

Multi-Agentic Automation for Evaluating Property Claims in Underwriting

Muhammad Imran Sajid

Al Rajhi Takaful, Riyadh, Saudi Arabia

Email: engrm.imran050@gmail.com

How to cite this paper: Sajid, M.I. (2025) Multi-Agentic Automation for Evaluating Property Claims in Underwriting. *Open Journal of Applied Sciences*, 15, 819-833. <https://doi.org/10.4236/ojapps.2025.154055>

Received: March 12, 2025

Accepted: March 28, 2025

Published: March 31, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Insurance underwriting for property damage demands extensive human labor efforts while consuming large amounts of time and presents challenges from document variety and complexity. This paper develops a Multi-Agentic Automation Framework by using Multi-Agent Systems with Large Language Models to improve property claim underwriting. Each individual task within the underwriting procedure receives autonomous AI processing from self-modular agents working in coordinated teams to process documents for risk assessment insights generation. Furthermore, the framework offers three main components which are its modular design along with its task split and coordination features and its use of LLMs for context-driven document interpretation. The outcomes of our experiments prove that with this framework we can achieve an accuracy of up to 92.9% while getting responses in under a minute. This comes with consistency and proning to deviation of about 94% while handling complex and edge-cased scenarios. We test this in a commercial and production-ready setting using Streamlit and CrewAI alongside Pydantic. This shows the practical usage of the framework to enhance scalability and underwriting workflow efficiency and cost reduction. The insurance industry is envisioned to benefit from LLM-driven multi-agent systems creating opportunities for quicker and better underwriting processes that build lasting and reliable insurance solutions.

Keywords

Multi-Agentic Automation, Property Claim Underwriting, Large Language Models (LLMs), Insurance, AI Agents, Document Analysis, Risk Assessment

1. Introduction

In the insurance industry, due to the rapid advancements in Artificial Intelligence;

especially in Large Language Models and Agentic AI, we are witnessing a transformative shift that is driving automation of key processes. One of the most critical of these is the sector of underwriting. Underwriting involves the evaluation of risks that are associated with insurance claims such that the expert evaluator is able to determine the coverage and premiums. Until now, in what is known to be “conventional”, the underwriting industry has largely relied on manual processes undertaken by human experts; who engaged in analyzing complex documentations, making assessments of inherent risks and making informed decisions. This manual process, however, is very time-consuming and is always prone to inconsistencies due to the high overload of documents, and the volume at which the claims can get complex, adding to the possibility of human errors.

In our particular discussion for this paper, the property damage claims are uniquely challenging. There is a higher diversity of data and documents, as compared with other insurance categories. For property damage, documents like insurance policies, forensic reports, photographic evidence and video analyses are a few to mention, out of the many. Each of these documents requires a specialized field of study and knowledge to fully understand the inherent information. This is where underwriting of claims for property damage becomes severely labor-intensive and error-prone. As there is now a growing demand for quicker and more performance friendly underwriting, we require an up-to-speed solution that can automate parts of this process.

The emergence and the wave of Large Language Models (LLMs), from the ChatGPT initial launch in November 2022, has sparked remarkable growth and has achieved significant milestones in AI automations. This wave has enabled machines to not just understand but also generate and reason with text that is completely human-like. But the potential does not solely lie in language being generated, it lies in the opportunity to make actionable decisions that are based on the input understanding. This is where we are seeing a new buzz, that started in 2024, AI Agents!

AI Agents are autonomous systems that can perform tasks, as specific as per the business logic, and collaborate with other specialized agents to make decision-making possible in dynamic environments. AI Agents are a shift from stand-alone, traditional AI models operating in silos. AI Agents are designed to be oriented towards their goals, they are meant to be context-aware and also interactive (where needed) making them the ideal solution for complex and multi-step processes like that of underwriting.

In the 2020s era, where LLMs have become prevalent. AI Agents are increasingly deployed for solving real-world problems that are in need of a blend putting together cognitive reasoning and execution capabilities. For instance, in the healthcare industry, the AI agents assist in diagnosis of diseases through analysis of medical records and recommendation of treatments. In the financial realm, the AI agents can automate strategies for trading through the interpretation of market trends. Similarly in our domain, insurance and underwriting, the AI Agents are

set to make underwriting a quicker, more efficient and trustable process.

In our work, we introduce a Multi-Agent Automation Framework that is designed for addressing the challenges of property claim underwriting by using Multi-Agent Systems (MAS) built on top of Large Language Models (LLMs). Our proposed system is able to decompose the underwriting process into distinct, specialized and modular tasks that are each handled by autonomous AI agents. These agents are tasked to collaborate for claim-related document analysis, extraction of insights and generation of detailed assessment reports listing out risks. Through our modular and agentic approach, the system not just enhances overall performance efficiency but also leads to scalability when adapting to diverse underwriting scenarios.

In our work, some of the key contributions are:

1) Novel Multi-Agent Architecture: We develop a modular system with specialized agents for analysis of specific types of claim documents, such as insurance policies, or forensic evidence.

2) Task Decomposition and Coordination: The underwriting process is divided into multiple defined tasks. Each of these tasks have clear objectives and output formats. Our central mechanism for coordination ensures that agents work together seamlessly.

3) Integration of Large Language Models for Document Analysis: Our system uses state-of-the-art Large Language Models to process and read through unstructured textual data, and documents, for enabling a context aware and memory-aware analysis.

4) Real-World Applicability: We build our framework by implementing it using an up-to-date and production-ready technology stack that includes Streamlit for user interaction, and CrewAI for orchestration of AI Agents. Pydantic is used for validation of data for ensuring accuracy and reliability for real-world scenarios of underwriting.

2. Background & Related Work

2.1. Challenges in Property Claim Underwriting

The property claim underwriting is a very intricate process of the insurance sector. It requires synthesis of diverse sources of data that assess risk accurately. It is not like health insurance or life insurance. Property claims have a wide range of document types that require sufficient specialized expertise. Some examples are:

- **Insurance Policies:** These documents contain dense jargon in the legal space, including exclusions and conditional clauses. Even a slight oversight in interpretation of these terms can lead to huge financial losses. For instance, in 2021, a major insurance company faced a huge lawsuit of about \$10 million due to the misinterpretation of a flood damage exclusion clause in the policy of the homeowner [1] [2].
- **Forensic Reports:** Such technical documents assess the overall cause and extent of the damage including structural failure or origins of fire. Analysis of

these forensic reports requires expertise in the fields like engineering or even environmental science. In a 2020 case study, the authors highlighted how a misinterpretation of a single forensic report led to a huge mistake with a wrong payout of several million dollars for a commercial property claim [3].

- **Photographic and Video Evidence:** Visual data should be analyzed in a meticulous manner, this is done to verify the damage claimed. Disputes can occur due to the inconsistencies in photo timestamps or missing some angles. In a 2022 report by Deloitte, 25% property claim disputes were linked to inadequate visual evidence [4].
- **Financial Records:** Repair estimates along with invoices and loss estimates should be cross-referenced for ensuring consistency. In such documents, discrepancy is a common fraud source. According to the Coalition Against Insurance Fraud, about 10% of property claims have some sort of financial misrepresentation [5].

Manual processing of these documents is not just labor-intensive but it is also prone to any human error. In Deloitte's 2022 case study it was revealed that over 30% of underwriting delays in property claims were caused by inconsistent document analysis. Moreover, the increased volume of claims, driven by climate change and urbanization, have caused scalability issues. One insurance company reports 20% year-on-year growth in property claims volume due to extreme weather events between 2018 and 2023.

2.2. The Evolution of Artificial Intelligence Models in Insurance Underwriting

Throughout its existence the insurance sector has used technology for maximizing operational efficiency and accuracy improvement. Every innovation in time since the 19th century actuarial tables has added new functionalities through to the predictive analytics of the 21st century. The adoption of AI in underwriting processes started recently since machine learning (ML) and natural language processing (NLP) technology has become more advanced.

AI systems were introduced in underwriting in its early stages, where it used only rule-driven Machine Learning for developing automated processes that detected incomplete applications and also processed premiums by algorithmic ways. These systems made operations more efficient but they could not process complex unstructured data in a sufficient manner. The McKinsey report from 2018 showcased how a system using rules processed insufficiently with regional storm risk information which resulted in incorrect premium estimations for coastal properties [6].

Insurers achieved unprecedented unstructured text data processing accuracy due to the introduction of large language models GPT and BERT. Lemonade, the leading insurtech company, employs LLMs to determine suspicious activity within claims descriptions. Through their AI system Lemonade succeeded in decreasing fraudulent payouts by 15% during 2023. Though advanced models exist today

they still face problems dealing with complex decisions while processing documents from various sources.

2.3. The Rise of Multi-Agent Systems for Complex Decision-Making

The approach of MAS serves as a robust solution for complex distributed problems. Multiple autonomous agents within MAS make up this distributed system design which enables agents to work together through negotiation and adaptation when operating in changing environments. Such systems show excellent capability for handling underwriting activities because they can handle specialized task requirements together with coordination needs.

MAS applications launched in supply chain management help optimize logistics through establishment of agents who handle inventory management and transportation execution and demand estimation functions. A 2021 research by Wooldridge *et al.* established that a MAS system operated in global retail production reduced both delivery time by 20% and spending by 15%. MAS applications have been used in healthcare facilities through agents which play essential roles in clinical diagnosis and treatment planning and monitoring. The application of MAS methods at a hospital facility led Jennings and Bussmann to record a 25% improvement in patient outcomes through their 2020 case study.

In the context of insurance underwriting, MAS offers several advantages:

- The individual agents should specialize in distinct document categories which ensure maximum accuracy together with superior efficiency. This specialized agent can detect arson signals which would escape basic automation because of its expertise in forensic report analysis.
- Agent systems possess the capability to exchange related information and assessment details in unison for producing complete claim interpretations. A cross-check between the analysis of photographic evidence by an agent with the data in the forensic report provides validation of damage extent.
- MAS scalability emerges from its modular structure which enables flexible handling of bigger amounts of claims alongside diverse document types.

2.4. Gaps in Existing Solutions

Numerous impediments continue to affect advancements in underwriting systems that use artificial intelligence.

Most current systems handle exclusively one specific data format at a time such as text or images so they cannot perform a well-rounded evaluation. Data collection from forensic reports to policy terms through manual processes continues to affect insurer efficiency according to a McKinsey study from 2023 because 40% of providers maintain this approach.

The processing strength of LLMs regarding single documents does not extend to relating insights between multiple documents. The research published by Lee and Kim in 2022 demonstrates that when LLMs try to link payment records to

photos they fail 30% of the time which produces wrong reimbursement results.

Rule-based systems together with various diverse AI models can lead to challenges in adapting to new claim types and regulatory changes because they are not very flexible. Insurers told PwC in a 2021 survey that inflexibility stood as their primary challenge to AI adoption in underwriting operations.

2.5. Related Work

The literature containing research about AI in underwriting underlines the specific difficulties within property claims assessment without sufficient detail.

For example:

- Researcher-developed NLP models collect essential policy terms from insurance contact yet they inadequately process policy exclusions and conditional factors. An NLP model processed a flood damage exclusion clause wrong according to Smith's 2024 study and triggered a \$5 million payment error [7].
- Image examination for damage evaluation through computer vision offers insight from photo data yet does not provide the context which forensic reports and other documentation would deliver.
- ML Algorithms identify fraudulent claims yet operate in a more reactive (instead of proactive) fashion for detecting fraud. The underwriting stage fraud detection models succeed in recognizing only 60% of the fraudulent claims yet the remaining 40% are found after payout occurs.

The proposed multi-agentic architecture utilizes modern advancements but solves existing weaknesses. The system addresses property claim underwriting through specialized document agents that enable smooth collaboration to supply an enhanced accurate solution.

3. Proposed Multi-Agent System Framework

3.1. System Architecture Overview

The designed framework uses a Multi-Agent System structure to build a system for automatically handling property damage claims underwriting operations. The framework organizes its basis as individual operational components that focus on distinct underwriting operations. The system integrates specialized AI agents with central coordination through a user-friendly interface for completing its operations smoothly. Key components make up the architecture alongside each additional component designed to perform a distinct aspect in the workflow.

- Streamlit creates the **User Interface (UI)** which provides underwriters with a web platform to upload claims as well as view analysis updates during real-time sessions.
- The **Document Preprocessing Module** checks and retrieves vital content from uploaded files to protect data quality before analytical procedures.
- The **Multi-Agent System (MAS)** constitutes the main operational component through agents which specialize in analyzing claim documentation. The task coordination engine controls task execution as it enables agents to work to-

gether effectively for consistent results.

- **Output Aggregation Module** is used to synthesize the results from individual agents into a comprehensive risk assessment report.

The system architecture implements modular elements which enable flexible capability expansion and allows the processing of different claim types along with the support of evolving underwriting algorithms.

3.2. Agent Design and Specialization

The MAS incorporates multiple independent agents which carry out individual functions needed for underwriting processes. These agents use domain-specific knowledge as well as context-based capabilities which allow them to review claim documents effectively. This agent serves as an Insurance Policy Coverage Expert that correctly understands insurance policies to discover boundaries together with excluded elements and policy conditions. Within insurance coverage assessments the Forensic Report Analyst agent makes technical forensic report evaluations to establish damage origins as well as severity levels while the Photographic Evidence Interpreter agent analyzes visual materials to confirm damage occurrences. These agents use NLP abilities to scan text information through GPT along with BERT which represents advanced language modeling technology. The agents processing visual data through photos or videos operate computer vision algorithm systems which detect and categorize damage. The agents include capabilities for teamwork and insight exchange which permits complete assessments of claims. The Forensic Report Analyst agent distributes assessment conclusions to the Photographic Evidence Interpreter agent to confirm that both types of evidence match appropriately. Specialized distribution across team members combined with teamwork enables the system to manage various types of property claims within complex situations.

3.3. Task Definition and Coordination

Underwriting operations are divided into specific sequences of tasks which different agents receive for execution. These tasks have been designed to perform atomic functions that operate within context-aware environments and maintain interdependence with each other to ensure the creation of accurate system results. A Task class defines the responsibilities of each assignment which contains essential information about objectives, step-by-step instructions, output specifications as well as dependencies on other tasks. When examining insurance policy documents the Insurance Policy Coverage Expert agent processes the policy to extract coverage terms along with their meaning through a JSON object output. The Task Coordination Engine uses directed acyclic graphs (DAGs) to control task execution as it manages the procession of tasks along with dependency scheduling. The Photographic Evidence Interpreter agent needs the forensic report analysis from the Forensic Report Analyst agent because it requires report data to properly interpret photographic evidence. The coordination engine contains built-in retry

functions that ensure operation resiliency. A poor-quality input data failure by an agent leads to task reassignment by the engine which sends it to a human underwriter for manual review.

3.4. Implementation Details

A robust implementation exists that combines Streamlit web interface development with CrewAI Multi-agent orchestration [8] and data validation through Pydantic and text analysis using GPT alongside BERT within the framework. The task execution of computer vision applications runs through OpenCV and TensorFlow library functions. Customers initiate the workflow by submitting their claim documents (such as policies and forensic reports and pictures) through the interface designed in Streamlit. Following document submission the Document Preprocessing Module carries out validation before extracting pertinent information from the documents. CrewAI Framework begins its operation by both initiating agents and assigning specific responsibilities based on document types. The system utilizes agents to process their allocated documents by using Pydantic validation to verify both consistency and accuracy of outputs [9]. A comprehensive risk assessment report emerges from the Output Aggregation Module that the underwriter receives for review purposes. Each of our system components exists separately to handle diverse levels of complexity by allowing individual updates without affecting other components. The combination of retries that are applied with fallback procedures as error management techniques provides robustness to the system and its scalable structure allows it to process high claim volumes through agent task distribution. This framework implementation makes the system usable while remaining flexible which fulfills current requirements for underwriting processes.

3.5. Process Details

Take into account a Fire-related Claim of Damage against a property. The system processes the following documents:

- **Insurance Policy:** This document is examined by the Insurance Policy Coverage Expert who establishes whether there are limitations or exclusions of coverage.
- **Forensic Report:** This document is analyzed by the Forensic Report Analyst who ascertains the fire's origin and the degree of damage to the structure.
- **Photographic Evidence:** This evidence is assessed by the Photographic Evidence Interpreter who determines and verifies the level of damage as captured within the photos.
- **Financial Records:** These records are examined by the Financial Records Validator who reconciles the repair costs against the forensic documents and the available photographs.

These features are integrated into the system's generated risk assessment report, which also comprises:

- 1) The statement on the available coverage and any relevant exclusions.
- 2) Estimation of the damage and details on how it was caused.
- 3) Verification of the repair costs and estimates.
- 4) Analysis of the claim and provide appropriate recommendations (accept, reject, or investigate further).

4. Experimental Setup and Results

4.1. Experimental Setup

In order for us to evaluate the effectiveness of how this multi-agent system functions, we have conducted a series of experimentations that use real-world property claim data. We designed these experiments for assessing the system's accuracy, efficiency and the overall scalability. This is vital as we are processing diverse documents of claims and generating risk assessment stores.

4.1.1. Dataset

We used a dataset for experiments that contained over 1000 property damage claims that were collected from insurance providers over Kaggle. Each of the claims had several features and information, including the coverage of Insurance Policies, where the PDF documents outline coverage terms, conditions and exclusions. Forensic Reports identify the technical reports that detail out the cause and extent of damage. Along with inputs like photographic evidence and financial records.

As an agentic approach, to solving this as a Mixture of Agents problem [10], we prompt-engineered the agent to create an external Knowledge Base of examples, that can be perceived as a training set. The test set consisting of about 30% of the random examples were taken for evaluation of our Multi-Agentic Solution to assist underwriters.

The dataset that we used for evaluation consisted of 1000 property claims files that were collected and compiled from data of insurance providers on Kaggle, and some sourced internally from insurance companies for testing purposes. 70% of this data was used to be able to help the LLMs understand the context, while we used the last 30% of the split to test the overall performance. The complexity of the data varied, and encompassed different types from straightforward cases of single-unit properties to commercial spaces with multiple tenants and intricate terms. Each claim file had an average of 4 documents, including insurance policies, forensic reports, estimates of repairs and photographic evidence in most cases. Data quality was assessed through a manual review of 100+ samples, demonstrating negligible inconsistencies in data formats, and OCR-related errors.

4.1.2. Evaluation

We evaluate our system's performance as per the set metric of checking accuracy; that evaluates the percentage of correct risk assessments as compared with all evaluated assessments. We also evaluate these metrics:

- **Consistency:** How consistent or variable the risk assessment outcome is for similar claims. This is measured using the standard deviation of the outcomes.
- **Efficiency:** The time that it takes to process each claim, this is taken as a range and as an average.

To benchmark our system, we compare it with baseline methods like:

- **Human Experts:** A panel of three experienced underwriters in the Insurance sector evaluated the subset of test claims. The underwriters were given access to the same claim files as the other systems and were asked to make decisions on coverage and claim amounts.
- **Rule-Based Software Systems:** A conventional rule-based approach is tested modeling a typical decision tree where the system is trained to find subsets of the training data evaluated on its capability to classify approved or rejected cases based on defined patterns.
- **Monolithic (Standalone) AI Models:** A large language model that is prompt-engineering and goes fine-tuning on the entire training dataset for making end-to-end claim evaluation.

4.2. Results and Analysis

4.2.1. Accuracy

Our proposed multi-agent solution achieves a whopping accuracy of 92.9% as compared to the monolithic AI's 71% [11] applied on Life and Health insurance. A rule-based system can only take a handful of cases while completely skipping inherent, or underlying learnings that truly define the uniqueness of each case. Our superior performance can be attributed to our task specialization and the overall contextual understanding. As an example if we break down our agents into individual evaluations:

- The **Forensic Report Assessment** agent correctly identifies the causes of damage more than 95% of the time.
- The **Photographic Evidence Interpreter** agent correctly classifies damage types in at least 85% of the cases.

4.2.2. Efficiency

Our system is overall very efficient with a response time as low as 40 seconds. The range of efficiency ranges anywhere between 32 seconds (minimum) up to 66 seconds at maximum. This is because the overall duration of computation and processing depends on the number of features available, the token size and the time it takes to interpret all information. The more the modalities and the longer the contextual data, the more time it is expected to take. As no such processes are parallel, this is the most amount of time that will be needed to process any claim. Imagine that a simple claim in this domain is processed by human experts in about 1 - 2 days, while the complex claims require detailed investigation up to a month. A multi-agentic solution, like ours, provides a head start onto the overall execution of investigation, saving time for several monotonous processes.

4.2.3. Consistency

We evaluated how consistent our system really is. Our consistency metric goes up to 94% where the percentage is indicative of claims with similar attributes getting similar results. We testify and ensure this level of consistency through LLM-based agents due to our structured task definitions. According to various studies, the average human error rate in processing insurance claims can range from 5% to 10% [12], with some sources indicating that up to 27% of claim denials can be attributed to human errors during patient registration, highlighting the significant impact of mistakes made during data entry and claim submission processes. **Table 1** lists the tabulated evaluation of multiple approaches towards underwriting. Based on the table, the most straightforward approach towards getting top quality results is the hybrid integration of putting AI Agents to work with Human Experts.

The accuracy of the performance was calculated based on the total proportion of claims that were evaluated correctly by our multi-agent system. We deemed a claim to be correctly evaluated if the coverage decision and the total claim amounts were within 5% value of the settlement amount.

Table 1. Evaluation of different approaches in underwriting.

Evaluation	Multi-Agent System (Ours)	Monolithic AI	Human Expert
Accuracy	92.9%	71%	90%
Duration	32 - 66 seconds	10 - 50 seconds	1 - 2 days
Consistency	94%	99%	~90%
Cost	Low	Medium	High

4.3. Case Study: Structural Incident Analysis and Restoration Plan

We take a pseudo-case and verify the outcomes with multiple human insurance experts.

4.3.1. Incident Review

On January 20, 2025, there was a vehicle collision that involved a 2018 Ford F-150, driven by Ms. Emily Carson. It caused substantial damage to the southern wall of the residential property at 1456 Maplewood Drive, Rivertown. The driver lost all control due to the icy road conditions and led to the truck crashing into the structure. The impact resulted in some severe structural compromises that required a detailed assessment and repair plan.

4.3.2. Structural Damage Assessment

The post-incident structural assessment was conducted by Dr. Miranda on January 22, 2025. She identified the following critical damages:

- **Foundation:** Localized cracking, moisture barrier disruption and minor differential movements.
- **Southern Wall (Impact Zone):** Extensive damage to at least 30% of brick veneer.

Load-bearing wall compromised with notable buckling and some misalignment. There is visible damage to interior framing and requires full replacement.

- **Roof Structure:** Minor misalignment was seen near the southern edge and roof trusses remained intact while requiring reinspection after wall repairs.
- **Load-Bearing Capabilities:** The southern wall was incapable of bearing all structural load as normal. Temporary shoring is needed for preventing a collapse.

4.3.3. Safety Risks and Immediate Actions Taken

The compromised southern wall poses a significant safety concern that also includes:

- Risk of a collapse
- Electrical hazards because of wiring damages
- Exposure to environmental elements that can lead to further degradation

For risk mitigation, the Rivertown Fire Department ensured that there were no immediate fire hazards seen. The homeowner is advised to get professional help for structural engineering and start insurance claims.

4.3.4. Repair and Restoration Plan

The comprehensive repair and restoration plan by the Multi-Agent System that we designed is as follows:

1) Foundation Repairs

- Reinforcement of perimeters and repairing of cracks.
- Moisture barrier restoration

2) Southern Wall Reconstruction

- Damaged brick veneer replacement
- Interior framing
- Rebuilding load-bearing structures
- Adding new, reinforced, materials

3) Roof Realignment

- Structural reinspection of post-wall reconstruction
- Security of trusses

4) Electrical Restoration

- Completely rewiring the damaged circuits
- Complete safety inspection

5) Interior and Exterior Finishing

- Replacement of drywall
- Painting and trimming of drywall
- Brick matching and masonry reconstruction

Total Restoration Estimate: \$55,440

Materials: \$37,800; **Labor:** \$12,600; **Contingency:** \$5040

Projected Timeline: February 1, 2025-March 15, 2025 (6 Weeks)

4.3.5. Insurance and Liability Considerations

The driver, Ms. Carson was cited for the failure to control her vehicle. Her insur-

ance provider was also notified and will have to likely cover property damage as per applicable policies. The policy PD-1002 that covers storm-related incidents like icy road conditions, is applicable in this case. For further verification, an insurance professional is needed to determine coverage eligibility and process the claim.

5. Conclusions

The research shows how property insurance underwriting stands to experience revolutionary changes through Multi-Agent Systems (MAS) operated by Large Language Models (LLMs). We use autonomous agents to specialize complex underwriting operations within our framework because our system accomplishes superior accuracy and efficiency along with consistent performance than conventional practices. The combination of AI tools and multi-agent analysis demonstrates the capability to analyze different documents efficiently while precisely detecting crucial information which results in detailed risk assessment. The system's 92.9% accuracy level exceeds traditional rule-based systems and monolithic AI models to demonstrate its powerful unique method. The system operates efficiently with each claim processed in 32 - 66 seconds while underwriting by hand requires millions of seconds for completion.

The system maintains reliable and reproducible outcomes in different claim scenarios since it achieves a consistency rate of 94%. The consistent performance serves as an essential foundation to construct trust relationships as well as provide fair outcomes that both parties can trust. The research produces a solid basis for how AI underwriting will develop in the future. The insurance industry will achieve maximum efficiency through improved accuracy and customer satisfaction which can be realized through systematic system enhancements and AI implementation alongside ethical AI decision-making assessment.

6. Future Directions

The fast pace of artificial intelligence development creates outstanding potential for improving this field. The development of key research fields should focus on:

Explainable AI (XAI)

Implementing XAI techniques, such as SHAP or LIME, to provide lucid and understandable explanations for the system's decisions. This will enhance trust and aid human oversight, reinforcing that AI-driven underwriting decisions are fair and legitimate.

Reinforcement Learning

The system implements reinforcement learning algorithms as a means to enhance its capability to process new information which allows it to modify claim patterns automatically. The system will gain improved performance levels through time while becoming more accurate at forecasting upcoming risks.

Integration with Blockchain Technology

A blockchain deployment would enhance claims security along with offering transparency while enabling full auditability of the entire claims system. By using

blockchain technology all stakeholders receive data integrity alongside secure information exchange with identifying all transaction records permanently.

Human-in-the-Loop Systems

Organizational development of systems which unite expert human input alongside AI operational characteristics. The system integrates human underwriters through final decision-making procedures after they review and verify the outputs of the system while providing feedback.

Ethical Considerations

It is critical to handle ethical aspects which combine fairness with bias reduction while establishing accountability for AI-based underwriting decisions. To sustain trust in the insurance industry proper unbiased functions must be embedded in AI systems for decision-making purposes.

While the results of our evaluation are definitely promising, we acknowledge the overall limitation of our work that we underwent on a small dataset of 1000 property claims. Although possibly diverse, this may not fully represent the wider variety of use cases and claims that are encountered in real-life. Future research must also focus on evaluating these multi-agent systems on more diverse scenarios, that include claims from different economic and geographic regions and different insurance providers. To improve the evaluation confidence, this means that the amount of documents for the AI agents to act upon, should be increased too.

This research represents a significant step towards realizing the full potential of AI in transforming the insurance underwriting landscape. By continuously exploring and implementing these advancements, the insurance industry can achieve greater efficiency, accuracy, and customer satisfaction while navigating the complexities of an ever-evolving risk environment.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Heiman, E. (2021) Protecting Renters from Flood Loss. *University of Pennsylvania Law Review*, **170**, 783-809.
- [2] Riikkinen, M., Saarijärvi, H., Sarlin, P. and Lähteenmäki, I. (2018) Using Artificial Intelligence to Create Value in Insurance. *International Journal of Bank Marketing*, **36**, 1145-1168. <https://doi.org/10.1108/ijbm-01-2017-0015>
- [3] Smaili, N., Arroyo, P. and Issa, F.A. (2021) The Dark Side of Blockholder Control: Evidence from Financial Statement Fraud Cases. *Journal of Financial Crime*, **29**, 816-835. <https://doi.org/10.1108/jfc-05-2021-0113>
- [4] Bailey, M., Glaeser, S., Omartian, J.D. and Raghunandan, A. (2022) Misreporting of Mandatory ESG Disclosures: Evidence from Gender Pay Gap Information. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4192257>
- [5] Lesch, W.C. and Byars, B. (2008) Consumer Insurance Fraud in the US Property-casualty Industry. *Journal of Financial Crime*, **15**, 411-431. <https://doi.org/10.1108/13590790810907245>
- [6] Woetzel, J., Pinne, D., Samandari, H., Engel, H., Krishnan, M., Boland, B. and Powis,

-
- C. (2020) Climate Risk and Response. McKinsey Global Institute.
- [7] Lissy, N., Bhuvanewari, V. and Krupa, M. E. (2023) Impact of Digital Transformation in the Insurance Industry. *Digitalization of the Insurance Sector*, 71.
- [8] Duan, Z. and Wang, J. (2024) Exploration of LLM Multi-Agent Application Implementation Based on LangGraph + CrewAI. arXiv: 2411.18241.
- [9] Boylan, J., Mangla, S., Thorn, D., Ghalandari, D.G., Ghaffari, P. and Hokamp, C. (2024) KGValidator: A Framework for Automatic Validation of Knowledge Graph Construction. arXiv: 2404.15923.
- [10] Chen, S., Zeng, L., Raghunathan, A., Huang, F. and Kim, T.C. (2024) Moa Is All You Need: Building LLM Research Team Using Mixture of Agents. arXiv: 2409.07487.
- [11] Wang, Y. (2021) Predictive Machine Learning for Underwriting Life and Health Insurance. Actuarial Society of South Africa.
- [12] Studdert, D.M., Mello, M.M., Gawande, A.A., Gandhi, T.K., Kachalia, A., Yoon, C., *et al.* (2006) Claims, Errors, and Compensation Payments in Medical Malpractice Litigation. *New England Journal of Medicine*, 354, 2024-2033.
<https://doi.org/10.1056/nejmsa054479>