Actuarial Large Language Model Proposal

Contents

[1) Introduction 2](#_Toc157170837)

[2) General Benefits 2](#_Toc157170838)

[3) Use cases 3](#_Toc157170839)

[3.1) Training of new staff 3](#_Toc157170840)

[3.2) Report writing/editing 3](#_Toc157170841)

[3.3) Code optimization and generation 3](#_Toc157170842)

[3.4) Business as usual productivity 4](#_Toc157170843)

[4) Model Requirements 4](#_Toc157170844)

# Introduction

The purpose of this document is to propose the investigation and possible development of a Sanlam actuarial specific large language model by outlining potential use cases and associated benefits. The suggestion is to create a large language model trained on Sanlam specific information to enhance efficiency of the actuarial teams and reduce the risk of using publicly available large language models (E.g. ChatGPT).

# General Benefits

Large language models have become integral tools across various industries, contributing to increased operational efficiency and higher output quality. However, there are two considerable shortcomings of these publicly available large language models. The first concern is that these models are publicly accessible, posing a potential risk to Sanlam, as sensitive information may inadvertently become public. The second limitation is that the models have insufficient knowledge of Sanlam specific structures, products, and processes, as illustrated in the screenshot below:

A screenshot of a message

Description automatically generated

Therefore, a Sanlam actuarial specific large language model will meet the following objectives:

1. **Enhanced confidentiality:** Reduce the risk of sensitive information entering the public domain, whilst retaining the existing benefits of using publicly available large language models.
2. **Tailored information retrieval:** Enable users to access Sanlam specific information to assist them with day-to-day work, improve their efficiency and their quality of output.

# Use cases

## Training of new staff

A Sanlam actuarial specific large language model will assist with training of new staff by offering immediate responses to queries related to Sanlam-specific products and processes. A new staff member will have immediate user-friendly access to relevant information that will significantly enhance the onboarding process. Examples of questions that new staff members might ask are listed below:

* What do SRA, SRM and SC stand for?
* What is an echo bonus and how does it work?
* What are the main products for SRA, SRM and SC respectively?
* What IFRS 17 measurement model does Matrix fall into?
* What are the coverage units for our SRA annuities?

## Report writing/editing

A Sanlam actuarial specific large language model will be able to assist with report writing on numerous fronts:

1. **Draft reports:** It will provide the capability to immediately generate draft reports on historical events. For example, drafting a report that explains the IFRS 17 transition methods used for Millenium.
2. **Secure review and revision:** It provides a safe environment to review and suggest changes to Sanlam draft reports. For example, sections of the GEV report can be pasted into the model for revision, without having to spend time removing sensitive information.
3. **Secure summarization of key findings:** It will provide the capability to summarize key findings from various reports and working documents that has sensitive information (Investor reports, analyst reports, actuarial reports/memos etc).
4. **Secure executive summary generation:** Similarly, it will provide the capability to generate professional executive summaries for formal reports and memos that includes sensitive information.

## Code optimization and generation

A Sanlam actuarial specific large language model will allow users to optimize and write Sanlam specific code for, for example, Prophet models and Data Conversion Systems (DCS). The significance lies in the fact that the code of our Sanlam models often contains sensitive information, rendering the direct pasting of existing code into publicly available large language models impractical without the need for meticulous code review and the removal of sensitive information.

## Business as usual productivity

In day-to-day work, significant amounts of time are spent to find information on historical events, product features, processes, and assumptions, to name only a few. A large language model trained on Sanlam actuarial specific information will be able to significantly improve resource efficiency by providing readily available information and eliminating the need to sift through past emails, reports, memos, and workbooks to locate pertinent data. Examples of questions that staff might ask as part of day-to-day activities include:

* What were the lapse assumptions for Millenium in December 2022?
* What is the IFRS 17 fair value date for GSB?
* What was the final reported VIF for SRA at December 2023?
* In what year did we reduce the required capital from R9bn to R7bn?

# Model Requirements

For the use cases above to be feasible and the model employed successfully, the following model requirements will need to be adhered to:

1. **Sanlam Actuarial training data:** The model will have to be trained on Sanlam actuarial specific information in order to answer Sanlam specific questions (E.g. valuation information, GEV information, Sanlam products and Sanlam abbreviations.)
2. **Temporal competence:** The model will have to be able to give time-dependant answers (E.g. before and after IFRS17, knowing when to use SPF/SRA or Sky/SRM and when these have changed).
3. **Security protocols:** The model will have to be safeguarded and suitable to accept sensitive company information.