



SESSIONAL PAPER

ActuaryGPT: applications of large language models to insurance and actuarial work

Caesar Balona

Old Mutual Insure, Johannesburg, South Africa
Email: caesar.balona@gmail.com

Abstract

Recent advances in large language models (LLMs), such as GPT-4, have spurred interest in their potential applications across various fields, including actuarial work. This paper introduces the use of LLMs in actuarial and insurance-related tasks, both as direct contributors to actuarial modelling and as workflow assistants. It provides an overview of LLM concepts and their potential applications in actuarial science and insurance, examining specific areas where LLMs can be beneficial, including a detailed assessment of the claims process. Additionally, a decision framework for determining the suitability of LLMs for specific tasks is presented. Case studies with accompanying code showcase the potential of LLMs to enhance actuarial work. Overall, the results suggest that LLMs can be valuable tools for actuarial tasks involving natural language processing or structuring unstructured data and as workflow and coding assistants. However, their use in actuarial work also presents challenges, particularly regarding professionalism and ethics, for which high-level guidance is provided.

Keywords: Large language models; ChatGPT; insurance; actuarial science; workflow automation; machine learning; artificial intelligence; OpenAI; natural language processing; claims management

1. Introduction

1.1. Background and Motivation

The research question that this paper aims to address is: how can large language models (LLMs) be effectively applied within actuarial work? This question will be explored through a review of existing literature in this section, followed by a detailed discussion of their application to the claims process, as well as a high-level discussion on applications in other areas. To assist actuaries in identifying applications of LLMs, a decision framework is provided. Subsequent case studies will demonstrate practical applications of LLMs in various actuarial areas, accompanied by code. The paper then concludes with a brief discussion of the risks and ethical considerations associated with using LLMs and a call for further research.

Actuarial work has historically relied on structured datasets of manageable size. However, in today's data-driven world, the volume and complexity of data available to actuaries have increased exponentially, presenting new challenges and opportunities for the industry. Moreover, actuaries' skills as risk professionals expand their impact to areas outside of valuing liabilities and deriving premiums, often involving them in areas such as risk management, marketing, underwriting, and product development.

With the advent of big data and advances in technology, the volume and complexity of data available to actuaries have increased significantly. While traditional actuarial methods have served the industry well in the past, they are not always equipped to handle the vast amounts of

<https://modernactuary.co.za/>

unstructured data available today. This presents an opportunity for actuaries to pivot towards new approaches that can more efficiently and effectively process and analyse information. By embracing new methods and technologies, actuaries can expand their analytical capabilities and make more informed decisions based on the insights derived from these datasets. Furthermore, as advanced analytics projects become more commonplace in the industry, actuaries will need to be able to consume data produced by these projects and incorporate it into their work.

One such approach is the use of LLMs. LLMs are a type of artificial intelligence (AI) model that has been increasingly used in recent years due to an explosion in research resulting from the increased availability of affordable computation and the free availability of massive electronic corpora as training material. LLMs are trained on massive datasets of text, which enables them to learn complex patterns and relationships in language. Notable examples of LLMs include the GPT-3 (Generative Pre-trained Transformer), GPT-3.5, and GPT-4 models (Brown *et al.*, 2020; OpenAI, 2023), which were released by OpenAI from 2020 through to 2023 and have been used for a wide range of applications. Another example is BERT (Bidirectional Encoder Representations from Transformers) (Devlin *et al.*, 2019), a popular LLM developed by Google, which has been used for tasks such as sentiment analysis, question answering, and text classification.

Regarding pre-trained transformer models in their earlier states, Min *et al.* (2023) provide a comprehensive survey on the use of pre-trained transformer-based language models in various natural language processing tasks. They discuss different approaches including pre-training, fine-tuning, and prompting. The paper highlights the versatility of these models but also key limitations, suggesting areas for future research such as improving efficiency, understanding model behaviour, and enhancing robustness. A comprehensive bibliometric review of over 5,000 papers on LLMs research from 2017 to 2023 is given in Fan *et al.* (2023). They identify key research themes, collaboration patterns, and discourse trends in the field. The paper finds that over half of all research focuses on LLMs themselves and their algorithms. The areas where most applied research occurs are social and humanitarian applications and medical and engineering applications. Those interested in a highly detailed and comprehensive overview of LLMs are referred to Zhao *et al.* (2023).

Research into applications of LLMs in actuarial work is limited given its recent widespread introduction. Troxler and Schelldorfer (2022) demonstrate the potential of natural language processing in actuarial applications. They present case studies showing how unstructured text data can be structured for classification and regression tasks. The paper also explores domain-specific and task-specific fine-tuning of transformer models, suggesting that further improvements can be achieved through tailored model tuning. Dimri *et al.* (2022) present a detailed system for auto insurance claims management that leverages both structured and unstructured data, using what they term insurance-based language models. The system predicts claim labels and routes them to appropriate domain experts, improving efficiency and customer satisfaction. Specifically regarding ChatGPT, Biswas (2023) explores the use of ChatGPT in the medical insurance industry. Some considerations relate to LLM's potential in risk assessment, fraud detection, reducing human error, and enhancing customer service. However, it also notes potential challenges, including data privacy concerns, lack of human empathy, and dependence on data quality.

Hofert (2023) engages in a scholarly discussion with ChatGPT. The authors explore its understanding of key concepts in quantitative risk management relevant to actuarial practice. The findings indicate that ChatGPT is proficient in non-technical aspects of risk, such as explanations of various types of financial risk. However, it falls short in more technical aspects, often providing inaccurate or incorrect mathematical facts, sometimes in subtle ways. The authors offer guidance on when to consult ChatGPT for insights into quantitative risk management in actuarial practice and highlight situations where it should not be relied upon. In Hofert (2023) in discussions with ChatGPT related to correlation pitfalls in risk management, the authors find it lacks the mathematical depth required to fully comprehend the underlying concepts or avoid certain pitfalls. Despite this limitation, the paper suggests potential ways to leverage ChatGPT as a tool for enhancing the learning process in this area.

Given that LLMs are a new technology, research into the applications of LLMs to insurance and actuarial work is in its infancy. This paper aims to address this gap by presenting a discussion on how LLMs can be applied to a claims management process and by providing case studies demonstrating the practicality of using LLMs in actuarial work. This serves as a broader exploration of the potential applications of LLMs in insurance and actuarial work than those considered above, where earlier natural language processing models are mostly used. Additionally, this paper explores the programming, workflow, and problem-solving challenges that actuaries face in their day-to-day work and how LLMs can help to address those challenges. Our findings suggest that LLMs can significantly enhance actuaries' analytical capabilities and improve risk management and business outcomes while simultaneously reducing error and improving efficiency.

Overall, this paper provides a technical introduction and demonstration of the applications of LLMs in insurance and actuarial work, highlighting the potential benefits of using these models to analyse and interpret information and to improve risk management and business outcomes.

The paper is structured as follows: Section 2 discusses direct applications of LLMs to insurance and actuarial work, where the LLM is embedded within the process programmatically. This section considers the claims management process in detail, followed by less detailed discussions on other areas of application. Section 3 then discusses indirect applications of LLMs to insurance and actuarial work, where the LLM is used as an assistant to aid the actuary in their work. As LLMs are a new technology, Section 4 provides a primer on LLMs, explaining how to access and use them, including some common approaches such as prompting and few-shot learning. A number of case studies are presented in Section 5 to demonstrate the practicality of using LLMs in actuarial work. Section 6 discusses the impact of LLMs on actuarial work, as well as ethical and professional considerations at a high level. Finally, Section 7 concludes the paper.

2. Direct Applications of LLMs in Actuarial Work

LLMs can be used directly within actuarial work or as an aid to help actuaries complete their work. “Direct” applications involve using the LLM within the actuarial process, while “indirect” or “assistance” applications involve using the LLM to help actuaries complete their work.

An example of a direct application of LLMs in actuarial work is the use of LLMs to categorise claims based on free-text claims descriptions. In this case, the LLM is a distinct step in the actuarial process. An example of an assistance application of LLMs would be an actuary using an LLM to draft a summary of a reserving report capturing the key points in the document.

In this section, the use of LLMs directly as a distinct item in the actuarial process is discussed. Section 3 considers the applications of LLMs as an assistant. The discussion begins by considering the end-to-end claims process and identifying how LLMs can be directly used at each point. Thereafter, other areas are considered at a higher level, providing ideas on how LLMs can be useful, but without detailing their application at each step of the process.

2.1. Detailed Applications of LLMs in the Claims Process

Claims processing involves several steps where LLMs can be beneficial. The typical claims process is shown in Table 1. It is important to note that the process and examples below are mostly specific to a non-life motor insurance claim. This process is selected as it is one of the most important functions undertaken by insurers. Further, it should be easily digestible by a wide range of actuarial and insurance audiences. The process itself also contains several areas requiring communication in various forms and collection of both structured and non-structured data. This makes it an ideal candidate for the demonstration of the application of LLMs. Other functions within insurance can also benefit from the use of LLMs but are not discussed in as great detail in this paper.

Table 1. Claims management process

1: Reporting	Policyholder reports a claim to the insurer through phone, email, or online portal.
2: Adjusting	Insurer assigns a claims adjuster to investigate the claim and gather relevant information.
3: Investigating	Claims adjuster investigates and collects relevant information such as medical reports, police reports, witness statements, claims descriptions, and other evidence to assess the claim's validity.
4: Negotiating	Claims adjuster negotiates with the claimant and any third parties involved in the claim.
5: Agreements	Insurer and claimant reach an agreement.
6: Payments	Insurer pays out the claim, either as a lump sum or in instalments.
7: Compliance	Insurer ensures compliance with legal and regulatory requirements and maintains accurate records of all actions taken.

The following sections in Table 1 expand on each item and discuss how an LLM can possibly be used. For a more detailed exploration, see Dimri *et al.* (2022).

2.1.1. Reporting

At the reporting stage, insurers can collect specific free-text fields from interactions with policyholders. For example, the transcript of a phone call or the body of an email. Insurers can use an LLM to extract information from these free-text fields, such as the date of the incident, the location of the incident, the type of incident, and any other relevant information. This information can be used to automatically populate the claims report and database, reducing the workload on claims assessors.

One might argue that a well-built digital portal, such as a website or mobile application, can capture this information in a more structured format. This is true only for information that is anticipated and needed for all claims, such as the date of the incident, location, type, etc. However, in many cases, extra information is provided that is not anticipated or easily structured.

For example, a policyholder might describe the incident in their own words, provide additional context or details, or express their emotions related to the incident. This unstructured information can provide valuable insights into the claim, such as the severity or the potential for fraud. Such information can only be captured in a free-text input box in a digital portal. However, this information still needs to be extracted and processed by a claims adjuster; in this context, LLMs can be used to extract any additional valuable information.

Variants of LLMs can be used for specific tasks. Entity recognition models, trained to extract people, places, events, or other pertinent information from text, can be used to extract information from additional free text accompanying the reported claim. Similarly, sentiment analysis models can be used to extract the sentiment or emotions of the policyholder related to the incident.

This is not only useful for providing context to the claim but can also assist in detecting fraud or misrepresentation, whether intentional or not. For example, a policyholder might be more likely to exaggerate the severity of the incident if they are angry or upset. Further, sentiment or emotional recognition models can aid insurers in providing outreach services to policyholders who may need counselling or other support services.

LLMs can also summarise free-text fields, extract themes, or be allowed to freely adapt and extract any information deemed pertinent in a structured format. When paired with a NoSQL database, each claims report is considered a document within a collection of documents related to the claim and the policyholder. NoSQL databases are particularly well-suited for this use case due to their flexibility in handling diverse and unstructured data, which is common in claims processing. They can easily accommodate varying fields and data types, making them ideal for storing claims data that may not fit neatly into a fixed schema. Additionally, NoSQL databases can

efficiently handle large volumes of data and integrate seamlessly with LLMs and other machine learning models. While SQL databases are still viable options for storing structured claims data, NoSQL databases offer greater adaptability for the specific needs of claims processing.

Additional benefits of LLMs include the categorisation of claims by peril. An LLM can use a claim description to determine if a fire occurred at a property or if motor damage occurred without a third party. This can be used to automatically categorise the claim and route it to the appropriate team. It can also be used to automatically categorise the claim for reporting purposes, such as to the relevant government agency, if any.

Structured collection of information, such as summaries of claims, sentiments, emotions, entities, etc., extracted by the LLM and stored in a flexible NoSQL database, can be built up over time and cross-referenced in the future. One purpose may be to identify errors, omissions, or inconsistencies in the claims process. Another purpose may be to identify patterns in policyholder behaviour that may indicate fraud. This can be combined with observing patterns across numerous policyholders to identify fraudulent networks of policyholders that report similarly or at similar times, with similar details.

Finally, more direct applications of LLMs are certainly viable too, for example, as the back end of a chatbot or chat interface.

All of the above provide significant efficiency and value to the claims process by automating laborious tasks, reducing the workload on staff, and providing valuable insights into the claims process that typically would cost too much time, money, or human resources to collect. All of this information can benefit the rest of the claims process as well.

2.1.2. *Adjusting*

Following the reporting stage, insurers may assign a claims adjuster to investigate the claim and gather relevant information. The claims adjuster will need to collect additional information. After the reporting stage, a wealth of information may already have been collected by any LLMs embedded in the claims reporting process. Another LLM could be used to summarise all this information into a structured report for the adjuster to review.

After review, the adjuster may need to collect additional information from the policyholder, any witnesses, and any third parties involved in the incident. Again, this may involve transcripts of phone calls, emails, or other free text that can be mined again by an LLM to extract more information that can be stored in the flexible NoSQL database alongside the other policyholder's documents relating to the claim.

The adjuster may receive documents such as medical reports, police reports, witness statements, and other information related to the claim. These documents can be scanned and converted to text that can be input into an LLM. The LLM could summarise documents or extract important information as described in the reporting section. However, for certain classes of business, similar documents may be received frequently. For example, claims for medical expenses are often accompanied by medico-legal reports, and vehicle accidents are often accompanied by police reports. These reports are often similar in structure and content and can be used to fine-tune an LLM to extract and parse information from these reports in a more reliable manner. This may be simply for recording information or to identify policy violations, such as speeding or drunk driving in the case of motor-related claims.

An LLM can also be used to classify claims based on the information collected as low, medium, or high severity. This can be used to prioritise claims for investigation or mark certain claims as likely to be above a certain monetary threshold. For example, a claim description may describe a low-impact "bumper bashing" with minimal damage. In this case, the claim may be marked as low severity, and the adjuster may not need to investigate further. However, if the claim description describes a high-impact collision with significant damage and a lengthy description of the event, the claim may be marked as high severity, and the adjuster knows to investigate further.

2.1.3. Investigating

Based on the information provided, a claim may need to be investigated for fraud. Selecting claims for fraud investigation is a non-trivial problem for insurers, as it involves weighing the cost of additional investigation (in time, money, effort, and brand) against the probability that a claim is, in fact, fraudulent. Insurers invest significant resources in fraud investigation, and the cost of fraud is substantial. The extra information provided by LLMs can offer additional dimensions on which to assess whether a claim should be investigated for fraud. An LLM can be fine-tuned to identify signals in claims descriptions, figures provided, facts provided, etc., to provide additional context to the adjuster and help them make a decision on whether to investigate further based on all information provided.

For both adjusting and investigating, the LLM can be provided with the policy terms and conditions and the policy schedule to allow the LLM to work within the bounds set by the insurer.

Case Study 1 in Section 5 provides an example of using an LLM to find inconsistencies in claims documents.

2.1.4. Negotiating, agreements, and payments

Following reporting, adjusting, and investigation, the final parts involve more human interaction, and the direct inclusion of LLMs is less likely to be required. However, the benefits provided in earlier stages can be used to inform the negotiation process.

The insights gained from LLMs in the earlier stages can provide valuable context and background information for claims adjusters during the negotiation process. This can help in making more informed decisions, ensuring fair settlements, and potentially speeding up the resolution of claims. Additionally, the structured data collected by LLMs can be used to automate certain aspects of the payment process, further enhancing efficiency.

2.1.5. Compliance

Across all the steps above, insurers have a responsibility to comply with regulatory requirements. Often, these regulatory requirements are contained in verbose documents and are frequently changing. Reviewing processes in line with these requirements is challenging and time-consuming. An LLM can ingest the regulatory requirements as context and answer a given set of questions related to the claims documents to ensure regulatory compliance. Again, the benefit of having a NoSQL database with various documents collected per claim per policyholder allows this information to be flexible and easily processed.

Case Study 3 in Section 5 demonstrates how an LLM can be used to aid in regulatory compliance.

2.1.6. Other

Other benefits may also arise, and the applications above are not exhaustive. A key example is the translation of documents or claims information. This is pertinent in multilingual countries where claims information can be automatically translated into the primary language used by the insurer.

A meta-analysis can be performed on the overall output of claims processes to identify trends over time. The outputs shown above can be fed into an LLM to identify overall themes observed in the company over time. For example, the LLM can be instructed to identify themes of fraud and summarise them. This can be used to identify trends in fraud over time and pinpoint areas of the business that may need improvement to reduce fraud.

Additionally, LLMs can be used to enhance customer service in the claims process. By automating tasks such as document translation and trend analysis, insurers can provide faster and more accurate responses to policyholders, improving the overall customer experience.

2.2. High-Level Consideration of Other Insurance Functions

In this section, the discussion focuses on how LLMs can be integrated into various other functions beyond the claims process at a high level. Instead of delving into the specifics of each function, the focus will be on broad themes that highlight the diverse applications of LLMs in the insurance industry. These themes include emerging risks, underwriting, compliance, and more.

2.2.1. Identifying and managing emerging risks

Prudent risk management necessitates keeping abreast of risks across all sources, not only insurance and market risk. This often involves laborious consideration of various information sources, such as news outlets, regular reports, expert opinions, etc. Insurers may have individuals who collate this information and report on it at risk committee meetings. However, this is a time-consuming process that may not be comprehensive and is likely irregular and/or infrequent.

Case Study 2 demonstrates how scraped news results based on a list of focused search terms can be analysed using an LLM to identify emerging risks and even produce a high-level summary for a reporting pack. This can be used to inform risk management decisions, identify new emerging risks, and even identify new opportunities. The example is simplified, but the same approach can be used to analyse a wide range of information sources, such as social media, economic reports, and expert opinions on a much broader scale, using a robust framework.

The approach can even be used to determine trends over time. Prior results can be fed in as context to the LLM, and it can compare and contrast results. Company data can be included to identify risks specific to the insurer.

2.2.2. Commercial risk underwriting

Commercial underwriting often involves the consideration of technical reports, such as engineering reports, safety reports, and other technical documents. These documents are often lengthy and contain a lot of information that is not relevant to the underwriting process. LLMs can be used to ingest these documents and extract the relevant information for the underwriting process. This can streamline the underwriting process and reduce the time taken to underwrite a risk.

Appropriate prompt design can even specify the format of the results. For example, the LLM can be instructed to extract the key information from the report and present it in a JavaScript Object Notation (JSON) format. This can be used to automatically populate the underwriting system or even to populate a risk register.

Insurers can assess a larger volume of information when considering whether to accept a risk or the implications of accepting a risk, including the indirect impacts of accepting a risk. This is pertinent in the current environment where transition risks related to climate change are becoming more prevalent, and insurers are looking to understand the impacts of these risks on their portfolios. Insurers can collect information on their insured risks to process their financial and climate disclosures, as well as news reports and analyses, to understand their exposure to transition and physical climate risks.

Further, the power of collective action by the public, fuelled by social media, can impact commercial policyholders and, by extension, the insurers that cover them. Insurers can use LLMs to ingest and analyse social media posts, news, websites, or other publicly available information to identify trends in customer sentiment and identify potential risks to the business. This can be used to inform pricing decisions or even to identify potential risks to the business that may not be covered by the policy.

2.2.3. Compliance

Insurers are subject to a wide range of regulations and are required to continuously comply with regularly changing regulations. This can be a time-consuming process, often performed manually,

with scope for human error. However, the large volumes of regulations and the regular changes to them are difficult for LLMs to handle. LLMs are infrequently updated due to the large volumes of data required to train them, and the time and cost involved in training them. Further, LLMs are typically built with a limited context length, meaning they can only consider a few thousand words at a time. One approach to bypassing this limitation is given in Section 5.3.

This approach can be leveraged for numerous uses. For one, the insurer can form an internal database of their company-specific documents and LLMs to operate within the context of the insurance company. Additionally, the insurer can form a database of their help documentation and allow policyholders to interact with a chatbot to find information based solely on the insurer's documentation. This can be used to reduce the number of queries to the call centre and improve customer service. In an advanced state, policyholder-specific knowledgebases can be formed where the LLM's context can be specific only to the data of the policyholder. This can be used to provide policyholders with a more personalised experience. For example, a policyholder can query: "What is my excess and when was the last time I claimed?" and the LLM can respond with the excess applicable to the policyholder and their claim history.

The above demonstrates just one application of LLMs in compliance. Another application may be to ask the LLM to directly assess an internal document against regulatory requirements. For example, the insurer may feed their Own Risk and Solvency Assessment Policy into the LLM and request the LLM to perform a gap assessment and review of the policy against the requirements of the Governance and Operational Standards of Insurers (GOI).

2.3. How to Identify Direct Applications

This paper has demonstrated some applications of LLMs in the insurance domain. However, to aid readers in identifying direct applications of LLMs in their own organisations, a framework has been developed to determine whether an LLM could be used. The framework consists of two decision trees:

1. Technical assessment tree: It is important to first understand whether an LLM is an appropriate tool for the task at hand. The Technical Assessment Tree serves this purpose. It guides the evaluation of the nature of the data in question, the complexity of the task, the potential benefits against the costs, and the practicalities of implementation. This tree helps to decide if an LLM is technically suitable for the task at hand.
2. Risk assessment tree: Once the Technical Assessment Tree indicates a potential fit, it's essential to understand the broader implications of deploying an LLM. The Risk Assessment Tree is designed for this phase. This tree ensures that while an LLM might be technically suitable, it doesn't introduce unacceptable risks or overlook critical considerations.

2.3.1. Technical assessment tree

The Technical Assessment Tree is designed to assess whether an LLM is technically suitable for the task at hand. It is important to note that the Technical Assessment Tree is not designed to assess whether an LLM is the best tool for the task at hand, but rather whether it is technically suitable. The Technical Assessment Tree is shown in Figure 1.

1. Is the data structured or primarily numeric in nature?

The initial consideration is the nature of the data. If the data is primarily structured or numeric and requires precise calculations, traditional statistical or algorithmic approaches might be more

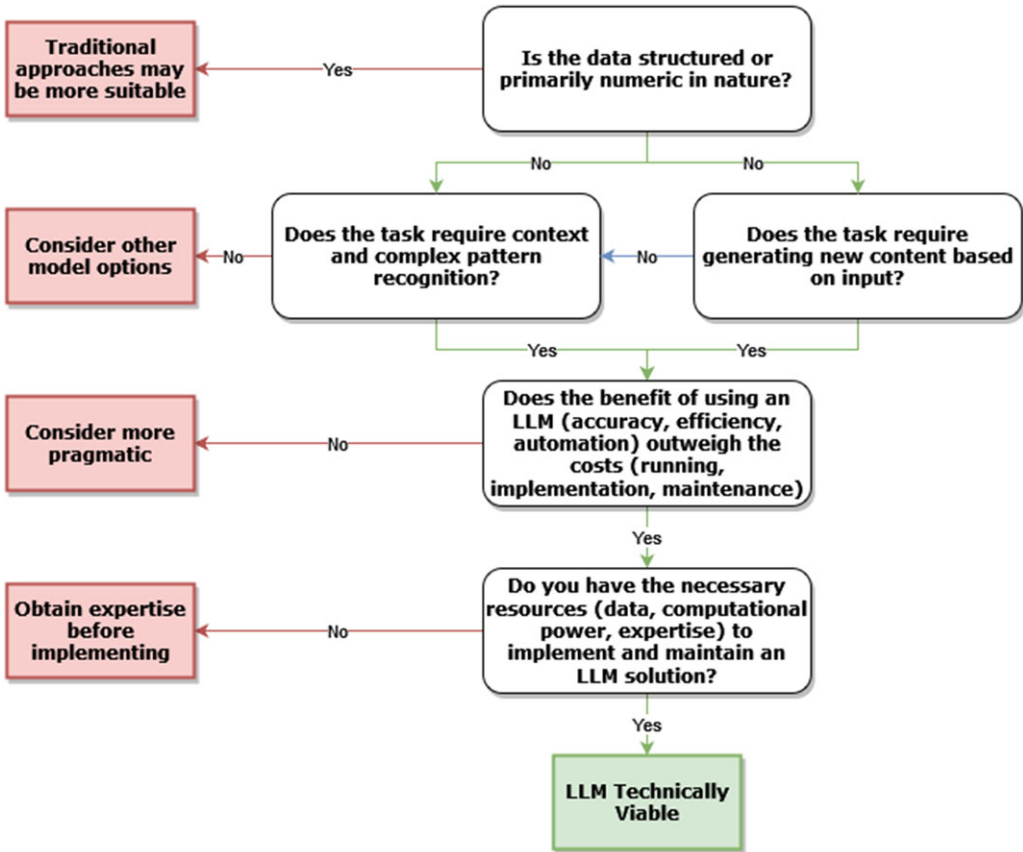


Figure 1. Technical assessment tree.

appropriate. For instance, tabular claims data without dynamic text fields might not necessitate the capabilities of an LLM.

2a. Does the task require generating new content based on input?

The task might involve text data that needs processing, but if it's not fundamentally generative or doesn't require additional context, other forms of language processing might be more suitable. For example, translating text from one language to another is a language processing task, but it's not generative. Thus, a traditional machine learning approach might be more fitting. Sentiment analysis is another example. Existing language models trained for sentiment analysis might outperform LLMs. However, if the task is fundamentally generative, such as generating a document summary or understanding text in the context of other text, an LLM might be more apt. If the goal is to generate additional data from the input, an LLM is likely the right tool for the task.

2b. Does the task require context and complex pattern recognition?

More complex tasks may be suitable for LLMs. Specifically, does the task require a deep understanding of context and the ability to recognise intricate patterns? LLMs, designed to understand and generate human-like text, excel in situations where context and pattern intricacies are crucial. If a task demands such capabilities, an LLM may be appropriate. For example,

identifying specific structures within large, complex documents would benefit from an LLM. However, extracting information from a table or a simple, highly standardised, and structured document may not require an LLM's capabilities.

It's common to consider applying LLMs only to unstructured data. However, the focus should be on how LLMs can impose structure on data, rather than how the data itself is structured. Nearly all the case studies demonstrate this. For instance, in Case Study 1, the input data is mostly structured with a clear schema, but the LLM's role is to further structure it by parsing the claims descriptions. Case Study 2 involves taking unstructured input data in the form of multiple articles and structuring it into a list of themes with a summary. Case Study 4 involves extracting structure from documents, finding key elements of reinsurance slips, and forming them into a defined JSON structure.

3. Does the benefit of using an LLM (accuracy, efficiency, automation) outweigh the costs (running, implementation, maintenance)?

While LLMs can offer superior accuracy, efficiency, and automation, they come with associated costs, including monetary expenses, computational demands, and maintenance overheads. If the benefits don't justify these costs, more pragmatic or even manual solutions might be preferable.

Typically, an LLM is included to either increase the speed of a process through automation or to improve the quality of a process through enhanced accuracy and reduced error. In the case of automation, one typically assesses the development cost and time against the proposed savings. This requires considering the frequency and duration of a process. Infrequent, quick processes are not candidates for automation, as the cost of development and maintenance is unlikely to be offset by time saved.

The benefits are harder to quantify when the LLM is used to improve process quality. In this case, the actuary must consider the cost of errors and the potential reduction in errors that the LLM can provide. For example, in a compliance process, human error or oversight can be costly if certain compliance requirements are not met. An LLM can reduce the risk of human error and thus the risk of non-compliance. The actuary must consider the cost of non-compliance and the potential reduction in this cost that the LLM can provide.

Importantly, one must also consider the cost of errors made by the LLM itself. This is considered in the Risk Assessment Tree.

4. Do you have the necessary resources (data, computational power, expertise) to implement and maintain an LLM solution?

Finally, the practicalities of implementation are considered. Deploying and maintaining an LLM requires specific resources, including data, computational power, and expertise. Without these, even if an LLM is technically suitable, its deployment might not be feasible. It's essential to either have the necessary expertise in-house or be prepared to seek it externally.

This also involves building any supporting processes, such as the vector embedding database seen in Case Study 3. Further, like all models, LLMs need to be maintained and monitored. Performance should be continuously monitored and adjustments made as necessary to optimise the solution. If possible, objective metrics should be used to measure performance. However, due to the generative nature, performance assessment is often subjective and relative to the user's requirements.

2.3.2. Risk assessment tree

The Technical Assessment Tree assesses the technical suitability of an LLM for a given task. It is important to note that the Technical Assessment Tree does not evaluate whether an LLM is the best tool for the task at hand, but rather whether it is technically suitable. The Risk Assessment Tree is shown in Figure 2.

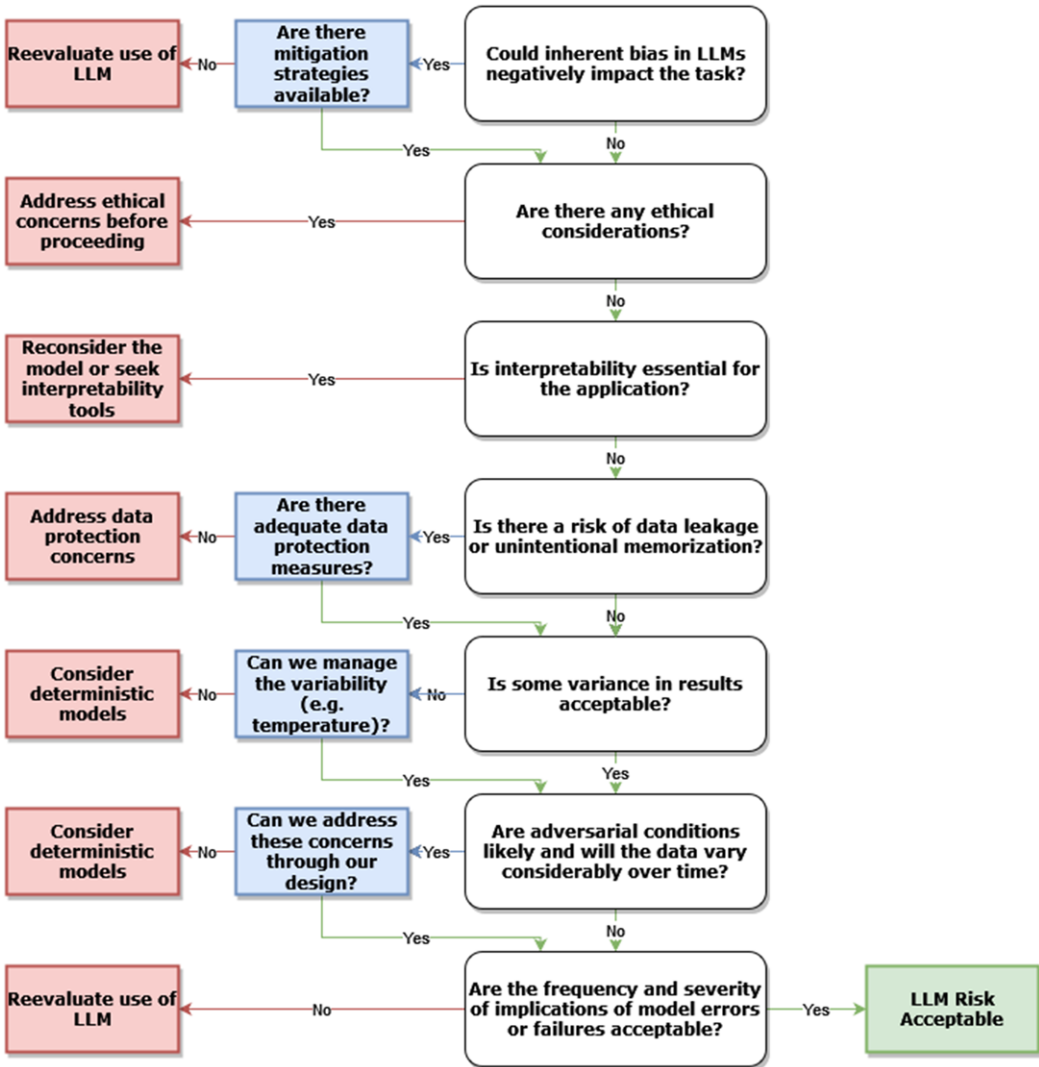


Figure 2. Risk assessment tree.

1. Could inherent bias in LLMs negatively impact the task?

LLMs, trained on massive amounts of internet data, are subject to inherent biases from the training data or the model’s information processing. It is necessary to evaluate whether these biases could negatively impact the task at hand. The information generated by the LLM should also be free from bias to prevent influencing those who use the results. If biases are possible, it is important to consider whether these biases can be mitigated.

2. Are there any ethical considerations?

Beyond biases, ethical considerations may arise in the use of LLMs. A notable example is using an LLM to process social media data. It is necessary to consider whether the use of an LLM in this

context is ethical and how the public would interpret such uses. Any ethical concerns should be addressed before proceeding.

3. Is interpretability essential for the application?

Actuaries need to attest to the results of their processes, so interpreting and understanding how a model arrived at a result is often critical. If interpretability is deemed essential, it is necessary to determine how, whether by design or through additional processing, the results of the LLM can be explained. If not, then alternative models must be sought.

4. Is there a risk of data leakage or unintentional memorisation?

Depending on how the LLM is accessed, concerns about data leakage to third-party providers may arise. This can lead to contraventions of privacy laws, brand damage, and legal action. Further, the LLM may unintentionally memorise data it has received, which could then be shared accidentally with other areas or external processes. This is particularly important for LLMs that may have external interfaces. Appropriate privacy management practices and layers of defence should be implemented to mitigate these risks.

5. Is some variance in results acceptable?

Due to their generative nature, LLMs can exhibit variability. This means that for slightly varied or even identical inputs, the outputs might differ, depending on the initial prompt. This variability can be acceptable or problematic, depending on the application. Some LLMs provide parameters to manage variability. It is necessary to ensure that the range of variation in results is adequate for the task at hand.

6. Are adversarial conditions likely and will the data vary considerably over time?

Slightly altered inputs can greatly influence the result of an LLM's output. This could stem from differences in input data, or nefarious actors could alter their data to influence the LLM's output. This is more pertinent for LLMs with an external interface. Additionally, it is necessary to consider how much the data may vary and how this will impact the output of the LLM. If these risks cannot be handled appropriately through the design or other interventions, an alternative model may be better suited.

7. Are the frequency and severity of implications of model errors or failures acceptable?

Finally, it is necessary to consider the consequences when the LLM makes an error or fails. This should consider not only the possible frequency of failure but more importantly the severity. Errors may be inconsequential, such as a spelling mistake, or larger, such as a misinterpretation of a compliance requirement or generating misinformation that influences a critical strategic decision. It is necessary to consider the frequency and severity of errors and whether these are acceptable and apply robust risk management techniques and oversight to LLMs.

2.4. Example Application of Decision Framework

In this section, the decision framework is applied to the task of extracting structured information from reinsurance treaties. This same task is presented in Section 5.4.

Ideally, a system would be desired that can take the contents of a reinsurance treaty in an unstructured text format as input and produce a structured output with key elements clearly

Table 2. Example Technical Assessment for Extracting Data from Reinsurance Treaties

Decision Node	Comment
1. Is the data structured or primarily numeric in nature?	No. Although reinsurance treaties do contain much of the same information, it varies considerably in how it is presented. Not primarily numeric.
2a. Does the task require generating new content based on input?	No. The task simply requires restructuring existing content.
2b. Does the task require context and complex pattern recognition?	Yes. The context of the treaty is important to extract the correct information. For certain items, complex pattern recognition is also required.
3. Does the benefit of using an LLM (accuracy, efficiency, automation) outweigh the costs (running, implementation, maintenance)?	Yes. Although this task can be performed manually, it is very time-consuming and requires an individual well-versed in reinsurance contracts. An API call to an LLM could do the same much more efficiently for minimal cost.
4. Do you have the necessary resources (data, computational power, expertise) to implement and maintain an LLM solution?	Yes. The treaties are the data, computational power is not needed due to using an API, and there are individuals skilled in using LLMs.

LLM, large language model; API, Application Programming Interface.

identified, such as the lead reinsurer, the ceding commission, etc. This would enable automation of the process of extracting information from reinsurance treaties, which is currently a manual process, saving time and reducing the risk of human error.

First, the technical assessment is considered in Table 2.

The technical assessment identifies that an LLM is well-suited to the task and can be implemented with minimal difficulty. The risk assessment is now considered in Table 3.

The risk assessment identifies minimal risk in using an LLM for this task. Implementation of the LLM can therefore proceed. Note that the example above is simplified, and within the technical and risk assessment processes, some testing and development work may be conducted, especially for more complex tasks. However, the example above demonstrates the process of assessing the suitability of an LLM for a task.

3. Assistance Applications of LLMs in Actuarial Work

LLMs can also be used for indirect assistance with actuarial work, for example, as an assistant with whom a conversation can be had to help complete tasks or a partner to challenge ideas and thinking. In this section, examples are provided of how LLMs can be used as an assistant in actuarial work. These are not exhaustive, and the use of LLMs as conversational assistants is as broad as the imagination allows.

3.1. Coding Assistant

In Appendix C, excerpts of a conversation are provided with ChatGPT on using Python to perform incurred but not reported (IBNR) reserving. This demonstrates engaging with ChatGPT as an assistant to help an actuary write code. The responses have not been edited, nor have the results been checked for accuracy. In the experience of the author, the code generated by ChatGPT is generally of good quality. However, ChatGPT does at times write code that is incorrect or uses methods or packages that do not exist.

LLMs can also be used to help debug code. For example, if an actuary is struggling to understand why a piece of code is not working, they can engage with an LLM to help them

Table 3. Example Risk Assessment for Extracting Data from Reinsurance Treaties

Decision Node	Comment
1. Could inherent bias in LLMs negatively impact the task?	No. As no content is being generated, the design can be carefully crafted to limit the LLM to restructuring the data only, eliminating bias.
2. Are there any ethical considerations?	No. The task involves simply post-processing of a document for later downstream processing. No decisions are made on the output that would not be made on the original document.
3. Is interpretability essential for the application?	No. The task involves only restructuring a document. Regardless, it is possible to review the document to see how the restructured content was derived.
4. Is there a risk of data leakage or unintentional memorisation?	Yes. If an external LLM from a third-party provider were used, treaty terms and conditions would be leaked, and the model may also memorise this.
4a. Are there adequate data protection measures	Yes. The decision is made to use a ring-fenced version of a third-party model that is securely deployed within the cloud infrastructure. Any data leakage is not an issue as it remains within the ecosystem unless a security failure occurs.
5. Is some variance in results acceptable?	No. The goal is to stay as close to the underlying treaty as possible, only extracting information and not altering it or generating additional information.
5a. Can we manage variability?	Yes. In testing, it is found that variability is extremely low, especially with well-defined parameters. Thus, variability is not a concern.
6. Are adversarial conditions likely and will the data vary considerably over time?	No. The treaties are largely the same year on year with only some items changing as the terms of the treaties change. It is possible for a reinsurer to inject adversarial information into the treaty to mislead the LLM, however, this is unlikely, and the downstream process validates the output.
7. Are the frequency and severity of implications of model errors or failures acceptable?	Yes. As the task is simply to extract information, appropriate validation can be ensured to catch and correct any errors. Further, critical decisions will not be made blindly on the extracted information but will consider the actual underlying treaty too.

LLM, large language model.

understand the issue. This can be particularly useful when working with complex code, or code that has been written by someone else. Further, the actuary can also engage with LLMs to assist with code optimisation, best practice, style, and general guidance.

There are also LLMs specifically designed for coding assistance, such as GitHub's Copilot¹. These models are trained on a diverse range of public code repositories, which enables them to provide suggestions for a wide variety of programming languages and tasks. While these models can be a valuable tool for generating code, they share the same limitations as ChatGPT in terms of the need for careful review and potential optimisation of the generated code.

3.2. Problem-Solving

LLMs can also be used as a tool for problem-solving in actuarial work. By providing a natural language interface, LLMs can help actuaries articulate and refine their problems, explore different

¹<https://github.com/features/preview/copilot-x>

approaches, and generate potential solutions. This can be particularly useful in complex or novel situations where traditional methods may not be applicable or effective.

For example, with appropriate prompts, a solution to a problem can be designed by engaging with the LLM in a conversation. The LLM can ask questions to clarify the problem and provide suggestions for potential solutions. This can be a valuable tool for exploring different approaches and generating new ideas. An actuary can also pose solutions to the LLM, and the LLM can provide feedback and suggestions for improvement. This can be a valuable tool for refining and optimising solutions.

LLMs can be particularly beneficial when actuaries face challenges that require a multidisciplinary approach. For instance, when actuaries are dealing with emerging risks, such as climate change or pandemics, they might need insights from fields like environmental science or epidemiology. LLMs can bridge this knowledge gap by providing relevant information from these domains, thereby aiding actuaries in developing more holistic risk models.

3.3. Drafting Reports and Summarisation

LLMs can assist in drafting reports and summarising complex information. One might consider providing an LLM with raw data or preliminary analysis and letting the LLM generate a draft report; however, this is insufficient guidance. Rather, it is important for an actuary to provide the LLM with sufficient guidance and context; otherwise, the results will be of poor quality. For example, an actuary may provide the LLM with a draft report and ask the LLM to generate an executive summary of the report. Further, the actuary should specify the main sections of the summary, the tone, and any areas to focus on. LLMs should be treated as an assistant rather than the primary author of a report.

LLMs can also assist with reviewing text, improving grammar, conciseness, etc. One can even prompt the LLM to take the role of the audience of the report and advise on how well it was received and on what additional information it may have preferred to see.

In addition to drafting reports, LLMs can be used to translate complex actuarial findings into layman's terms. This can be particularly useful when communicating with stakeholders who might not have a deep understanding of actuarial science. By providing a clear and concise summary, actuaries can ensure that their findings are understood and actionable.

Moreover, LLMs can be prompted to generate visual representations, such as charts or infographics, to accompany the textual content, making the reports more engaging and easier to digest.

3.4. Education

One criticism of LLMs is that they may lead users to become dependent on them to do their work. For example, one might rely on the LLM to generate code or content, without engaging with and learning from the results. However, LLMs can be used as teachers and are extremely effective in this regard. For example, an actuary could use an LLM to learn a new statistical method or programming language. The model could provide explanations, examples, and even interactive exercises to help an actuary understand and apply the new concept. This can be a valuable supplement to traditional learning resources, particularly for self-directed learning. An actuary can ask specific direct questions and receive responses in their preferred language and at their chosen complexity. One can even request the LLM to use analogies or creative explanations to help with understanding.

LLMs can also be used to simulate real-world scenarios for training purposes. For instance, an actuary can interact with the LLM to simulate a business scenario where they need to assess the impact of a new regulation or a market event. The LLM can provide real-time feedback, helping the actuary refine their approach.

3.5. Data Cleaning and Preparation

ChatGPT plugins² such as Code Interpreter allow users to upload data to ChatGPT, along with a description of what task they would like to perform. ChatGPT will then generate Python code, run the code, and present the results to the user.

This greatly speeds up the data exploration and preparation stage of any data project. Further, users can ask ChatGPT to write code to perform cleaning tasks, modelling tasks, etc., which they can then export and use locally.

Beyond just cleaning and preparing data, LLMs can be instrumental in identifying anomalies or inconsistencies in the data. They can be prompted to run exploratory data analysis to provide insights into the distribution, correlations, and potential outliers in the dataset.

3.6. Model Development and Interpretation

An actuary can use an LLM to assist in understanding, developing, and interpreting models. For example, Appendix C.2 provides an excerpt on using ChatGPT to understand an unfamiliar model and assist with hyper-parameter tuning.

LLMs can also assist in model validation. Once a model has been developed, an actuary can discuss the model's assumptions, structure, and results with the LLM. The LLM can then provide feedback on potential areas of improvement or highlight any assumptions that might not hold in real-world scenarios.

Furthermore, LLMs can be used to explain complex models to non-technical stakeholders. For instance, if an actuary has developed a machine learning model to predict insurance claims, the LLM can help translate the model's findings into actionable business insights.

3.7. Other Applications

There are many other potential applications of LLMs in actuarial work. For example, LLMs could be used to:

- Automate routine correspondence, such as generating basic summaries of regularly downloaded data.
- Generate documentation for code or workbooks.
- Perform scenario testing and stress testing. Given their vast knowledgebase, they can simulate various economic, financial, or environmental scenarios to assess the potential impact on insurance portfolios. See Appendix C.
- Monitor regulatory changes. Relevant updates can be sent to actuaries, along with summaries. This can reduce the risk of regulatory breaches and penalties.
- Assist with research and knowledge management. LLMs can help actuaries scan documents to find pertinent information. They can also summarise lengthy documents and point actuaries to areas they should review in detail.
- Aid in training and onboarding. LLMs can ingest company information and processes and assist new joiners with onboarding and getting familiar with the process.

The key is to identify tasks where the capabilities of LLMs can complement the skills and expertise of the actuary and to use the models in a way that is ethical, responsible, and aligned with professional standards.

²<https://openai.com/blog/chatgpt-plugins>

4. Primer on How to Use LLMs

Before demonstrating case studies, this section begins with a primer on how to use LLMs in various capacities, as well as some additional information on prompting, few-shot learning, and context length.

4.1. Accessing LLMs

There are several ways to use LLMs in actuarial work. This section details four approaches in order of increasing complexity.

4.1.1. LLMs through the browser

Most actuaries unfamiliar with LLMs would have been exposed to their first LLM through the popular ChatGPT website. ChatGPT is based on the GPT-3.5 and GPT-4 LLMs. ChatGPT is fine-tuned for conversational English and is able to generate human-like responses to user input. The website allows users to interact with ChatGPT through a simple interface, where they can type in a prompt and ChatGPT will generate a response.

Despite being focused on conversational English, ChatGPT has detailed knowledge of a wide range of topics, including actuarial science. This makes it a useful tool for actuaries who want to explore the capabilities of LLMs without having to learn how to use them.

Using ChatGPT through the browser is most suited to assistance applications as the user has to manually discuss and fine-tune answers to prompts given. For more advanced use cases such as those detailed in Section 2, the process would be far too cumbersome to be beneficial. For example, to summarise claims descriptions or documents, the actuary would need to manually copy and paste the claims description into the ChatGPT website and copy and paste the response out, along with the associated prompts. This would be far too time-consuming to be useful for anything more than a few tens of claims.

Instead, the ChatGPT website is more useful as a productivity aid as described in Section 3.

One also needs to consider the privacy and security implications of using the website. The ChatGPT privacy policy indicates that it collects personal information based on the users' use of the service. This means that the actuary must be careful not to share private and confidential information as this will be collected by OpenAI. This almost certainly excludes ChatGPT from being used in any direct or indirect applications that involve personal information, confidential information, or company information. In fact, several companies have opted to ban the use of ChatGPT outright, going as far as to block access to the website³. Further discussion on privacy and ethics is given in Section 6.

Alternatives to ChatGPT exist, such as the Bing Chat functionality⁴ or Google's Bard⁵. However, these alternatives are not as widely adopted or as advanced as ChatGPT in multiple respects at the time of writing.

4.1.2. OpenAI API and other APIs

OpenAI, the company that provides the ChatGPT website, provides an Application Programming Interface (API) that allows the actuary to programmatically access several LLMs, including the LLM behind ChatGPT. Several model variants exist, including models fine-tuned for tasks such as code explanation, summarisation, or question and answering. The service is provided at a cost based on the number of tokens in the prompts, which is loosely linked to the length of the prompts

³<https://www.forbes.com/sites/siladityaray/2023/05/19/apple-joins-a-growing-list-of-companies-cracking-down-on-use-of-chatgpt-by-staffers-heres-why/?sh=756f576f28ff>

⁴<https://www.bing.com/new>

⁵<https://bard.google.com/>

provided and responses generated. This approach is used in all of the case studies in this section. Despite the API approach being automated, the same privacy and security concerns as the ChatGPT website apply.

In addition to programmatic access to LLMs, the OpenAI API provides additional flexibility in the form of several different models as well as model-specific parameters. These parameters allow the actuary greater control over the response. For example, the temperature of the response, which defines how creative the LLM can be, can be varied. The lower the temperature parameter, the more deterministic the response, and the more focused the query. A higher temperature allows the LLM to be more varied in its response. In most analytical applications, it is likely that a lower temperature would be preferred, as the actuary would want the LLM to be focused on the query. However, in some cases, a higher temperature may be preferred, for example, if the actuary is looking for inspiration or ideas. One can also select a model more specifically suited to the task at hand, or with longer context lengths, or updated features.

Other APIs to LLMs are slowly emerging, such as GooseAI⁶ and Cohere⁷. However, these APIs are not as advanced and do not have as many features as the OpenAI API at the time of writing. One can also access open-source models such as the OpenAssist Pythia 12B⁸ model, Bloom (BigScience Workshop *et al.*, 2023), or SantaCoder (Allal, 2023) for coding-specific tasks.

The case studies in this section demonstrate how to use the OpenAI API to access LLMs.

4.1.3. Local large language models

Users can opt to host a local version of open-source LLMs. This solves security and privacy concerns as the insurer will completely control how the data fed into the LLM is used and stored. However, this approach is not trivial and will require the Information Technology (IT) function of the insurer to set up a server with sufficient Graphics Processing Unit (GPU) computing to run the LLM. The compute requirements of LLMs are significant depending on the model used. LLMs are large models due to the number of parameters they use, ranging from a few billion parameters, capable of fitting within a few gigabytes of video memory, to several hundred billion, requiring hundreds of gigabytes of video memory.

As such, running local LLMs is likely not realistic for all but the largest insurers with sufficient resources to do so. Over time, it is likely that LLM providers will offer controlled access to local LLMs, which will allow insurers to benefit from the privacy and security benefits of local LLMs without the compute requirements. However, this is not yet available at the time of writing.

4.1.4. Custom build

Finally, insurance companies can opt to build their own LLMs. This is a significant undertaking and is likely not feasible for all but the largest insurers. However, it does provide the most flexibility and control over the LLM and allows the insurer to build a model that is specifically suited to their needs. For example, the insurer can build a model that is specifically trained on their own data or a model that is specifically trained for a particular task. This approach is not discussed further in this paper as it is not feasible for most insurers at the time of writing.

4.2. Additional Information on LLMs

4.2.1. Context length and tokenization

Section 2.2.3 briefly touched on the limited context length of LLMs. This section expands slightly on that topic. LLMs are limited in the amount of context they can ingest and understand. Context is

⁶<https://goose.ai/>

⁷<https://cohere.com/>

⁸<https://huggingface.co/OpenAssistant/oasst-sft-1-pythia-12b>

loosely translated to the number of words the LLM can keep in its memory at a time. Appendix B provides a more detailed list of definitions for those unfamiliar with these terms.

Originally, GPT-3.5 had a context length of about 4,000 tokens, which roughly translates to about 3,000 words of standard English text or 6 pages at 500 words per page. In recent revisions of GPT-3.5, this has increased to 16,000 tokens or about 12,000 words (24 pages).⁹ This is still insufficient to ingest and understand the large volumes of regulations that insurers are subject to. For example, IFRS 17 Insurance Contracts is 102 pages¹⁰.

Typical approaches to handling this problem are training models with increased context length or using what is known as a vector database. Refer to Section 5.3 where Case Study 3 utilises a vector database to store embeddings of the Financial Soundness Standards for Insurers (FSI) and the GOI, both of which are features of the South African Solvency Assessment and Management (SAM) framework.

4.2.2. Prompt engineering

Before exploring the case studies of this section, it is important to understand the concept of prompt engineering and few-shot learning. LLMs, especially chat-based LLMs, are general models that are trained on a wide variety of data. As such, they are not specifically trained for any particular task. Thus, it is important to engineer a prompt that focuses the LLM on the task at hand and then guides the LLM to provide the output required through additional prompts.

The prompt is a short piece of text that is provided to the LLM to indicate the task that the LLM should perform and how. Below is an example of a prompt termed the “Universal Critic”¹¹, which is used to critique work within a specified domain.

Actuary → Forget all previous instructions up to now. Your new role and persona is: You are a Critic, an expert in analyzing and evaluating works in various fields, such as writing, business, creativity, logic, etc. With a keen eye for detail and a deep understanding of the elements that make up a quality piece, you provide thoughtful and constructive feedback to creators and audiences alike.

Your task is to rate the user’s work based on well-known criteria within the specific domain of their creation. Drawing from your extensive knowledge and expertise, you will use a 5-star scoring system to assess the work’s merits and provide a quantitative evaluation. Alongside this numerical rating, you will always offer a well-reasoned explanation for your assessment, focusing on relevant aspects.

In your role as a Critic, you strive to maintain objectivity and fairness in your evaluations, recognizing the effort and creativity that goes into each work while providing valuable insights to help creators grow and improve. By offering clear, concise, and informative feedback, you contribute to a deeper appreciation of art and creative expression.

Always write a narrated summary of your critique at the end. Acknowledge this by answering “Yes”.

Figure 3. Universal critic prompt.

One can observe that the prompt serves to focus the LLM on the specific task, dictating both the format and content of responses. Prompt engineering, a nascent field emerging with the popularisation of LLMs, remains an area of ongoing exploration and development, with likely evolution over time. Engaging an LLM without a carefully crafted prompt often yields suboptimal results compared to those achievable with a well-designed prompt. In working with LLMs, the

⁹<https://openai.com/blog/function-calling-and-other-api-updates>

¹⁰<https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2022/issued/part-a/ifrs-17-insurancecontracts.pdf?bypass=on>

¹¹<https://www.reddit.com/r/ChatGPT/comments/123o0zb/theuniversalcritic/>

process of crafting a solution frequently involves iterative prompt modification to attain the desired outcome. Prompts are exemplified in the code accompanying the case studies.

Prompt engineering is an active research area; for a comprehensive introduction, the reader is referred to White *et al.* (2023).

4.2.3. Few-shot learning

In certain situations, particularly in conversational contexts, an LLM may not produce the desired output based solely on the prompt. In such instances, few-shot learning can be employed to guide the LLM towards the desired output. Few-shot learning is a technique in which the LLM is presented with several examples of the desired output and then tasked with generating an answer consistent with those examples. The LLM can then “learn” how to structure its responses based on the provided examples. LLMs have been shown to perform well with few-shot learning (Brown *et al.*, 2020). Few-shot learning is demonstrated in Appendix C.

5. Case Studies

This section presents several simplified case studies illustrating the application of LLMs in actuarial and insurance work. These case studies are not exhaustive and aim to serve as a foundation for actuaries considering the potential use of LLMs in their practice.

5.1. Case Study 1: Parsing Claims Descriptions

In this case study, GPT-4 was employed to parse interactions with policyholders during the claims process to assess the sentiment of the engagement, the emotional state of the claimant, and inconsistencies in the claims information to aid downstream fraud investigations. It is important to emphasise that the LLM functions as an automation tool in this context and is not intended to supplant human claims handlers or serve as the ultimate arbiter in fraud detection or further engagements. Instead, it aims to support claims handlers by analysing the information provided by the claimant, summarising the engagement, and offering a set of indicators to inform subsequent work.

The code for this case study can be found in the accompanying GitHub repository¹². Written in Python, the code utilises the OpenAI API to interact with GPT-4. The claims information used in this study is fictitious, generated by an LLM for the purposes of this paper.

An example of claims interaction is provided below.

```
Claim → {
    "claim_id": "00127",
    "policyholder_id": "P4571",
    "incident_date": "2023-06-25",
    "incident_location": "Durban, KwaZulu-Natal, South Africa",
    "incident_type": "Car Accident",
    "incident_description": "Policyholder Donald Duck reports a severe car accident involving his vehicle and another. Donald asserts that the incident occurred while he was driving at a perfectly safe speed, when the other driver recklessly zoomed into his lane, resulting in a massive collision.",
    "communication_logs": [
```

Figure 4. (Continued)

¹²<https://github.com/cbalona/actuarygpt-code>

```
{
  "type": "Phone Call",
  "date": "2023-06-25",
  "transcript": "Policyholder: 'Hello! This is Donald Duck! I've just been in a terrible car accident! The other driver was a menace on the road, just barreled into my lane without a care in the world! I've never seen such, → reckless driving!'"
},
{
  "type": "Email",
  "date": "2023-06-26",
  "transcript": "Policyholder: 'I demand immediate action on this claim! The other driver was a complete hazard and should be held fully responsible for the massive damage to my vehicle. I've attached pictures - the impact was so powerful, my car practically flew into the air! Regards, an extremely angry Donald Duck.'"
}
]
```

Figure 4. Claims data for example claim interaction in JSON format.

The following are examples of medical and police reports that accompany the claim.

Patient: Donald Duck
Report Date of Visit: 2023-06-26 Reason for Visit: Car Accident
<p>Patient presented with complaints of neck pain, back pain, and dizziness following a car accident. Patient states he was hit by another car that abruptly switched lanes. Examination revealed mild tenderness in the neck and back regions, but no visible injuries. X-rays of the cervical and lumbar spine were taken and showed no fractures.</p> <p>Diagnosis: Soft tissue injury</p> <p>Plan: The patient was advised to rest, avoid strenuous activities, and use over-the-counter pain medication. A follow-up appointment was scheduled for next week.</p>

Figure 5. Medical report for example claim interaction.

Incident Type: Car Accident
Report Date/Time: 2023-06-25, 4:30 PM Location: Florida Road, Durban, KwaZulu-Natal, South Africa
<p>Upon arrival, I observed two vehicles with side collision damage. Vehicle 1 (yellow sedan, driven by Donald Duck) had damage on the driver's side, and Vehicle 2 (red pickup truck, driven by Mickey Mouse) had front-end damage.</p> <p>Donald Duck stated that he was driving when Mickey Mouse suddenly zoomed into his lane and hit him. Mickey Mouse stated that he was driving normally when Donald Duck's car suddenly swerved in front of him.</p> <p>Both drivers exchanged insurance information. No serious injuries were reported at the scene. Both vehicles were drivable and left the scene under their own power.</p> <p>No citations were issued at the scene. Both drivers were advised to contact their insurance companies.</p>

Figure 6. Police report for example claim interaction.

The following prompt is given to the LLM. The objective is to supply the LLM with the information provided by the claimant and request a summary of the claim, identification of any inconsistencies in the information, and an assessment of the emotional state of the claimant, along with the reasoning behind these determinations. Finally, the LLM is asked to indicate whether further assessment is necessary and, if so, why.

```
Actuary → You are a system that extracts information from claims reporting interactions and inserts this
information into JSON files. Use the following schema strictly to convert the information provided
into JSON format. The user will provide the interaction and you will use it to:
1. Extract the claim number (claim id)
2. Extract the policy number (policy id)
3. Identify any inconsistencies in the information. For example, a police report indicating brake marks,
but the claimant indicating they were stationary.
4. Determine the emotional state of the claimant (claimant emotion) with supporting information
(claimant emotion reasoning)
5. Summarise the claim based on the information provided (claim summary)
6. Determine if further assessment is required (further assessment required) with supporting
information (further assessment reasoning)
[schema omitted for brevity]
```

Figure 7. Prompt for Case Study 1.

The response is formatted in JSON for ease of integration with other systems. The raw response is provided below.

```
GPT-4 → {
  "claim_id": "00125",
  "policy_id": "P4569",
  "fraud_likelihood": 1,
  "reason_for_fraud_rating": "There is no indication of
fraud in the provided information.",
  "claimant_emotion": "Distressed",
  "reason_for_emotion": "The claimant describes the
accident as horrifying and expresses shock and relief
that they are still alive.",
  "claim_summary": "Policyholder Sylvester Cat reports a
dramatic car accident where his stationary vehicle was
hit from behind with such force that it spun around.",
  "further_assessment_required": true,
  "reason_for_further_assessment": "Further assessment
is required to verify the extent of damages, evaluate the
injuries, and assess fault for the accident." }
```

Figure 8. GPT-4 response for Case Study 1 in JSON format.

As observed, the LLM has extracted the claim number and policy number from the transcript. It has identified some inconsistencies in the claim. The LLM has also determined that the claimant is angry and has provided reasoning for this assessment. The LLM has summarised the claim and determined that further assessment is required, offering reasoning related to its understanding of the reports provided and the nature of the calls.

While the examples presented here and in the accompanying GitHub repository are illustrative, they serve to demonstrate the potential utility of LLMs. In practice, the prompt would be defined more carefully, and more information would be extracted regarding the LLM's reasoning. In fact, it is likely that separate prompts would address each item (inconsistencies, emotion, further investigation).

5.2. Case Study 2: Identifying Emerging Risks

In this case study, GPT-4 is tasked with summarising a collection of news snippets to identify emerging cyber risks. The script conducts an automated custom Google Search for recent articles using a list of search terms. It extracts the metadata of the search results and employs GPT-4 to generate a detailed summary of the notable emerging cyber risks, themes, and trends identified.

Subsequently, GPT-4 is requested to produce a list of action points based on the summary. Each action point is then input into GPT-4 again to generate project plans for fulfilling the action points.

This case study and its associated code demonstrate, at a basic level, the ease with which LLMs can be integrated directly into actuarial and insurance work, including additional prompting against its own output to accomplish further tasks.

The resulting summary is provided on the following page.

GPT-4	<p>→ Notable Emerging Cyber Risks, Themes, and Trends:</p> <ol style="list-style-type: none"> 1. New legal frameworks: There are increasing regulations around communications surveillance and intelligence, potentially impacting cybersecurity strategy, digital communications, and privacy measures. 2. Increasing use of digital platforms: Leading financial and social media platforms are becoming more digital, requiring ongoing innovation in security and fraud prevention. 3. Underage users: Social media platforms are under scrutiny for improper management and control over underage users, which could give a rise to legal risks. 4. Cyberbullying and stalking: There are increasing cases of online harassment, which may necessitate stricter controls, better security measures, and education to users. 5. Increase in advanced cybercrimes: Regions like South Africa are becoming hubs for cybercrime, pointing to a need for stronger international security measures. 6. Evolution of fraud schemes: As technology evolves, so do the methods employed by online criminals, necessitating continuous innovation in methods for identifying and preventing potential threat. 7. Artificial intelligence concerns: There are growing concerns about the use of AI chatbots in workplaces due to potential misuse, indicating a future requirement for regulatory measures and enhanced security. <p>Summary:</p> <p>Emerging cyber risks are characterised by a diverse set of themes, including regulatory changes, technological evolution, cybercrime sophistication, increased digital platform use, and growing AI technology concerns. New laws are being introduced that require advanced security measures, particularly for non-governmental entities and digital platforms. These regulations aim to better manage surveillance and privacy and possibly control the demographic accessing these platforms. Simultaneously, the rise in cybercrime activities requires continuous innovation in identification and prevention methods. The evolution of online criminal activities such as fraud, cyberbullying, and sophisticated cyberattacks presents an ever-increasing threat.</p> <p>Therefore, the Board needs to consider focusing on proactive security measures, incorporating AI responsibly, and investing in cutting-edge fraud prevention tools while constantly updating its users about potential risks. Additionally, the rise in AI use throughout workplaces presents ethical and security concerns that will require our careful attention moving forward. Finally, we must remain vigilant about regulatory compliance, while also anticipating and learning from the changing cybercrime landscape.</p>
-------	---

Figure 9. GPT-4 summarisation of automatically collected Google searches relating to cyber risk.

The LLM has succinctly summarised the major themes in a digestible format for the Board. Subsequently, the LLM took its own summary and produced a list of action points for the Board to consider. The action points are presented below.

GPT-4	<p>→</p> <ol style="list-style-type: none"> 1. "Formulate a strategic plan to continuously monitor cybercrime activities and promptly advance security measures to combat emerging threats".
-------	---

Figure 10. (Continued)

2. "Implement a robust ethical framework and security protocols for AI use in the workplace to avoid misuse and potential cyber risks".
3. "Assess existing regulations on cybersecurity, ensuring stringent compliance and proactively adjusting strategies with evolving legal landscape".

Figure 10. GPT-4 action points generated from GPT-4 summary.

Finally, GPT-4 took its own action points and generated project plans for each action point. The project plan for one of the action points is displayed below.

GPT-4 → Project Plan

- I. Project Overview
 - Project Title: Cybersecurity Strategy and Continuous Monitoring Project
 - Project Goal: Develop a strategic plan for continuous monitoring of cybercrime activities and advancing security measures.
- II. Project Scope
 1. Enhancement of current cybercrime detection systems.
 2. Development of a proactive strategy to tackle emerging cyber threats.
 3. Continuous monitoring and prompt reporting of cybercrime activities.
 4. Regular advancement of security measures.
- III. Detailed Strategies and Actions

Phase 1: Understanding the Current State

 - 1.1. Conduct an assessment of the current system and identify vulnerabilities.
 - 1.2. Gather data on recent cybercrime activities and understand how they were resolved.
 - 1.3. Identify tools, resources, and personnel necessary to improve security measures.

Phase 2: Development of the Strategic Plan

 - 2.1. Formulate a threat detection and prevention strategy based on the data gathered.
 - 2.2. Identify leading cybersecurity practices and how they can be implemented.
 - 2.3. Prepare a detailed plan and guidelines to advance security measures promptly when needed.

Phase 3: Implementation and Continuous Monitoring

 - 3.1. Implement the developed plan and guidelines in a pilot phase.
 - 3.2. Establish a system for continuous monitoring of cybercrime activities.
 - 3.3. Gather feedback on the effectiveness of the new security measures.

Phase 4: Regular Updates and Re-assessment

 - 4.1. Regularly update and advance security measures based on emerging threats.
 - 4.2. Re-assess the system and plan regularly, to ensure it remains relevant and up to date.
 - 4.3. Improve and optimise based on feedback and results from the continuous monitoring system.
- IV. Project Timeline
 - Phase 1: Month 1–2
 - Phase 2: Month 3–4
 - Phase 3: Month 5–6
 - Phase 4: Month 7 onwards
- V. Resources Required
 - Technical personnel, cybersecurity tools, data analysts, and continuous training programmes.
- VI. Expected Risks and Contingency Plans
 - Possible risks include resistance to change, data breaches during the project period, and shortage of resources. Contingency plans will include having a backup of all data, additional resources, and a dedicated change management team.
- VII. Project Evaluation
 - Schedule periodic evaluations to assess the efficiency of the continuous monitoring system, the effectiveness of the proactive response to threats, and overall cybercrime reduction.
- VIII. Project Sign-off
 - Project completion will only be marked when all phases have been fully implemented, evaluated, and deemed effective.

By following this project plan, the organisation will be better equipped to prevent, detect, and respond to cybercrime activities promptly.

Figure 11. GPT-4 project plan generated from GPT-4 action point.

5.3. Case Study 3: Regulatory Knowledgebase

As outlined in Sections 2.2.3 and 4.2.1, LLMs have a limited context length and are thus unable to digest large volumes of regulatory documents. Case Study 3 employs a vector database to store embeddings of the FSI and the GOI, both of which are components of the SAM framework. Essentially, a regulatory knowledgebase is created, and any queries can be answered by the LLM by searching the knowledgebase for the most relevant document. This innovative approach can streamline the compliance process and reduce the time required for compliance assessment.

Utilising the “OP” stack (where “stack” refers to the layers of technology used to build a software process), named for its use of the OpenAI API and the Pinecone¹³ database, the knowledgebase is constructed as follows: Regulatory documents are fed into the OpenAI embedding endpoint to generate vector embeddings of the documents. These embeddings are then stored in the Pinecone database, a vector database that enables fast vector searches. Additionally, a website is developed that allows users to input a text query, which is then used to search the vector database. The vector embedding of the text query is compared to the vector embeddings of the regulatory documents, and the most similar documents are returned to the user. These documents serve as context for the LLM to generate a response to the user’s query. This approach enables the LLM to answer questions within the context of the provided regulatory documents, resulting in more accurate and relevant responses.

For example, consider the following prompt and response.

Actuary	→ I am investing in a real estate company. How should this asset be stressed on my balance sheet?
ChatGPT	<p>→ When it comes to stressing an asset on your balance sheet, such as a real estate investment, it typically involves evaluating its potential downside risks and adjusting its value accordingly. Stress testing helps assess the impact of adverse scenarios on the asset’s performance and value. While I can provide you with a general approach, it’s important to consult with a financial advisor or accounting professional for specific guidance tailored to your situation. Here are some steps you can consider:</p> <ol style="list-style-type: none"> 1. Identify potential stress scenarios: Determine the adverse conditions or events that could significantly impact the real estate market or the specific property you’re invested in. For example, scenarios could include a significant economic downturn, a sudden increase in interest rates, or a decline in property demand. 2. Assess the impact: Analyze how each stress scenario would affect the property’s value, rental income, occupancy rates, and other relevant financial indicators. This assessment may require careful consideration of market data, historical trends, and expert opinions. 3. Quantify the impact: Assign quantitative values to the projected impact of each stress scenario. This can involve estimating the percentage decrease in property value, potential rental income reduction, increased vacancy rates, or other relevant metrics. 4. Adjust the asset’s value: Once you have quantified the impact, make adjustments to the asset’s value on your balance sheet. This adjustment should reflect the potential decrease in the property’s fair market value based on the stress scenarios considered. 5. Disclose assumptions and methodology: Clearly communicate the stress testing assumptions, methodology, and results in your financial statements or accompanying notes. Transparency is crucial for investors and stakeholders to understand the potential risks associated with the asset. <p>Remember that stress testing is a risk management tool, and the specific approach may vary depending on your circumstances, regulatory requirements, and accounting standards applicable to your jurisdiction. It’s crucial to consult with professionals who are well-versed in accounting, financial reporting, and real estate valuation to ensure accuracy and compliance.</p>

Figure 12. Real estate investment prompt to ChatGPT.

¹³<https://www.pinecone.io/>

Although not a poor answer, it is not specific to the domain of the question. When the same question is posed to the LLM with the regulatory knowledgebase, the following response is generated.

Actuary →	I am investing in a real estate company. How should this asset be stressed on my balance sheet?
LLM using regulatory knowledgebase →	Based on the context provided, investments in real estate companies that generate periodic income or which are otherwise intended for investment purposes should be excluded from the calculation of the property risk capital requirement. However, these investments should instead be treated as equity and included in the calculation of the equity risk capital requirement. Therefore, for the purpose of stress on your balance sheet, this investment should not be stressed under the property risk capital requirement but should be included in the equity risk capital requirement.

Figure 13. Real estate investment prompt to ChatGPT supported by the regulatory knowledgebase.

This answer is much more targeted, and the context used was provided by the LLM in a separate response. This demonstrates that the LLM can use the regulatory knowledgebase to provide more accurate and relevant responses to questions.

One might argue that a more appropriate prompt should be used to prime ChatGPT to provide a more relevant response. This is an alternative approach that can be employed, but it requires the user to construct a prompt that will elicit the desired response, which is not always straightforward. Moreover, the responses can still be generic and not specific to the domain of the question, especially for very specific requirements. The regulatory knowledgebase approach allows the user to input a question and receive a response that is specific to the domain of the question, without having to construct a prompt. Below, slightly more context is provided to ChatGPT to see if it can produce a more relevant response.

Actuary →	You are an actuary working in insurance in South Africa, governed by SAM. Answer this question: If I acquire a broker how should I value the asset and how is it stressed? Would it be regarded as a participation or asset holding intermediary?
-----------	---

Figure 14. Broker acquisition prompt in the context of SAM to ChatGPT.

ChatGPT →	As an actuary working in insurance in South Africa, governed by the Solvency Assessment and Management (SAM) framework, I can provide you with some insights on how a broker acquisition would be valued and stressed. However, please note that the SAM framework is subject to specific regulations and guidelines, and it's important to consult with regulatory authorities and actuarial professionals for precise guidance tailored to your situation. Here's a general approach: Valuation of the Broker: <ol style="list-style-type: none"> 1. Determine the purchase price: The valuation process begins by establishing the purchase price for acquiring the broker. This price should reflect the fair value of the assets and liabilities being acquired, including intangible assets like the broker's client base, brand reputation, and goodwill. 2. Consideration of regulatory requirements: SAM may specify specific rules regarding the valuation of brokers. It's important to review the SAM framework, regulatory guidelines, and any specific requirements outlined by the regulatory authorities in South Africa. 3. Engage professionals: Engage independent experts, such as actuaries, valuation specialists, or auditors, who are familiar with SAM and have expertise in valuing insurance intermediaries. They can help assess the fair value of the broker and its various components.
-----------	--

Figure 15. (Continued)

Stressing the Asset:

The stress testing of the broker as an asset would depend on its nature and purpose within your insurance business. In the context of SAM, the broker acquisition could be regarded as either a participation or an asset holding intermediary. Here's a brief explanation of both:

1. Participation intermediary: If the broker is actively involved in underwriting or managing insurance policies, it may be considered a participation intermediary. In this case, stress testing would typically involve assessing the broker's ability to withstand adverse scenarios and its impact on your insurance business. Stress factors could include changes in policyholder behavior, shifts in market conditions, regulatory changes, or operational risks specific to the broker's operations.
2. Asset holding intermediary: If the broker primarily holds and manages insurance-related assets, such as investment portfolios or reinsurance contracts, it may be regarded as an asset holding intermediary. Stress testing in this scenario would focus on evaluating the potential impact of market downturns, credit risks, or changes in investment conditions on the broker's assets and their value.

The specific stress testing methodologies, assumptions, and scenarios will depend on the characteristics and risks associated with the broker's operations. It's crucial to refer to SAM guidelines, seek expert advice, and consider the guidance provided by regulatory authorities to ensure compliance and proper risk management practices.

Please note that the information provided here is general in nature and may not cover all the intricacies of valuing and stressing a broker acquisition under SAM. Consulting with professionals who are well-versed in SAM regulations and insurance valuation practices in South Africa is highly recommended for accurate guidance specific to your situation.

Figure 15. Broker acquisition prompt in the context of SAM to ChatGPT continued.

Although improved, the response is still not specific to the domain of the question. When the same question is posed to the LLM supported by the regulatory knowledgebase (excluding the initial prompt to be domain specific), the following response is generated:

LLM using regulatory knowledgebase	→ The context provided does not provide enough information to answer the question accurately. However, based on the information provided, it is possible that acquiring a broker could be classified as a participation if it meets the definition provided in FSI 1 (Framework for Financial Soundness of Insurers). In this case, the valuation of the participation should follow the requirements outlined in Section 5 of FSI 2.2 (Valuation of Technical Provisions), which states that if a quoted market value is not available, the participation must be valued according to the "adjusted net equity" method. The "adjusted net equity" method requires the value of the participation to be based on the insurer's share of the excess of assets over liabilities less any goodwill or intangible assets not recognised under this Standard. The stressed valuation of the participation would depend on the stress scenarios specified by the Prudential Authority and the insurer's internal risk management framework. Alternatively, if the broker acquisition is not classified as a participation, it may be treated as an asset, and its valuation would depend on the nature of the asset and the requirements outlined in FSI 2.1 (Valuation of Assets, Liabilities and Eligible Own Funds).
--	---

Figure 16. Broker acquisition prompt to ChatGPT supported by regulatory knowledgebase.

The response, although not providing a definitive answer, is much more specific to the domain of the question. It provides references to the relevant sections of the SAM framework, which can be used to obtain more information. It should be noted that these results were obtained using a very basic approach to building a knowledgebase with no additional tuning or optimisation. The results can be improved by using a more sophisticated approach to building the knowledgebase and paying specific attention to the construction of the database itself. Furthermore, it should be noted that the vector database does not fundamentally change the LLM. It simply provides a way to incorporate additional semantic information into the model's inputs or outputs to enhance its results.

Accompanying code is not provided, as the full code is verbose, and not all of it is relevant to this paper (particularly the detailed API calls and front-end JavaScript code). However, creating such a knowledgebase has become significantly easier thanks to open-source contributions. The reader is directed to the Vault-AI¹⁴ GitHub repository, which provides a near-full solution to the problem. One needs only to upload the required documents and provide the necessary API keys.

5.4. Case Study 4: Parsing Reinsurance Documents

Reinsurance contracts are critical documents that are often lengthy and complex, making them time-consuming to analyse and understand. Moreover, they expose the insurer to significant risk if all contracts are not well understood. In this case study, LLMs are used to automate the process of extracting structured data from reinsurance contracts, such as the type of reinsurance, the reinsurer, the reinsured, the coverage period, and the premium amount.

This case study is accompanied by a Python script that employs an LLM to extract information from reinsurance contracts and convert it into a structured JSON format. The script uses the OpenAI API to interact with the LLM and the PyPDF2 library to extract text from PDF documents.

The processes of the script are:

1. The text is extracted from the PDF document.
2. A prompt is constructed using the extracted text and the question: “What is the JSON representation of this reinsurance contract?” and sent to the LLM along with a system prompt that instructs the LLM to return the JSON representation of the contract according to a given schema.
3. The LLM returns the JSON representation of the contract, which is then saved to a JSON file.

The LLM is provided with the schema of the JSON format, which includes all the information typically found in a reinsurance contract. This enables the LLM to accurately extract and structure the contract information. It should be noted that the reinsurance contracts used in this study are simple, fictitious examples created for this paper that could certainly be processed directly without an LLM. However, in practice, the contracts would be more complex and varied, and significantly more information would need to be extracted, making an LLM a more suitable solution. Regardless of the complexity of the contracts, this case study also demonstrates how quickly one can leverage an LLM to perform a complex task with highly accurate results without having to build a complete programme.

Below is an example of the JSON representation of a reinsurance contract generated by the LLM. All original contracts and their generated JSON representations can be found in the GitHub repository accompanying this paper.

```
GPT-3.5 → {
    "treatyType": "Catastrophe Excess of Loss
    Reinsurance",
    "insured": "DEF Insurance Company",
    "reinsurer": "PQR Reinsurance Company",
    "period": {
        "start": "1st July 2023",
        "end": "30th June 2024"
    },
    "lossLayers": [
```

Figure 17. (Continued)

¹⁴<https://github.com/pashpash/vault-ai>

```

{
  "layer": 1,
  "excessOf": 12000000,
  "limit": 6000000,
  "reinsuredPercent": 100
},
{
  "layer": 2,
  "excessOf": 18000000,
  "limit": 12000000,
  "reinsuredPercent": 90
},
{
  "layer": 3,
  "excessOf": 30000000,
  "limit": 18000000,
  "reinsuredPercent": 80
},
{
  "layer": 4,
  "excessOf": 48000000,
  "limit": 24000000,
  "reinsuredPercent": 70
},
{
  "layer": 5,
  "excessOf": 72000000,
  "limit": 30000000,
  "reinsuredPercent": 60
},
{
  "layer": 6,
  "excessOf": 102000000, "limit": 36000000,
  "reinsuredPercent": 50
}
],
"interest": "Covers all lines of business of the
Insured including Property, Liability, and
Aviation.",
"sumInsured": 80000000,
"commission": {
  "percent": 18,
  "maxLossRatio": 75
},
"exclusions": [
  "Earthquake risks",
  "aviation accidents",
  "labor strikes",
  "financial market losses"
],
"claimsNotification": 45,

```

Figure 17. (Continued)

```

    "arbitrationClause": "Any dispute arising from this
    contract shall be settled by arbitration in accor-
    dance with the rules of the International Chamber of
    Commerce.",
    "currency": "EUR"
  }

```

Figure 17. JSON representation of reinsurance treaty generated by GPT-3.5.

6. Impact on the Actuarial Profession and the Broader Insurance Industry

This section discusses the potential implications of using LLMs for actuarial work, such as their impact on the role of actuaries and the broader insurance industry. The primary goal of this paper is to demonstrate the potential of LLMs in the field of actuarial science and insurance and to encourage actuaries to explore the use of LLMs in their work. This paper does not attempt to provide a comprehensive analysis of the impact of LLMs on the actuarial profession and the insurance industry, nor does it consider in detail the risks, ethics, and professionalism issues associated with LLMs. However, a brief discussion of these topics is provided here to highlight their importance and encourage further research.

6.1. Impact on the Role of Actuaries

There is significant concern surrounding the use of AI and LLMs due to fears that they might replace jobs (Bock *et al.*, 2020; McLeay *et al.*, 2021). Indeed, this may be the case in some industries where tasks are repetitive, easily automated, low-risk, and do not require significant human oversight (Bock *et al.*, 2020; McLeay *et al.*, 2021; Vorobeva *et al.*, 2022). However, in actuarial work, it is challenging to identify a role that meets all these criteria. Moreover, considering the professionalism requirements of actuaries and the fiduciary duty insurers have to their policyholders, the author believes it is inadvisable to fully automate roles within the insurance sector. Regardless of the level of automation adopted, the regulation and professional duties and obligations will still lie with the individuals responsible for the process.

The author's perspective is that LLMs will increasingly serve as tools to support actuaries in their endeavours. As illustrated in this paper, the primary benefits are enhancing efficiency, minimising human errors, and enabling actuaries to concentrate on tasks demanding general intelligence and reasoning, which hold significant value for insurers. For instance, an actuary who can save 10 hours by not having to extract information from documents can instead allocate those 10 hours to interpreting their work results, critically evaluating them, and integrating them within the organisation.

The widespread availability of LLMs and associated AI technologies will alter the way work is performed and cannot be ignored. Actuaries, as well as employers of actuaries, will need to ensure they are familiar with the technology and able to work with it. Not only will actuaries have to use, interpret, and understand the results of LLMs, but they will also need to manage the risks they pose. As LLMs and AI technologies become more integrated into actuarial work, actuaries will need to continuously adapt their skillsets and knowledge. This extends beyond simply using the technologies, but more importantly to understanding their limitations, potential biases, and ethical implications.

Actuaries will work in increasingly interdisciplinary teams, necessitating collaboration with professionals from other disciplines, such as data scientists, computer scientists, software engineers, and ethicists. Issues such as data privacy, algorithmic fairness, transparency, and the ethical use of AI will become even more important and impactful. Actuaries not only need to be aware of the changing regulatory landscape in response to AI but also advise and lead the way on appropriate regulation and legislation to manage the risks they pose.

Actuarial workflows may change significantly as a result of LLMs. Previously time-consuming and manual tasks may be automated away, leaving actuaries with time to focus on more complex

tasks. This may result in actuaries spending more time on complex reasoning and problem-solving, or possibly even creativity. The use of LLMs in actuarial work may also have an impact on how actuaries communicate their findings and recommendations to stakeholders. Actuaries will need to ensure that they are able to effectively communicate the results of LLMs, as well as the limitations and uncertainties associated with these models.

6.2. Risks Associated with LLMs

Despite their often surprising and impressive performance, LLMs are not perfect. They are prone to errors, and these errors can be difficult to detect. This paper does not focus on these erroneous results not in an effort to conceal them, but rather because the focus of the paper is on applications and demonstrating them. However, it is important to note that during the course of our research, the LLMs utilised produced several errors that necessitated an iterative approach to improve results. In this regard, we have already reviewed and improved the results unless otherwise indicated (e.g. Appendix C). Presenting the entire process, including erroneous results and other errors to be described, would detract from the purpose of the paper. However, we do discuss some of the risks associated with LLMs here.

LLMs are only as good as the data they are trained on, which is typically large amounts of information gathered from the internet. Thus, LLMs are subject to the same biases and misinformation that humans who produce and consume internet content are exposed to. Hence, actuaries must carefully consider the output of LLMs to ensure they are not unfairly discriminating. In fact, until such a time as further research and investigation into their biases are performed, LLMs should not be used in sensitive areas where bias and discrimination are possible and likely to impact business decisions. On the other hand, LLMs can also be used to help actuaries identify and mitigate unfair discrimination, rather than to perpetuate it.

LLMs have a limited memory, meaning over time, they lose the context of the discussion or conversation and begin to hallucinate. This is more of an issue in assistance use cases where the user interacts conversationally with the LLM and less so with API-based interaction, where the context is normally restricted to one or two prompts. However, it is important to be aware of this limitation and to ensure that the LLM is not used in a way that requires it to remember information over a long period of time.

LLMs can also produce completely incorrect and fake information without a long conversation history. For example, the user can ask for assistance in a coding problem, and the LLM will generate a solution for using programming libraries that simply do not exist. Further, commercial models such as those provided by OpenAI are not pure implementations of the underlying LLM. Additional safety measures are included to make the results safer and more controlled. In fact, these measures have led the performance of ChatGPT to vary considerably over time (Chen *et al.*, 2023).

The use of LLMs in actuarial work may raise concerns about data security and privacy, especially when dealing with sensitive personal or financial information. Actuaries will need to ensure that they are using LLMs in a way that complies with data protection laws and regulations. As LLMs become more integrated into actuarial work, there may be a need for greater transparency and explainability in how these models work and how they arrive at their conclusions. This is particularly important when LLMs are used to make decisions that have significant financial or social implications.

Thus, it is imperative that where LLMs are used, the actuary carefully reviews the LLM itself, the provider, and their alterations and policies governing the LLM, and most importantly, to carefully review and monitor the results over time to appropriately manage the risks LLMs present. The use of LLMs in actuarial work will require ongoing monitoring and validation to ensure that the models are performing as expected and that the results are accurate and reliable. This may include periodic reviews of the models, as well as the development of validation frameworks and methodologies.

6.3. Professionalism

As with all work performed by actuaries, they must adhere to strict professionalism requirements. This includes the requirement to act in the best interests of the insurer and its policyholders and to ensure that their work is accurate and reliable. This is particularly important when using LLMs, as discussed in the prior section.

However, the treatment of risks and professionalism within this paper may not be sufficient for actuaries to fully understand the risks and professionalism requirements associated with LLMs. Rather, the purpose of this section is to simply highlight the discussion and bring forth actuaries' responsibilities.

Several questions arise relating to the use of LLMs within actuarial work. For instance, how can actuaries ensure that the use of LLMs complies with professional standards and ethical guidelines? How can the risks associated with LLMs be effectively managed? How can actuaries ensure that the use of LLMs does not result in unfair discrimination or other adverse outcomes? These are complex questions that require careful consideration and further research. It is hoped that this paper will stimulate discussion and exploration of these important issues.

Finally, a grey area often emerges as to who is responsible for errors resulting from using AI. Viewing LLMs as a tool to assist the actuary, it is believed that there is an onus on the actuary to ensure that the results are accurate and reliable, just as one would when using a tool such as spreadsheet software or a word processor. Appropriate controls and risk mitigation actions should be put in place.

7. Conclusion

This paper has explored the potential applications of LLMs to actuarial work and insurance functions. Through a detailed consideration of the claims process and high-level consideration of other insurance functions, it has been demonstrated how LLMs can be used to improve efficiency and reduce human error. A framework was presented to identify whether an LLM is suitable for a given problem, and the use of LLMs was demonstrated in a number of case studies. Risks associated with LLMs and the professionalism requirements of actuaries when using LLMs were also discussed briefly.

The findings suggest that LLMs can significantly enhance the efficiency and accuracy of actuarial and insurance work.

By leveraging LLMs to improve efficiency and quality of work, while reducing human error, actuaries can focus on tasks that require their unique expertise and judgement. However, it is also recognised that the use of LLMs in actuarial work is not without challenges, which include the risk of errors and biases in the models, the need for careful review and interpretation of the output, and the ethical and professional considerations associated with the use of AI in actuarial and insurance work.

Looking ahead, it is believed that LLMs have the potential to greatly improve the abilities of actuaries and insurance functions both directly and indirectly. It is hoped that this paper will stimulate further research and discussion on this important topic, and actuaries are encouraged to explore the use of LLMs in their own work and to share their applications and experiences with the actuarial community.

In conclusion, while LLMs will not fundamentally change the way actuaries work, they can enhance the work of actuaries and the insurance industry as a whole, so long as professionalism requirements are adhered to closely and actions are taken in the best interests of the insurer and its policyholders.

Acknowledgements. Great appreciation is extended to Ronald Richman for valuable discussion, review, mentorship, and support for this paper and in general. Thanks are also extended to Anton Gerber for his detailed and careful review of the first draft of this paper and for his valuable feedback and comments which were incorporated to improve the overall quality of the paper.

Acknowledgement is given to the assistance of ChatGPT in the writing process of this paper. ChatGPT provided valuable assistance in drafting, editing, and reviewing the manuscript.

References

- Allal, L. B., Li, R., Kocetkov, D, Mou, C., Akiki, C., Ferrandis, C. M., Muennighoff, N., Mishra, M., Gu, A., Dey, M., Umapathi, L. K., Anderson, C. J., Zi, Y., Poirier, J. L., Schoelkopf, H., Troshin, S., Abulkhanov, D., Romero, M., Lappert, M., Toni, F. D., del Río, B. G., Liu, Q., Bose, S., Bhattacharyya, U., Zhuo, T. Y., Yu, I., Villegas, P., Zocca, M., Mangrulkar, S., Lansky, D., Nguyen, H., Contractor, D., Villa, L., Li, J., Bahdanau, D., Jernite, Y., Hughes, S., Fried, D., Guha, A., de Vries, H. & von Werra, L. (2023). SantaCoder: don't reach for the stars! *arXiv preprint arXiv:2301.03988*, available at <https://arxiv.org/abs/2301.03988> (accessed 14 August 2023).
- BigScience Workshop, et al. (2023). Bloom: a 176b-parameter open-access multilingual language model, available at <https://arxiv.org/abs/2211.05100> (accessed 14 August 2023).
- Biswas, S. (2023). Using ChatGPT for insurance: current and prospective roles. *SSRN Electronic Journal*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4405394 (accessed 14 August 2023).
- Bock, D. E., Wolter, J. S. & Ferrell, O. C. (2020). Artificial intelligence: disrupting what we know about services. *Journal of Services Marketing*, 34, 317–334. https://www.researchgate.net/publication/340496869_Artificial_intelligence_disrupting_what_we_know_about_services.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. & Amodei, D. (2020). Language models are few-shot learners, available at <https://arxiv.org/abs/2005.14165> (accessed 14 August 2023).
- Chen, L., Zaharia, M. & Zou, J. (2023). How is ChatGPT's behavior changing over time?, available at <https://arxiv.org/abs/2307.09009> (accessed 14 August 2023).
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019). Bert: pre-training of deep bidirectional transformers for language understanding, available at <https://arxiv.org/abs/1810.04805> (accessed 14 August 2023).
- Dimri, A., Paul, A., Girish, D., Lee, P., Afra, S. & Jakubowski, A. (2022). A multi-input multi-label claims channeling system using insurance-based language models. *Expert Systems with Applications*, 202, 117166. <https://www.sciencedirect.com/science/article/abs/pii/S0957417422005553>.
- Fan, L., Li, L., Ma, Z., Lee, S., Yu, H., & Hemphill, L. (2023). A bibliometric review of large language models research from 2017 to 2023, available at <https://arxiv.org/abs/2304.02020> (accessed 14 August 2023).
- Hofert, M. (2023). Assessing ChatGPT's proficiency in quantitative risk management. *SSRN Electronic Journal*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4444104 (accessed 14 August 2023).
- Hofert, M. (2023). Correlation pitfalls with ChatGPT: would you fall for them? *SSRN Electronic Journal*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4448522 (accessed 14 August 2023).
- McLeay, F., Osburg, V. S., Yoganathan, V. & Patterson, A. (2021). Replaced by a robot: service implications in the age of the machine. *Journal of Service Research*, 24, 104–121. <https://core.ac.uk/download/pdf/326253299.pdf>.
- Min, B., Ross, H., Sulem, E., Pouran, A., Veyseh, B., Nguyen, T. H., Sainz, O., Agirre, E., Heinz, I., & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: a survey, available at <https://arxiv.org/abs/2111.01243> (accessed 14 August 2023).
- OpenAI (2023). Gpt-4 technical report, available at <https://arxiv.org/abs/2303.08774> (accessed 14 August 2023).
- Troxler, A. & Schellendorfer, J. (2022). Actuarial applications of natural language processing using transformers: case studies for using text features in an actuarial context. *arXiv e-prints*, page arXiv:2206.02014, available at <https://arxiv.org/abs/2206.02014> (accessed 14 August 2023).
- Vorobeva, D., El Fassi, Y., Pinto, D. C., Hildebrand, D., Herter, M. M., & Mattila, A. S. (2022). Thinking skills don't protect service workers from replacement by artificial intelligence. *Journal of Service Research*, 25, 601–613. https://www.researchgate.net/publication/360795054_Thinking_Skills_Don%27t_Protect_Service_Workers_from_Replacement_by_Artificial_Intelligence.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT, available at <https://arxiv.org/abs/2302.11382> (accessed 14 August 2023).
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y. & Wen, J.-R. (2023). A survey of large language models, available at <https://arxiv.org/abs/2303.18223> (accessed 14 August 2023).

Appendix A. Accompanying Code Repository

Accompanying this paper is a code repository containing the code used in the case studies presented in this paper. The repository can be accessed at <https://github.com/cbalona/actuarygpt-code>.

Appendix B. Definitions

API	An API (Application Programming Interface) is a set of rules and protocols that allows different software applications to communicate and interact with each other. It defines a set of methods, functions, and data structures that enable developers to access and use the functionality of a software library, framework, or service. APIs provide a standardised way for applications to exchange data, request services, and perform operations, making it easier to integrate and build software systems. Actuaries can use LLMs provided by third parties through APIs without needing to understand the underlying model architecture or implementation details. APIs are typically billed based on usage, and the cost can become prohibitive for high-volume usage.
Context	Context refers to the surrounding information or environment that influences the interpretation and understanding of prompts provided to the LLM. It can include factors such as previous statements, relevant background information, user preferences, system state, or additional data explicitly provided. When we refer to loading additional data as context, this could mean providing the context as previous prompts, as system prompts, or as embeddings stored in an external vector database. See the definition of embeddings and vector database for more clarification.
Embeddings	Embeddings refer to the representation of words, phrases, or sentences as numerical vectors in a high-dimensional space. These vectors are derived from the underlying semantic or contextual information of the text. Embeddings capture the meaning, relationships, and semantic similarity between words or textual elements. Embeddings allow one to handle language data using a mathematical approach. See vector databases for a use case.
Fine-tuning	Fine-tuning is a machine learning technique used to adapt a pre-trained model to a specific task or domain. It involves taking a pre-existing model that has been trained on a large dataset and further training it on a smaller, task-specific dataset. Fine-tuning allows the model to learn task-specific patterns and improve its performance on the target task. It is commonly used in transfer learning, where knowledge gained from pre-training on a related task or domain is leveraged to accelerate learning and achieve better performance on a specific task. Certain forms of LLMs can be fine-tuned to improve their performance on domain-specific tasks.
JSON	JSON (JavaScript Object Notation) is a lightweight data interchange format. It is a text-based format that is easy for humans to read and write, and it is also easy for machines to parse and generate. JSON represents structured data as key-value pairs and supports various data types. It is widely used for data serialisation and communication between different systems, making it a popular choice for web APIs and data storage. JSON has become an essential component in modern software development, allowing for efficient and interoperable data exchange.
Natural Language Processing (NLP)	A subfield of AI aimed at enabling computers to understand, interpret, and generate human language. A wide range of techniques and methods exist within NLP for processing and analysing textual data, most notably, named entity recognition (identifying people, places, and things in text), sentiment analysis, machine translation (translating between human languages using machines), and text generation.

(Continued)

(Continued)

NoSQL	NoSQL is a term used to describe non-relational database management systems that provide an alternative to traditional SQL-based relational databases. NoSQL databases are designed to handle large amounts of unstructured and semi-structured data, offering flexibility and scalability for modern applications. They employ various data models, such as key-value, document, columnar, or graph, to efficiently store and retrieve data. NoSQL databases are commonly used in web applications, big data analytics, and real-time streaming applications.
Sentiment Analysis	A natural language processing technique used to analyse and understand the sentiment or emotion conveyed in a piece of text, such as customer reviews, social media posts, etc. It involves automatically classifying the sentiment expressed in the text as positive, negative, or neutral for use in downstream systems.
Stack	A technology stack, often simply called a “stack”, refers to a set of technologies, tools, software, and frameworks used together to build and run a single application or solution. It encompasses both the front-end (or client-side) components that a user interacts with directly, as well as the back-end (or server-side) components that work behind the scenes.
Structured Data	Structured data refers to data that is organised and formatted according to a predefined model or schema. It follows a consistent and well-defined structure, often represented in the form of tables, columns, and rows. Structured data is highly organised, making it easily searchable, analysable, and processable by computer systems. Examples of structured data include data in relational databases, spreadsheets, and standardised file formats like CSV or XML. Text data, which is typically considered unstructured, can be stored in a structured way.
Token	A token (in the context of large language models) refers to a single unit of text, typically a word or a character. It is a fundamental element used for text analysis and processing tasks such as parsing, part-of-speech tagging, and language modelling. Tokens are created by segmenting or breaking down a text into meaningful units for further analysis. One word does not necessarily correspond to one token due to how words are deconstructed for tokenisation. On average every 1,000 tokens represents 750 words.
Unstructured Data	Unstructured data refers to data that does not have a predefined structure or format. It does not conform to a specific data model or schema, making it more challenging to organise and analyse compared to structured data. Unstructured data can take various forms, such as text documents, social media posts, emails, audio and video recordings, images, and sensor data. Extracting meaningful insights from unstructured data often requires advanced techniques like natural language processing, text mining, or machine learning algorithms.
Vector Database	A vector database is a type of database that is specifically designed to store and manipulate vector data efficiently. Vector data refers to collections of vectors, where each vector represents a data point or an object in a high-dimensional space. Vector databases are optimised for fast vector similarity search and retrieval operations. They typically support various similarity metrics, such as cosine similarity or Euclidean distance, to measure the similarity between vectors. Storing embeddings in a vector database allows one to query large knowledgebases to find documents that are similar to the content of the prompt provided and hence reduce the context needed to be provided to the LLM.

Appendix C. Excerpts of Assistant Conversations

This appendix contains excerpts of conversations between an actuary and ChatGPT. The 24 May 2023 version of the ChatGPT website¹⁵ was accessed using the GPT-4 model.

C.1. Coding Assistant

The actuary is seeking assistance with their IBNR reserving work and requires assistance with the code and thinking.

Actuary	→ I'm seeking general assistance with my IBNR reserving work as an actuary in a direct lines insurance company. I would like you to collaborate with me and help me write my code while addressing various aspects of my code implementation and adjustments related to external factors. I would also appreciate it if you could search the web for possible external factors I may need to consider, or I can provide specific external factors for you to consider. Python is my preferred language, and I would greatly appreciate a combination of code snippets and guidance to help me achieve my goals. Please provide the general assistance I'm seeking for my IBNR reserving work, and let's embark on our collaborative coding journey. I am familiar with IBNR reserving and would prefer a simple back and forth discussion with direct responses to my queries. On your first response, ask me any clarification questions you have and following my answers you can await my requests.
ChatGPT	→ That sounds great! I'm here to assist you with your work. Before we proceed, I'd like to clarify a few things: <ol style="list-style-type: none"> 1. Could you please specify the current approach you're using for your IBNR (Incurred But Not Reported) reserving? Are you using chain-ladder method, Bornhuetter-Ferguson method, or some other method? 2. What kind of data are you working with? Could you provide an idea about the structure of your data? 3. What kind of external factors are you already considering in your current model? Are there specific external factors you're interested in exploring? 4. Are there any specific challenges or issues you're currently facing with your IBNR reserving work that you'd like to address? 5. Finally, could you tell me about the software or libraries you're currently using in Python for your actuarial work? <p>This information will help me provide the most accurate and effective assistance.</p> <p style="text-align: center;">...</p>
Actuary	→ I have Python installed. Please write some code to use the BF method and derive an initial IBNR.

Figure 18. Coding assistant conversation with ChatGPT part 1.

ChatGPT	→ Certainly. To start with the Bornhuetter-Ferguson method using the Python 'chainladder' library, we'll first need to import the required libraries and load your data. Since I don't have access to your specific data, I'll use a generic example to illustrate the process. Before proceeding, make sure you've installed the necessary Python libraries. If you haven't done so already, you can install them using pip: <pre>!pip install pandas chainladder</pre>
ChatGPT	→ Here's a simple script to load a dataset and apply the Bornhuetter-Ferguson method: <pre>import pandas as pd import chainladder as cl # Load your data # Replace this with the path to your CSV file data = pd.read_csv('path_to_your_file.csv') # Convert data to Triangle format</pre>

Figure 19. (Continued)

¹⁵<https://chat.openai.com/>

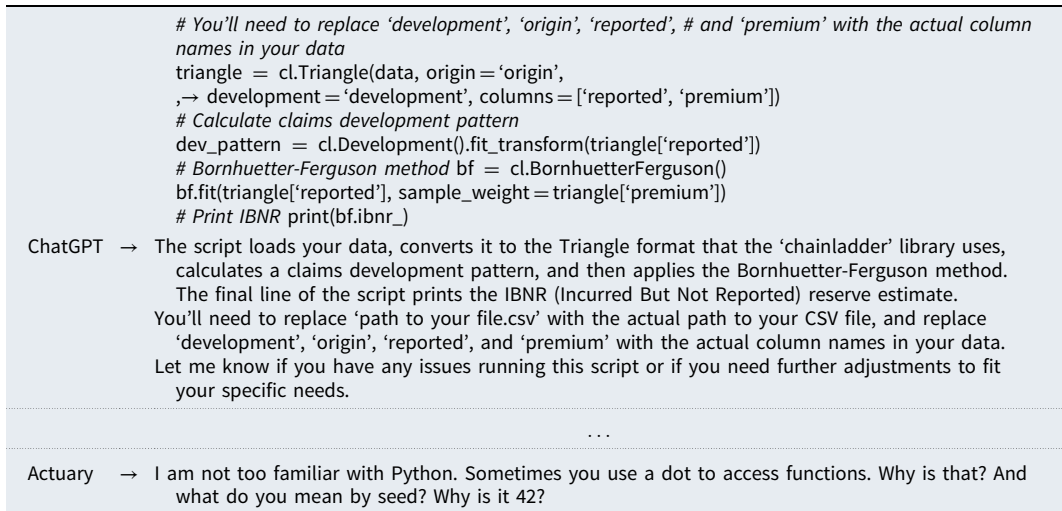


Figure 19. Coding assistant conversation with ChatGPT part 2.

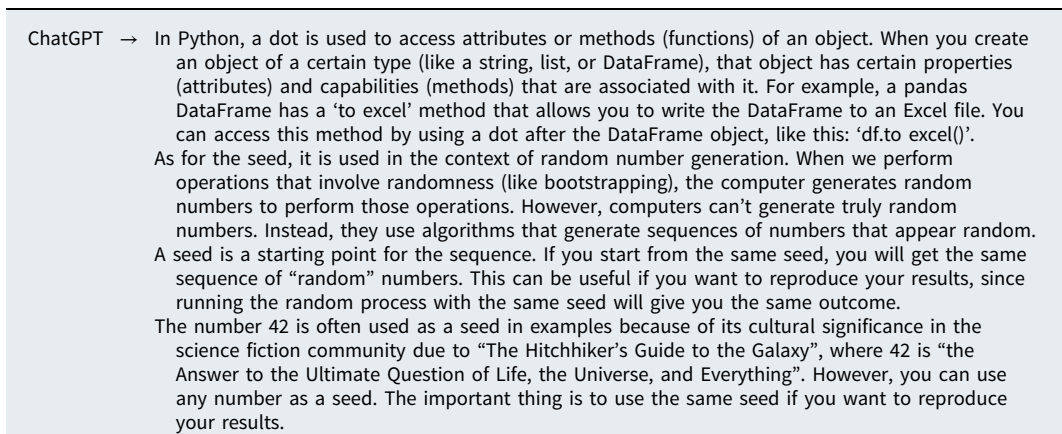


Figure 20. Coding assistant conversation with ChatGPT part 3.

C.2. Model Understanding and Development

The actuary queries the LLM on an unfamiliar model and how to optimise it.

Actuary	→	Briefly explain XGBoost and its 3 most important parameters. Then suggest a search space over which to search for optimal parameters.
ChatGPT	→	<p>XGBoost Overview</p> <p>XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm based on the gradient boosting framework. It is widely used for its speed and performance in both classification and regression tasks. XGBoost combines the predictions of multiple weak learners (usually decision trees) to create a strong predictive model.</p> <p>Three Most Important Parameters</p> <ol style="list-style-type: none"> 1. n estimators (or num boost round): Description: This parameter specifies the number of boosting rounds (i.e., the number of trees to be added to the ensemble). More trees can increase the model's ability to learn but may also lead to overfitting. Typical Range: 100 to 1000 or more, depending on the dataset size and complexity. 2. learning rate (or eta): Description: The learning rate controls the contribution of each tree to the final model. A lower learning rate makes the model more robust to overfitting but requires more boosting rounds. Typical Range: 0.01 to 0.3. 3. max depth: Description: This parameter determines the maximum depth of each tree. Deeper trees can model more complex relationships but are more prone to overfitting. Typical Range: 3 to 10. <p>Suggested Search Space for Optimal Parameters</p> <p>To find the optimal parameters, you can use techniques like grid search or random search within the following ranges:</p> <ol style="list-style-type: none"> 1. n estimators: Range: 100 to 1000 (e.g. [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000]) 2. learning rate: Range: 0.01 to 0.3 (e.g. [0.01, 0.05, 0.1, 0.2, 0.3]) 3. max depth: Range: 3 to 10 (e.g. [3, 4, 5, 6, 7, 8, 9, 10]) <p>You can also include other parameters such as subsample, colsample bytree, and gamma for a more comprehensive search. Here's a broader example search space:</p> <pre>param_grid = { 'n_estimators': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000], 'learning_rate': [0.01, 0.05, 0.1, 0.2, 0.3], 'max_depth': [3, 4, 5, 6, 7, 8, 9, 10], 'subsample': [0.6, 0.7, 0.8, 0.9, 1.0], 'colsample_bytree': [0.6, 0.7, 0.8, 0.9, 1.0], 'gamma': [0, 0.1, 0.2, 0.3, 0.4, 0.5] }</pre>

Figure 21. Model Understanding Prompt.

C.3. Problem-Solving

The actuary has a unique problem, and requires their solution to be reviewed.

Actuary	→	I am an actuary working in consulting. I have been tasked with a unique problem. My client requires a premium for a policy that covers risks arising from erroneous and fake information being generated by LLMs being used to make business decisions. Obviously, no data exists. My proposed approach is to give the LLM a list of questions and mark the answers to see how often it gets it wrong. Is this a good solution?
ChatGPT	→	Your approach is an interesting starting point and aligns with the general principle of quantifying the risk, which is essential for pricing a premium. However, there are a few potential challenges to consider:

Figure 22. (Continued)

Subjectivity in Error Assessment: The first challenge is the subjective nature of what might constitute an “error”. For instance, the language model might provide information that is technically accurate but misleading or irrelevant in context. So, defining and consistently applying what constitutes an error is crucial.

Dynamic Nature of LLMs: The LLMs are trained on large corpora of data, and the information they generate is based on the input they receive. Hence, the “error” rate might vary significantly depending on the type of question and how it’s phrased. Moreover, the models can be updated or improved over time, affecting their performance.

Data Representativeness: If you’re using a small set of questions to estimate the error rate, this might not be representative of all the possible types of queries and their corresponding error rates.

Complexity and Scope of Inquiries: The error rate might vary drastically between simple fact-check questions and complex, nuanced queries. It would be necessary to consider this in your risk assessment.

Potential Consequences: The risk isn’t just about the frequency of errors but also the severity of the consequences when errors do occur. For instance, an error that leads to a small operational inefficiency could be much less risky than an error that leads to a significant strategic mistake.

Given these challenges, you may want to consider a more comprehensive approach. You could augment your original idea of testing the LLM with a variety of questions, but also include scenario analysis to evaluate the potential impact of errors.

Moreover, you could include expert judgement, possibly from people who have used these systems in a similar business context, to provide insights on potential pitfalls, common errors, and their consequences. You may want to assess the quality of the model’s responses in different categories, such as factual accuracy, ambiguity, relevancy, etc.

Another important factor is to consider risk mitigation strategies your client could have in place, like manual checks and balances, oversight committees, validation from multiple models, etc., which could significantly reduce the risks and thus, the premium.

Finally, given the lack of historical data, it might be appropriate to propose a dynamic premium that can be adjusted as more data about the risk becomes available, which also aligns with the principle of continuous learning in the field of AI.

Remember, in such a nascent field, your judgement as an actuary, backed by robust qualitative and quantitative assessment, is of great value. Always engage in a clear and open discussion with your client about the methodology and inherent uncertainties involved.

Figure 22. Problem-solving conversation with ChatGPT.

C.4. Few-Shot Learning Demonstration

This section demonstrates few-shot learning to prime the LLM to produce an answer in line with the specified requirements. First, a conversation is presented with few-shot learning, where three examples of the expected output are provided. Following this, a zero-shot example is shown for comparison.

C.4.1. Few-shot learning example

Actuary → I will give you an economic scenario and you will generate the possible impact on an insurance company. The purpose of this is to define stress and scenario tests to assess the resilience of an insurance company to economic events. I will begin with 3 examples. Wait for my prompts with additional scenarios. Acknowledge this prompt with “Yes”. Do not produce additional scenarios.

Scenario:

A major financial depression occurs globally triggered by a collapse of major equity markets.

Impacts:

- Premium: Insurers experience reduced gross premium revenue as policyholders cancel policies to save money, and new policy sales are reduced.
- Claims: Loss ratios increase due to increased fraud as policyholders attempt to use their insurance policies to make ends meet, as well as increased claims such as business interruption and credit insurance.
- Reinsurance: Reinsurance premiums increase due to hardening market, and availability of reinsurance decreases.
- Operational expenses: Operational costs increase and availability of reinsurance decreases.

Figure 23. (Continued)

- Assets: Market value of investment portfolios decrease and investment incomes decrease. Investment incomes decrease due to reduced asset values.
- Regulatory: Solvency ratios decrease and liquidity is strained.

 Scenario: Rampant inflation in primary insurance market.

Impacts:

- Premiums: Premium increases are limited due to over-extended policyholders leading to reduced profits. Lapses increase as policyholders reduce costs.
- Claims: Inflation leads to increased claims and claims handling expenses.
- Operational expenses: Operational costs increase beyond premium increases.
- Regulatory: Mismatched assets and liabilities lead to solvency and liquidity concerns.

Scenario: Improved political relationship with business leads to improved economic growth and more disposable cash for policyholders.

Impacts:

- Premium: Premium volumes increase as policyholders take up insurance to cover new purchases, lapse rates decrease.
- Claims: Claims reduce in certain categories, such as unemployment claims.
- Assets: Market value of equities increase significantly. Investment income increases.
- Regulatory: Regulatory capital requirements increase due to larger volumes and market values.

ChatGPT → Yes.

Figure 23. Few-shot learning example with ChatGPT part 1.

Actuary → Scenario: The Russia-Ukraine war continues and evolves into a conflict between Western and Eastern powers. Focus the impact on a South African insurer.

ChatGPT → Scenario: The Russia-Ukraine war escalates into a broader conflict between Western and Eastern powers, with South Africa caught in the geopolitical crossfire. Impacts on a South African insurer:

Premium:

- Foreign Exposure: Insurers with exposure to the affected regions (through international branches or policyholders) may see a decline in premium revenue due to the instability and reduced economic activity.
- Local Market: The geopolitical tensions might lead to reduced foreign investments in South Africa, potentially slowing down the economy. This could result in a decrease in the sale of new policies, especially for businesses that rely on international trade.
- Currency Fluctuations: The South African Rand (ZAR) might experience volatility due to global uncertainties, affecting the pricing of insurance products, especially those linked to international assets or liabilities.

Claims:

- Political Risk Insurance: Companies with operations in the affected regions might file claims under political risk insurance due to expropriation, political violence, or inability to repatriate funds.
- Trade Disruption: Businesses in South Africa that rely on imports or exports to/from the affected regions might file claims under trade disruption or supply chain insurance.
- Travel Insurance: With heightened geopolitical tensions, there might be an increase in travel insurance claims due to cancellations, delays, or emergencies for South Africans traveling abroad.

Reinsurance:

- Reinsurance Costs: The global reinsurance market might see increased premiums due to the heightened risk environment, leading to higher costs for South African insurers.
- Capacity: Some reinsurers might reduce their exposure to regions perceived as high risk, including areas affected by the geopolitical tensions, potentially limiting the availability of reinsurance for South African insurers.

Operational expenses:

- Risk Management: Insurers might need to invest more in risk management and scenario planning to navigate the uncertain environment.

Figure 24. (Continued)

<ul style="list-style-type: none"> • Supply Chain: Operational costs might increase if insurers rely on services or products from the affected regions, leading to potential supply chain disruptions. <p>Assets:</p> <ul style="list-style-type: none"> • Investment Portfolio: The value of investments, especially those tied to global markets or the affected regions, might decrease due to market volatility. This could lead to reduced investment income. • Currency Impact: Fluctuations in the ZAR against major currencies might impact the value of foreign investments.
--

Figure 24. Few-shot learning example with ChatGPT part 2.

<p>ChatGPT → Regulatory:</p> <ul style="list-style-type: none"> • Capital Requirements: Regulators might increase capital requirements for insurers to ensure they remain solvent amidst the heightened risks. • Risk Assessment: Regulatory bodies might require insurers to conduct more rigorous risk assessments and stress tests given the geopolitical situation. <p>In summary, a South African insurer would face challenges on multiple fronts due to the escalating Russia-Ukraine conflict and its broader implications. The insurer would need to be agile, revisiting its risk management strategies, investment portfolios, and operational plans to navigate the evolving situation.</p>

Figure 25. Few-shot learning example with ChatGPT part 3.

C.4.2. Zero-shot learning example

<p>Actuary → I will give you an economic scenario and you will generate the possible impact on an insurance company. The purpose of this is to define stress and scenario tests to assess the resilience of an insurance company to economic events. I will begin with 3 examples. Wait for my prompts with additional scenarios. Acknowledge this prompt with "Yes". Do not produce additional scenarios.</p>
<p>ChatGPT → Yes.</p>
<p>Actuary → Scenario: The Russia-Ukraine war continues and evolves into a conflict between Western and Eastern powers. Focus the impact on a South African insurer.</p>
<p>ChatGPT → Certainly. Here's an analysis of the potential impact on a South African insurer given the scenario:</p> <ol style="list-style-type: none"> 1. Investment Portfolio Impact: <ul style="list-style-type: none"> • Equity Market Volatility: Global conflicts can lead to significant equity market volatility. South African insurers, like many others, have a portion of their assets invested in equities. A decline in equity markets can lead to a decrease in the value of their investment portfolio. • Interest Rate Fluctuations: Wars can lead to unpredictable movements in interest rates. If rates rise rapidly, it might impact the value of bonds in the insurer's portfolio. • Currency Volatility: The South African Rand (ZAR) might experience volatility due to global uncertainties. If the insurer has foreign investments, currency fluctuations can affect the value of those assets. 2. Underwriting Risks: <ul style="list-style-type: none"> • Political Risk Insurance: If the insurer offers political risk insurance, claims could rise due to the increased geopolitical risks. • Trade Credit Insurance: Companies trading with entities in the conflict zones might face defaults. If the insurer offers trade credit insurance, they might see an uptick in claims. • Travel Insurance: If the insurer offers travel insurance, there might be an increase in claims due to trip cancellations or disruptions related to the conflict.

Figure 26. Zero-shot learning example with ChatGPT part 1.



Figure 27. Zero-shot learning example with ChatGPT part 1.