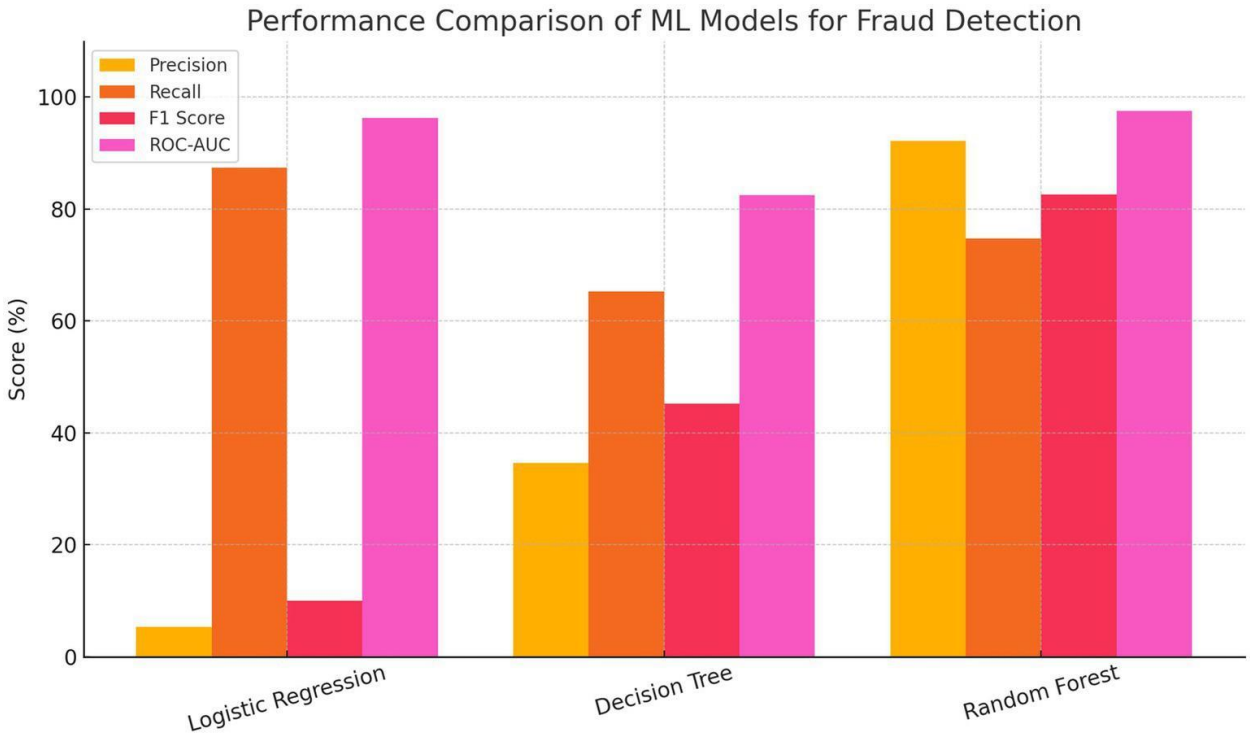
ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 02 ISSUE: 02
RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

- Creating an audit trail allowing a final-user attribution of the final decision made by the ML system, indicating what components were involved and when. Attention is devoted in particular to the design of procedures for dealing with incidents in which the model judgment is overruled by the business expert.

7.3. Regulatory Landscape and Compliance Frameworks

AI systems for risk-sensitive domains must comply with the relevant laws, regulations, and policies that govern the targeted sector. Depending on factors such as the geographical location of the financial services provider and any applicable service distribution networks and agent partnerships, these rules can vary in scope and complexity. In markets with a lower degree of regulatory oversight, such as the United States, the categories and nature of these rules differ from those in Europe. Consequently, it is essential to consider the requirements pertinent to the specific context of implementation and to demonstrate legal compliance at all levels of the AI system development.

In the EU, the General Data Protection Regulation (GDPR) is often cited as one of the most stringent data privacy regulations globally. Designed to provide data subjects with greater control over the processing of their personal information, the GDPR incorporates principles of data use statement, storage limitation, data-minimization, and purpose limitation into all data-processing activities. However, in the absence of legal provisions to facilitate trade commissions on insurance sale contracts, the business activity of selling insurance policies remains prohibited in Cuba.



ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

8. Conclusion

Despite their substantial contributions to the field, prior works lack integration of both Generative and Agentic AIs within a single system framework. Consequently, the thesis proposes an intelligent fraud detection architecture leveraging synergies between the two approaches. A Generative framework detects suspicious activities based on corroborative evidence, while a decision-centric Agentic setting guides each stage in its own specialized way. Such an integrated, hybrid method overcomes pitfalls intrinsic to specialized models.

Despite rich insights from these previous works, neither has yet explored a combined Generative and Agentic AI architecture for the fraud detection domain that truly synergizes the two paradigms. Accordingly, the present study proposes an intelligent fraud detection framework that employs synergies between Generative and Agentic AIs. At any point in the process, Generative technology is used to detect fraudulent activities based on corroborative evidence across data sources, while an Agentic approach services decision-making guidance, driving every stage toward realization of claimed objectives in the most cost-effective way. Such an integrated hybrid method overcomes pitfalls intrinsic to specialized models, ensuring best-of-both-worlds performance.

8.1. Future Trends

Many InsurTech companies are striving toward the goal of "completely digital insurance"—automated decision-making, real-time predictions, no human interactions (except possibly when purchasing); often even beyond no human interactions, for example full claim handling based solely on the automated evaluation of digital data such as accident pictures. This utopia will inevitably take a long time to realize and furthermore address the problem if not impossibility of a fully error-free digital system.

Truly zero-defect is the target of traditional being or action (realized by the technology-centered AI approach). Therefore, two future directions at the other end of the digital error spectrum are conceivable. In the first the gap to zero-defect operation in totally digital systems is closed more and more (there are only a few, mostly minor errors remaining) mostly by generative AI (e.g., claims fraud detection). In the second direction the target of a totally digital insurance company is radically relaxed, with high error rates in low-probability regions compensated for by low-cost and fast error handling. In such a set-up agentic AI (e.g., real-time detection of fraud initiators) enables the back end of major systems to operate at a larger distance to zero-defect, while efficiently reducing or eliminating business risks. Combining both approaches then leads to a truly hybrid architecture: real-time control via agentic AI and operational decisions mostly by generative AI.

9. References

1. Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. Research in International Business and Finance, 62, 101744.
2. Gomes, C., Jin, Z., & Yang, H. (2021). Insurance fraud detection with unsupervised deep learning. Journal of Risk and Insurance, 88(3), 591–624.
3. Kolla, T., & Kolla, S. K. (2023). FHIR-Based Real-Time Healthcare Analytics using Unsupervised Learning. International Journal of Future Innovative Science and Technology (IJFIST), 6(6), 11751.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

4. Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Computer Science Review, 40, 100402.
5. Bommasani, R., Hudson, D. A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. arXiv.
6. Brown, T. B., Mann, B., Ryder, N., et al. (2021). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.
7. Mattaparthi, R. (2023). Connected Fleet Intelligence: Edge-Centric Analytics and Computer Vision for Predictive Manufacturing and Asset Resilience. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 6(5), 9077-9088.
8. Chowdhery, A., Narang, S., Devlin, J., et al. (2022). PaLM: Scaling language modeling with pathways. arXiv.
9. Touvron, H., Lavril, T., Izacard, G., et al. (2023). LLaMA: Open and efficient foundation language models. arXiv.
10. OpenAI. (2023). GPT-4 technical report. arXiv.
11. Mangalampalli, B. M. Intelligent Data Profiling for Healthcare Data Lakes Using AI-Enhanced Analytics.
12. Wei, J., Wang, X., Schuurmans, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824–24837.
13. Yao, S., Zhao, J., Yu, D., et al. (2023). ReAct: Synergizing reasoning and acting in language models. International Conference on Learning Representations.
14. Pereira, K., Vinagre, J., Alonso, A. N., Coelho, F., & Carvalho, M. (2022, September). Privacy-preserving machine learning in life insurance risk prediction. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 44-52). Cham: Springer Nature Switzerland.
15. Schick, T., Dwivedi-Yu, J., Dessi, R., et al. (2023). Toolformer: Language models can teach themselves to use tools. Advances in Neural Information Processing Systems.
16. Wang, L., Ma, C., Feng, W., et al. (2023). A survey on large language models. arXiv.
17. Wu, Q., Bansal, G., Zhang, J., et al. (2023). AutoGen: Enabling next-generation LLM applications via multi-agent conversation. arXiv.
18. Guo, T., Chen, X., Wang, Y., et al. (2024). Large language model based multi-agent systems: A survey. arXiv.
19. Reddy, V. A. R. (2023). Orchestrating the Future Autonomous Healthcare Data Pipeline Management through Agentic AI Architectures. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 6(2), 7979-7992.
20. Xi, Z., Chen, W., Guo, X., et al. (2023). The rise and potential of large language model based agents: A survey. arXiv.
21. Wang, X., Wei, J., Schuurmans, D., et al. (2023). Self-consistency improves chain of thought reasoning in language models. International Conference on Learning Representations.
22. Aitha, A. R. (2023). Cloud-Native Big Data AI/ML Framework for Risk Intelligence and Fraud Control in Banking and Insurance Ecosystems. Available at SSRN 6157967.
23. Mialon, G., Dessì, R., Lombardi, M., et al. (2023). Augmented language models: A survey. Transactions on Machine Learning Research.
24. Liu, P., Yuan, W., Fu, J., et al. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1–35.
25. Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706.
26. Ziegler, D. M., Stiennon, N., Wu, J., et al. (2021). Fine-tuning language models from human preferences. arXiv.
27. Vaswani, A., Gomez, A. N., Jones, L., et al. (2021). Scaling transformer architectures for large language models. arXiv.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

28. Wei, J., Bosma, M., Zhao, V. Y., et al. (2022). Finetuned language models are zero-shot learners. International Conference on Learning Representations.
29. Mangala, N. (2024). Leveraging Microsoft Fabric lakehouse as an AI-ready data platform for enterprise analytics. Journal of Information Systems Engineering and Management.
30. Madaan, A., Tandon, N., Clark, P., et al. (2023). Self-Refine: Iterative refinement with self-feedback. Advances in Neural Information Processing Systems Workshops.
31. Park, J. S., O'Brien, J., Cai, C. J., et al. (2023). Generative agents: Interactive simulacra of human behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.
32. Inala, R. AI-Powered Investment Decision Support Systems: Building Smart Data Products with Embedded Governance Controls.
33. Wang, C., Zhao, S., Wang, M., et al. (2023). Large language models for financial applications: A survey. arXiv.
34. Fan, S., Li, Y., Chen, X., & Zhang, H. (2024). Large language models for financial intelligence: Opportunities and challenges. IEEE Access, 12, 41025–41049.
35. Banulescu-Radu, D., & Yankol-Schalck, M. (2024). Practical guideline to efficiently detect insurance fraud in the era of machine learning: A household insurance case. Journal of Risk and Insurance, 91(4), 867–913.
36. Soni, H., Perla, S., Maddela, S., & Kumar, U. (2025, November). Generative AI in Cloud CRM: Securing Intelligent Workflows in Multi-Cloud Environments. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1-9). IEEE.
37. Debener, J., Heinke, V., Kriebel, J., & Schiller, J. (2024). Detecting insurance fraud using supervised and unsupervised machine learning. Journal of Risk and Insurance.
38. Mudigonda, S. S., Baruah, P. K., Kandala, P. K., Gupta, R. Y., Hegde, S. N., & Chebrolu, S. (2024). Using interpretable machine learning methods: An application to health insurance fraud detection. Society of Actuaries Research Report.
39. Nagabhyru, K. C., & Engineer, S. D. (2023). Unifying Data Engineering and Machine Learning Pipelines: An Enterprise Roadmap to Automated Model Deployment.
40. Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730–27744.
41. Bubeck, S., Chandrasekaran, V., Eldan, R., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv.
42. Amistapuram, K. (2024). Smart Decision Support Systems For Dynamic Tax Policy Optimization Using Reinforcement Learning. Available at SSRN 6143426.
43. Liu, Y., Wang, Z., Li, H., et al. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. Meta-Radiology, 1(2), 100017.
44. Zhao, W. X., Zhou, K., Li, J., et al. (2023). A survey of large language models. arXiv.
45. Davuluri, P. N. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems.
46. Pan, S. J., & Yang, Q. (2021). A survey on transfer learning for intelligent systems. IEEE Transactions on Knowledge and Data Engineering.
47. Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

48. Kolla, S. K. (2023). Learning Health Systems Machine Intelligence for Clinical Prediction and Healthcare Optimization. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 6(2), 7955-7966.
49. Bommasani, R., Liang, P., Hudson, D. A., et al. (2023). Ecosystem, risks, and opportunities of foundation models. Communications of the ACM, 66(7), 84–90.
50. Li, Y., Bubeck, S., Eldan, R., et al. (2023). Textbooks are all you need. arXiv.
51. Gao, L., Biderman, S., Black, S., et al. (2021). The Pile: An 800GB dataset of diverse text for language modeling.
52. Yandamuri, U. S. (2024). AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology. Metallurgical and Materials Engineering.
53. Advances in Neural Information Processing Systems, 34, 18719–18740.
54. Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. Research in International Business and Finance, 62, 101744.
55. Inala, R. Advancing Group Insurance Solutions Through Ai-Enhanced Technology Architectures And Big Data Insights.
56. Van der Merwe, D., Pretorius, P., & Van Vuuren, G. (2024). Automobile insurance fraud detection using data mining: A systematic literature review. Intelligent Systems with Applications, 21, 200340.
57. Akinola, O., & Nieuwenhuizen, C. (2022). Automobile insurance fraud detection in the age of big data: A systematic and comprehensive literature review. Journal of Financial Regulation and Compliance, 30(4), 503–523.
58. Kolla, S. K. (2022). Engineering Healthcare Data Infrastructures for Predictive Clinical Analytics and Evidence-Based Decision Making. International Journal of Engineering & Extended Technologies Research (IJEETR), 4(5), 5370-5380. |
59. Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730–27744.
60. OpenAI. (2023). GPT-4 technical report. arXiv:2303.08774.
61. Touvron, H., Lavril, T., Izacard, G., et al. (2023). LLaMA: Open and efficient foundation language models. arXiv:2302.13971.
62. Peddi, R. K. (2024). AI-Based Workforce Analytics for SLA Governance and Uptime Assurance in Data Centers. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 8589-8601.
63. Yao, S., Zhao, J., Yu, D., et al. (2023). ReAct: Synergizing reasoning and acting in language models. International Conference on Learning Representations.
64. Schick, T., Dwivedi-Yu, J., Dessì, R., et al. (2023). Toolformer: Language models can teach themselves to use tools. Advances in Neural Information Processing Systems.
65. Kolla, S. H. (2024). Retrieval-Augmented Enterprise Intelligence: Enhancing Accuracy, Trust, and Operational Decision-Making. International Journal of Future Innovative Science and Technology (IJFIST), 7(2), 12425.
66. Wu, Q., Bansal, G., Zhang, J., et al. (2023). AutoGen: Enabling next-generation LLM applications via multi-agent conversation. arXiv:2308.08155.
67. Xi, Z., Chen, W., Guo, X., et al. (2023). The rise and potential of large language model based agents: A survey. arXiv:2309.07864.
68. Guo, T., Chen, X., Wang, Y., et al. (2024). Large language model based multi-agent systems: A survey of progress and challenges. arXiv:2402.01680.
69. Mangalampalli, B. M. Generative AI Applications In Healthcare Data Mart Design And Optimization.

ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 02 ISSUE: 02

RECEIVED: APRIL 17 REVISED: MAY 01 ACCEPTED: MAY 22 PUBLISHED: JUNE 07

70. Mialon, G., Dessì, R., Lombardi, M., et al. (2023). Augmented language models: A survey. Transactions on Machine Learning Research.
71. Liu, P., Yuan, W., Fu, J., et al. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1–35.
72. Wang, L., Ma, C., Feng, W., et al. (2023). A survey on large language models. arXiv:2303.18223.
73. Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. power, 9(12).
74. Zhao, W. X., Zhou, K., Li, J., et al. (2023). A survey of large language models. arXiv:2303.18223.
75. Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys.
76. Brown, T. B., Mann, B., Ryder, N., et al. (2021). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.
77. Bandi, V. D. V. K. (2023). MLOps frameworks for reliable model deployment in cloud data platforms. Journal of Artificial Intelligence and Big Data, 3(1), 84-101.
78. Chowdhery, A., Narang, S., Devlin, J., et al. (2022). PaLM: Scaling language modeling with pathways. arXiv:2204.02311.
79. Wei, J., Wang, X., Schuurmans, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824–24837.
80. Wang, X., Wei, J., Schuurmans, D., et al. (2023). Self-consistency improves chain of thought reasoning in language models. International Conference on Learning Representations.
81. Reddy, V. A. R., & Kolla, S. K. (1984). Infrastructure-As-Code Practices For Regulated Healthcare Cloud Environments. Metallurgical and Materials Engineering, 30 (4), 1028–1042.
82. Gomes, C., Jin, Z., & Yang, H. (2021). Insurance fraud detection with unsupervised deep learning. Journal of Risk and Insurance, 88(3), 591–624.
83. Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. Research in International Business and Finance, 62, Article 101744.
84. Kolla, T. (2024). Graph Neural Networks for HCC Risk Adjustment and Interoperability. International Journal of Science, Research and Technology, 7(6), 13244-13255.
85. Schrijver, G., Sarmah, D. K., & El-hajj, M. (2024). Automobile insurance fraud detection using data mining: A systematic literature review. Intelligent Systems with Applications, 21, Article 200340.