

# Untruthful salespeople and detecting them using Gen AI

[DRAFT]

James Coulson

## Overview

Transparency and accurate information disclosure are essential for insurance companies to ensure that their customers have a clear understanding of the products they obtain from them. The primary tool used to outline the various features of an insurers different policies is the Product Disclosure Statement (PDS) which provides comprehensive details about a given insurance product. However, a fault of these PDSs is that they can often be too complex for the average individual to understand and as such customers often rely on direct communications with an insurer (or more likely a representative of the insurer) to determine if a product is right for them. This reliance on an insurer's representatives to disclose information about their different products can instead lead to a different problem where salespeople mistakenly make claims/promises about products which deviate from what is said in its PDS. This paper explores the use of generative AI (particularly ChatGPT) to detect such deviations in claims from the PDS using a transcript of the communications between a customer and an insurer's representative.

## Model Inputs

The inputs of the model comprise of three distinct inputs with them being,

- A meta prompt
- An overview of the PDS
- A transcript

## Meta Prompt

The meta prompt is used to give the model context on what its role is and how it should be performed. In this model we used the following text as the meta prompt,

"You are a machine that is responsible for processing telephone call transcripts between salespeople and potential customers. You will be given an overview of a product disclosure statement and a transcript of a phone call and will need to determine if the salesperson has made any promises or claims that do not align with the product disclosure statement. The transcript will follow the structure of the conversation with any sentences said by the salesperson being prefaced by "SALESPERSON" and similarly any sentences said by the potential customer being prefaced by "CUSTOMER:". The PDS may contain some information that is not able to be shared with you, this information will be replaced by "(REDACTED)" and should be ignored by you. If you do detect any variance in what is said from the product disclosure statement, you will respond by saying "POTENTIAL BREACH" followed by the reasoning for why you believe that there has been a variance by the salesperson. If you do not detect any variance you will respond with "COMPLIANT"."

This meta prompt outlines the model is responsible for processing transcripts of telephone calls between a salesperson and a potential customer. It then outlines that its role is to detect if any claims made by the salesperson deviate from what is said in the product

disclosure statement. It then details the structure of the transcript that it will be inputted and how it should format its output.

You will also note that the meta prompt alerts the model that there will be information that will not be shared with it and instead replaced with “(REDACTED)”. This has been added as in real world applications we would want to limit the amount of personal information that is given to the model.

Additionally, the meta prompt does not state that the PDS is in reference to an insurance product. Instead, the model will have to determine this from the PDS itself. This is done intentionally.

### Overview of the PDS

For each transcript the model is provided with an overview of the PDS for the product that is mentioned within the transcript. It will then use this description to gauge whether the salesperson has made a claim or promises about the product that differs from the PDS. An example overview of a PDS used is shown below.

“This PDS is for a car insurance policy which covers the policyholder from any third party liabilities in the case that they have an accident while driving. The policy does not include cover for the driver's vehicle only third-party damage caused in the incident. The product includes a premium of \$100 paid monthly by the user to the insurance company and has a lock in period of 2 years. If the customer decides that they wish to exit the contract early they will have to pay an exit fee equivalent to two months of premiums. If by the end of the two years, the user has not made any claims they will be refunded half of the premiums they paid.”

While uncomplicated this example overview does present several challenges to the model. For example, it mentions that in the case of early termination an “exit fee equivalent to two months of premiums” will need to be paid instead of the exact amount (\$200). This means that the model will have to perform its own reasoning if a claim such as “\$150 will be charged as an exit fee” is made by a salesperson as the model is not directly presented the numerical amount.

### Transcript

Finally, the transcript is used to provide the actual details of what was communicated between the salesperson and customer. The meta prompt outlines that the format of this input will have each passage spoken be prefaced by either “CUSTOMER:” or “SALESPERSON” depending on who was speaking. Example transcripts are given in the next section.

### Example Outputs

This section outlines select examples that were encountered during the testing of this model. All testing examples can be found in the separate Excel file. For consistency all the below examples have used the same meta prompt which is shown below.

You are a machine that is responsible for processing telephone call transcripts between salespeople and potential customers. You will be given an overview of a product disclosure statement and a transcript of a phone call and will need to determine if the salesperson has made any promises or claims that do not align with the product disclosure statement. The transcript will follow the structure of the conversation with any sentences said by the salesperson being prefaced by "SALESPERSON" and similarly any sentences said by the potential customer being prefaced by "CUSTOMER:". The PDS may contain some information that is not able to be shared with you, this information will be replaced by "(REDACTED)" and should be ignored by you. If you do detect any variance in what is said from the product disclosure statement, you will respond by saying "POTENTIAL BREACH" followed by the reasoning for why you believe that there has been a variance by the salesperson. If you do not detect any variance you will respond with "COMPLIANT".

### Example One – Third-Party Car Insurance Policy

In this example, a customer wishes to find more information about a third-party car insurance product which includes a fixed premium and policy length in addition to a fee for early cancellation of the policy. However, during the exchange between the customer and salesperson the salesperson mistakenly claims that the policy does not have a cancellation fee.

#### PDS Overview

"This PDS is for a car insurance policy which covers the policyholder from any third-party liabilities in the case that they have an accident while driving. The policy does not include cover for the driver's vehicle only third-party damage caused in the incident. The product includes a premium of \$100 paid monthly by the user to the insurance company and has a lock in period of 2 years. If the customer decides that they wish to exit the contract early they will have to pay an exit fee equivalent to two months of premiums. If by the end of the two years, the user has not made any claims they will be refunded half of the premiums they paid."

#### Transcript

CUSTOMER: Hello, (REDACTED) speaking?  
SALESPERSON: Hello (REDACTED). This is (REDACTED) calling from car insurance 101 about your recent inquiry into our third party car insurance policy.  
CUSTOMER: Oh yes, hi (REDACTED) I saw an advertisement for the policy on my way to work and saw that it included a good no claims discounts and wanted to learn about it.  
SALESPERSON: Great, well the policy covers any third-party damage that may occur excluding any damage to your car if you have an accident and lasts for two years. The discount you saw would have been that if you do not make any claims during the two years you receive half of the premium you paid back.  
CUSTOMER: Yes, that is the one I was looking at. One detail I wasn't sure on though was what the monthly premiums were and if there were any penalties for ending the policy early?

SALESPERSON: Ok, well we charge a flat \$100 premiums paid monthly which is in line with our competitors and do not charge any additional fees if you decide to cancel to policy early.  
CUSTOMER: Oh wow that sounds like a great deal.

#### Model Output

POTENTIAL BREACH: The salesperson has misrepresented the penalty for ending the policy early. According to the product disclosure statement, there is an exit fee equivalent to two months of premiums if the customer decides to cancel the policy early. The salesperson stated that there are no additional fees for early cancellation.

#### Notes

From the above output we can see that the model has correctly identify that the salesperson has incorrectly stated that there is not fee for early cancellation of the policy and has followed the formatting outlined

#### Example Two – Life Insurance Policy with Conditional Discounts

In this example, the product discussed is a life insurance policy where a youth discount applies to any policy holder who takes out a policy prior to the age of 35. An additional oddity is that if the customer lapses on the policy and then takes out a new policy once passing the age of 35, they are still applicable for the 'youth discount'. This example shows a case where the salesperson initially is mistaken in how this youth discount is applied and is later corrected.

#### PDS Overview

This PDS is for a life insurance policy which provides a payout in the case of death equal to their age multiplied by 1,000 up until the age of 95. However, in the case that the individual dies between the ages of 70 and 80 not inclusive they will receive an additional \$10,000 on top of their base payout. The policy holder will be charged a monthly premium \$120 from when they first take the policy out until they reach an age of 95 at which point, they will be refunded all of their premiums paid. A special discount is if the user is under the age of 35 when they first take the policy out they will only have to pay a monthly premium of \$100 instead of \$120, however if the policy holder lapses on their payments while having this discount and then at a later stage they take the policy out once again after they have surpassed the age of 35 this discount will still apply.

#### Transcript

CUSTOMER: Hello, I'm (REDACTED). I'm calling as I wish to take out a life insurance policy.  
SALESPERSON: Hello (REDACTED). That's great thanks for choosing us to provide you with this policy. Before we can proceed, I will first need to ask you some questions.  
CUSTOMER: Ok, sounds good.  
SALESPERSON: Would you mind providing me with your age?  
CUSTOMER: I am 56  
SALESPERSON: Thank you, now as you are 56 you are not applicable to the youth discount and as such your monthly premiums will be \$120.

CUSTOMER: Oh, I thought that as I had this policy previously while under the age of 35 that I would still be eligible for the discount?

SALESPERSON: Sorry (REDACTED) but that is incorrect. You can only receive the youth discount if you are under the age of 35 when you first take out the policy.

CUSTOMER: Oh, that is disappointing I must have misremembered the terms of the product. Would you mind checking with in the PDS for me?

SALESPERSON: Yes certainly .... Oh, it appears I was mistaken you actually are applicable for the youth discount apologies for my mistake

CUSTOMER: No worries, in that case I'd like to continue and take out this policy.

### Model Output

POTENTIAL BREACH: The salesperson initially stated that the customer is not eligible for the youth discount because they are 56 years old. However, according to the product disclosure statement, if the policyholder had previously taken out the policy while under the age of 35, they are still eligible for the discount. The salesperson later corrected themselves after checking the PDS, but there was initially a variance between what was stated and what is outlined in the PDS.

### Notes

This case demonstrates a more complex scenario to the previous example not only due to the conditional nature of the youth discount but also as the salesperson is initially mistaken but later corrects themselves. We can see in the model output that it detects the initially incorrect assertion that the customer is not applicable to receive the discount but also notes that this mistake was corrected later in the conversation. As the model has detected there was a deviation in the conversation it has labelled the transcript as having a potential breach. This output, while initially may seem incorrect, is the desired response as it shows a more cautious approach favouring to label it as potentially non-compliant.

### Conclusion and Caveats

The examples provided demonstrate the effectiveness of this model in detecting deviations between salesperson statements and the information in a PDS, even in complex scenarios with very little fine (or course) tuning. However, it is important to consider the limitations of this study, such as the small sample size and the simplified nature of the testing scenarios compared to what would be observed in the real world. As a result, it remains unclear whether this approach can be effectively applied in real-world situations. Further research is needed to evaluate the practicality and effectiveness of this model in more complex real-life scenarios.

The source code for this report can be found at <https://github.com/James-Coulson/detecting-untruthful-salespeople>.