Proceedings Paper:
Mostert, R.W. and Altahhan, A. (2025) Self-Refining Language Models to Assess Insurance Claims. In: 2025 International Joint Conference on Neural Networks (IJCNN). 2025 International Joint Conference on Neural Networks (IJCNN), 30 Jun - 05 Jul 2025, Rome, Italy. IEEE. ISBN: 979-8-3315-1043-5. ISSN: 2161-4393. EISSN: 2161-4407.

https://doi.org/10.1109/ijcnn64981.2025.11228917

Self-Refining Language Models to Assess Insurance Claims

1st Rayno Willem Mostert
School of Computing
University of Leeds
London, United Kingdom
rmostert@me.com

2nd Abdulrahman Altahhan
School of Computing
University of Leeds
Leeds, United Kingdom
a.altahhan@leeds.ac.uk

***Abstract*—This research investigates the application of Large Language Models (LLMs) for automating the assessment of bicycle insurance claims. The study aims to determine the effectiveness of LLMs in interpreting insurance policy documents and evaluating the validity of claims. The ultimate objective being to enhance consistency and efficiency in claims processing by leveraging LLM capabilities to understand complex legal language. The methodology involves training and testing various models on a bicycle insurance policy and hypothetical claims against the policy. Key achievements include the introduction of an agent-based approach that significantly improves the acceptability of written claims handler responses. The findings indicate that, by using a combination of agents with crafted prompts, LLMs can produce acceptable claims handler responses to claim descriptions.**

***Index Terms*—language models, insurance technology, automated claims processing, customer service automation, artificial intelligence**

I. Introduction

Claims Assessment involves evaluating the validity of insurance claims, relying on dense legal language and requiring precise understanding. It's an expensive bottleneck for insurers, often handled by third-party administrators in the UK. Assessors must interpret policy wording to determine how policy conditions apply to claims.

Large Language Models (LLMs) could understand insurance policy wording and claim descriptions. This research assesses whether LLMs can automatically evaluate claims and provide responses to claimants. Leveraging LLMs could automate claim validity assessment, leading to more consistent decisions, reducing human variability, and decreasing training time for new assessors. This project investigates how LLMs can enhance and expedite this process, benefiting the insurance industry, customers, and potentially other contractual disputes.

II. Literature Review

A. Language Models for Contract Understanding

LLMs have effectively informed decisions based on clinical records for health insurance claims [1] and can generate relevant responses to extensive legal texts [2]. These capabilities are crucial for insurance claim assessment, requiring an understanding of claim circumstances, policy nuances, regulatory frameworks, and legal precedents to accurately assess claim validity.

Similar studies in healthcare insurance have been successful. Aracena et al. used NLP for automated decisions on patient coverage for labour-related illness [3], finding fine-tuned Pre-trained Language Models (PLMs) particularly effective.

In other domains, LLMs have demonstrated their ability to navigate complex legal documents. For example, Arbel and Hoffman showed these models can obtain "the same answers at lower costs and with greater certainty" compared to legal practitioners [4].

B. Libraries and Frameworks

At the time, Langchain [5] was the primary library for building LLM systems, offering prompt-chaining tools crucial for the proposed agent-based system. Combined with Python [6], it provided a robust framework for implementing and testing LLMs. Additionally, the study used Ollama [7] to run smaller language models locally.

III. Methodology

A. Language Models

The study compared four models for claims handling:

- **GPT 4o** [8]: Excels in large-scale text processing.
- **GPT 4o-mini** [9]: A smaller, cost-effective version of GPT 4o.
- **Llama 3.1 8B** [10]: An open-source model runnable locally in its 8B form.
- **Mistral Nemo** [11]: A top-performing small open-source model runnable locally.

B. Agent-Based System

To enhance response quality, we implemented a self-improvement scheme assigning PLMs roles that mimic a real-world quality assessment process. This involves three main actors: the claim handler (processes the initial claim), the disgruntled customer (provides critical feedback), and the independent adjudicator (evaluates and suggests improvements).

1) Agent: Claims Handler:
The claims handler agent design is influenced by Chain-of-Thought prompting [12]. Single prompts were insufficient with the tested PLMs, so a series of chained prompts was utilized. The agent generates arguments both for and against a claim's validity to reach a final verdict.

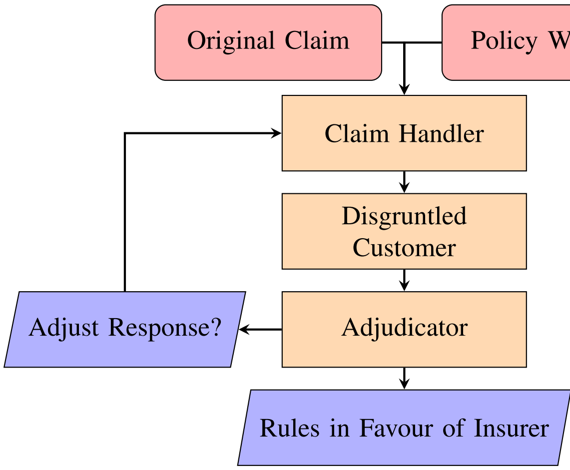

Fig. 1. A schematic providing a high-level overview of the agent-based system.

- Claim descriptions and policy documentation are provided to the model.
- The PLM first generates reasons the claim might be invalid (Prompt 1).
- Next, it produces counterarguments supporting the claim's validity based on the previous output (Prompt 2).
- These arguments are used to decide the claim's validity and formulate a response (Prompt 3).
- Finally, the model composes an email to the customer based on the verdict.

If the adjudicator has recommended that the insurer should change their response to the customer, then this is incorporated into the prompt.

Prompt 1

Write a response based on the policy documentation provided below:
<context>

A claimant has submitted the following claim:
<claim>

Explain if there are any reasons why the claim may not be valid.

Prompt 2

<context, claim>

Another agent has come up with a set of reasons for why the claim might not be valid. These are those reasons:
<invalid claim arguments>

Carefully consider these reasons, and the policy documentation and then write up a set of counterarguments as to why the claim should be considered valid.

Prompt 3

<context, claim, invalid claim arguments>

Another agent has come up with a set of counterarguments as to why the claim should be considered valid. These are their reasons:

<valid claim counterarguments>

IF ADJUDICATOR FEEDBACK AVAILABLE:

Separately, an adjudicator has made the following recommendations:
<additional feedback>

END IF

Carefully consider the reasons for and against the claim's validity. Compare these against the policy documentation and then write up a final verdict. If there are any reasons why the claim may not be valid, explain them. If there are any further documentation or information that you need to process the claim, request it.

Prompt 4

<context, claim>

An agent has come up with a final verdict on the current status of the claim:

<verdict>

Write a professional email body to the claimant informing them of the final verdict, and requesting them to provide additional information if the verdict requires it.

2) Agent: Disgruntled Customer:
A single disgruntled customer prompt is used to respond to the claims handler's email, aiming to identify any elements in the email or verdict that a potential customer might object to.

Prompt 5

Write a conclusion based only on the policy documentation provided below:
`<context>`

You have submitted the following claim:
`<claim>`

The insurance company has sent you the following response:
`<response>`

You are annoyed with the response from the insurer. Write a response to the insurance company, explaining which parts of their request you either do not agree with or you think is unreasonable.
Do not include any new information in your response, or provide any new documents. Only use the information provided in the policy documentation and the claim.

3) Agent: Adjudicator:
The independent adjudicator is an agent designed to provide feedback to the claims handler in order to improve the design of the response. This is a single prompt designed to identify elements in the disgruntled customer response that should be used to improve the response.

Prompt 6

You are a financial ombudsman adjudicator. The relevant policy documentation is provided below:
`<context>`

A claimant has submitted the following claim:
`<claim>`

An insurance company has sent the customer the following email:
`<response>`

The claimant has not responded positively, with the following email:
`<customer response>`

Decide whether the claimant has valid grievance and write a recommendation to the insurance company. The recommendation should be based on the policy documentation and the emails provided.

IV. Experimental Results

The results discussion follows a representative sample of the generated results from the implementation described in the Methodology.

1) Claim:
A simple claim description was selected, representative of what can be expected as a message from a potential claimant:

Example Submission

`<claim>`:
I want to claim for a stolen bike insured on my policy with you. It was securely attached to a bike rack at my office.

`<context>`:
The **Sundays Insurance Policy Wording** [13] was provided to the model.

2) Agent: Claim Handler:
The result of Prompt 3 (generating a final verdict) is included as Example Response 1.

Generated Response 1

`<final verdict>` (*model-generated*):

Final Verdict: At this stage, your claim's validity hinges on your ability to provide additional documentation and evidence that meet the policy's requirements.
Required Information and Documentation To proceed with the assessment and processing of your claim, please provide the following:

- **Proof of Secure Locking**: Evidence that the bike was securely locked through the frame and all quick release wheels to an immovable object. This could include photographs or witness statements.
- **Evidence of Forced Entry**: The remains of the approved lock or other evidence indicating that the lock was forcibly broken. A police report indicating the nature of the theft would be beneficial.
- **Proof of Ownership**: The original purchase receipt or similar proof of purchase for the bike.
- **Confirmation of Unattended Time**: Confirmation that the bike was not left unattended for more than 18 hours. Details on when the bike was last seen would help clarify this.

The response correctly identifies the requirements set out in the policy wording and lists the evidence required to process the claim. The response is generally well-considered, but has one problematic element. The agent response asks for photographs or witness statements as evidence that the bicycle was locked correctly. From a practical perspective, this would be impossible for the claimant to obtain.

3) Agent: Disgruntled Customer:
Next, the disgruntled customer agent is prompted (Prompt 5) to devise reasons not to comply with the email generated by the claims handler.

The generated example response highlights the impracticality of providing photographs or witness statements as evidence that the bicycle was locked correctly. It also takes issue with the ask for the remains of the approved lock - which is a policy requirement. There are some other points raised, particularly around the clarity of the evidence requirements confirming the time the bicycle was left unattended.

Generated Response 2

<customer response> (*model-generated*):

I am writing to express my concerns regarding your response to my recent claim for my stolen bike. I find several aspects of your request for additional information to be unreasonable.

- **Proof of Secure Locking**: Your request for evidence that the bike was securely locked to an immovable object is noted. However, it is important to point out that the policy documentation does not explicitly state that photograph evidence or witness statements are mandatory for claims processing. While I understand the need for verifying compliance with the security conditions, expecting photographic evidence may not always be practical or feasible.
- **Evidence of Forced Entry**: You have asked for the remains of the approved lock or evidence indicating that the lock was forcibly broken, along with a police report. While I acknowledge the importance of such evidence, the requirement for the remains of the lock can be overly stringent, especially if the remains are not recoverable. Furthermore, the policy documentation mentions providing a crime reference number, which should suffice in corroborating the theft incident.
- **Confirmation of Unattended Time**: You have asked for confirmation that the bike was not left unattended for more than 18 hours. While the policy documentation does specify the unattended time limit, the exact nature of the required evidence is not clearly defined. It would be more reasonable to accept a statement of facts regarding the timeline of the incident.

4) Agent: Independent adjudicator:
The independent adjudicator (Prompt 6) compares the response sent by the claims handler with that generated by the disgruntled customer, and prepares a set of recommendations.

The main purpose of the agent is to moderate feedback from the customer's perspective, keeping feedback to the claims handler agent consistent with the policy wording, while also treating the customer fairly.

In the example presented, the adjudicator acknowledges the impracticality of requiring witness statements or photographic evidence of the lock requirement, and waives the request for the remains of the lock - instead urging the claims assessor to rely on the police report.

Generated Response 3

<additional feedback> (*model-generated*):

Based on the analysis of the policy documentation and the issues raised by the claimant, I recommend the following.

- **Proof of Secure Locking**: While photographic evidence or witness statements can be helpful, they are not explicitly mandated by the policy. The claimant should be allowed to provide alternative forms of evidence, such as a detailed written statement, explaining how the bike was securely locked in compliance with the policy requirements.
- **Evidence of Forced Entry**: Given the practical challenges of recovering the remains of the lock, the claimant's provision of a police crime reference number should be considered sufficient, especially if the police report details the nature of the theft and indicates that forced entry occurred.
- **Confirmation of Unattended Time**: The claimant should be allowed to provide a statement of facts, detailing the timeline of the incident, to confirm that the bike was not left unattended for more than 18 hours.

5) Agent: Claim Handler:
This is the updated verdict generated by the claim handler incorporating the feedback from the independent adjudicator. It uses Prompt 3 to regenerate a verdict.

It corrects the main issue identified in the original verdict - namely the unpractical requirement for a witness statement or photographic evidence of how the bicycle was locked.

Generated Response 4

<response> (*model-generated*):

Required Documentation and Information
To proceed with the claim, the claimant must provide the following documentation and information:

- **Detailed Written Statement**: A detailed written statement explaining how the bike was securely locked through the frame and all quick release wheels to an immovable object by an approved lock.
- **Police Crime Reference Number**: A police crime reference number and, if available, a copy of the police report detailing the nature of the theft and indicating that forced entry occurred.
- **Proof of Ownership**: The original purchase receipt or similar proof of purchase for the bike.
- **Evidence of Forced Entry**: The remains of the approved lock, if available. If not, the police report should suffice as evidence of forced entry.
- **Timeline of the Incident**: A statement of facts detailing the timeline of the incident to confirm that the bike was not left unattended for more than 18 hours.

V. Evaluation

From the experimental results obtained, it is clear that - without any fine-tuning - LLMs can reason about policy wording and the validity of a claim description against the wording.

However, initial responses can fall short of the wholistic response one would expect of a claims assessor. By applying the agent-based iteration model discussed, a more appropriate verdict can be achieved.

For example, as discussed in the example displayed above, the first iteration of the verdict required ”photographs or witness statements” to prove that the bicycle was locked correctly. This is an impractical and onerous requirement - no cyclist takes a photo of how they've locked up their bike, nor do they keep track of witnesses for this purpose. By applying the agent iteration approach, a much more sensible requirement was achieved in the final verdict, rather asking for a ”written statement” detailing how the bicycle was locked.

A. Evaluation Methodology

To evaluate the performance of the proposed approach, seven claim descriptions were selected as input. These descriptions tested the model performance on three different perils, as well as varying claim validity. All descriptions were taken from actual claims submitted to the Sundays insurance company. Descriptions have, however, been adjusted to remove personal details. The claim descriptions are as follows:

1. **Accidental Damage**: *I was riding in a race, coming into a sprint finish with around 5000m to go, another rider rode into me and tripped me out, we both managed to stay up but my wheel had buckled and 12 spokes had broken, I believe something must have gone into the spokes.*
2. **Theft Invalid Abandoned**: *I had my bike parked up using my two usual secure D locks (Kryptonite and one Abus) - insuring [sic] the frame and wheels were locked. I parked my bike on the main road in Camberwell close to 'The Nunhead Gardener' on Sunday 2nd June in the afternoon. I went to cycle it home on Friday 7th June at roughly 6pm and the bike and locks were gone.*
3. **Theft Live Usage**: *I cycled with my partner and stopped opposite Sainsbury's for a quick food grabbing. We parked on the pavement, I turned the bike computer off and crossed the road to Sainsbury's, Mary my partner stood next to our bikes and was waiting for me. While she was waiting, a young guy came over and asked her a few questions about my bicycle, she answered him shortly, put her hands on both bikes and then told him she did not want to keep taking [sic] to him. This guy left and a minute after he crept up behind her along with another bigger guy, dressed in face covered. The bigger guy went for my partner's bike grabbing its handles and pushing her away whilst the young one went for my bike, there was nothing she could do to prevent the theft in this moment, she said that it hsappened so fast, she was shocked and frightened of the big muscular guy. The big guy run [sic] away in one direction and the young one with my bike the other direction. My partner immediately cycled in the direction of the young one, but he was very fast, reckless and disappeared. All of this took about one minute, when I came out of the shop I saw Mary, my partner coming down the road after cashing the guy and my bike was not there.*
4. **Theft Invalid Locking**: *I left my bike locked to the rack behind my house. I came down the next morning to find that the wheels and front fork had been stolen, while the frame was left behind still chained to the rack.*
5. **Theft Valid**: *I want to claim for a stolen bike insured on my policy with you. It was securely attached to a bike rack in my garage at home. I left it there this morning, and returned in the afternoon to find only the remains of the lock. See attached proof of purchase of the lock, a photo of the lock key, as well as a police report. I have also attached a photo of the bike, and the original receipt from the factory detailing its serial number.*
6. **Theft Invalid Usage**: *I was picking up a customer's UberEats delivery and bike was stolen outside the McDonald's. It was securely attached to the bike rack with a gold-rated lock.*
7. **Personal Injury**: *While cycling through Richmond Park, my bike slipped due to the wet roads, crashing into the embankment on my left. The fall impacted the left side of my body. My left shoulder is sore, and I'm also feeling some soreness in my head.*

A passable response satisfies three main criteria:

- **Criteria A**: It correctly detects claim invalidity. It does not approve a claim where there exists evidence that the claim falls outside the requirements of the policy wording.
- **Criteria B**: It identifies every piece further evidence that is required to establish claim validity.
- **Criteria C**: It does not place impossible or unnecessarily onerous evidence requirements on the claimant.
- **Criteria D**: The response is logically consistent. It does not contradict itself.

For every model and every claim scenario, two of the outputs were tested against the criteria:

- **First Loop Output**: The first iteration of the customer email generated by the claims handler (Prompt 4)
- **Second Loop Output**: The second iteration of the customer email generated by the claims handler, after incorporating feedback from the adjudicator in the claim verdict (Prompt 3)

The tables below show the model outputs that pass the criteria, on the first loop, and the second loop, respectively.

TABLE I
GENERATED RESPONSES THAT PASS ALL RESPONSE CRITERIA IN THE FIRST LOOP OUTPUT.

Model **Claim Submission**	**GPT-4o**	**GPT-4o Mini**	**Llama 3.1**	**Mistral Nemo**
Accidental Damage	PASS	PASS	FAIL	FAIL
Abandonment	FAIL	FAIL	FAIL	FAIL
Theft Live Usage	PASS	PASS	FAIL	FAIL
Theft Invalid Locking	PASS	PASS	FAIL	FAIL
Theft Valid	FAIL	FAIL	PASS	FAIL
Theft Invalid Usage	PASS	FAIL	FAIL	FAIL
Personal Injury	FAIL	FAIL	FAIL	FAIL

TABLE II
GENERATED RESPONSES THAT PASS ALL RESPONSE CRITERIA IN THE SECOND LOOP OUTPUT.

Model **Claim Submission**	**GPT-4o**	**GPT-4o Mini**	**Llama 3.1**	**Mistral Nemo**
Accidental Damage	PASS	PASS	FAIL	FAIL
Abandonment	PASS	FAIL	FAIL	FAIL
Theft Live Usage	PASS	PASS	FAIL	FAIL
Theft Invalid Locking	FAIL	PASS	FAIL	FAIL
Theft Valid	PASS	FAIL	FAIL	FAIL
Theft Invalid Usage	PASS	PASS	FAIL	FAIL
Personal Injury	PASS	PASS	PASS	FAIL

Table 3 summarises the results. The GPT models saw a significant improvement in the pass rate of the generated result on the second loop, compared with the first. Llama 3.1 and Mistral Nemo did not produce acceptable results in either pass. An improvement in the second pass demonstrates the benefits of an agent-based feedback loop.

TABLE III
THE PASS RATE OF THE MODEL OUTPUTS ON THE FIRST AND SECOND LOOP.

Model	**GPT-4o**	**GPT-4o Mini**	**Llama 3.1**	**Mistral Nemo**
First Loop Pass Rate	57.14%	42.86%	14.29%	0.0%
Second Loop Pass Rate	85.71%	71.43%	14.29%	0.0%

B. Shortcomings

While the agent-based approach does provide higher quality outputs on the OpenAI GPT models, it falls short on some specific areas.

1) Insurance Terms and Conventions:

Often the models will list a requirement for an excess to be paid. A quote from a particular response reads: "As per the policy terms, you will need to pay the applicable excess. Please confirm your willingness to pay this amount so we can proceed with the claim.". In reality, this is not how insurance payments are accounted for. Excess payments are simply deducted from the claimed amount before they are paid out to the customer.

2) Missing a Link to External Data:

The agent-based scheme in its current form does not allow the agents to look up necessary data on a policy or a claim. This makes the generation of an appropriate response difficult. For example, an insurer will often already have data on the resident-status of a policyholder, yet the models would often try to confirm this in the response. Similarly, models don't have access to the date at which the claim was notified, and would request this.

This seems like a relatively simple shortcoming to solve, by adding all known and necessary facts to the context for the model.

3) Introduction of False Information:

The agent-based approach can introduce false information into the response.

For example, the disgruntled customer agent, running on GPT-4o-mini, on the Invalid Theft (Abandonment) case, made the following statement during one iteration: "The policy stipulates an unattended duration of 18 hours, and I believe that my bicycle was left for a time frame well within this limit."

This contradicts the original facts presented in the claim description, which explicitly states "I left it there last week Thursday, and returned on Monday".

VI. CONCLUSION AND FUTURE WORK

The study demonstrated that an agent-based approach could enhance the quality of responses generated by language models for the purposes of insurance claims handling. For example, the iterative process of refining the response to a theft claim resulted in a more practical and relevant request for supporting evidence from the customer. Despite these successes, certain limitations were noted, such as the variability in model performance, the possible introduction of false information and a lack of understanding of key insurance terms and conventions.

A. Future Work

This study leaves many avenues for future work.

1) Improve generated responses for smaller models.: There is evidence that both Llama 3.1 and Mistral Nemo have the ability to reason about claims and policy wording. However, they performed particularly poorly in this study. One suspected reason is due to the size of the input document. A very lengthy policy wording document was used, with all coverage terms included. One approach to mediating this might be to first classify the peril to which a claim relates, and then only feed the appropriate sections from the policy wording to the model. For example, if it is a theft claim, one can reduce the size of the policy wording by roughly two thirds, just by excluding the sections that aren't relevant.

2) Fine tune the models to improve performance.: This study did not perform any fine tuning to improve performance. It seems reasonable to attempt to improve the response quality by tuning the models to output responses that are similar to an existing database of customer interactions.

B. Ethical and Legal Considerations

1) Data Privacy and Security: is a significant concern for the application of this study. Real-world claim descriptions will include sensitive personal details of the claimant. This data needs to be protected against unauthorised access (or being passed to a third party language model and added to their training data).

2) Model Bias.: LLMs are trained on vast datasets that may contain biases (Gallegos, Rossi, Barrow, *et al.*, 2024). It is crucial to ensure that the models do not perpetuate or exacerbate biases in claims assessment. This includes being vigilant about biases related to race, gender, socio-economic status, and other protected characteristics. Implementing fairness audits and bias mitigation techniques can help in producing equitable outcomes.

3) Regulatory Compliance.: Insurance companies must comply with industry regulations and standards. Any system that performs claims assessment should adhere to regulatory requirements set by bodies such as the Financial Conduct Authority in the UK.

References

[1] S. Wadhwa, S. Amir, and B. C. Wallace, "Revisiting relation extraction in the era of large language models," *Proceedings of the conference. Association for Computational Linguistics. Meeting*, vol. 2023, pp. 15 566–15 589, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258564662.

[2] J. J. Nay, D. Karamardian, S. Lawsky, *et al.*, "Large language models as tax attorneys: A case study in legal capabilities emergence," *ArXiv*, vol. abs/2306.07075, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259138525.

[3] C. Aracena, N. Rodríguez, V. Rocco, and J. Dunstan, "Pre-trained language models in Spanish for health insurance coverage," in *Proceedings of the 5th Clinical Natural Language Processing Workshop*, T. Naumann, A. Ben Abacha, S. Bethard, K. Roberts, and A. Rumshisky, Eds., Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 433–438. DOI: 10.18653/v1/2023.clinicalnlp-1.46. [Online]. Available: https://aclanthology.org/2023.clinicalnlp-1.46.

[4] Y. A. Arbel and D. Hoffman, "Generative interpretation," *New York University Law Review*, vol. 99, 2024. eprint: 2308.06907 (cs.CL). [Online]. Available: https://ssrn.com/abstract=4526219%20or%20http://dx.doi.org/10.2139/ssrn.4526219.

[5] Langchain Inc., 2024. [Online]. Available: https://python.langchain.com/.

[6] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009, ISBN: 1441412697.

[7] Ollama, 2024. [Online]. Available: https://ollama.com/.

[8] OpenAI, *Hello gpt-4o*, Jul. 2024. [Online]. Available: https://openai.com/index/hello-gpt-4o/.

[9] OpenAI, *Gpt-4o mini: Advancing cost-efficient intelligence*, May 2024. [Online]. Available: https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/.

[10] Meta, *Introducing llama 3.1: Our most capable models to date*, Jul. 2024. [Online]. Available: https://ai.meta.com/blog/meta-llama-3-1/.

[11] MistralAI, *Mistral nemo*, Jul. 2024. [Online]. Available: https://mistral.ai/news/mistral-nemo/.

[12] J. Wei, X. Wang, D. Schuurmans, *et al.*, *Chain-of-thought prompting elicits reasoning in large language models*, 2023. arXiv: 2201.11903 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2201.11903.

[13] Sundays, 2023. [Online]. Available: https://cdn.sundaysinsurance.co.uk/uk/docs/Sundays_UK_Policy_Wording.pdf.

[14] I. O. Gallegos, R. A. Rossi, J. Barrow, *et al.*, *Bias and fairness in large language models: A survey*, 2024. arXiv: 2309.00770 [cs.CL]. [Online]. Available: https://arxiv.org/abs/2309.00770.