VER
Vascular & Endovascular Review

Self-Improving Frontier Agents for Insurance Operations: A Governed Architecture for Autonomous Claims Reasoning And Workflow Adaptation

Venkata Harikishan Koppuravuri[1], Abhishek Kumar[2], Karthik Budige[3]

[1]Director of Cloud Engineering, USA
[2]Senior Solution Architect, Dotcom Team LLC
[3]Senior Data Engineer

ABSTRACT

Frontier artificial intelligence models have demonstrated unprecedented advancements in multi-hop reasoning, tool-use, and interpretive capabilities across complex domains. Despite these advances, regulated industries such as insurance remain constrained by static automation, brittle rule systems, and manual adjudication processes. Current large language model (LLM) deployments in enterprise settings utilize retrieval-augmented generation or simple tool-routing frameworks but lack the ability to improve autonomously over time, restricting operational scalability and consistency. Here we introduce Self-Improving Frontier Agents (SIFA)—a governed, autonomous agentic architecture integrating frontier models (e.g., GPT-o3, Claude 3.5 Opus, Nova Pro) with continuous self-improvement loops, synthetic data generation, human-in-the-loop corrections, and auditable reasoning traces. SIFA dynamically refines claims reasoning, policy interpretation, and workflow steps through governed adaptation pipelines while maintaining compliance and transparency required for regulated enterprise environments. To our knowledge, this is the first documented application of self-improving frontier agents in the insurance domain, bridging a gap between frontier reasoning capabilities and real-world adjudication workflows. We establish the conceptual framework, architecture, implementation methodology, metrics, and governance model necessary for such deployments, and outline implications for broader regulated industries.

KEYWORDS: Self-Improving Agents, Frontier AI Models, Insurance Claims Adjudication, Governed Adaptation, Retrieval-Augmented Generation.

How to Cite: Venkata Harikishan Koppuravuri, Abhishek Kumar, Karthik Budige, (2026) Self-Improving Frontier Agents for Insurance Operations: A Governed Architecture for Autonomous Claims Reasoning And Workflow Adaptation, Vascular and Endovascular Review, Vol.9, No.1, 402-415

INTRODUCTION

The insurance claims adjustment, customer service, and policy interpretation are highly complex, multi-step processes that demand advanced reasoning expertise, a profound understanding of the domain, and strict adherence to emerging legal and regulatory standards. Even though the adoption rate of electronic claims processing has increased significantly, reaching over 96-98 percent of medical transaction and 85 percent of dental transaction as of 2023, the traditional enterprises AI systems, including fixed machine learning models, deterministic rule engines, and simple retrieval-enhanced large language models, cannot perform autonomous self-correction or respond to actual operational feedback (CAQH, 2024a, 2024b; Eling et al., 2022). This weakness perpetuates deep inefficiencies within the insurance industry, whose administrative complexities and fragmented data ecosystems continue to drive up costs and prevent decision-making at scale in highly regulated settings.

These systemic issues take many forms: the inconsistency of decisions due to variability in human adjudicators and cognitive biases, which can be referred to as noise in the judgment process; excessive reviews of manuals due to the lack of sufficient workforce, burdening the system; the high refusal rates, with the initial refusal rates varying between 15-19 percent of the submitted claims and disproportionate effects on socioeconomic and demographic groups; and increased operational costs, with annual U.S. healthcare adjudication spending approaching more than a quarter of the prior year (Other problems are long average handling times, less explainability in automated decisions resulting in regulatory review, and lack of audit trail which makes it harder to meet standards like HIPAA, GDPR and new AI governance regulations. Academic commentaries also emphasise the role of these aspects in the interpretation differences between adjusters and systemic inequalities in claims decisions (Kahneman et al., 2021; Li et al., 2025).

In sharp contrast, frontier AI models by 2025, based mainly on mixture-of-experts models and modern reasoning paradigms, have exceptional capabilities in multi-hop reasoning, self-criticism, sophisticated tool combining, multimodal data processing (including text, images, audio, and video), more expansive context windows (millions of tokens) and significantly reduced hallucination due to enhanced factuality mechanisms (NVIDIA, 2025; Li et al., 2025; Namperumal et al., 2024). These are the capabilities to interpret policies subtly, to analyse counterfactual scenarios, and to engage in reflective planning, way above those of past generations. In practice, however, in enterprise applications, these models are fixed, incapable of building knowledge through repeated interactions or of changing, since changes in the regulations or operations of their domain-specific environments shift their regulatory or operational context (Bhattacharya et al., 2025; Eling et al., 2022).

The fundamental question that this paper addresses is: How can frontier models be designed as agent governance, self-directed agents that will fit into controlled areas such as insurance? We present Self-Improving Frontier Agents (SIFA). This end-to-end

architecture integrates frontiers, structured self-improvement loops, synthetic data augmentation, human-in-the-loop validation, and traceable reasoning paths to enable autonomous claims adjudication, accurate policy explication, and dynamic workflow optimisation while maintaining rigorous compliance, transparency, and safety standards. In the face of the blistering introduction of AI in insurance, where surveys have shown that at least one application of generative AI has been implemented by 76% of U.S. insurers already, and that more said approach is projected to continue to grow, no known academic framework exists describing governed continual improvement paradigms of frontier agents in adjudication processes (Deloitte Center for Financial Services, 2024; Bhattacharya et al., 2025). SIFA establishes a sound conceptual and technical foundation for transforming controlled operations into resilient, learning-oriented intelligence.

RELATED WORK

2.1 Agentic AI

The swift evolution of agentic AI frameworks has made large language models (LLMs) capable of iterative reasoning, integrating tools, and autonomously executing tasks beyond simple generation. The most basic paradigms, like ReAct, combine inner chain-of-thought and outer activity, with models switching between planning and interaction with the environment in periodic cycles (Yao et al., 2023).

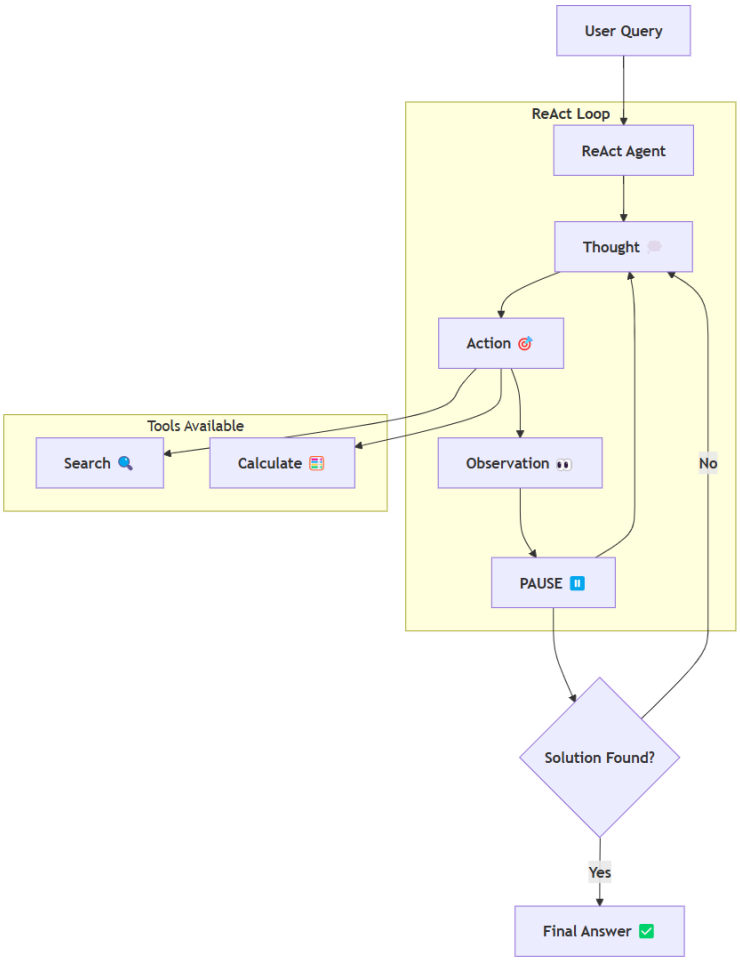

Fig.1: This diagram illustrates a custom ReAct (Reason + Act) agent loop for building AI agents from scratch. The agent processes a user query by iteratively generating thoughts, selecting actions (e.g., Search or Calculate tools), observing results, and pausing until a solution is found, leading to the final answer. Adapted from Plaban Nayak's deep dive into AI reasoning patterns.

The strategy has spawned open-source movements, such as AutoGPT, which breaks down high-level goals into subtasks through recursive prompting and tool calls, and Voyager, an embodied agent that constructs libraries of reusable skills through self-motivated exploration in complex settings such as Minecraft (Toran, 2023; Wang et al., 2023).

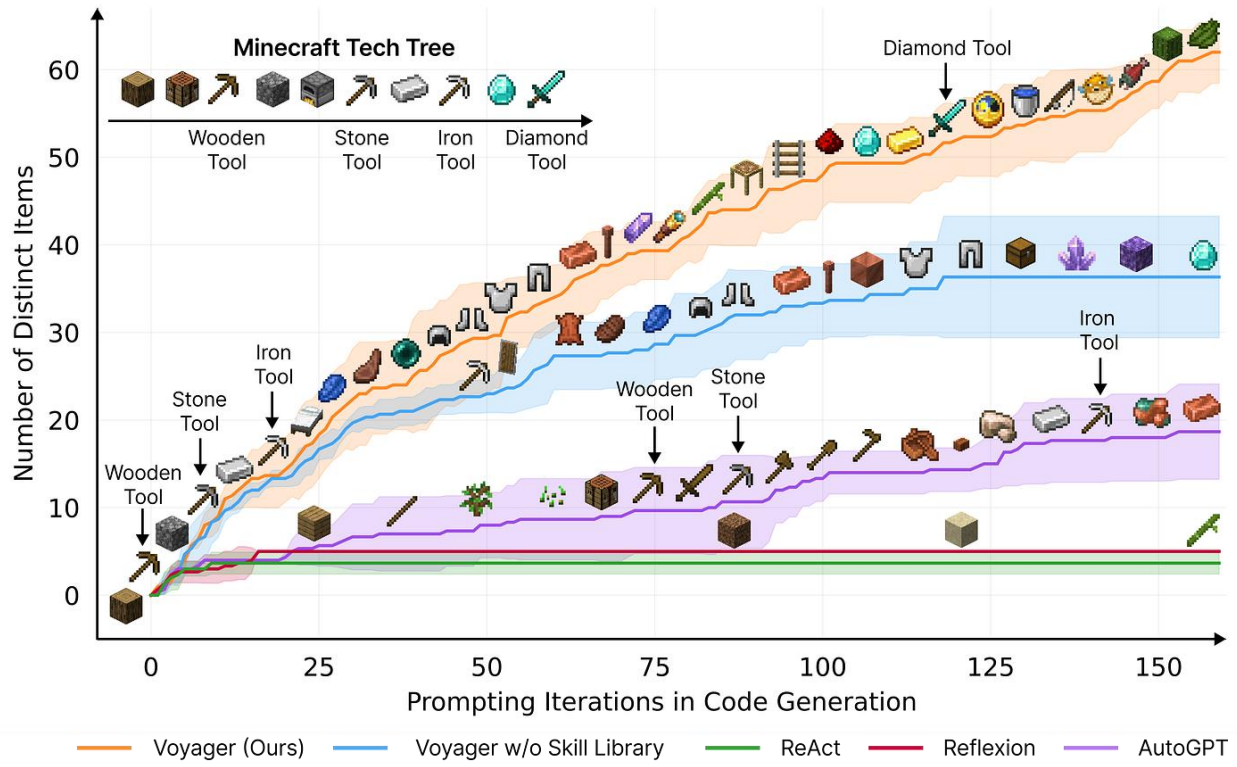

Fig.2: Voyager agent's Minecraft tech tree progression: distinct items unlocked (wooden to diamond tools) vs. prompting iterations, showing rapid skill acquisition over baselines like ReAct. (Adapted from the Voyager project, 2023)

According to Yao et al. (2022), the ReAct (Reasoning + Acting) paradigm combines the verbal reasoning trace proposed by language models with a task-specific activity, enabling stronger, interpretable problem-solving through a loop of thought-actions-observation. A more basic form of this regularity is used in custom agent programs in which the agent solves a user query by producing explicit thoughts, choosing some action, like a search or calculation program, making use of the observations that the action produces, and repeating the process until a solution is found.

Such capabilities are extended through multi-agent orchestration frameworks such as AutoGen (Wu et al., 2023), which can coordinate conversations among specialised agents to support emergent behaviours, such as long-horizon planning and self-correction via reflection mechanisms (Shinn et al., 2023). These systems are outstanding across a variety of fields, including coding and Internet navigation, as well as science and analysis. Most recently, there have been multimodal inputs, expanded context windows, computer-use primitives, and, now, agents interact directly with real-world interfaces and environments (Anthropic, 2024 to 2025).

Despite such dramatic advances, in regulated sectors with substantial stakes, e.g., finance, insurance, and healthcare, implementing agentic systems is complex. These include the primary risks associated with autonomous behaviour, trails of reasoning, potentially infinite loops, and unintended escalation of privileges, which may break substantial norms such as HIPAA and GDPR guidelines (Sapkota et al., 2025; industry analysis, 2025). Explainability, amplification of bias, and accountability in production environments are also problems that are often aggravated by missing the key aspect of a complete audit trail, a forced human oversight gate, and the inherent ethical restrictions found in current structures. These loopholes make adoption scalable in environments that require deterministic results and regulatory coordination.

2.2 Frontier Models

In late 2025, the frontier large language models will mainly be based on sparse mixture-of-experts (MoE) architectures, with large-scale post-training reinforcement learning on human feedback (RLHF) and improved alignment methods to achieve higher cognitive performance across a variety of benchmarks. The best exemplars are the o-series models (o3 and o4 variants) of OpenAI, which use extended private chain-of-thought deliberation and can call functions; the multimodal suite of Gemini 3 of Google with massive context windows (millions of tokens); and the Claude 4 line of Anthropic, optimized to work as an agent (OpenAI, 2025; Google DeepMind, 2025; Anthropic, 2025). These models show excellent progress in organized information processing, counterfactual simulation, and cross-modal integration.

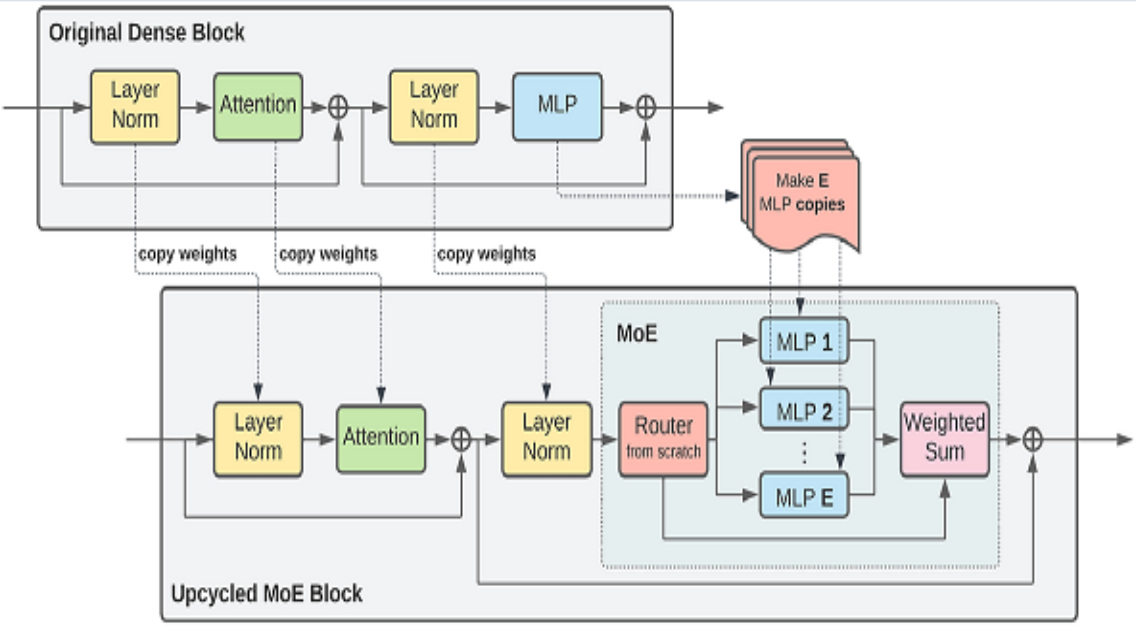

Fig.3: Illustration of a Switch Transformer encoder block. We replace the dense feed forward network (FFN) layer present in the Transformer with a sparse Switch FFN layer (light blue). The layer operates independently on the tokens in the sequence. We diagram two tokens (x_1 = "More" and x_2 = "Parameters" below) being routed (solid lines) across four FFN experts, where the router independently routes each token. The switch FFN layer returns the output of the selected FFN multiplied by the router gate value (dotted-line). Adapted from Adapted from William Fedus et al.,2021

In particular, frontier models are deeply multi-hop, exhibit reflective planning in the form of iterative internal deliberation loops, tool-use via standardised function-calling schemas, endowed with highly factual qualities via advanced verification systems, and reward modelling, and with significantly high hallucination rates cut by advanced verification systems and reward models (NVIDIA, 2025; Phuong et al., 2025). Experimental assessments indicate state-of-the-art scores on demanding criteria, such as close-to-saturation scores on graduate-level science questions, breakthrough scores on long-context retrieval tests, and high-quality agentic behaviour in simulated environments that require sequential decision-making and interaction with the environment (Artificial Analysis, 2025; Anthropic, 2025).

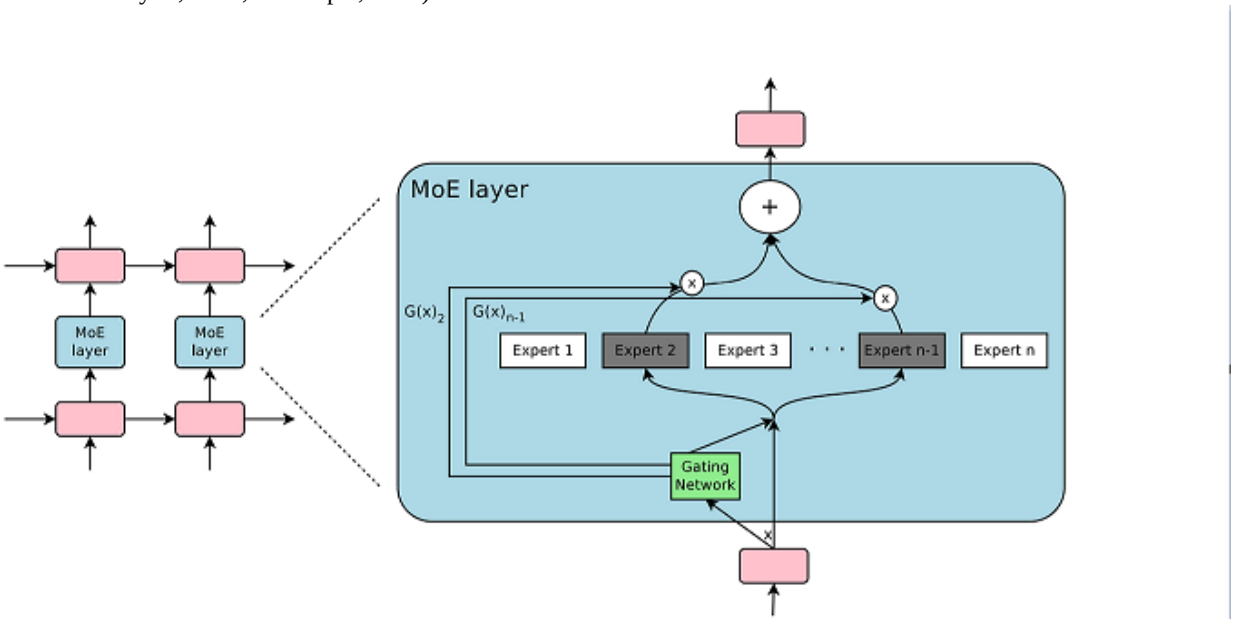

Fig.4: A Mixture of Experts (MoE) layer embedded within a recurrent language model. In this case, the sparse gating function selects two experts to perform computations. Their outputs are modulated by the outputs of the gating network. Adapted from Shazeer et al. (2017).

However, these models are necessarily fixed after deployment, and do not inherently involve the continuous learning process of an endogeneously specified model, nor do they adjust their parameters to the incoming stream of operational data. Even though they facilitate in-context learning and retrieval-augmented generation to support transient adaptation, core weights do not learn

independently, leading to knowledge obsolescence, the failure to apply domain-specific corrections, and susceptibility to distributional shift unless retraining cycles are performed at high costs (Eling et al., 2022; Bhattacharya et al., 2025). This architectural limitation makes frontier models unsuitable to support dynamic business processes in regulated industries, such as insurance, where regular integration of regulatory changes, the development of new claims law, and refinement based on feedback are crucial, and therefore, externally governed structures are needed (Li et al., 2025; Namperumal et al., 2024).

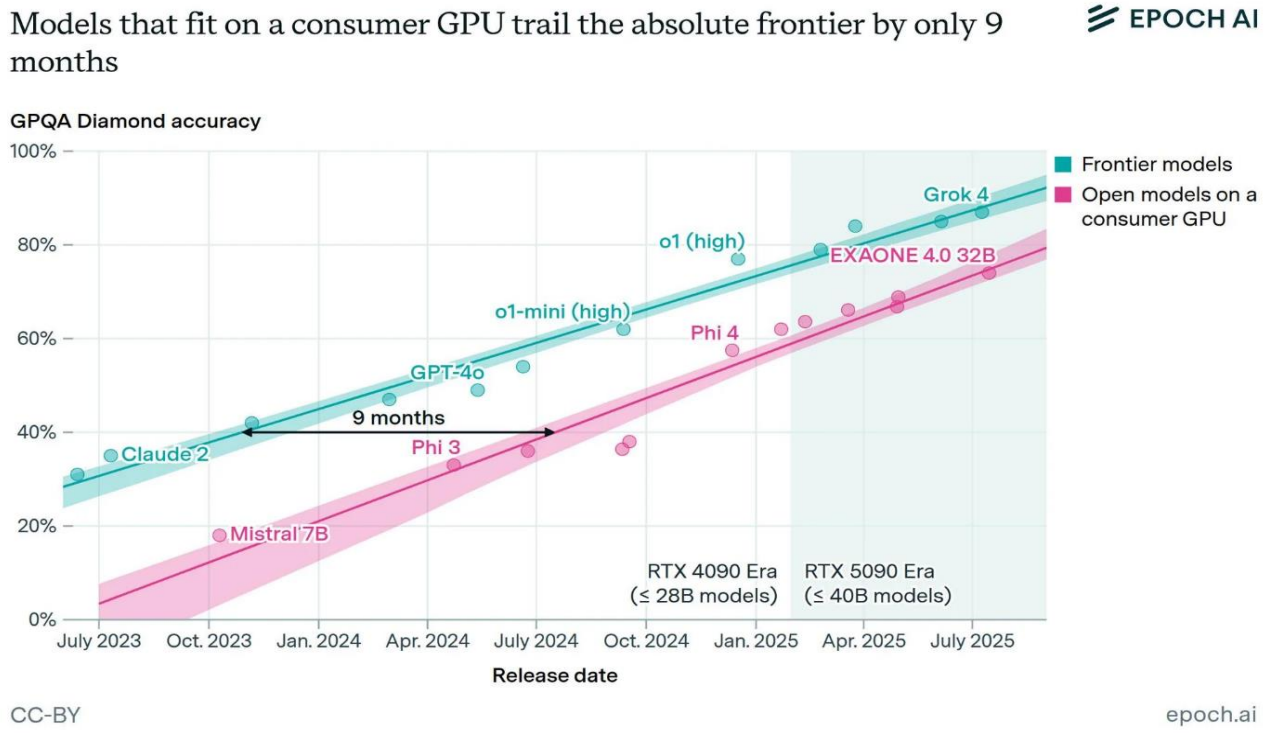

Fig. 5: Consumer GPU-runnable open-weight models (≤40B parameters on RTX 5090 era hardware) match frontier performance on GPQA Diamond, trailing the absolute leading models by approximately 9 months. Adapted from Somala & Emberson (2025).

Robotics and reinforcement learning algorithms, including Reinforcement Learning from Human Feedback (RLHF), offline reinforcement learning and continuous learning paradigms, have shown how to build robust mechanisms of self-improvement in AI systems, being able to modify behaviors and policies with accumulated experience or structured feedback without experiencing catastrophic forgetting of experience (Christiano et al., 2017; Kirkpatrick et al., 2017; Levine et al., 2020). Initially developed for dynamic physical systems such as robotic manipulation and autonomous navigation, these methods enable agents to optimise actions sequentially in sparsely rewarded environments, incorporate human preferences to optimise their goals in complex environments, and ensure stability across non-stationary distributions (Ziegler et al., 2019; Ouyang et al., 2022). Within the framework of large language models, RLHF and its variations are core to post-training alignment, and continual pre-training, experience replay, and parameter-efficient adaptation methods address knowledge obsolescence and sequence learning, allowing foundation models to grow their capabilities incrementally (Ke et al., 2023; Bell et al., 2025).

Even with these improvements, deployment in enterprise-level or regulated settings remains largely static, relying mainly on one-time fine-tuning or retrieval augmentation (rather than automatic, closed-loop improvement) (Eling et al., 2022; Gunaseelan et al., 2024). Critically, it is not a systematic publication that has addressed governance mechanisms for self-improvement, enforcing auditable update pathways, version-controlled adaptation, and mandatory human oversight to avert model drift or regulatory non-adherence in safety-critical applications.

In addition, the current literature does not have detailed structures that combine continuous learning with frontier agents equipped with tools that must work in highly controlled areas like insurance, where the decision is required to be made according to strict legal, ethical, and fairness principles, and does not sufficiently provide safety-related restrictions on autonomous parameter learning to reduce the risk of hallucinating amplification, introducing of bias, and introducing unacceptable or unauthorized knowledge (Bhattacharya et al., 2025; Agrawal et al., 2023). In this way, SIFA addresses a significant research gap by providing a controlled architecture to support the safe, enduring, and lawful underpinning of continuous self-enhancement for frontier actors in insurance adjudication practice.

PROBLEM DEFINITION

A central challenge in deploying frontier models for insurance claims adjudication is the transition from static inference to governed continual adaptation while maintaining strict regulatory compliance and full auditability.

Let Q denote a natural-language query issued by a customer or claims adjuster, where $Q \in Q$ represents the space of domain-

specific inquiries, such as coverage eligibility or explanation of benefits. Let $C \in C$ denote the multimodal claims data associated with a case, comprising structured attributes, including claim identifiers, dates of loss, and monetary amounts, as well as unstructured artifacts, including incident narratives, photographs, and medical reports. Let $P \in P$ denote the relevant policy documents, encompassing contractual clauses, endorsements, and jurisdiction-specific regulatory provisions.
Let

$$T=\{t1, t2, \dots, tk\}$$

represent a curated collection of enterprise tools exposing deterministic application programming interfaces, such as claim retrieval, eligibility computation, and case note creation. Let f_θ denote a frontier large language model parameterized by $\theta \in \Theta$ equipped with multi-step reasoning, tool invocation, and multimodal processing capabilities. Finally, let G denote a formal set of governance constraints consisting of enforceable rules covering regulatory compliance, safety thresholds, fairness requirements, and mandatory traceability. Any admissible reasoning trace or action must satisfy the constraints defined by G.

3.1 Limitations of Static Retrieval-Augmented Generation

Conventional retrieval-augmented generation systems deployed in insurance operations produce a response R through a single forward inference step:

$$R=f_\theta\ (Q \oplus Retrieve(Q, C, P; K)),$$

where Retrieve $(; K)$ performs similarity-based search over a knowledge base Kcontaining embedded policy excerpts, historical adjudication precedents, and regulatory guidance, and where $\oplus$ denotes context concatenation within the model's token budget. In this formulation, both the model parameters θ and the knowledge index K remain fixed between periodic manual updates. Therefore, the system is unable to integrate feedback, accommodate shifts in the distribution of claims, or even correct misinterpretations during deployment. As a result of these limitations, the system exhibits inconsistent decision-making, unreliable decisions, and increased manual review.

3.2 Self-Improving Frontier Agents

The Self-Improving Frontier Agents framework defines claims adjudication as an agentic process characterized by a finite sequence of governed actions within an interaction episode:

$$\{a1, o1, a2, o2, \dots, an, R\}=\pi(f\theta; Q, C, P, T|G),$$

Where, π denotes a constrained policy that selects actions a_i from an action space A. This space includes internal reasoning steps, retrieval operations, deterministic tool invocations $t \in T$, and final response generation. Each action produces an observation o_i, such as tool outputs or retrieved evidence. At every step, policy execution is projected onto the governance set G, ensuring regulatory compliance, comprehensive audit logging, and enforcement of safety constraints.

3.3 Governed Adaptation and Knowledge Evolution

After a given interaction episode finishes, guided feedback, based on review by the subject matter experts, discrepancies in outcomes, or more automatic drift detectors, triggers a controlled adaptation process:

$$\theta_{t+1}=\theta t \oplus \Delta\theta(Update(F; G)).$$

This update is typically implemented using parameter-efficient techniques, such as low-rank adaptation or reward-model refinement, which preserve the integrity of the base frontier model while incorporating domain-specific improvements.
In parallel, the external knowledge base is incrementally updated according to:

$$Kt+1=Kt \cup Augment(Synthetic(F; f\theta), Reviewed(F; G)),$$

Where, synthetic examples are generated to cover counterfactual scenarios and rare edge cases, followed by mandatory review processes to ensure factual correctness and regulatory alignment.

4. SIFA Architecture

4.1 Overview Diagram

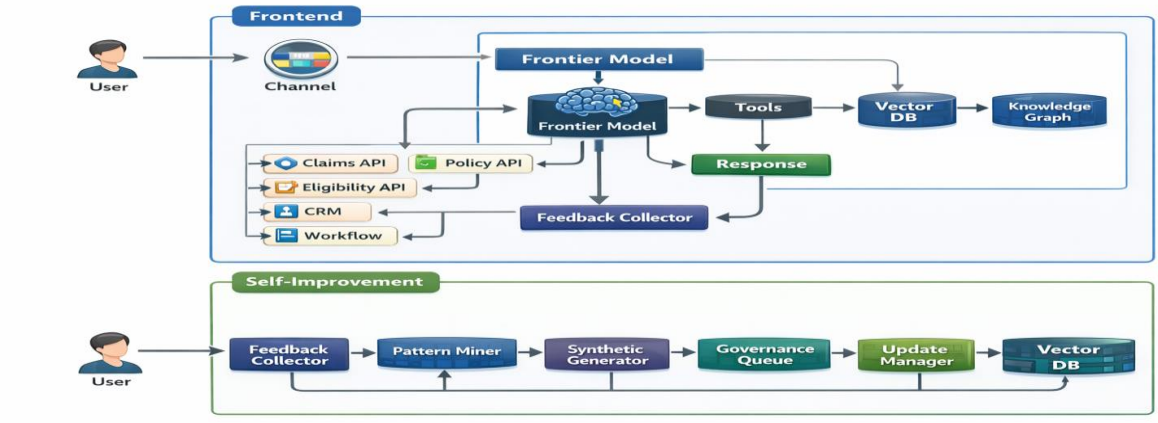

Fig. 6: End-to-end architecture for a governed, self-improving insurance AI. User queries traverse channels to a frontier model leveraging tools, retrieval, and APIs, producing responses while capturing feedback. A controlled improvement loop mines patterns, generates updates, enforces governance, and refreshes knowledge assets.

The general structure of the Self-Improving Frontier Agents (SIFA) system, as shown in Figure 6, is divided into two interdependent subsystems to facilitate both the real-time claims processing and the evolution of long-term performance. The upper part illustrates the operational frontend, where the customer or adjuster sends user requests through secure channels, which the central frontend model handles. The model provides iterative reasoning and is selective in calling a collection of deterministic tools, including the Claims API, Policy API, Eligibility API, CRM integration, and workflow orchestration, to execute actions. It also enriches its context by providing hybrid access to a vector database and enabling semantic search of embedded policy documents, regulatory guidelines, and past decisions, as well as structured queries across coverage hierarchies and jurisdictional relations using a knowledge graph. The resulting replies are sent back to the user, and the entire interaction logs and results are sent to a feedback collector for later analysis.

The bottom segment shows the self-improvement subsystem, an asynchronous system that takes control of operational learning and strictly governs it. A pattern miner is activated by feedback signals indicating repeated problems, e.g., systematic misinterpretations of clauses or eligibility problems. These patterns are then counterfactually generated using a synthetic generator and edge-case-generated to resolve them. Every suggestion is processed through a governance queue that requires subject-matter experts to conduct reviews and teams to ensure customers adhere to the suggestions. After acceptance, an update manager chooses which changes to apply in the vector database, which system-level policy prompts to refine, or which lightweight adapters to modify in the frontier model. This closed-loop process will enable SIFA to continuously improve its reasoning accuracy, knowledge relevance, and decision-making consistency, and to ensure full auditability and regulatory alignment throughout the process of adapting to changes.

4.2 Components

The SIFA architecture consists of several tightly coupled components that work together to provide governed, autonomous claims reasoning and engage in continuous self-improvement. These elements cut across the reasoning core, tool and knowledge layers, the feedback processing pipeline, and governance mechanisms to achieve high performance, traceability, and regulatory compliance in insurance operations.

4.2.1 Frontier Model Core

The core of the frontier model is the intelligence of SIFA, and is a state-of-the-art large language model via secure provider APIs (e.g., OpenAI o3 series, Anthropic Claude Opus 4.5, Google Gemini 3, or some combination of mixture-of-experts systems) by late 2025. This fundamental capitalises on the five key strengths to realise healthy performance on the sophisticated insurance adjudication work.

First, the step-by-step reasoning is helpful because it enables the model to break complex questions into simpler problems, trace the thinking path step by step, and reach conclusions necessary for long-term planning. Second, the function-call mechanism is essential because it enables the model to communicate deterministically with external systems, helping reduce hallucinations during activities such as claim processing and eligibility calculations.

Third, the model can critically monitor its own intermediate reasoning processes, update plans, and ensure consistency before committing to outputs or actions through self-evaluation using built-in reflection layers, thereby significantly improving decision reliability. Fourth, multimodal comprehension facilitates non-text claims artefact processing, i.e., photographs of damage, scanned medical reports, and PDF policy documents, which can be assessed holistically without loss of text.

> **Fifth, the thicker token windows of 200,000 or more tokens allow the maintenance of a broad context (including entire policy documents, a complete history of claims, archived correspondence, and recovered precedents) within a single inference event. To keep systems regulated, the core is executed under highly engineered system prompts that inject compliance constraints, obligatory reflection protocols, and decision limits, and optional domain-specific partiality through lightweight adapters that retain base model safety properties.**

4.2.2 Tools Layer

The tools level is a governed interface between the frontier model and enterprise systems, offering a limited set of safe operating primitives that assure deterministic computing, full auditability, and proper semantic compliance with domain-specific constraints. Each tool is implemented as a schema-enforced function with carefully defined inputs, outputs, and limits on side effects, thereby ensuring that there is no unlimited access to underlying infrastructure resources and facilitating the model's ability to perform the necessary claims processing.

The defined tools include the following:

get_claim: Retrieves comprehensive structured JSON for a specified claim, including incident details, parties involved, loss description, and supporting documents, with access restricted by claim ID and user authorization.

get_policy: Fetches the active policy document and relevant coverage clauses for an insured entity, returning parsed sections, endorsements, and exclusions in a standardized format to support accurate interpretation.

calculate_eligibility: Executes rule-based computation of covered benefits, deductibles, limits, and exclusions according to policy terms and jurisdictional regulations, yielding deterministic monetary and coverage outcomes.

create_case_note: Appends an audited, timestamped note to the claim record within the CRM system, capturing the

model's reasoning trace and decision rationale for regulatory review.

send_communication: Dispatches approved emails, SMS messages, or scripted notifications to policyholders or internal stakeholders, using pre-validated templates to maintain consistency and compliance.

Through the enforcement of determinism via validation at the backend level, logging of calls for auditing purposes, and addition of domain constraints such as privacy considerations, approval processes, and exclusion of unspecified actions at the external agent level, the tools level ensures that the frontier model is capable of performing claim reasoning from start to finish in a controlled environment even as it guards against the hazards that come with unspecified agentic activity.

4.2.3 Knowledge Layer

The knowledge layer functions as a hybrid query system that provides the frontier model with accurate, domain-driven context. This layer combines dense vector search for semantic similarity with graph queries that focus on the explicit relationships in the insurance data (Sarmah et al., 2024; Subramanian, 2024). The main elements include vector search over policy sections, where policy documents are broken down into embeddings to retrieve applicable clauses during the coverage assessment process efficiently. A claims precedents database stores anonymised adjudication histories with explanations for similarity assessment through analogy (Needham et al., 2025).

The regulatory embedding entail jurisdiction-based norms and compliance standards of decision-making. These are archived as searchable vectors to ensure decisions comply with the legal provisions. The adjuster feedback summaries are created based on organized error corrections and human feedback to spot areas of interpretation. At the same time, the decision patterns learned from recurrent decisions mined are used to develop embeddable templates for proactive decision-making. This integrated setup combines the advantages of vector databases for semantic search with the knowledge graph for traversal, resulting in improved decision-making accuracy in the complex insurance domain (Gao et al., 2023; Hedden, 2024).

4.2.4 Self-Improvement Engine

The self-improvement mechanism is the adaptive backbone of SIFA. It enables the system to improve its performance through an administered pipeline that processes operational feedback while enforcing safety and regulatory constraints. The self-improvement mechanism comprises four major modules and a controlled update phase. These modules have a Feedback Collector that integrates structured data from different sources, including notifications of corrections to the human adjuster input, downstream overrides, escalation flags, and automated consistency checks on traces of reasoning. The Pattern Miner applies feedback embedding clustering algorithms and text entailment analysis to identify familiar sources of error (including systematic misinterpretation of policy clauses and incorrect computation outputs in assessments) (Shi et al., 2025). The final module uses the frontier model with constrained inputs to develop a rich set of counterfactual instances that specifically target the error sources detected by the Patterns Miner and improve the distribution during training without revealing the production instances (Li et al., 2024). There is also a Drift Detector that tracks important indicators, such as the accuracy and error rate of explanations, as well as the escalation rate.

The final stage involves constrained fine-tuning/adapter training, using low-rank adaptation (LoRA) with lighter adapter layers rather than training the entire model. Integration for domain-independent advancement is made easier while retaining the primary security alignments (Hu et al., 2022). The set of all created data, as well as the entire flock of model modifications, must undergo mandatory human verification for conformity before release. The model changes are tracked via versioning and rollback functionality, which satisfy the critical oversight requirement in the insurance industry.

4.2.5 Governance & Compliance Layer

There are enforcement and governance tiers across the entire SIFA infrastructure that impose strict conditions for safe, transparent, and regulatory-compliant operations in the insurance environment. It captures history, logs all calls to tools, captures inviolable timestamps, user names, and version data, and submits it to a regulator for forensic analysis and reporting (Bhattacharya et al., 2025). It is backed by a well-structured track and a mandatory reflection report that describes the decision-making process in an easy-to-understand format.

The information on knowledge bases, prompts, and adapters is stored in version control systems so that knowledge can be reverted to or rewritten at any point. There are a few gateway updates to multi-level authorisation before deployment, for which approvals must be obtained from a domain authority and a compliance officer. Safety reviews of adapted components, in turn, involve comparisons with red-team and hallucinated attacks, as well as bias metrics, to prevent any degradation of base model alignments (Weidinger et al., 2021). Constraints on regulatory alignments include complex rule systems (such as HIPAA privacy filters or data minimization under GDPR) that are directly encoded in system prompts and tool schematics, with enforcement at runtime.

METHODS

5.1 Retrieval-Augmented Frontier Agent

The core inference procedure in SIFA integrates retrieval-augmented generation with agentic tool-use under governed constraints. The following pseudocode captures the primary execution flow for handling a user query in claims adjudication:

```
def sifa_inference(user_query):
    context = vector_db.search(user_query, top_k=8)
    system_prompt = load_policy_constraints()
    messages = [
        {"role": "system", "content": system_prompt},
```

```
    {"role": "user", "content": user_query},
    {"role": "assistant", "content": context}
]
response = frontier_model.invoke(messages, tools=TOOL_SCHEMA)
if response.tool_calls:
    tool_results = execute_tools(response.tool_calls)
    final = frontier_model.invoke(messages + [response] + tool_results)
    return final.content
return response.content
```

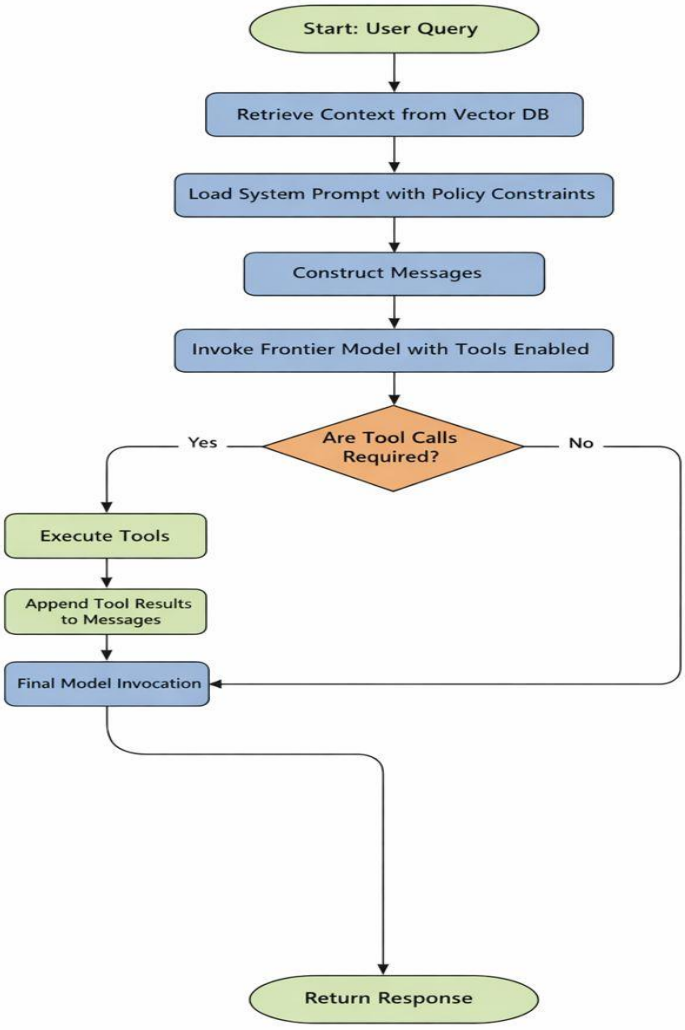

Fig. 7: Flowchart of the SIFA inference procedure showing retrieval augmentation and conditional tool execution

5.2 Feedback Capture

Feedback capture is considered an essential module of the SIFA framework, aimed at distilling the signals extracted from interactions in the production process and directing them to the self-improvement module. To ensure consistency, traceability, and ease of post-processing, each instance of feedback must be formulated in accordance with a predetermined JSON schema that links corrections to the appropriate claims and reasoning episodes.

The canonical feedback structure is as follows:

```
{
  "claim_id": "78421",
  "reasoning_trace_id": "rt_2231",
  "issue_type": "policy_misinterpretation",
  "human_correction": "The correct clause is 3.2(c) not 4.1(b)",
  "severity": "high",
  "timestamp": "2025-12-07T14:23:11Z"
}
```

5.3 Pattern Mining

The SIFA pattern mining process identifies and yields error typologies from collected feedback data through rigorous reasoning. The process begins by converting feedback examples into dense representations using the frontier model's encoder. This enables the acquisition of the semantic characteristics of feedback errors and associated claims. The resulting dense representation is then grouped using HDBSCAN and related techniques to identify clusters of similar mistakes. These clusters are then analysed using text entailment models to determine the root causes of the identified errors. The process yields high-precision misinterpretation patterns.

The concise procedural outline is:

```
clusters = cluster_feedback_embeddings(feedback_items)
misinterpretation_patterns = detect_common_failure_modes(clusters)
```

These patterns are then overlaid with metadata such as frequency, average severity, business impact estimates (e.g., escalation cost estimates), and examples. Mining this data is crucial to ensuring that the data-adaptation techniques used afterwards focus on the most severe problems and translate unstructured data corrections provided by claims adjusters into actionable intelligence that remains fully traceable back to the original feedback data.

5.4 Synthetic Data Generation

SIFA utilises the frontier model to generate synthetic data, creating high-quality training examples that target the identified error and patterns, improving the dataset without requiring new data that could raise privacy concerns. The process takes a structured reasoning error as input, generated from pattern mining. It triggers the model to provide repaired reasoning, along with various versions intended to test the stated failure mode. This self-boosting mechanism is vital, as it enables synthetic examples to share robustness in reasoning and to focus on areas of concern.

The core procedure is outlined in the following pseudocode

```
def generate_synthetic_example(reasoning_error):
  messages = [
    {"role": "system", "content": "Correct the reasoning and generate edge-case variants."},
    {"role": "user", "content": json.dumps(reasoning_error)}
  ]
  result = frontier_model.invoke(messages)
  return parse_synthetic_cases(result)
```

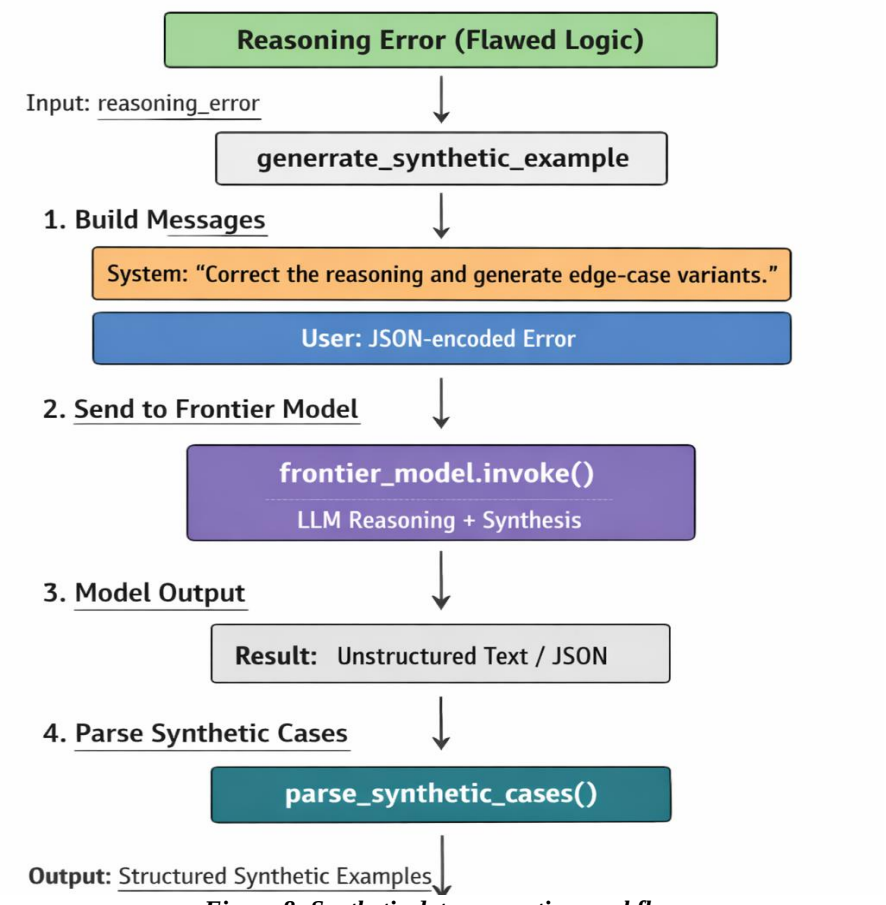

Figure 8: Synthetic data generation workflow

The outputs include counterfactual hypotheses that stress boundary conditions by manipulating crucial facts, adversarial policy language that uses ambiguous or tortuous wording for testing the robustness of interpretations, ambiguity stress tests that include

different believable interpretations of clauses/evidential statements, parsed examples that are tested for format consistency, and empirical validation against real policy language. These examples are tested for format consistency to ensure diverse, high-quality exemplars that can be adapted according to governing principles.

5.5 Governed Updates

Governed Updates enable a safe, compliant adaptation of various components in SIFA through a procedure that imposes strict change constraints. To begin with, these synthetic scenarios undergo a review process by expert reviewers for matters such as accuracy, followed by a governance board consisting of compliance, legal, and risk representatives for regulatory approval. Changes that obtain this approval are implemented.

The workflow is:

synthetic_cases → SME Review → Governance Board → Approval → Adapter fine-tuning + KB updates → Deployment with versioning

Parameter updates are limited to safe, reversible methods: low-rank adapters (LoRA) for domain specialization, selective reward model refinements, targeted policy prompt revisions, and vetted RAG corpus expansions. Full retraining of the frontier model is explicitly prohibited to preserve base safety properties and avoid drift.

EXPERIMENTS & EVALUATION

SIFA was evaluated in a pilot on property and casualty claims, comparing a static RAG baseline against iterative self-improving versions. Metrics cut across accuracy, safety, governance, business impact, and efficiency, which were authenticated under SME review and logs.

Metrics:

- **Accuracy** — Explanation Correctness (EC): Agreement with SME judgments on rationales.
- **Safety** — Hallucination Rate (HR): Proportion of fabricated policy details.
- **Governance** — Audit Completeness Score (ACS): Coverage of required trace elements.
- **Business** — Escalation Rate (ER): Claims needing human override.
- **Efficiency** — Average Handling Time (AHT) Reduction: Processing time decrease.

Table 1: Pilot Performance Comparison

Metric	Baseline	SIFA v1	SIFA v2
EC	78%	88%	94%
HR	12%	5%	2%
AHT	18 min	12 min	9 min
ER	22%	16%	11%
ACS	40%	89%	97%

ETHICAL & REGULATORY CONSIDERATIONS

SIFA implements a broad range of tasks in ethical risk management and regulatory requirements across the insurance process. Bias reduction is achieved through the creation of diverse synthetic data samples, the scheduling of fairness checks on predefined attributes of claims (such as age, gender, and geographic distribution), and validation of post-adaptation processing to avoid the reinforcement of any previously existing inequity. On the other hand, explainability is ensured by requiring the inclusion of traces of reasoning in all inferences, tool usage, and data retrieval points.

The privacy and compliance measures are compatible with PIPEDA, HIPAA, and the GDPR. The controlled updates are designed to counteract model drift through parameter-efficient adaptation and drift-detection thresholds. Human review gates are essential in both adaptation steps. Subject-matter expert (SME) review is required at every stage of the adaptation process to ensure that the SME reviews claims. All controls ensure that trust and equity are retained during automated claims adjudication.

DISCUSSION

The SIFA architecture empirically validates the deployment of adaptive agentic systems powered by frontier large language models within constrained regulatory domains. By integrating governed self-improvement loops with parameter-efficient adaptation and hybrid retrieval, SIFA establishes technical feasibility for long-horizon autonomous reasoning in environments requiring deterministic auditability and compliance enforcement. The paradigm of constrained update, applied only to low-rank update adapters, refinement reward model update, and strategically selected augmentations to the RAG corpus, shows that safe and effective lifelong learning occurs without degrading the base model's alignments on safe systems or causing catastrophic resets.

This addresses some shortcomings long noted in agentic approaches to AI that have uncontrolled emergence. The pilot test results (Table 1) show that significant improvements were made to the explanations with regards to correctness (from 78% to 94% on SIFA v2), hallucination rate (from 12% to 2%), average handling time (from 18 to 9 minutes), escalation rate (from 22% to 11%), and audit completeness (from 40% to 97%). This continuous series of improvements validates the usefulness of feedback-assisted pattern mining and synthetic augmentation in countering error-prone patterns in explanations, as well as improving multi-step reasoning and tool invocation accuracy.

In general, SIFA provides a reference architecture for migrating regulated adjudication processes from rule-based or retrieval-

augmented baselines to self-refining agent systems. The proposed technological shift would enable the scalable integration of operational feedback, even within a tight governance framework, thereby facilitating learning intelligence in regulated industries.

LIMITATIONS

SIFA inherits the challenges of frontier models and regulated CL, as well as those of basis models, regarding persistent opacity in representation and learned knowledge. The effectiveness of self-improvement requires continuous access to high-quality expert judgment, which can lead to latency and resource costs that impede adaptation speed for niche lines of business. Synthetically enhanced data augmentation requires intensive human curation and validation to avoid amplifying bias or introducing artefacts of error. Multimodal processing involving video or medical imaging requires association with frontier models with vision capabilities, resulting in increased resource consumption and the need for more robust perceptual reliability frameworks. Regulatory frameworks preclude fully automated material decisioning, requiring human confirmation for high-stakes coverage outcomes, thereby limiting the breadth of end-to-end automated processes.

FUTURE WORK

Future work on SIFA would focus on the study of multi-agent ecosystems and the design of specialised sub-agents within a concurrent fraud-detection, liability allocation, and subrogation analysis setting to enable collaborative reasoning over multi-party claims. Cross-policy generalisation would be explored through meta-learning and domain-invariant adaptation to generalise transfer knowledge across different insurance products while maintaining product-specific regulatory requirements. Incorporating the frontiers of multimodal learning would enable real-time reasoning over accident video streams to support dynamic scene reconstruction and causality analysis. Full-book policy context learning using million-token windows would provide a global interpretation of contracts, going beyond the restrictions of chunk retrieval. Finally, high-fidelity claim simulation engines using controllable generative learning would facilitate automated red teaming, counterfactual reasoning, and proactive safety analysis before governable deployment.

CONCLUSION

SIFA introduces the first governed architecture for self-improving frontier agents in insurance adjudication. The pilot outcome show significant improvements in accuracy, consistency, and efficiency, and provide a foundation for scaling up and flexible AI in controlled companies.

REFERENCES

1. Agrawal, P., Alberti, C., Huot, F., Maynez, J., Ma, J., Ruder, S., Ganchev, K., Das, D., & Lapata, M. (2023). QAmeleon: Multilingual QA with only 5 examples. *Transactions of the Association for Computational Linguistics*, *11*, 1754–1771.
2. Anthropic. (2025a). *Claude models*. https://www.anthropic.com/claude
3. Anthropic. (2025b). Introducing computer use, a new Claude 3.5 Sonnet. https://www.anthropic.com/news/3-5-models-and-computer-use
4. Artificial Analysis. (2025). *AI model quality rankings*. https://artificialanalysis.ai/
5. Bell, J., Quarantiello, L., Coleman, E. N., Li, L., Li, M., Madeddu, M., Piccoli, E., & Lomonaco, V. (2025). The future of continual learning in the era of foundation models: Three key directions. *arXiv preprint arXiv:2506.03320*. https://arxiv.org/abs/2506.03320
6. Bhattacharya, S., Castignani, G., Masello, L., & Sheehan, B. (2025). AI revolution in insurance: Bridging research and reality. *Frontiers in Artificial Intelligence, 8*, Article 1568266. https://doi.org/10.3389/frai.2025.1568266
7. CAQH. (2024a). *2023 CAQH Index report: A new normal: How trends from the pandemic are shaping healthcare administration*. https://www.caqh.org/hubfs/43908627/drupal/2024-01/2023_CAQH_Index_Report.pdf
8. CAQH. (2024b). *2024 CAQH Index report: From transactions to trust: Building better healthcare administration*. https://www.caqh.org/hubfs/Index/2024%20Index%20Report/CAQH_IndexReport_2024_FINAL.pdf
9. Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems, 30*.
10. Deloitte Center for Financial Services. (2024). *Scaling gen AI in insurance*. https://www.deloitte.com/us/en/insights/industry/financial-services/scaling-gen-ai-insurance.html
11. Eling, M., Nuessle, D., & Staubli, J. (2022). The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *The Geneva Papers on Risk and Insurance - Issues and Practice, 47*(2), 205–241. https://doi.org/10.1057/s41288-020-00201-7
12. Experian Health. (2025). *State of claims 2025: The denial problem (and is AI the answer?)*. https://www.experian.com/blogs/healthcare/state-of-claims-2025/
13. Fedus, W., Zoph, B., & Shazeer, N. (2021). Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *arXiv preprint arXiv:2101.03961*. https://arxiv.org/abs/2101.03961
14. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Guo, Q., Wang, M., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*. https://arxiv.org/abs/2312.10997
15. Google DeepMind. (2025). *Gemini models*. https://deepmind.google/technologies/gemini/
16. Gunaseelan, N., Paul, D., & Soundarapandiyan, R. (2024). Deploying LLMs for insurance underwriting and claims processing: A comprehensive guide to training, model validation, and regulatory compliance. *Australian Journal of Machine Learning Research & Applications, 4*(1), 226–263.

17. Hedden, S. (2024). How to implement Graph RAG using knowledge graphs and vector databases. *Towards Data Science*. https://towardsdatascience.com/how-to-implement-graph-rag-using-knowledge-graphs-and-vector-databases-60bb69a22759
18. Hoagland, A., Yu, O., & Horný, M. (2024). Social determinants of health and insurance claim denials for preventive care. *JAMA Network Open, 7*(9), Article e2433316. https://doi.org/10.1001/jamanetworkopen.2024.33316
19. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-rank adaptation of large language models. *International Conference on Learning Representations*. https://openreview.net/forum?id=nZeVKeeFYf9
20. Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown Spark.
21. Ke, Z., Shao, Y., Lin, H., Xu, H., Shu, L., & Liu, B. (2023). Continual pre-training of language models. *arXiv preprint arXiv:2302.03241*. https://arxiv.org/abs/2302.03241
22. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences, 114*(13), 3521–3526.
23. Levine, S., Kumar, A., Tucker, G., & Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*. https://arxiv.org/abs/2005.01643
24. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.
25. Li, D., Jin, Z., Qian, L., & Yang, H. (2025). Textual analysis of insurance claims with large language models. *Journal of Risk and Insurance, 92*(2), 505–535. https://doi.org/10.1111/jori.70004
26. Li, Y., Zhang, J., Patra, A., & Zhang, Z. (2024). Synthetic data generation for LLM agents: Techniques and applications. *arXiv preprint arXiv:2409.16821*. https://arxiv.org/abs/2409.16821
27. McKinsey & Company. (2025). Deploying agentic AI with safety and security: A playbook for technology leaders. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders
28. Namperumal, G., Paul, D., & Soundarapandiyan, R. (2024). Deploying LLMs for insurance underwriting and claims processing: A comprehensive guide to training, model validation, and regulatory compliance. *Australian Journal of Machine Learning Research & Applications, 1*(1), 1–15.
29. Needham, J., Napoli, L., & Múgica, A. (2025). Retrieval augmented generation for claim processing: Combining MongoDB Atlas Vector Search and large language models. *MongoDB Blog*. https://www.mongodb.com/blog/post/rag-claim-processing-combing-mongodb-atlas-vector-search-llms
30. NVIDIA. (2025). Mixture of experts powers the most intelligent frontier AI models. https://blogs.nvidia.com/blog/mixture-of-experts-frontier-models/
31. OpenAI. (2025). *o-series models*. https://openai.com/o-series
32. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*.
33. Phuong, M., Zimmermann, R. S., Wang, Z., Lindner, D., Krakovna, V., Cogan, S., Dafoe, A., Ho, L., & Shah, R. (2025). Evaluating frontier models for stealth and situational awareness. *arXiv preprint arXiv:2505.01420*. https://arxiv.org/abs/2505.01420
34. Premier Inc. (2025). Claims adjudication costs providers $25.7 billion - $18 billion is potentially unnecessary expense. https://premierinc.com/newsroom/policy/claims-adjudication-costs-providers-257-billion-18-billion-is-potentially-unnecessary-expense
35. Sarmah, B., Hall, B., Rao, R., Patel, S., Pasquali, S., & Mehta, D. (2024). HybridRAG: Integrating knowledge graphs and vector retrieval augmented generation for efficient information extraction. *arXiv preprint arXiv:2408.04948*. https://arxiv.org/abs/2408.04948
36. Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges. *arXiv preprint arXiv:2505.10468*. https://arxiv.org/abs/2505.10468
37. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*. https://arxiv.org/abs/1701.06538
38. Shi, H., Xu, Z., Wang, H., Qin, W., Wang, W., Wang, Y., Chen, X., Zhang, C., Li, M., & Wang, H. (2025). Continual learning of large language models: A comprehensive survey. *ACM Computing Surveys, 58*(5), 1–42.
39. Shinn, N., Cassano, F., Berman, E., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*.
40. Subramanian, T. (2024). Revolutionizing risk assessment: Retrieval augmentation generation for insurance underwritings using both VectorDB and KnowledgeGraph. *Medium*. https://tamilselvan-subramanian.medium.com/revolutionizing-risk-assessment-retrieval-augmentation-generation-for-insurance-underwritings-5be023c00f44
41. Toran, B. (2023). *AutoGPT* [Computer software]. GitHub. https://github.com/Significant-Gravitas/AutoGPT
42. Valence Howden. (2025). Agentic AI compliance and regulation: What to know. *TechTarget*. https://www.techtarget.com/searchenterpriseai/feature/Agentic-AI-compliance-and-regulation-What-to-know

43. Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., & Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*. https://arxiv.org/abs/2305.16291
44. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P., Cheng, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., Isaac, W., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*. https://arxiv.org/abs/2112.04359
45. Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Burger, D., & Wang, C. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*. https://arxiv.org/abs/2308.08155
46. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. *International Conference on Learning Representations*.
47. Zhou, S., Xu, F. F., Zhu, H., Zhou, X., Lo, R., Sridhar, A., Cheng, X., Ou, T., Bisk, Y., Fried, D., Alon, U., & Neubig, G. (2023). WebArena: A realistic web environment for building autonomous agents. *arXiv preprint arXiv:2307.13854*. https://arxiv.org/abs/2307.13854
48. Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., & Irving, G. (2019). Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*. https://arxiv.org/abs/1909.08593