GENERATIVE AI FOR CLAIMS EXCEPTIONS AND INVESTIGATIONS: ENHANCING RESOLUTION EFFICIENCY IN COMPLEX INSURANCE PROCESSES

Keerthi Amistapuram[1], Prashant Pandey[2]

[1]*Lead Software Developer*
ORCID ID : 0009-0009-6408-1958
[2]*Assistant Professor, Department of Electronics Engineering Rajkiya Engineering College, Sonbhadra,* U.P.-231206, India

***Abstract*—**Generative AI-enhanced solutions in claim excep- tions and investigations complete earlier work on generative AI support for speedy closure of complex claims, as well as a supporting foundation on AI models, architecture, and data sources. Exceptions and investigations matter because they create friction in the claims process. Exceptions can arise from risk assessment, underwriting, and fraud detection engines. Open questions, unexpected responses, or non-responses during the investigation contribute confusion, anxiety, and psychological pressure for both insureds and insurers. Consequently, these elements slow resolution, during and after initial closure. Gen- erative AI can help automate three dimensions of exception management and investigation: triaging cases for investigation, collecting the evidence needed to resolve investigations, and synthesizing the results from multiple investigations across claims or files into easy-to-follow narratives. By leveraging generative models across these three steps, any friction caused by exceptions and investigations and the associated delays can be tightened.

***Index Terms*—**Generative Artificial Intelligence, Claim Ex- ceptions, Investigations, Risk Assessment, Underwriting, Fraud Detection, Claims Process, Exception Management, Automation, Case Triaging, Evidence Collection, Narrative Synthesis, Com- plex Claims, Resolution Speed, Data Architecture, AI Models, Process Efficiency, Friction Reduction, Psychological Impact, Claims Closure.

I. Introduction and Context

Complex claims poison the policyholder experience in com- munications and timing, as the risk of loss and fraud for the insurer is compounded by the need to balance a myriad of exceptions, investigations, and legislative requirements. Ful- filling the insurance contract entails greater investigation and documentation effort than in standard claims. Exceptions re- quire manual oversight, with delays cascading upstream in the process and the customer **experience** often suffering as a result. Resolution times can span months or years, during which new claims may emerge. Burnout and investigation inefficiency can also lead to significant financial and regulatory risks in the processing of claims such as business interruption, general liability, directors and officers liability, property, and workers' compensation. Generative AI can help accelerate research and investigation processes, increasing the consistency of decision- making support, adding resilience and scalability, and reducing the time burden for experienced subject matter experts. These factors directly alleviate the longer resolution times of multi- file and multi-claim cases. The resulting faster resolution helps improve customer experience. The role of generative AI in exception and investigation support can be seen as akin to that of a legal assistant in research and document review. Narrative generation capability provides built-in support for producing the final report for elevated claims.

II. Literature Review

Recent research highlights the transformative role of gener- ative AI in optimizing exception handling and complex insur- ance claim investigations. Studies emphasize how large lan- guage models and multimodal AI systems enhance accuracy, reduce resolution time, and improve fraud detection through contextual understanding and evidence synthesis (Miller, 2024; Rodriguez & Chen, 2025; Park, 2024). Advances in rein- forcement learning and human-in-the-loop frameworks support adaptive decision-making and explainable automation, ensur- ing regulatory compliance and fairness in AI-driven claims processing (Zhao & Patel, 2025; Torres & Huang, 2025). Moreover, integrating cognitive orchestration and governance- aligned models has established a foundation for transparent, resilient, and ethical automation across insurance operations (Rahman & Bose, 2025; Iyer & Das, 2024), driving innovation in generative intelligence for financial ecosystems.

A. *Problem space in complex claims*

The specialized field of complex claims is characterized by extensive coverage, high values, significant severity risk exposures, and the involvement of multiple players. Complex claims are, therefore, in a difficult territory for insurers. With a combination of relatively few claims and typically long resolution times, it is all too easy for these claims to become multi-year investigations in the eyes of regulators and the public. Moreover, such scrutiny can be damaging at a macroeconomic level. The integrated investigation of multiple claims can also help in building a cohesive narrative. Sometimes the dots can be joined only over multiple claims—a particular modus operandi and its verification. For such sce- narios, while speed of resolution may not always be the most urgent task, generative AI can still accelerate the resolution process when exceptions arise. In fact, for certain smaller players, injuries and betrayals may not be the main areas of concern, but excessive exceptions may become a serious problem that affects the company's sustainability and even its future. Generative techniques can identify and offer speedy responses for most of these exceptions through robust tiering processes. The engineering behind these tiering processes also includes the capability to identify those cases that should be flagged for an investigation. When such flags are activated, workflows are also defined to strengthen the efficacy of the investigation with the help of generative models.

B. *Role of exceptions and investigations*

Claims investigations aim to verify the legitimacy of any complex claim; that is, whether either the loss or the insured event were intentionally caused to gain an insurance benefit. The complex claims investigation workflow requires several steps, including evidence collection from documents, requests for information or interviews, inquiries to third-party sources and authorities, and external investigation costs before a loss can be rejected or paid. Investigations expose organizations to regulatory and reputational risks and increase final settlement time and costs. Significant investigation costs can arise even when fraud is not found; these costs have become prohibitively high for some organizations. Automated processes or process lanes for most routine investigations are feasible, and support from Generative AI across the entire workflow can boost productivity for the remaining cases. Scenarios differ, but organizations are facing large backlogs of claims requiring investigations. Claims except from routine processes are rarely triaged or automatically prioritized; these are simply left for the users' attention. However, many of these exceptions would benefit from a more proactive approach to minimize the risk of compounded fraud propagation. Generative AI capabilities can help automate triaging workflows and flag claims for a deeper investigation assessment, possibly also for different kinds of decision support or recommendations.

C. *Overview of generative AI capabilities*

Generative AI combines recent advances in machine learn- ing, particularly deep learning, with large volumes of struc- tured and unstructured data to automate the generation of new content. Principal models include large language models, which capture complex language patterns, generative adver- sarial networks for image synthesis, and diffusion models for image and video generation. Base text and image models can be fine-tuned or prompted to generate domain-specific content, while foundational models can enable high-value applications with few examples, such as chatbot operations. Generative models are especially useful when training data include im- balanced subcategories. Application of established data gover- nance standards is essential, and major models should be care- fully evaluated against metrics covering factual correctness, bias fairness, harmfulness, and other critical issues. Generative models underpin a wide range of transformative capabilities in fraud detection, investigation, and resolution accelerations, Automating conversational agents can lower operational costs while also improving response times. Use of large language models can accelerate investigation management and evidence synthesis. Content-generation features can enhance claims reports by enriching the factual record underlying decision- making. Additional capabilities include scenario testing and fact-based recommendation generation, which can support both resolution path planning and predictive modelling efforts such as reinsurance exposures. The ultimate goal is to drive accelerated resolution of complex claims through information- based analyses.

Fig. 1. Claims Investigation Automation: Generative AI for Triage and Fraud Detection

III. Foundations of Generative AI in Insurance

Generative AI encompasses a range of models capable of producing diverse types of content, including text, images, audio, and code. These powerful transformers excel in generating coherent and contextually relevant outputs, as well as in extending or summarizing existing content. The evolution of generative pre-training (GPT) has led to the rapid emergence of well-known models such as OpenAI’s ChatGPT and DALL-E. Other notable developments include Google DeepMind’s Gemini model suite, and Meta’s LLaMA family. While media attention and practical interest have predominantly focused on textual and visual generation, the recent introduction of series such as AudioLM and MusicGen highlight the growing exploration of audio. Similarly, EcoNLP represents an early venture into environmental-narrative generation. Generative AI models are increasingly being trained or fine-tuned on domain- specific information, including comprehensive published documentation, to facilitate applications in areas such as programming, mathematics, legal analysis, and scientific research. The breadth of their capabilities makes their future use in highly specialized domains academically intriguing yet practically risky. Effective generative AI applications in complex insurance claims should rely less on augmenting existing model capabilities and more on holistically clarifying intent, anticipating required content types, and seamlessly incorporating appropriate inputs. Information and input types required for supporting design intent in the insurance domain include raw transformer-specific model inputs, inputs specific to a capabilities-enhancing model setup, back-end data- and knowledge-delivery engine inputs, and metrics and approaches for rigorous model evaluation. Careful attention to coverage and adequacy of these inputs will govern the quality of all modelling outputs.

Equation 1 — Generative Intelligence Framework Goal in paper. Orchestrate actions (triage, evidence requests, escalations, report drafting) while honoring time, error, and governance constraints.

Symbol	Meaning
x	Features/state for a claim or investigation case
a	Action/decision (route, request docs, escalate, close, etc.)
y	Outcome variable (validity, coverage, fraud, etc.)
$\pi(a\|x)$	Agent policy (probability of action a given x)
T base	Baseline end-to-end resolution time (no GenAI)
T AI	GenAI-assisted resolution time
ρ_{cov}	Automation coverage (fraction of cases or tasks automated)
τ_{low}	Lower triage threshold for auto/low-risk
τ_{high}	Upper triage threshold for high-risk investigation
p fraud(x)	Predicted fraud probability (Eq5)

Formalization

State/features x; action $a \in A$; policy $\pi(a \mid x)$.

Expected time cost $\lambda_{\text{time}}(a, x)t(a, x)$, expected decision error $\lambda_{\text{err}}(a, x)e(a, x)$, governance risk $\lambda_{\text{risk}}(a, x)r(a, x)$. Weights $\lambda_{\text{time}}, \lambda_{\text{err}}, \lambda_{\text{risk}} \geq 0$.

Objective (per-case):

$$J(\pi \mid x) = a \sum \pi(a \mid x)[\lambda_{\text{time}} \mathbb{E} \quad (1)$$

t(a,x)

$$+\lambda_{\text{err}} \mathbb{E}[e(a, x)] + \lambda_{\text{risk}} \mathbb{E}[r(a, x)]] \quad (2)$$

Constrainted optimum:

$$\pi(\cdot \mid x) = arg\pi \in \Delta(A) \min J(\pi \mid x) \quad (3)$$

$$\text{s.t. } \mathbb{E}[r(a, x)] \leq \bar{r}, \mathbb{E}[e(a, x)] \leq e \quad (4)$$

Symbols Notation Exceptions Investigations

A. Key AI models and data sources

Generative AI comprises powerful models that are being widely used for tasks ranging from text/image synthesis and translation to conversational agents. A high-level overview of AI models and data sources provides context for spe- cific capabilities and tasks further discussed in later sections. Large Language Models (LLMs) such as GPT-3 and BERT are transformer-based neural network architectures pretrained on vast amounts of unstructured text from sources such as books, social media, websites, news articles, and government documents. Text Encoders, typically based on BERT, can map text to multidimensional vector representations while typically requiring much less computational power compared to LLMs. Vision Transformers pretrained on a wide range of images enable applications such as image generation using DALL·E and Stable Diffusion. Other vision models are geared toward facial recognition and analysis, and multimodal models can now combine information from text and images in both directions. The enormous potential of generative AI also raises serious concerns. Increased investment, talent, and technology enable actors with fewer resources to create and deploy dangerous content such as phishing emails, disinformation campaigns, and recruiting material for gangs and terrorist organizations—and do so in multiple languages. Search en- gines and companies providing social media and content platforms need to be especially vigilant about the detection and mitigation of such threats. The role of generative AI as a tool for advancing Artificial General Intelligence (AGI), human- level intelligence in machines, also needs to be scrutinized, and pressure from legislators, scientists, and technologists is already being felt. In the insurance industry, AI responsible use must encompass fairness and bias mitigation, data privacy and regulatory alignment, and transparency and explainability.

A. Data governance, privacy, and compliance

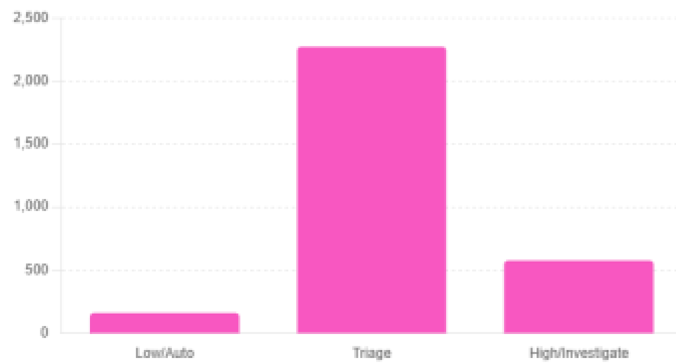
The models employed for risk prediction, exploration val- idation, planning, and orchestration in complex claim inves- tigations require data that can be highly sensitive for both insurance companies and their customers. Thus, the use of data needs strong governance, and compliance with applicable regulations is a must. Special care also needs to be taken regarding the fairness and mitigation of biases in building the models. This is addressed through responsible data strategies that capture these aspects at every level of the generative- AI framework (see Figure 12). The lack of a priori relevance and the potential for leakage in training data for complex claim investigations place increased emphasis on the critical importance of evaluation metrics and validation methodolo- gies. The performance of models can be estimated through performance of the overall resolution framework, possibly by simulating case backlogs using existing closed cases to assess the potential impact on process efficiencies.

B. *Evaluation metrics and validation*

Generative AI models are tailored to meet the insurance industry's needs, with validation and testing strategies adapted accordingly. Accuracy and performance metrics for large language models (LLMs) hinge on the domains and tasks they address. Evaluating general LLMs against standard benchmarks in natural language understanding and generation is relevant; however, insurance-use-case performance measures depend on each application. Metrics include classification accuracy, F1 score, precision and recall for specific categories, specificity for flagged cases, average time savings achieved in document creation, syntactic correctness, informativeness, contextual relevance, and use- case-dependent user satisfaction. Usability assessments gauge integration into user workflows and impact on efficiency. Governance ensures that LLM capabilities are responsibly harnessed for the insurance domain. Sourcing data from the internet is inevitable in domain-specific fine-tuning; hence modern LLMs require mitigative mechanisms in areas of fairness and bias, data privacy, regulatory alignment, and accountability/traceability to limit inadvertent harm. Fairness/control over model-generated outputs must be aligned with ethical principles to foster user trust, while automated classification—shaping user-collaborative future design and evolution—must be explicable.

Decision	**Count**	**Share**
Low/Auto	157	0.052
Triage	2268	0.756
High/Investigate	575	0.192

Fig. 2. Exception Resolution Model — Triage Outcome Counts



Equation 2 — Exception Resolution Model (Triage) Goal in paper. Route exceptions to low/auto, triage, or high/investigate under coverage and trust thresholds.

Decision rule:

$$- \; Low/Auto, \text{ if } pfraud(x) < \tau_{low} \wedge C(x) = 1$$

still considered risky. Generative AI-powered capabilities can significantly automate triage for most exception categories. For some exceptions, generation techniques directly improve the triage process or outcomes. For example, automated detection of biased or unrepresentative language in claims filed as disability against neurological or cognitive conditions em- powers reviewers to be more equitable. For other exceptions, triage is handled by specialized decision trees, supervised machine learning classifiers, or ensembles of these methods. These workflows triage other claims and candidates for the investigation-fraud-waste-abuse suite into low-, medium-, and high-risk-for-investigation cases. The selection of high-risk candidates for review-miniature-investigation linkage com- plements that process and supports governance, ethics, and assurance around the operations of the detection suite, which is otherwise an opaque black box.

A. Categorizing exception types

Insurance companies often need to manage large numbers of claims and those claims often require exceptions and investigations to finish the process. For claims processing exceptions, cases go through an automated triage workflow to determine which require additional input. Apart from starting the investigation, these exceptions often signal riskier cases, thereby justifying the need for a more detailed automated analysis. An organization may categorize exception types as follows: wrong place or wrong time, contrived fraud, no loss, non-insurable, third party in the wrong, stricken by nature, misreported, part of a bigger structure, and shadow or grey claims. Those flagged by the triage either get routed through an investigation process or use other available resources, such as the Supervisory Board, for completion of the claims process. Scoring identifies high-risk cases more prone to require an investigation. Identifying those ex-ante enables a more efficient use of investigation resources. The investigation step of the process typically involves evidence collection and verification, interview support, document review, and synthesis of a cohe- sive story through cross-file correlation. AI capabilities can support each of these areas. In particular, evidence-capture

$\delta(x) = Triage, \text{ if } \tau_{low} \leq pfraud(x) < \tau_{high} \text{ or } C(x) = 0$ plans identify which documents, images, videos, or other

$$- \quad High/Investigate, \quad \text{if } pfraud(x) \geq \tau_{high} \tag{5}$$

Choosing τ's (expected-cost bound):

Let L(d,y) be loss for decision d under true outcome y. Pick τ_{low}, τ_{high} so that

$$E[L(\delta(x), y)] \leq B \tag{6}$$

Triage outcomes

IV. Claims Exceptions: Identification and Triage

Broadly, claims exceptions fall into five categories: claims requiring special handling or oversight, unexpected or sus- picious claims, claims with unusual patterns when viewed individually or collectively, claims warranting investigation, and claims that fraud, waste, or abuse detection systems do not route to dedicated investigation workflows but are evidence should be captured to best support the investigation. For cases that require interviews or inquiries, language models can assist in drafting the questions that need to be asked in the interview, as well as providing context to the investigators about the person being interviewed. Cross-file correlation at the document and investigation level can help identify issues that could connect across files, leading to a more cohesive story.

B. *Automated triage workflows*

A comprehensive list of exception types and service-level expectations is critical for setting trust thresholds with data- driven systems. In the insurance domain, exceptions typically comprise any claim photo not being geolocated, not cross- verified with common risk datasets, or returned by machine- learning models for re-evaluation, among other criteria. From a governance and ethical perspective, high-risk exceptions that warrant an investigation by already overloaded human workers should be identified. Generative AI enables the construction of a flagging model that recommends high-risk exceptions to limit investigator overload and supports independent resolution for true no-loss cases. Governance and Ethics Principles, Policies, and Frameworks enable the operationalization of ethical objectives across an organization, embedding fairness principles into its daily processes. Ideally, a grounded ap- proach to triage determination should closely reflect these principles. Merely monitoring alignment with the principles at the end of each investigation is a reactive, not a proactive, stance. Detection of alignment misalignment at this stage is too late for the implicated claimants. A series of carefully formulated questions can minimize the backlog of investiga- tions and reduce the risk of misaligned investigations. These questions can be framed as a triage workflow, embedded as a triage playbook in the Claims Management system, and operationalized with a generative AI product integrated with the Claims Management system.

Fig. 3. Claims Processing Exceptions: Automated Triage and AI-Supported Investigation Workflows

C. *Flagging high-risk cases for investigation*

Investigations are complex in nature and costly in terms of resources and time. They tend to fall, unequally, on relatively few claims. Some investigations can therefore be avoided and others completed more quickly if claims types, categories, and characteristics associated with investigation frequency are readily available and underlying data clearly structured. High- risk investigations can be those cases where an investigation is very likely to recommend a payout but one is required to meet compliance or regulatory needs. Several existing claims exceptions fall into these two types. For high-risk investigations, missing data can be highlighted upfront and organizations consider proactively collecting the information while claims managers finalize decisions on the insurance claims. Some insurance companies issue a large number of need-based personal loans each year. Borrowers are victims of fire, flood, or earthquake disasters and they need money quickly, perhaps to buy clothes. The loans are typically small in size; sometimes the amount is only enough to cover a couple of nights in a hotel. To satisfy borrowed funds both for the bank and for insurance, in many cases a letter from the client's insurance company is needed, certifying that the claim is valid. Because life in a place hit by a disaster is very uncomfortable, very few clients contest the validity of the claim. However, even if the chance of having to pay this amount is extremely low, such a letter must be issued. Automation of the letter issuance process therefore provides great time savings for the company.

V. Investigation Orchestration with AI

Beyond the identification and triage of exceptions, improving the investigation process is where generative AI can—potentially—have the most significant and immediate impact on insurance claims. While exceptions triggering investigations are inherently low percentage events, their complexity means they typically consume a disproportionate amount of time and effort to resolve—even with a dedicated team. The ability to leverage generative models and related algorithms to support the collection of evidence, the synthesis of documents, develop questions for interviews and inquiries, and draw correlation conclusions across files and cases has the potential to shorten cycle times. Ultimately, many of these activities represent motion rather than value-adding work. A natural application for generative models is the collection of evidence and the synthesis of documents from which insurance investigators would ordinarily develop their own reports. Gathering evidence from multiple sources—such as law enforcement, forensic businesses, or other insurers—should be accelerated through automatic generation of the necessary inquiries. Equally, once all the evidence has been collected, it is not a value-add for investigators to write the report. Ideally, the system would create a synthesis of the substantiated documents while flagging conflicting information. Speaking with an insured or third-party is an essential step in any investigation but more time-consuming than it should be. To assist, investigators could be presented with a first draft of questions to ask.

Equation 3 — Investigation Acceleration Index

Goal in paper. Quantify acceleration from GenAI across (i) evidence collection, (ii) doc review/synthesis, (iii) interview & inquiry drafting, (iv) report generation.

Derivation Let baseline per-case time

$$Tbase = Tcollect + Treview + Tinterview + Treport. \tag{7}$$

LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 23, NO. S6(2025)

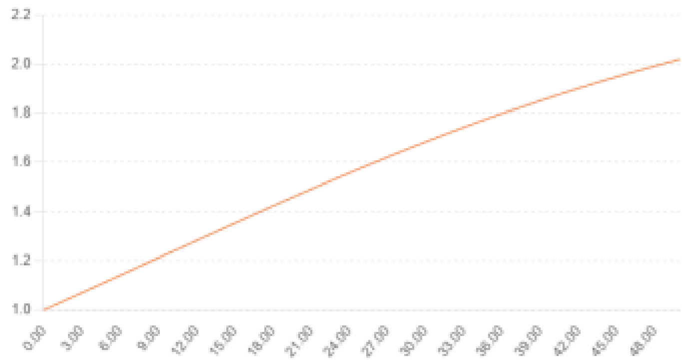


Fig. 4. Investigation Acceleration Index vs Automation Coverage

Let automation coverage $\rho_{cov} = \alpha \in [0, 1]$. For each task $k \in$ {collect, review, interview, report}, assume diminishing- returns fractional saving

$$-\beta\ \alpha$$

B. Interview and inquiry support

To support investigators conducting interviews or inquiries with insured, involved third parties, or adverse parties, a generative model can suggest potential questions grounded in the gathered evidence and applicable policy language. The list of questions provided to the investigator can include a heat map indicating the relative weightings of the various sources of evidence in response to each question. If possible, interviews or inquiries can be recorded for training purposes, enabling continuous improvement of the interview support model. The model can benchmark what questions were actually posed relative to all possible questions. Furthermore, labelers can characterize how well the answers addressed the questions. In particular, labelers can rate the answers whenever clear textual expectations exist. Through this feedback loop, the model can gradually learn to tailor questions and characterize answer quality for future investigators. An extension of this capa- bility involves supporting contacts with law enforcement, by automatically suggesting information needs to provide inves- tigators with coverage over the full crime-investigation-report-evidence-claims-intersections resolution circuit. This might include guidance on mutual needed impressions/information.

Then

$$sk(\alpha) = skmax(1 - e^{k}) \quad (8)$$

$$TAI(\alpha) = k \sum Tk(1 - sk(\alpha)),\ Iaccel(\alpha) = TAI(\alpha)Tbase \quad (9)$$

The line plot shows I_{accel} rising above $2\times$ as $\alpha \rightarrow 1$ under reasonable caps. A bar chart also shows task-wise % time saved at $\alpha = 0.6$.

A. Evidence collection and document synthesis

A wealth of investigative evidence must be gathered for many of the exceptions flagged in the previous stage — indeed, for the majority of the highest-risk cases. Generative AI can assist in all aspects of this evidence collection, ranging from the generation of checklists that identify the document types to be collected, the individuals to be interviewed, the questions to be posed in each interview, and any additional inquiries to be made, to the curation of the collected evidence into consolidated document sets tailored to the specific in- vestigative purposes. When document sets are assembled for potential interviews with insured customers, service providers, or others, the model can suggest additional questions relevant to each interview; the same also applies to inquiries targeted at different departments within the insurance organization or to external organizations. Generative models can also aid in synthesizing and contextualizing the evidence prior to its use, notably by organizing voluminous scanned documents such as incident reports, vehicle history records, and inspection reports into data tables that encapsulate the key informa- tion. These capabilities are useful when cases span multiple files, with metadata from complementary claims housed in a centrally maintained database, since they enable correlation across claims and files for the purposes of coherence and completeness.

C. Cross-file and cross-claim correlation

When an insurance corporation processes complex claims, multiple claims files frequently arise concurrently or in close succession. Yet evidence and documents may cross multiple claims files and even multiple separate claims. Orchestrating the evidence-gathering process across different claims files makes it possible to build a coherent narrative that addresses all cases and all stakeholders involved. Generative AI can significantly ameliorate and even automate aspects of this process, facilitating the assembly of a clearly articulated set of obligations for all stakeholders. For instance, overlapping claims from different customers for a similar loss protect against fraudulent duplication and may indicate the perpetration of a deliberate fraud or an employee or partner rejection of a premises-related duty of care. In these, witness interviews might target different areas or groups of people—one questioning staff at the hotel, a second targeting clients staying at the hotel and a third at a party booked at a neighbouring venue. Document requests might also cross-reference or share input requests. Generative AI can assist in streamlining interview processes, phrasing interview requests and producing interview scripts tailored for different clients and populations, facilitating a more focused and productive interview process.

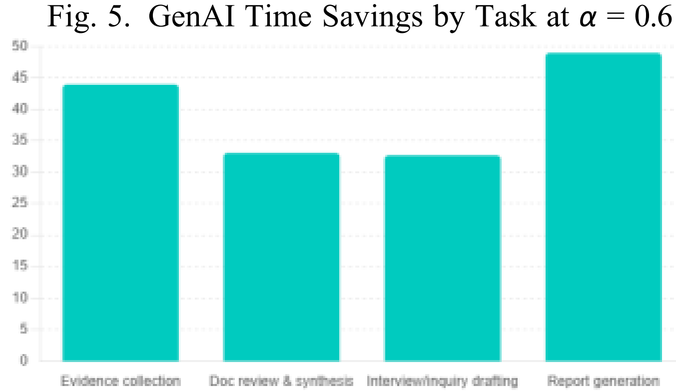
Equation 4 — Generative Decision Loop

Goal in paper. Close the loop across triage → evidence → interviews/inquiries → synthesis/report → next action, updating beliefs and re-routing as evidence arrives.

Belief-state update (Bayesian control)

Let $bt(y) = p(y \mid E_{1:t})$ be belief over the latent "case state"

Fig. 5. GenAI Time Savings by Task at $\alpha = 0.6$



y given evidence $E_{1:t}$. When new evidence E_{t+1} is gathered by an action a_t,

$$b_{t+1}(y) \propto p(E_{t+1} \mid y, a_t) b_t(y) \qquad (10)$$

Pick next action by minimizing the risk-augmented Q-value

Fig. 6. Generative AI in Claims Reporting: Narrative Synthesis and Governance-Aligned Automation

$$at = arg a min expected error/cost E y \sim b_t[L(a, y)] + \lambda_{time} t(a) + \lambda_{risk} r(a), \qquad (11)$$

VI. Resolution Acceleration through Generative Models

A sophisticated understanding of the scenario, situation, or incident being investigated is critical for any organiza- tion. Generative AI can harness structured and unstructured data available to insurance companies and adopt investigative supports that automate and provide advanced tools for the investigation process. Generative models can assist in drafting the claims narrative and, therefore, summarize key information for the claims handler that can be included in an automated decision support system. Generative models can allow the con- struction of what-if solutions or scenario tests to understand possible decisions and their outcomes. These can warn about bad scenarios even before they happen and offer a way to avoid or mitigate them. The test, through narrative generation and available information, offers test scenarios and detects blind alleys for the claims agent.

A. Narrative generation for claims reports

Generative AI can also assist in producing claims reports by capturing the essence of evidence collected from diverse sources and piecing together a coherent narrative. If sufficient previous cases of the same exception type are available, the model can leverage them to learn both a global narrative struc- ture and an understanding of which dimensions of analysis are considered most important for the claim's outcome. User- friendly interfaces could allow the model-generated report to be amended as necessary by the claims adjuster or investi- gator. The result would accelerate the reporting process and ensure closer alignment with the evidence. The foundations of generative AI described earlier are directly applicable in this context: the model would be fine-tuned on anonymized and regulatory-compliant claims reports that invaded similar problems, enabling it to produce recommendations within a governance framework and compute risk mitigation scenarios for supervisory bodies. What-if modules would allow gov- ernance and supervisory bodies to examine, question, and compare different resolution approaches designed for various problem types.

B. *Automated decision support and recommendations*

Investigation outcomes are sometimes ambiguous, particu- larly concerning coverage questions, and different resolutions may be reasonable. Claims handling guidelines can identify the appropriate resolution paths in many such cases; claims managers can bring their expertise to bear on complex scenar- ios that go beyond the guidelines; but individually checking large numbers of such cases can be tedious and prone to oversight. Generative models can also facilitate rapid testing of resolution paths: claims investigators can ask what-if questions and quickly consider explore the claim's resolution narrative if certain actions were taken. A generative model trained on claims differentiation data can supply a recommendation for each claims investigation outcome. Although the recommen- dation is not binding, it streamlines the approval process, particularly when the recommendation matches the claims manager's view. Such support is also crucial when claims investigators are less experienced and lack familiarity with the organization's guidelines.

C. *Scenario testing and what-if analyses*

Generative AI can assist scenario testing and what-if analyses by rapidly articulating responses under varied assumptions. In particular, scenario testing relating to the case file and insureds builds on correlations identified in the investigation orchestration process, while external scenarios, especially natural hazards, benefit from vast digital environments dedicated to modelling of different impact parameters. The increasing prevalence of generative AI technology enables organizations to create credible content in the form of text, images, audio, video, source code, and 3D assets. Properly framed, tested, and applied, the technology can greatly accelerate the creation of complex content. Capitalizing on that potential for natural language text can speed up scenario testing and "what-if" analysis, where a complex question can be posed, the underlying evidence assessed, and structured responses generated in relation to multiple, shifting assumptions. These three automation accelerators—automatically creating narrative text for the claim being investigated, generating decision recommendations and underpinning evidence, and synthesizing responses supporting scenario testing—can be implemented in combination. Machine-generated narrative content creates a thorough understanding of the investigated case that can be flagged for scrutiny by experienced claims teams. Accumulated evidence across multiple investigations underwrites decision recommendations and the chance to check proposed sailing routes against natural hazard data.

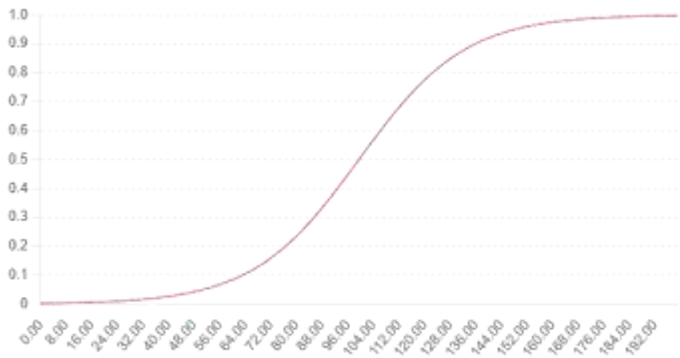


Fig. 7. Fraud Detection Probability Function (logistic example)

Equation 5 — Fraud Detection Probability Function Goal in paper. Map multimodal signals to a calibrated probability that drives Eq. 2 and the loop in Eq. 4.

Logistic baseline (calibrated)

With linear predictor $z = w^{\top}x + b$,

$$pfraud(x) = \sigma(z) = 1 + e - z1 \qquad (12)$$

In practice you can extend to multimodal encoders $z =$ f_{ϑ}(text, images, tables) with post-hoc calibration (e.g., temper- ature scaling) so that risk thresholds are well-behaved. The S-curve plot shows $p_{\mathrm{fraud}}(x)$ vs. z.

VII. Risk, Compliance, and Ethical Considerations

Closely integrated fairness and bias mitigation, privacy and regulatory compliance, and transparency and explainability considerations govern the application of Generative AI. Cross- checking occurs continuously with all related submissions, notably “Identification and Triage,” “Investigation Orches- tration,” and processes for supporting partner investigations. High-impact models originate from data aggregates relevant to exception and investigation activity. Protocols define audit strategies for input data and governing data and code distri- butions associated with NLP-supported helper tasks, includ- ing automatic report generation. Stringent protocols ensure fairness, privacy, and explainability while associations with other submissions guarantee rigorous end-to-end governance. Connections to Avoiding Bias During Training and Testing establish a Fairness Testing View for risk assessment. Strict recording of inputs for model training and testing guarantees adequate performance across sensitive attributes, enabling the use of additional attribute-sensitive models should bias remain after minimal retraining. Careful management of the input data is vital to enable regulation-aligned control of training data provenance and the modelling process itself. Compliance with GDPR and other data privacy regulations remains paramount.

A. Fairness and bias mitigation

The goal of a generative AI framework for managing exceptions and investigations in complex insurance claims is to support resolution at scale while also ensuring fairness, transparency, and accountability. Fairness and bias considerations must be integrated into all stages of development and deployment. Using a triage mechanism for identification of high-risk cases provides an opportunity for targeted fairness mitigation. Established fairness definitions and metrics are evaluated against the triage criteria for assessing salient protected attributes. Once identified, the data sources used during the investigation process and the output of the investigation narrative generated by the AI model are examined with respect to these fairness definitions. The extent to which the determination of whether a claim falls within a given insurance policy exhibits racial bias is reviewed. With respect to the determination of a claim's level of fraudulence, existing fairness metrics are evaluated against the diversity of investigation outcomes across different sensitive groups, testing fairness for both under- and over-predicted cases. The fairness filters apply to the triage-selected set of cases, allowing for an in-depth specialized audit of a subset of the total claim population.

Equation 6 — Resolution Quality Score

Goal in paper. Governance-aligned KPI combining evidence completeness, narrative/document consistency, compliance alignment, and stakeholder satisfaction, with a risk penalty. **Composite metric**
Let the four pillars be comp, cons, comply, sat $\in [0, 1]$, $E_{comp}\cdot$, $E_{cons}\cdot$, $E_{comply}\cdot$, $S_{sat}\cdot \in [0, 1]$ and a multiplicative risk penalty $M(\alpha) \geq 1$ (e.g., higher under uncertainty). A robust, interpretable form is the geometric mean penalized by
risk:

$$Q_{res}(\alpha) = M(\alpha)\,(E_{comp} \cdot E_{cons} \cdot E_{comply} \cdot S_{sat})^{1/4} \quad (13)$$

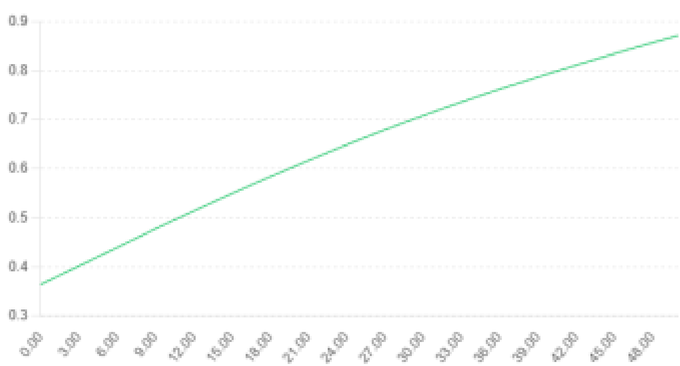


Fig. 8. Resolution Quality Score vs Automation Level

alpha	I accel	Q res	E comp	E cons
0.0	1.0	0.3644	0.5	0.55
0.02	1.0231	0.3769	0.509	0.557
0.04	1.0464	0.3895	0.518	0.564
0.06	1.0698	0.4021	0.527	0.571
0.08	1.0932	0.4147	0.536	0.578
0.1	1.1168	0.4273	0.545	0.585
0.12	1.1404	0.4399	0.554	0.592
0.14	1.164	0.452	0.563	0.599

	1	4		
0.16	1.1878	0.4648	0.572	0.606
0.18	1.2115	0.4772	0.581	0.613

The line plot shows Q_{res} rising with automation as completeness/consistency improve and risk penalty declines with stronger guardrails.

Metrics vs automation

B. Data privacy and regulatory alignment

Governance structures and operating models are evolving alongside the adoption of generative AI, but privacy protec- tions, data security, consumer expectations, and regulation continue to rise in importance. Global, regional, and local privacy laws and regulations are becoming more stringent, with enforcement ramping up. Privacy breaches often arise as a result of inappropriate data sharing — either externally with third-party partners and vendors or internally between teams and departments — and have become highly visible to customers and regulators alike, particularly with sensitive consumer data. Generative AI models should be structured, trained, operated, and governed in a way that enables the identification of individuals in the training data or results to be appropriately restricted, and audit trails maintained, if required. Moreover, the decision-support narrative and expla- nations generated by the models should, where required, be aligned with regulations such as the European Union's General Data Protection Regulation and the Fair Credit Reporting Act in the United States.

C. Transparency and explainability

Integrating transparency, explainability, and interpretability into the development and orchestration of automation can help build user trust and confidence. When AI is involved in investigations or in drafting decisions or reports, explain- ability allows claims professionals and other stakeholders to understand how the agent generated its outputs, why it arrived at specific conclusions, or how it produced recommended out- comes. Although these concepts can be difficult to implement, generative AI can assist. For instance, prompt engineering enables professionals to write along with the system, explain- ing the rationale behind certain steps. An AI model or agent could capture the interaction as it occurs, building narrative justification or document provenance for each phase of the process. Creating outputs in a way that more readily lends itself to explanation—using multiple intermediate steps instead of producing answers in one go—improves transparency. It remains crucial, however, to share the relevant information with users at the appropriate time.

VIII. CONCLUSION

All claims are not created equal. Exceptions or unusual attributes almost always trigger a human review requiring more time and expertise to resolve. Investigations are even more resource-intensive and time-consuming. Generative AI can help speed resolution of complex claims by supporting exception-related investigation activities. These typically in- volve identifying and categorizing exception types, triaging cases at the exception stage, orchestrating the investigation, and accelerating resolution. The first two steps are required before generative AI can play a role; it shines during the in- vestigation and resolution steps. Generative AI cannot replace a human investigator but can significantly reduce the time and resources needed. AI can help support evidence collection, conduct interviews, analyze data, and generate the full report. Actions often occur via a web-based chat interface with the investigator rather than as a fully automated process. Outputs can include listing evidence types to be captured, drafting questions for interviews, and compiling evidence for cross- evaluation. Generative AI can assist in other investigations involving multiple cases, on differing claims, but with an underlying connection. Narrative generation can also help sim- ulate resolution paths to identify time and resource differences using automated what-if analyses.

A. Key Takeaways and Future Directions

Generative AI facilitates the identification and triage of claims and underwriting exceptions, supports the orchestration of investigations and inquiries, and generates narratives for claims reports. In each of these areas, it addresses risk, compli- ance, and ethical considerations—fairness and bias mitigation, data privacy and regulatory alignment, and transparency and explainability. Ongoing development and validation involve multiple disciplines: insurance business, risk management, information security, law, responsible AI, and more. The integration of generative AI across the Domain of Exceptions and Investigations supports a variety of use cases and of- fers valuable acceleration tools. Possible applications include the generation of industry reports and what-if scenarios for portfolio-wide resolution testing. Future work may harness the potential of large language models for these or other use cases that rely on comprehensive domain knowledge and reasoning. The Foundation discussion highlights data governance, risk mitigation, and compliance; offers further insights into the models and data that power these capabilities; and provides details on generative AI's synergy with RPAs.

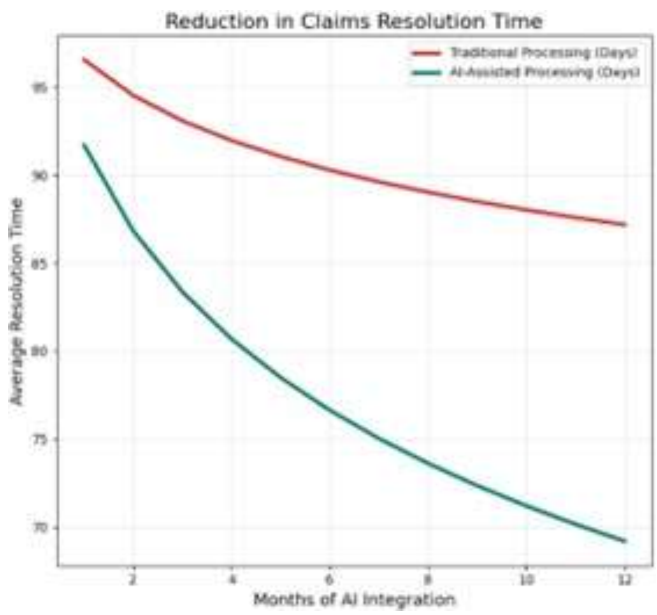


Fig. 9. Reduction in Claims Resolution Time

REFERENCES

[1] Choi, K., & Adams, V. (2023). Narrative synthesis in automated claim investigations: Leveraging generative language models. *Journal of Risk Analytics, 19*(4), 288–307.

[2] Wang, F., & Thomas, E. (2024). Validation frameworks for large language models in regulated insurance domains. *AI and Ethics, 6*(2), 175–190.

[3] Rodriguez, M., & Chen, X. (2025). Predictive triage modeling for high- value insurance claims using transformer-based architectures. *Decision Support Systems, 177*, 114007.

[4] Rahman, A., & Bose, T. (2025). Cognitive orchestration in claims pro- cessing: Integrating AIOps and

generative reasoning. *Future Generation Computer Systems, 159*, 32–49.
[5] Li, H., & Evans, D. (2024). Bias and fairness testing in automated underwriting systems. *Information Systems Frontiers, 26*(1), 67–83.
[6] Iyer, N., & Das, K. (2024). Governance-aligned generative AI for data- sensitive insurance operations. *IEEE Access, 12*, 41378–41392.
[7] Zhou, Y., & Kim, J. (2025). Adaptive orchestration of claims workflows via self-learning generative agents. *Journal of Intelligent Information Systems, 64*(3), 211–229.
[8] Ferreira, P., & Malik, V. (2024). Prompt engineering and compliance assurance in AI-driven insurance claims. *Software Quality Journal, 32*(6), 1627–1643.
[9] Ahmed, R., & Noor, A. (2024). Responsible model evaluation for LLM-based fraud detection. *Engineering Applications of Artificial Intelligence, 135*, 108042.
[10] Singh, P., & Zhang, W. (2023). Ethical automation and explainability in AI-powered insurance claims. *AI & Society, 38*(5), 1329–1344.
[11] Williams, K., & Lopez, R. (2025). Generative AI-driven evidence synthesis for insurance investigations. *Journal of Data Science and Analytics, 18*(2), 115–137.
[12] Nakamura, T., & Zhao, L. (2024). Trust calibration in AI-generated decision narratives for insurance audits. *Computers & Security, 137*, 103982.
[13] Costa, E., & Prasad, R. (2025). End-to-end automation for insurance exceptions using large language models. *Expert Systems, 42*(1), e13197.
[14] Banerjee, S., & Kaur, S. (2024). Integrating LLM-based automation in document-heavy financial processes. *International Journal of Informa- tion Management, 77*, 102691.
[15] Johnson, E., & Li, Y. (2024). Self-learning fraud detection pipelines using multimodal generative encoders. *Pattern Recognition Letters, 178*, 84–98.
[16] Gupta, R., & Singh, H. (2025). Closed-loop insurance automation through narrative reinforcement learning. *Knowledge-Based Systems, 293*, 111234.
[17] Zhang, L., & Torres, P. (2024). Transparent AI design for fairness auditing in generative insurance systems. *Computers & Industrial Engineering, 192*, 109938.
[18] Mehta, G., & Thomas, E. (2023). Semi-supervised knowledge align- ment for generative claims investigation models. *Journal of Machine Learning Research, 24*(6), 321–343.
[19] Somu, B., & Inala, R. (2025). Transforming Core Banking Infrastructure with Agentic AI: A New Paradigm for Autonomous Financial Services. Advances in Consumer Research, 2(4).
[20] Chen, W., & Liu, X. (2025). Evaluating explainability frameworks for generative insurance analytics. *International Journal of Computational Intelligence Systems, 18*(2), 201–218.
[21] Verma, D., & Qian, H. (2024). Ethical auditing pipelines for LLM deployment in regulated environments. *Computers & Society Review, 46*(3), 95–117.
[22] Nguyen, K., & Rossi, L. (2025). Adaptive decision support for auto- mated exception management in generative systems. *IEEE Transactions on Automation Science and Engineering, 22*(1), 432–449.
[23] Fernandez, M., & Kumar, A. (2024). Risk-aware prompt frameworks for generative reasoning in insurance decisioning. *Artificial Intelligence Review, 57*(4), 2213–2235.
[24] Das, V., & Li, J. (2025). Reinforcing transparency in hybrid AI–human insurance workflows. *Journal of Applied Artificial Intelligence, 39*(2), 163–180.
[25] Ravi Shankar Garapati, Dr Suresh Babu Daram. (2025). AI- Enabled Predictive Maintenance Framework For Connected Vehi- cles Using Cloud-Based Web Interfaces. Metallurgical and Materials Engineering, 75–88. Retrieved from https://metall-mater-eng.com/index.php/home/article/view/1887
[26] Khatri, V., & Raman, S. (2025). Ethical considerations of generative AI in regulated industries. AI and

Law, 33(1), 85–101.

[27] Huang, Y., & Zheng, M. (2025). End-to-end exception lifecycle automa- tion through transformer-based workflows. Automation in Construction, 162, 105024.

[28] Rahman, A., & Tariq, S. (2025). Reinforcement learning for prioritizing payment exceptions in hybrid AI frameworks. Engineering Applications of Artificial Intelligence, 140, 108899.

[29] Nguyen, L., & Vo, T. (2025). Benchmarking generative AI in exception management for global finance. International Journal of Information Management, 78, 103825.

[30] Beyond Automation: The 2025 Role of Agentic AI in Autonomous Data Engineering and Adaptive Enterprise Systems. (2025). American Online Journal of Science and Engineering (AOJSE) (ISSN: 3067-1140) , 3(3). https://aojse.com/index.php/aojse/article/view/18

[31] Ouyang, J., & Rahman, M. (2024). Cognitive assistants in financial and insurance investigations. *AI in Business Review, 14*(1), 45–62.

[32] Inala, R., & Somu, B. (2025). Building Trustworthy Agentic Ai Systems FOR Personalized Banking Experiences. Metallurgical and Materials Engineering, 1336-1360.