Institute and Faculty of Actuaries

IFoA Life Conference

Documenting spreadsheets using LLMs

Aniketh Pittea – Grant Thornton UK Advisory & Tax LLP

01 Background

What are Large Language Models (LLMs)?

LLMs are a type of neural network which is trained on a massive of text data. They are generally trained on data found online and consist of 3 main building blocks.

Data

LLM are trained on petabytes of data including web scraping, online books, articles, etc..

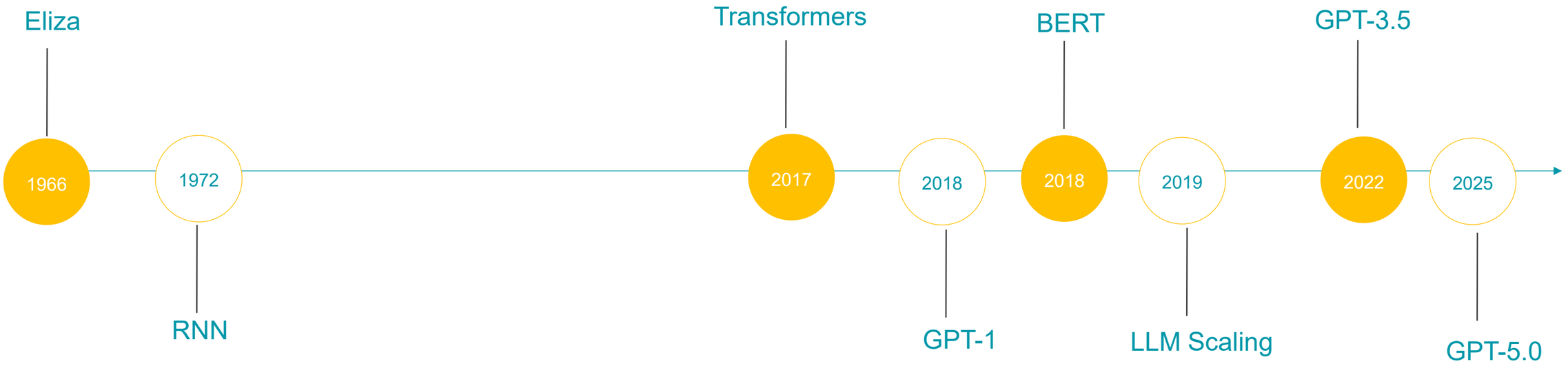
Architecture

LLMs consist of a series of algorithms which try to recognise patterns of data. The algorithms are focused on understanding natural language.

Training

Traditional programming is instructions based. In contrast, LLMs are example-based and their output relies on providing lots of examples.

History and evolution of LLMs



History and evolution of LLMs

- Eliza model (1966) – first Language model based on key-word.
- RNN (1972) – First model to predict the next word; the basis of an LLM.
- Attention is all you need (2017) – Breakthrough paper on Transformers.
- GPT-1 (2018) – Revolutionary model with 117 million parameters.
- BERT – 340 million parameters and bi-directionality.
- GPT-3.5-5.0 – The models we use when we use ChatGPT.

Features of LLMs

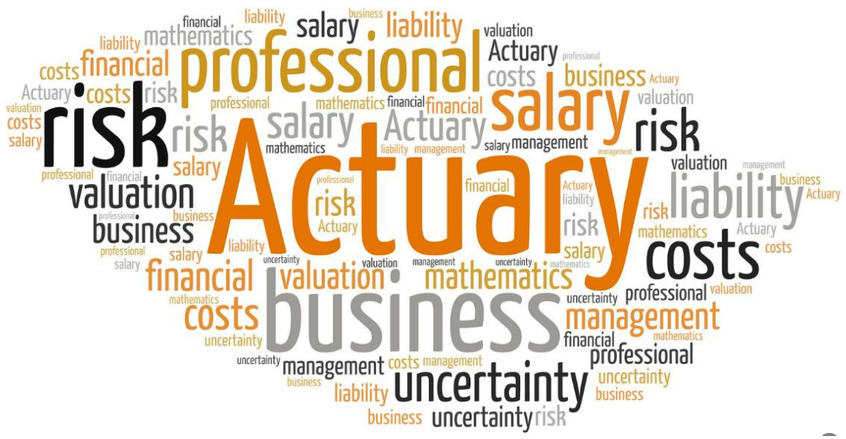
- More flexible and scalable than traditional programming
- Can be used for a wide range of tasks including summarisation, creative writing, QnA and Programming
- Rapidly evolving – we are still in the early days of LLMs and their capabilities will only continue to grow.

How LLMs work?

Tokenisation/ Embeddings → Positional Embedding → Attention

Word	Vector
Cat	[0.1, 0.2, 0.3, 0.4, 0.5]
Dog	[0.1, 0.2, 0.3, 0.4, 0.6]
Ball	[0.2, 0.4, 0.6, 0.8, 1.0]
House	[0.3, 0.6, 0.9, 1.2, 1.5]

- *They watched the frog **spring (verb)** out of the box.*
- *She has to replace a **spring (noun)** on her bicycle.*



Tuning of models

- Tuning is the process of further training a pre-trained LLM on a specific dataset or for a particular task to improve its performance in a specific area.
- It allows users to tailor the general capabilities of a pre-trained model to meet the unique needs of their application without having to train a new model from scratch.

What an off-the-shelf LLM can do

- Create a generic summary of a spreadsheet.
- Describe the general structure of a spreadsheet.
- Describe actuarial principles.
- Create Python codes based on a set of instructions.

What an off-the-shelf LLM cannot do (yet) without tuning

- Create spreadsheet documentation based on a certain set of standards.
- Identify the inputs, calculations and output reliably.
- Describe complex spreadsheet mechanics or logic.
- Convert spreadsheets into reliable Python codes.

02 AI powered spreadsheet tool

Context

Excel spreadsheets are used extensively by firms for data management, analysis, and reporting. These spreadsheets can be resource intensive to manage and prone to high-levels of control risk. Our AI-Powered spreadsheet management tool makes it easier to understand, review, audit, document and where required replace spreadsheets.

Challenges

- Reviewing spreadsheets is resource intensive
- Dealing with legacy spreadsheets which lack documentation
- Complex spreadsheet interdependencies
- Volume of skilled resource required to move code to Python or other alternatives

Our approach

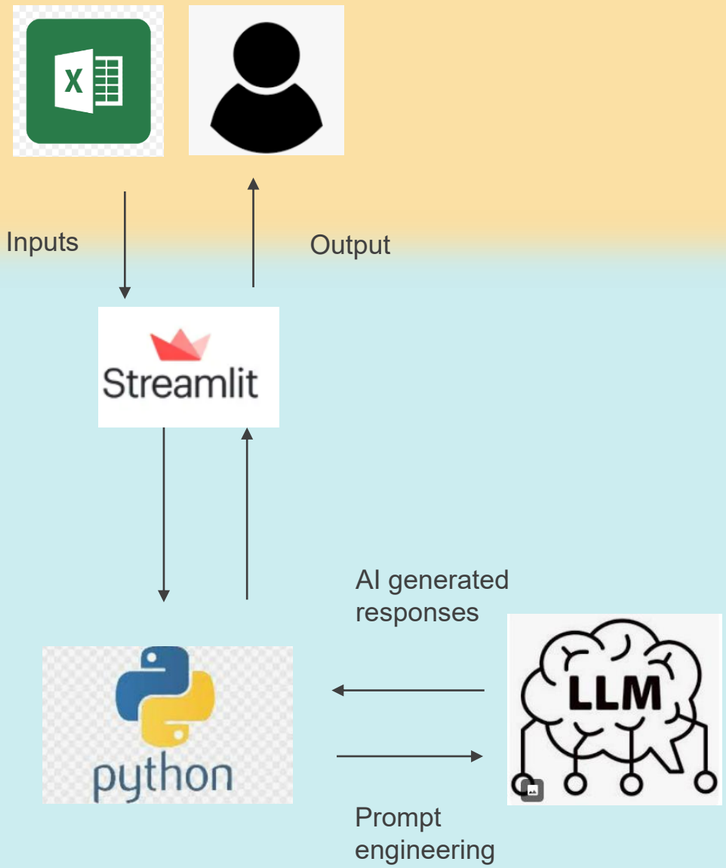
To create a bespoke AI-powered app to streamline spreadsheet:

- Analysis
- Documentation
- Validation
- Translation to programming codes (eg Python, R, Julia)

Benefits

- Auditability: Ensures transparency by automatically documenting spreadsheet structures and dependencies, making it easier to track data flows and formula logic.
- Cost Efficiency: Reduces the need for manual review and extensive consulting time, leading to significant cost savings.
- Speed: Automates spreadsheet analysis, cutting down time spent on documentation and dependency.
- Control: Provides users with an intuitive interface to review AI-generated documentation, allowing them to verify and refine the insights before implementation.

So how does it work?



Client and /or GT IT and model risk policy framework

Microsoft Azure environment

03 Live example

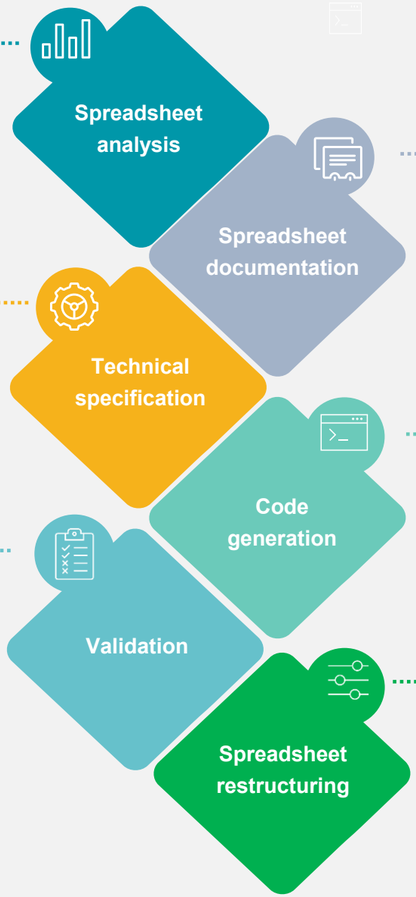
Our concept

Our analysis, combined with feedback from clients, has led us to a take a building block approach to enable bespoke solutions to client needs

Analysis of spreadsheets and production of data flow diagrams

Creation of specifications to support management of existing processes or migrations

Validation to support testing and review.

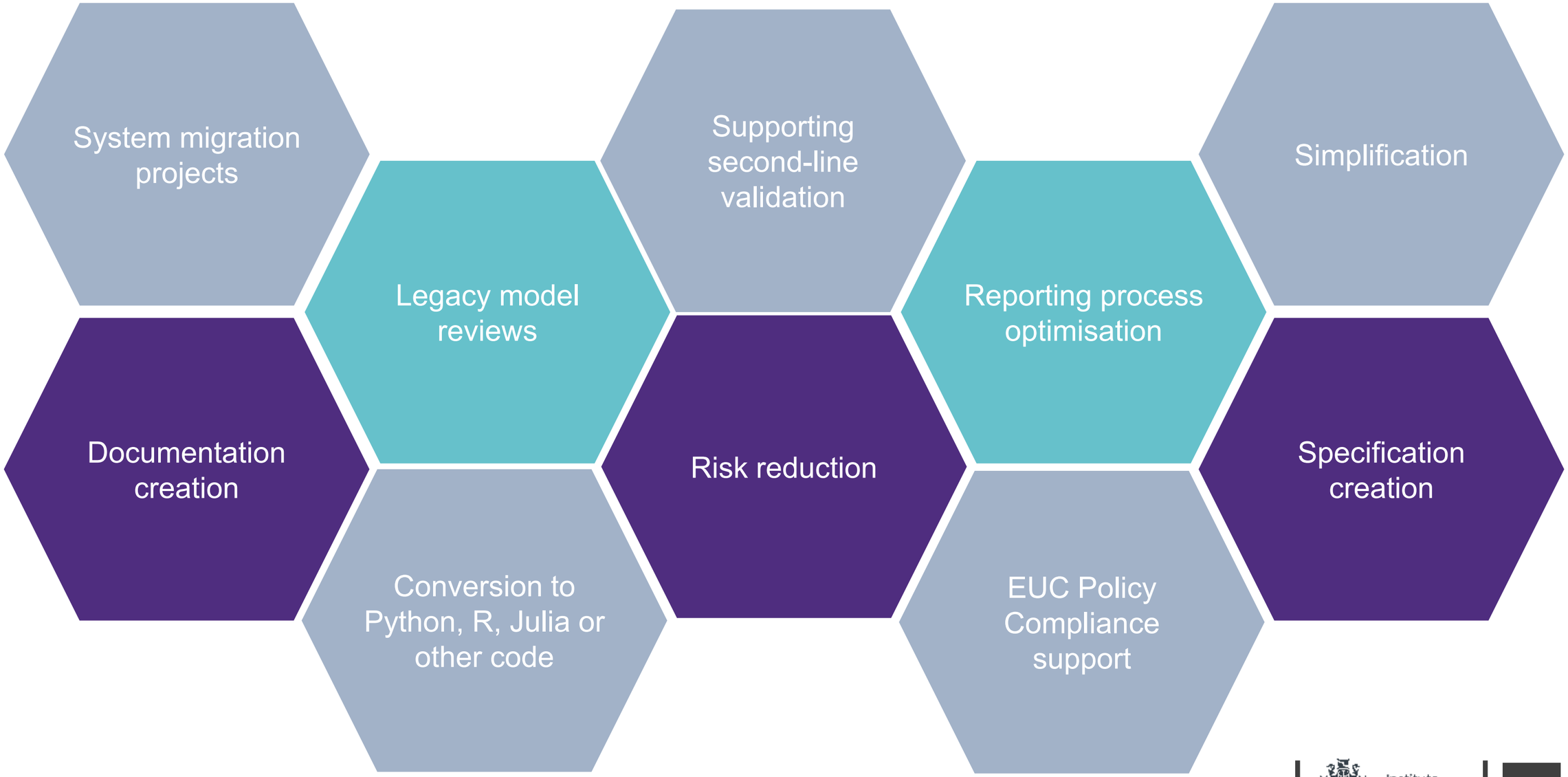


Documentation of existing functionality

Programming code creation (e.g. Python, R, Julia)

Restructuring of spreadsheets to align with EUC requirements.

Supporting the business



Model performance

Item	Positive Features	Can still be improved
General description of model	• Clearly states the model framework the calculations are based on. • Describes both inputs, calculations and outputs.	• Can be overly generic when describing application for actuaries.
Describing inputs and their purpose	• Each input is listed with its name, type, source, and a detailed explanation. • Descriptions link inputs to their role in the model.	• Can be more concise. • Complex technical inputs may be incorrect.
Describing calculations logic	• Each logic step is clearly numbered and named. • Descriptions explain dependencies between steps and how inputs are transformed.	• Some steps are overly detailed and could be summarised more concisely. • Could benefit from grouping steps into thematic blocks (e.g., trend projection, volatility, final output.
Describing output	• Output is clearly defined with its location and structure.	• Complex output may be incorrect.

Techniques to improve model performance

Prompt Engineering

Avoid vague prompts - *"Kim, can you get me something to drink?“*

Chain-of-thought prompting - *"Kim, could you please go to the kitchen, open the fridge, and bring me a cold can of cola? If there's no cola, a glass of chilled water would be fine.“*

Custom Instructions

Decomposition prompting - *solve part of the problem for the AI.*

Few shot prompting - *show the AI what good looks like.*

RAG

Giving more context to the AI

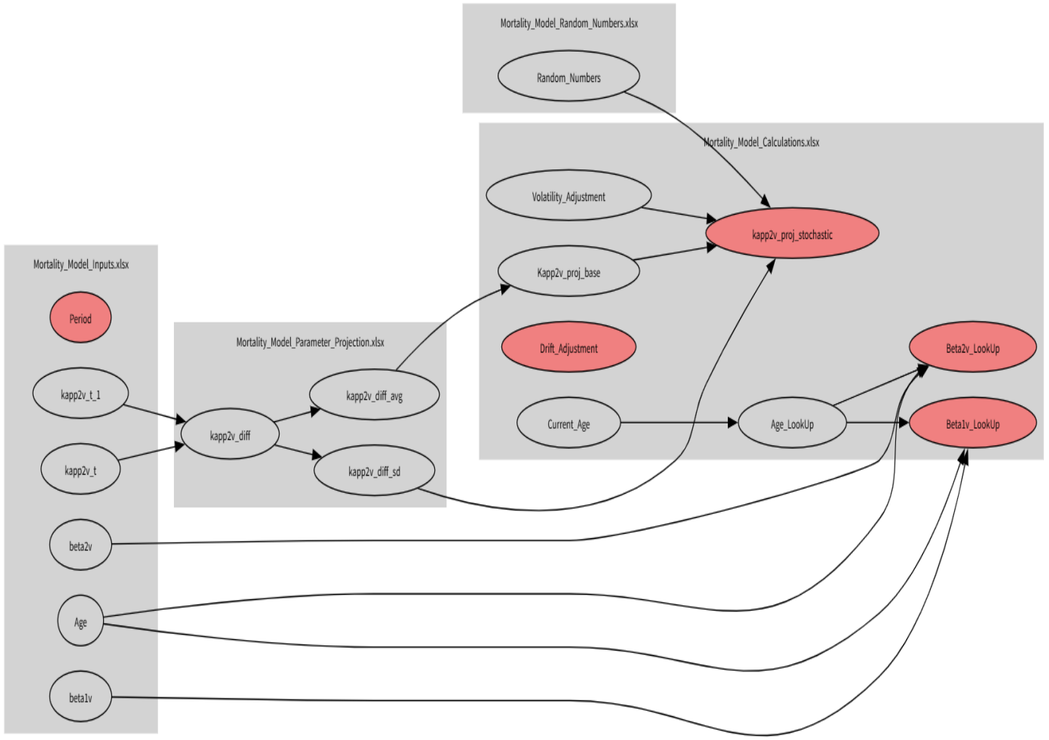
Model Selection

GPT-4, GPT-5o, Claude 3, LLaMA 3

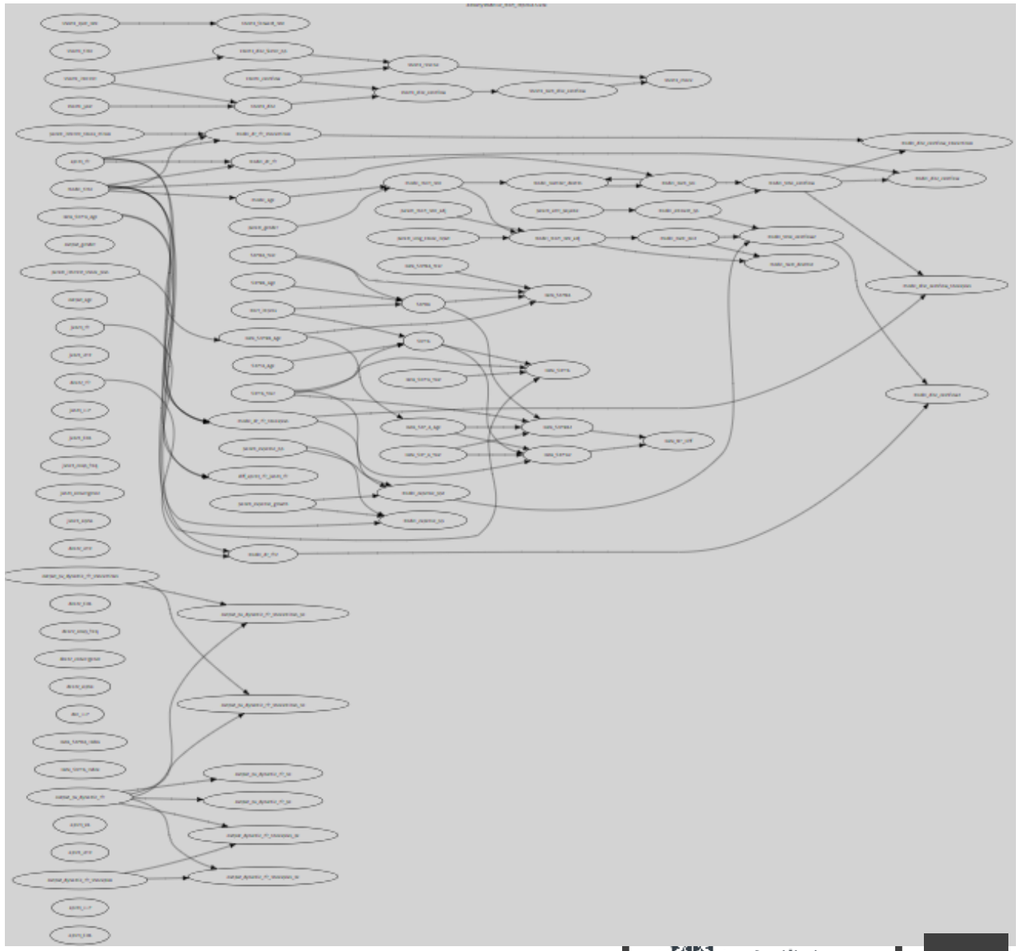
Test cases

We have tested the tool suite on multiple interconnected spreadsheets and on very large spreadsheets

Multiple spreadsheets



Large spreadsheet



Q&A

Institute and Faculty of Actuaries

Thank you

For more information, contact

Aniketh.Pittea@uk.gt.com