Revolutionising Financial Data Analysis

Generative AI for Investment Bank Equity Research

By Peter Griffiths, Murali Menon

Investment Bank (IB) Equity Research partnered with Calimere Point to conduct a proof of concept (POC) exercise. The goal was to assess the potential of Generative AI models in automating the extraction of key financial metrics from unstructured company filing data, transforming it into a structured Excel output for use in IB's research models.

Harnessing Generative AI for Automated Equity Research

The goal was to leverage Generative AI to automate the extraction of key financial metrics from unstructured quarterly reports of major insurance companies - a task traditionally requiring hours of manual work. The PoC aimed to transform these metrics into structured Excel outputs for IB's research models.

While not without challenges, the results were promising: the exercise demonstrated that AI could indeed automate this process, potentially revolutionising the speed and efficiency of equity research. By testing the limits of current AI technology in handling real-world financial data complexity, this project has implications that could reshape how financial analysis is conducted in the future.

The use of AI in this field is expected to grow exponentially, with estimates suggesting that by 2025, AI could save financial institutions up to $1 trillion annually by streamlining operations and reducing human error. Moreover, as AI continues to evolve, its capabilities in natural language processing and data interpretation are set to revolutionise how equity research is conducted, making it faster, more reliable, and ultimately more insightful.

This use case delves into a groundbreaking proof of concept PoC that showcases the transformative power of Generative AI in automating critical aspects of equity research, offering a glimpse into the future of finance.

Contents

1. Introduction

AI in Equity Research: Navigating complexity for greater efficiency

The main challenge in applying Generative AI to equity research is the variability in financial report formats and structures. This variability ranges from semi-structured tabular data to highly unstructured information in charts and narratives. The financial industry is actively exploring AI-driven solutions to automate data extraction, assist in analysis, and improve overall efficiency in equity research. Approaches include pure AI models, hybrid systems combining traditional analytics with AI, and sequential AI processes. These efforts aim to standardise data handling and accelerate research workflows.

It's important to note that while there's a clear trend towards automation and AI-assisted research, the complexities highlighted in this use case suggest that human expertise remains crucial in interpreting and validating AI-generated insights in equity research.

2. Overview and Context

Background

This PoC exercise was designed to explore the frontier of Generative AI applications in financial analysis, specifically focusing on the potential of Generative AI models in automating the extraction of key financial metrics from unstructured company filing data. The goal was to transform these metrics into a structured Excel output for use in IB's research models, potentially revolutionising the speed and efficiency of equity research.

Scope

The PoC focused on a carefully selected group of 5 insurance companies, each representing different geographical markets and reporting structures. **This diverse selection allowed the team to test the AI models against a variety of reporting styles and regulatory frameworks.**

Legal and General (UK): A leading provider of life insurance, pensions, and investment management services
Sampo Group (Finland and Nordics): A major player in the Nordic and Baltic insurance markets
Hannover Re (Germany): One of the world's largest reinsurance groups
Just Group (UK): A specialist in retirement income products and services
Allianz (Germany): A global leader in insurance and asset management

Project Structure:

The PoC was structured to maximise efficiency and minimise resource allocation while still providing robust insights:

Technology Stack

The project leveraged cutting-edge Generative AI technologies, complemented by traditional data analytics approaches. This hybrid approach aimed to combine the strengths of both paradigms.

Resource Allocation

Two members of Calimere Point's data science team were dedicated to the project, bringing specialised expertise in AI and financial data analysis.

Timeline

The development phase was time-bounded to 2.5 weeks, necessitating rapid iteration and focused problem-solving.

Infrastructure

The entire PoC was conducted on Calimere Point's Calimere Point's analytics-as-a-service platform (CPRA Cloud), providing a secure and powerful environment for AI model training and testing.

Cost Efficiency

Remarkably, this was executed as a cost-effective PoC exercise, valued at $25k, demonstrating the potential for high-impact results with a modest financial investment.

Summary: Structure of the PoC

Leverage cutting-edge Generative AI technologies combined with traditional data analytics approaches

$25k Cost

The PoC was time-bounded to 2.5 weeks of development and leveraged 2 of Calimere Point’s data science team

Conducted on Calimere Point’s data analytics architecture (CPRA Cloud)

3. Key Challenges in Equity Research

In the pursuit of automating financial data extraction through Generative AI, the team encountered several significant challenges. These obstacles highlight the complexity of working with real-world financial data and the nuances involved in applying AI to such tasks. The primary challenge was the high degree of variability in report formats and data structures, both between companies and between different quarterly reports for the same company.

1. Variability in Report Formats and Structures

The primary challenge stemmed from the high degree of variability in how different companies present their financial information. This variability manifested in two main forms:

Inter-company Variability:

Each of the five insurance companies in the study had its own unique way of structuring and presenting financial data. This inconsistency made it difficult to create a one-size-fits-all AI solution.

Intra-company Variability:

Even within the same company, the format and structure of reports could change from one quarter to another. This temporal inconsistency added another layer of complexity to the AI's task.

2. Types of Data Encountered

The team categorised the data they encountered into two main types, each presenting its own set of challenges:

Structured Unstructured Data

Lower Complexity = Greater Automation Potential

This category primarily consisted of tabular data, often presented in PDF format.

Challenges:

- Extracting data accurately from PDF tables
- Handling slight variations in table structures across reports

Opportunities:

- More amenable to programmatic approaches
- Higher potential for accurate data extraction
- Improved accuracy when data structures remain consistent across periods

SAMPO GROUP

Group key figures

EURm	Q1/2023	Q1/2022	Change, %
Profit before taxes (P&C operations)	359	692	-48
If	337	495	-32
Topdanmark	63	15	320
Hastings	10	21	-52
Holding	-45	164	-
Net profit for the equity holders	271	773	-65
- of which from life operations	28	199	-86
Underwriting result	292	242	21
EPS, EUR	0.53	1.42	-63
Operational result per share, EUR	0.51	-	-

Figures for 2022 restated for IFRS 17 but not for IFRS 9

Group overview	Q1 2022	Q1 2023c	Q1 2023	Delta	Q1 2023 DBe
- If profit before taxes (EURm)	495	281	337	20%	303
- Topdanmark profit before taxes (EURm)	15	58	63	8%	50
- Hastings profit before taxes (EURm)	21	15	10	-34%	23
- Holding profit before taxes (EURm)	164	-16	-45	-380%	-10
Group profit before taxes (P&C operations, EURm)	692	342	359	5%	367
Profit for the equity holders of the parent (including Life operations, EURm)	773	292	271	-7%	317
EPS (EUR)	1.42	0.57	0.53	-7%	
- Regular DPS (EUR)	NA	NA			
- Extra DPS (EUR)	NA	NA			
Total DPS (EUR)	NA	NA			
Buybacks (announced YTD, EURm)	228	400		-100%	400
Solvency II ratio (including div. accrual, %)	200	209	208	-84	

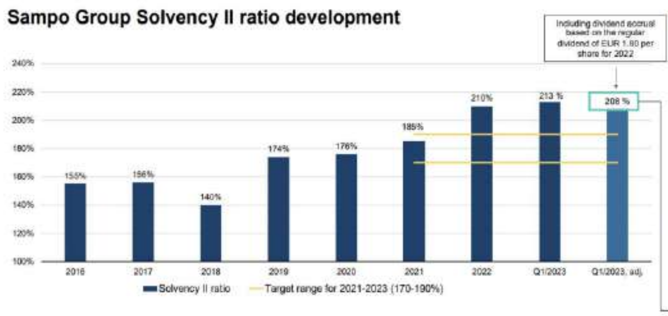
Unstructured Unstructured Data

Higher Complexity = Lower Automation Potential

This category included non-tabular data in PDFs, narrative text, and information embedded in charts or graphics.

Challenges:
- Extracting precise numerical data from charts or graphics
- Interpreting contextual information from narrative text
- Dealing with inconsistent presentation of key metrics

Example:
Sampo's Solvency II ratio was only available in PowerPoint charts in their quarterly reporting, making it particularly challenging to extract accurately.



Group overview	Q1 2022	Q1 2023c	Q1 2023	Delta	Q1 2023 DBe
- If profit before taxes (EURm)	495	281	337	20%	303
- Topdanmark profit before taxes (EURm)	15	58	63	8%	50
- Hastings profit before taxes (EURm)	21	15	10	-34%	23
- Holding profit before taxes (EURm)	164	-16	-45	-380%	-10
Group profit before taxes (P&C operations, EURm)	692	342	359	5%	367
Profit for the equity holders of the parent (including Life operations, EURm)	773	292	271	-7%	317
EPS (EUR)	1.42	0.57	0.53	-7%	
- Regular DPS (EUR)	NA	NA			
- Extra DPS (EUR)	NA	NA			
Total DPS (EUR)	NA	NA			
Buybacks (announced YTD, EURm)	228	400		-100%	400
Solvency II ratio (including div. accrual, %)	200	209	208	-84	

3. Balancing Accuracy and Automation

Another key challenge was finding the right balance between the level of automation and the accuracy of extracted data:

- Highly prescriptive AI models could achieve good accuracy but required significant manual setup and were vulnerable to changes in report formats.
- More flexible, general-purpose AI models offered greater adaptability but sometimes sacrificed accuracy or required more manual verification.

4. Handling Edge Cases and Exceptions

Financial reports often contain exceptions, footnotes, and special cases that are crucial for accurate analysis but challenging for AI to interpret correctly. Ensuring that the AI models could identify and handle these edge cases appropriately was a significant challenge.

Maintaining consistency across different AI approaches

As the team experimented with different AI models and approaches, ensuring consistency in results across these various methods became a challenge. This was particularly important for comparing the effectiveness of different AI strategies. By addressing these challenges, the team not only advanced their understanding of applying Generative AI to financial data extraction but also gained valuable insights into the broader implications of AI in financial analysis. These learnings form the foundation for future developments in this field, potentially transforming how equity research is conducted.

CALIMEREPOINT

4. Insights and Discoveries

Potential for Automation: A Critical Finding
These insights provide a nuanced view of the current state of Generative AI in financial data extraction. They reveal both the significant progress made and the challenges that lie ahead in fully realising the potential of AI in Equity Research.

Effectiveness of AI Assistants

Key Finding: AI Assistants demonstrated the capability to generate accurate and automated outputs for the 5 insurance companies in the PoC. This success suggests that AI Assistants have significant potential in automating routine data extraction tasks in equity research.
However, they are more programmatic in nature so contingent on carefully underlying defined and a high degree of consistency in input data formats.

Potential of sequential AI approaches

Key Finding: Experiments with sequential AI approaches showed promise in balancing flexibility and accuracy. By combining different AI models in sequence, the team found potential pathways to reduce reliance on prescriptive logic while maintaining accuracy.
This approach could offer a middle ground between highly structured and purely generative methods.

API vs. User Interface Discrepancies

Key Finding: The team observed variance in behaviour between user interface queries and API queries.
This discovery highlights the importance of consistent performance across different interaction modes with AI systems. It also suggests that the method of querying AI models can significantly impact results, a crucial consideration for real-world applications.

Data format variability impact

Key Finding: The high variability in report formats and data structures significantly influenced AI performance.
This insight underscores the need for robust, adaptable AI systems capable of handling diverse financial reporting styles. It also highlights a potential area for standardisation in financial reporting to facilitate AI-driven analysis.

Basic Questions and Challenges

Can you use Generative AI approaches to automate the extraction of key financial metrics from quarterly results publications from 5 insurance companies and populate a set of excel templates with the values for onward analysis?

Short Answer: Yes.
Long Answer: It's complicated.

What makes it complicated?

- High degree of variability in report formats, data structures between each of the POC population of companies and between different quarterly reports for the same companies
- Unstructured data – to paraphrase Donald Rumsfeld; we were faced with structured unstructured data

5. Experimentation and Approach

The PoC employed a multi-faceted experimental approach to thoroughly explore the potential of Generative AI in financial data extraction. Three primary methodologies were investigated, each offering unique advantages and challenges. This comprehensive experimentation strategy allowed the team to gain a nuanced understanding of how different AI approaches perform in the context of financial data extraction. The insights gained from these experiments formed the basis for the project's conclusions and recommendations for future development in this field.

Summary of Approaches

The PoC employed a multi-faceted experimental approach to thoroughly explore the potential of Generative AI in financial data extraction. Three primary methodologies were investigated, each offering unique advantages and challenges:

1. Pure Generative AI:

In our exploration of pure Generative AI models, we leveraged cutting-edge LLMs like Claude 3 Sonnet and ChatGPT4. This approach offered flexibility in handling diverse document formats and showed promising accuracy at 80%. While it required no prescriptive coding, the process still relied on manual data sourcing and provision. The models demonstrated an impressive ability to understand financial contexts, though they struggled with certain file formats like Excel. As LLM technology rapidly evolves, we anticipate significant improvements in accuracy and versatility, making this a promising avenue for future development in automated financial data extraction.

2. AI Assistants Models:

Our AI Assistants approach combined Generative AI with traditional data analytics, resulting in a highly accurate and automated system. This hybrid method achieved 100% accuracy in data extraction, with both sourcing and provision fully automated. The assistants excelled in structured data extraction and incorporated domain-specific financial knowledge. However, the system's reliance on prescriptive code made it vulnerable to changes in input data formats. Despite this limitation, the AI Assistants model stands out as a powerful tool for consistent, high-accuracy financial data processing, particularly in stable reporting environments.

3. Sequential LLM / Generative AI Models:

The Sequential LLM approach represented our most innovative experiment, utilising multiple AI models in a predefined sequence via API infrastructure. This method achieved a high degree of automation in both data sourcing and provision. However, it unexpectedly resulted in lower accuracy (40%) compared to other approaches, particularly when using API calls instead of user interfaces. While this method shows potential for handling complex financial data structures, the significant accuracy drop highlights the need for further investigation and refinement. The sequential approach offers valuable insights into the challenges of creating more sophisticated, multi-model AI systems for financial data extraction.

Approach 1: Pure Generative AI

Models Used

Claude 3 Sonnet by Anthropic
A highly advanced language model designed for nuanced understanding and ethical AI interactions, optimised for contextual comprehension and reliable responses.

ChatGPT-4 by OpenAI
An advanced generative language model known for its broad general knowledge, conversational abilities, and adaptability across various topics and tasks.

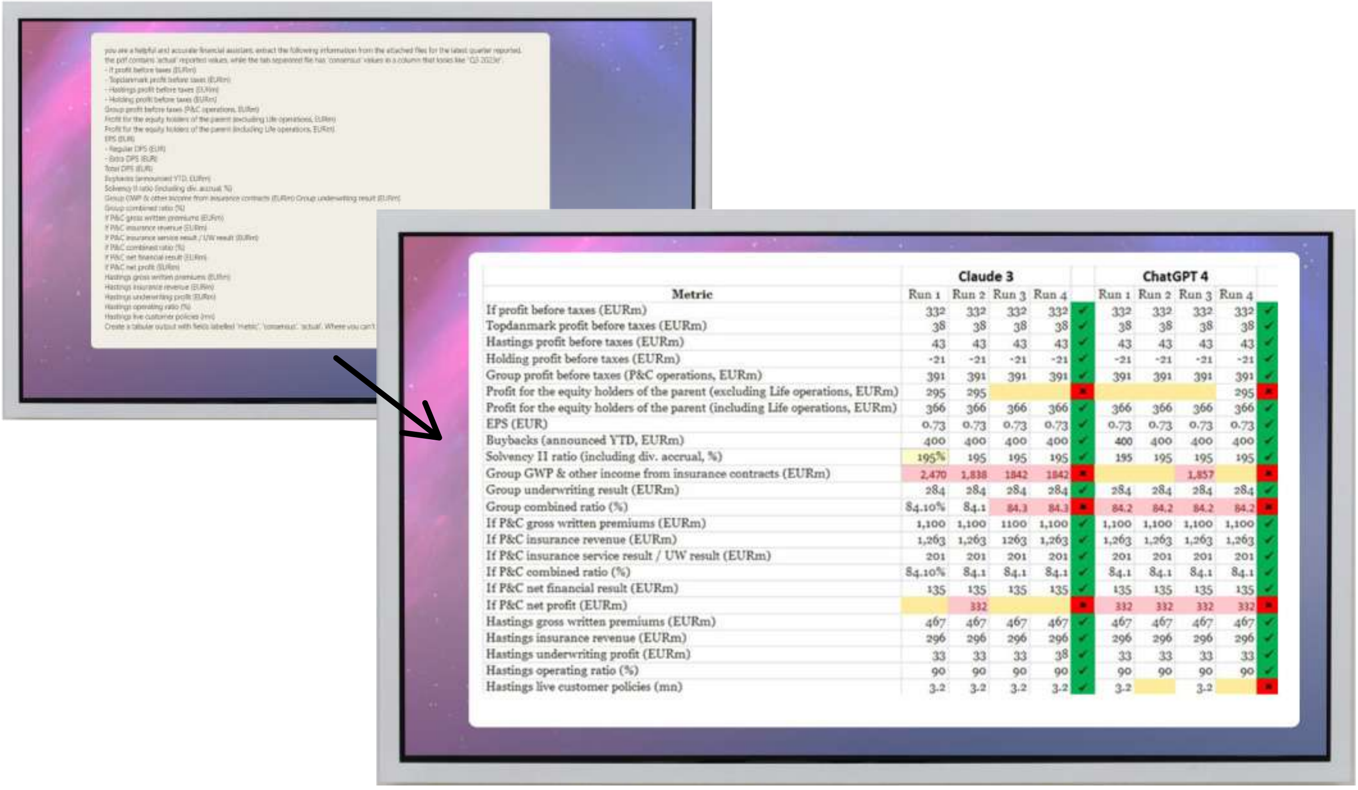
Methodology

In this approach, we examine the standard interaction model with Generative AI engines, focusing on the process of extracting and analysing financial data. The methodology includes:

1. **Data Collection:** Manually source and save quarterly reports and consensus information for each company of interest.
2. **Data Input:** Manually upload these quarterly reports and consensus files into the Generative AI interface.
3. **Query Development:** Develop and iteratively refine a set of queries for the language model (LLM) to execute. This iterative query refinement aims to optimize LLM performance specifically for financial data contexts, enhancing the precision and relevance of the responses.
4. **Results Export:** Export the generated results into a preferred Excel template format for further analysis and reporting.
5. **Query Repetition:** Conduct repeated query executions to assess the stability and consistency of responses from the LLM.

This methodology facilitates a thorough comparison between Generative AI models, such as Claude and ChatGPT, by evaluating their effectiveness in handling financial data, accuracy in generating insights, and overall performance stability.

Results for the Sampo Q3 2023 actuals presented in the table



Metric	Claude 3 Run 1	Claude 3 Run 2	Claude 3 Run 3	Claude 3 Run 4		ChatGPT 4 Run 1	ChatGPT 4 Run 2	ChatGPT 4 Run 3	ChatGPT 4 Run 4	
If profit before taxes (EURm)	332	332	332	332	✓	332	332	332	332	✓
Topdanmark profit before taxes (EURm)	38	38	38	38	✓	38	38	38	38	✓
Hastings profit before taxes (EURm)	43	43	43	43	✓	43	43	43	43	✓
Holding profit before taxes (EURm)	-21	-21	-21	-21	✓	-21	-21	-21	-21	✓
Group profit before taxes (P&C operations, EURm)	391	391	391	391	✓	391	391	391	391	✓
Profit for the equity holders of the parent (excluding Life operations, EURm)	295	295			✗				295	✗
Profit for the equity holders of the parent (including Life operations, EURm)	366	366	366	366	✓	366	366	366	366	✓
EPS (EUR)	0.73	0.73	0.73	0.73	✓	0.73	0.73	0.73	0.73	✓
Buybacks (announced YTD, EURm)	400	400	400	400	✓	400	400	400	400	✓
Solvency II ratio (including div. accrual, %)	195%	195	195	195	✓	195	195	195	195	✓
Group GWP & other income from insurance contracts (EURm)	2,470	1,838	1842	1842	✗			1,857		✗
Group underwriting result (EURm)	284	284	284	284	✓	284	284	284	284	✓
Group combined ratio (%)	84.10%	84.1	84.3	84.3	✗	84.2	84.2	84.2	84.2	✗
If P&C gross written premiums (EURm)	1,100	1,100	1100	1,100	✓	1,100	1,100	1,100	1,100	✓
If P&C insurance revenue (EURm)	1,263	1,263	1263	1,263	✓	1,263	1,263	1,263	1,263	✓
If P&C insurance service result / UW result (EURm)	201	201	201	201	✓	201	201	201	201	✓
If P&C combined ratio (%)	84.10%	84.1	84.1	84.1	✓	84.1	84.1	84.1	84.1	✓
If P&C net financial result (EURm)	135	135	135	135	✓	135	135	135	135	✓
If P&C net profit (EURm)		332			✗	332	332	332	332	✗
Hastings gross written premiums (EURm)	467	467	467	467	✓	467	467	467	467	✓
Hastings insurance revenue (EURm)	296	296	296	296	✓	296	296	296	296	✓
Hastings underwriting profit (EURm)	33	33	33	38	✓	33	33	33	33	✓
Hastings operating ratio (%)	90	90	90	90	✓	90	90	90	90	✓
Hastings live customer policies (mn)	3.2	3.2	3.2	3.2	✓	3.2		3.2		✗

Performance Metrics

Data Sourcing:
The data sourcing process is fully manual, requiring users to gather and save quarterly reports and consensus information independently.

Data Provision:
The data provision step is also manual, involving the direct upload of these reports and files into the Generative AI interface.

Accuracy:
The current accuracy of the model is rated as medium, with an accuracy level of approximately 80%.

Strengths

No Prescriptive Coding Required:
This approach does not require prescriptive coding, apart from the necessary query refinement, making it accessible for users without deep technical expertise.

Rapidly Improving LLM Accuracy:
The accuracy of the language models is increasing rapidly, enhancing their effectiveness over time.

Flexibility in Handling Diverse Document Formats:
The model demonstrates flexibility in processing and interpreting various document formats, which is crucial for handling diverse financial data.

Contextual and Nuanced Understanding:
The AI's ability to understand and interpret context and nuances within financial reports adds value to the analysis process.

Weaknesses

Fully Manual Data Provisioning Process:
The data provisioning process remains fully manual, which can be time-consuming and prone to human error.

Medium Accuracy:
Despite improvements, the current accuracy level is still medium, which may impact the reliability of some outputs.

Limited Document Format Acceptance:
The AI does not accept all document formats; while it handles PDFs well, it does not support Excel files, limiting its versatility.

Inconsistency in Query Outputs:
There is inconsistency in the outputs generated by the model across different queries, necessitating careful review.

Need for Manual Verification:
The results often require manual verification to ensure accuracy and reliability, adding an extra step to the process.
.

Outcomes of Approach 1: Pure Generative AI

Accuracy & Flexibility:
Provides moderate accuracy (80%) and handles diverse document formats well, but requires significant manual effort for data sourcing and query refinement.
User-Driven:
Manual data provisioning and verification mean more work for users, making this approach less efficient than the automated alternatives.
Consistency Issues:
The system occasionally struggles with consistency across queries, highlighting the need for further refinement.

Approach 2: AI Assistants Models

Methodology

This approach represents a fully automated solution that integrates AI assistants with traditional data analytics, embedded within ChatGPT's assistance layer. It's designed to streamline the entire process of sourcing, processing, and analyzing financial data.

Context and Process

Complete Automation:

This approach automates the sourcing and processing of raw quarterly results publications. The system can potentially source consensus data as well, though this requires user-defined parameters. Once the data is sourced, programmatic logic is employed to extract all key financial metrics automatically, ensuring a fully repeatable process.

100% Accuracy:

The automated approach ensures a high degree of accuracy, with data extraction achieving 100% accuracy, making it highly reliable for financial analysis.

Results Matrix:

After extracting the necessary metrics, the system automatically creates and populates a results matrix, which can be directly used by Investment Banking (IB) Equity Research teams for further analysis.

Performance Metrics

Data Sourcing:

The process of sourcing data is fully automated, significantly reducing the time and effort required to gather information.

Data Provision:

Data provision is also automated, with the system handling the extraction and processing of key metrics without manual intervention.

Accuracy:

The approach consistently delivers 100% accuracy in data extraction, making it a dependable choice for financial data analysis.

Weaknesses

Dependence on Input Data Formats:
The programmatic approach is highly dependent on the stability of input data formats. Any changes in the format can disrupt the process and require adjustments to the underlying logic.

Inflexibility to Change:
Due to its reliance on specific data formats, the system is less adaptable to changes in report formats or structures, which could pose challenges as companies update their reporting methods.

Initial Setup Complexity:
The initial setup of the underlying infrastructure can be time-consuming, particularly when analysing company publications. While feasible for a moderate number of companies, scaling the approach to handle hundreds of companies may be more challenging.

Limitations

Stability of Input Formats:
The effectiveness of this approach hinges on the stability of input data formats. Any significant changes could necessitate a reconfiguration of the system, impacting its overall flexibility.

Scalability Concerns:
While the system is highly effective for a smaller number of companies, scaling it to handle data for hundreds of companies requires careful analysis and infrastructure planning. This could introduce delays in implementation and reduce efficiency at larger scales.

Deep Dive – AI Assistants

Simplified Example of How AI Assistants Work – LLM Running Solo

Imagine you want to ask ChatGPT two simple questions:

1. Who is the President of the United States of America?
2. Extract the key financial metrics for Sampo Group from their Q3 2023 quarterly report.

Without an AI assistant, ChatGPT uses its algorithms to search its data repository for answers.

- For the first question, ChatGPT accurately identifies Joe Biden as the President of the USA.
- However, for the second question, when asked to extract specific financial metrics from a PDF report, ChatGPT's responses were inconsistent. While it could sometimes extract the data correctly, the results varied too much between attempts, making the answers unreliable.

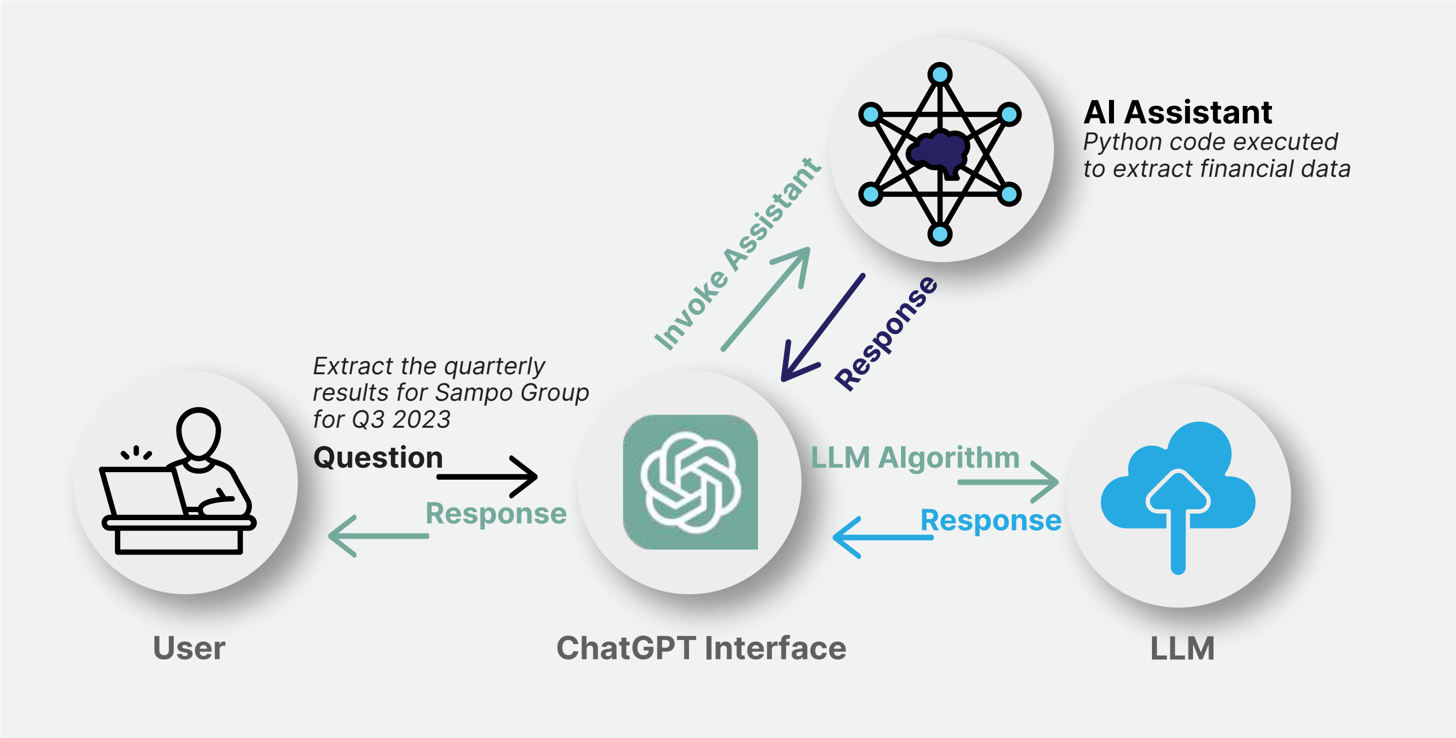
CALIMERE POINT

Simplified Example of How AI Assistants Help – Assistant Integration

As we saw, when ChatGPT worked alone, it struggled to consistently extract accurate financial data for Sampo Group from the Q3 2023 report, even when given the exact text to work from.

So, how does an AI assistant improve this process?

- **Assistant Integration:** The AI assistant allows us to integrate pre-scripted code behind the ChatGPT interface. This code is designed to handle specific tasks, like extracting financial data, with more precision.
- **User Interaction:** The user continues to interact with ChatGPT as usual. However, when you use specific prompts, the interface triggers the assistant to take over, rather than relying solely on the language model to generate the response.
- **Improved Responses:** As a result, ChatGPT provides responses that include both the output from the language model and the more reliable, consistent data extraction performed by the assistant, leading to more accurate and dependable results.



Outcomes of Approach 2: AI Assistants Models

High Accuracy & Automation:
Delivers 100% accuracy with fully automated data sourcing, processing, and results generation, making it highly efficient and reliable.
Dependence on Stability:
Relies on stable input data formats, which limits adaptability to changes and could affect scalability for larger datasets.
Scalability:
Effective for smaller-scale operations, but scaling up to handle hundreds of companies presents challenges due to the need for stable formats and infrastructure.

Approach 3: Sequential LLM / Generative AI Models

Methodology
This approach leverages multiple AI models in a predetermined sequence, utilizing OpenAI's API infrastructure to automate the process. The goal is to enhance the flexibility and efficiency of AI-assisted data analysis.

Context and Process
Enhancing Flexibility: This experimental approach was developed to build on the strengths of the AI Assistants model while addressing its limitations in flexibility. By combining different Generative AI models sequentially, the system aims to achieve more adaptable and robust data processing capabilities.

Data Handling:
The process begins by leveraging the GPT-3.5 API, which uses Python scripts to input data sets, such as quarterly reports and consensus files. The same queries used in Approach 1 are applied to identify and extract the most relevant pages from these files for further analysis.

Sequential AI Processing:
The specific pages identified by GPT-3.5 are then sent to GPT-4.0 for deeper analysis. GPT-4.0 processes these pages and produces a results matrix based on the data extracted.

Automation:
The entire workflow—from data sourcing and extraction to analysis and results generation—is automated using OpenAI's API infrastructure. This automation aims to reduce manual intervention and streamline the process.

Performance Metrics

Data Sourcing:
Fully automated, contingent on predefined file locations.

Data Provision:
Automated via API, ensuring a seamless flow of information between the models.

Accuracy:
Despite the advanced setup, the accuracy remains relatively low at 40%, which impacts the reliability of the outputs.

Strengths

Increased Automation: This approach offers a higher degree of automation compared to the purely generative AI methods, reducing the need for manual tasks.

Potential for Flexibility: By leveraging different AI models in sequence, the system can potentially combine their strengths, making it more adaptable to varied data structures and tasks.

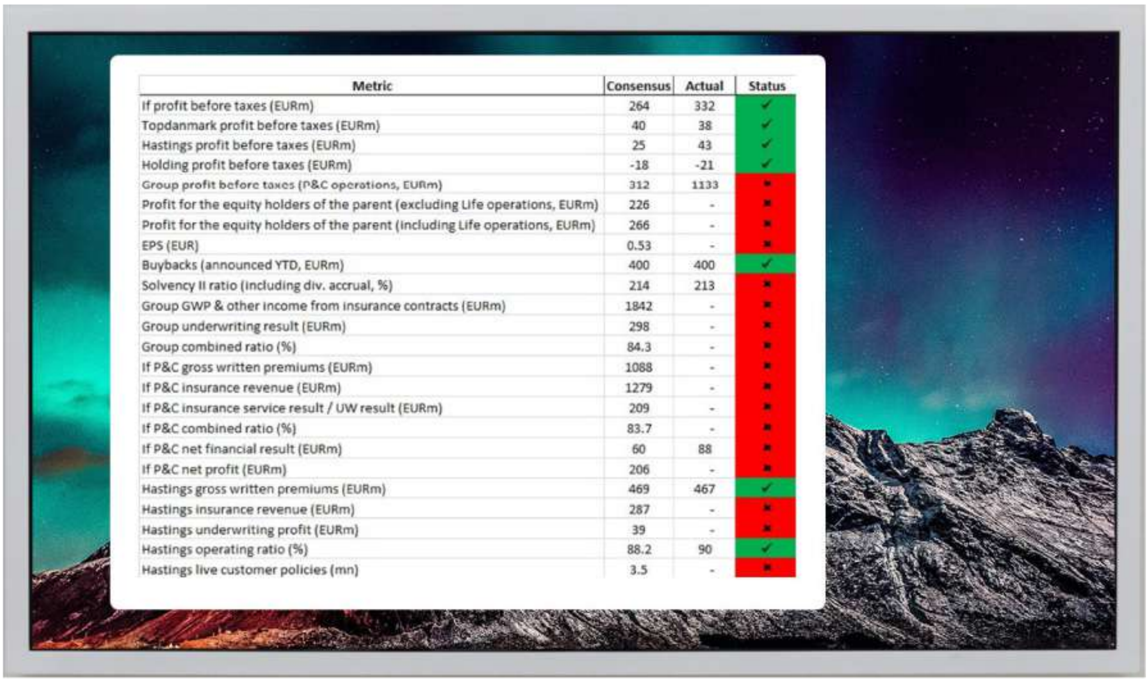
Robust Data Handling: The sequential use of models allows for more effective management of complex financial data, enhancing the system's overall robustness.

Weaknesses

Drop in Accuracy: The efficacy of results is lower than in Approach 1, with an inexplicable drop in accuracy when using APIs compared to direct user interface interactions. This discrepancy is currently under investigation.

Increased Complexity: The sequential process introduces additional complexity in both system design and maintenance, making it more challenging to manage and troubleshoot.

Error Propagation: The risk of error propagation increases with each sequential step, potentially affecting the quality of the final output.



Metric	Consensus	Actual	Status
If profit before taxes (EURm)	264	332	✓
Topdanmark profit before taxes (EURm)	40	38	✓
Hastings profit before taxes (EURm)	25	43	✓
Holding profit before taxes (EURm)	-18	-21	✓
Group profit before taxes (P&C operations, EURm)	312	1133	✗
Profit for the equity holders of the parent (excluding Life operations, EURm)	226	-	✗
Profit for the equity holders of the parent (including Life operations, EURm)	266	-	✗
EPS (EUR)	0.53	-	✗
Buybacks (announced YTD, EURm)	400	400	✓
Solvency II ratio (including div. accrual, %)	214	213	✗
Group GWP & other income from insurance contracts (EURm)	1842	-	✗
Group underwriting result (EURm)	298	-	✗
Group combined ratio (%)	84.3	-	✗
If P&C gross written premiums (EURm)	1088	-	✗
If P&C insurance revenue (EURm)	1279	-	✗
If P&C insurance service result / UW result (EURm)	209	-	✗
If P&C combined ratio (%)	83.7	-	✗
If P&C net financial result (EURm)	60	88	✗
If P&C net profit (EURm)	206	-	✗
Hastings gross written premiums (EURm)	469	467	✓
Hastings insurance revenue (EURm)	287	-	✗
Hastings underwriting profit (EURm)	39	-	✗
Hastings operating ratio (%)	88.2	90	✓
Hastings live customer policies (mn)	3.5	-	✗

Outcomes of Approach 3: Sequential LLM / Generative AI Models

Less Effective than Approach 1:
In terms of result accuracy, this approach is less effective than Approach 1. The automated process, while efficient, does not yet match the reliability of the earlier methods.

Ongoing Investigation:
Further investigation is ongoing to understand why the APIs produce different outputs compared to the user interface, with the goal of improving accuracy and consistency in future iterations.

6. Overall Conclusion and Analysis

Additional Experimentation

The team also conducted:
- Comparative assessments of different methods
- Stability testing of outputs
- Exploration of ways to reduce prescriptive logic while maintaining accuracy
- Investigation of discrepancies between user interface and API query results

This comprehensive experimentation strategy provided a nuanced understanding of how different AI approaches perform in financial data extraction. The varying results across accuracy, automation, and flexibility highlight the complexities involved in applying AI to real-world financial analysis tasks. These insights form the basis for future development and refinement of AI-assisted financial data processing systems.

Summary of Findings: Comparative Analysis of AI Model Performance and Capabilities

1. Pure Generative AI
Accuracy: Medium (80%)
Strengths: No prescriptive coding required – excluding query refinement; LLM Accuracy increasing rapidly
Weaknesses: Fully manual data provisioning process, Medium current level of accuracy; Doesn't accept all document formats (PDF – OK, Excel – not OK)

2. AI Assistants Models
Accuracy: High (100%)
Strengths: Highly accurate outputs; High degree of automation
Weaknesses: Prescriptive underlying code infrastructure; Vulnerable to input data format changes

3. Sequential LLM / Generative AI Models
Accuracy: Low (40%)
Strengths: Higher degree of process automation than via pure generative AI approaches
Weakness: Inexplicable drop in accuracy of results leveraging API versus user interface in Approach 1

Conclusion

The PoC demonstrated that Generative AI approaches can be used to automate the extraction of key financial metrics from quarterly results publications for the 5 insurance companies and populate Excel templates for onward analysis. However, the process is complex due to the variability in data structures and formats. While AI Assistants were able to generate accurate and automated outputs, they require considerable prescriptive logic. Pure Generative AI models are becoming more effective but still have limitations in accuracy and automation.

7. Opportunities with Generative AI

Based on the POC results, several opportunities for further exploration and development were identified:

1. Prototype the AI Assistant Approach across a broader population of companies to assess the consistency of quarterly reporting.
2. Continue testing hybrid approaches to find a balance between prescriptive AI Assistants and more flexible pure LLM approaches.
3. Align the technical capabilities demonstrated in the POC with the IB Equity Research team's processes to assess real-world feasibility.
4. Explore ways to handle the variability in report formats and data structures more effectively, possibly through adaptive AI models or improved data preprocessing techniques.
5. Investigate methods to improve the accuracy and automation of data extraction from highly unstructured sources, such as PDF'd PowerPoint charts.
6. Develop strategies to make the AI models more resilient to changes in input data formats, reducing the need for frequent updates to prescriptive logic.

By pursuing these opportunities, IB Equity Research could potentially streamline their data extraction processes, improve efficiency, and enhance the accuracy of their financial analysis models.

Contact

Peter Griffiths | CEO
Direct:+44 203 290 7817
Mobile: +44 7787 540764
peter.griffiths@calimerepoint.com

Murali Menon | Head of Data Science
Direct:+44 203 290 7817
Mobile: +44 7952 358194
murali.menon@calimerepoint.com

CALIMERE POINT

Calimere Point is a data analytics consultancy based in the United Kingdom. We provide our clients with value by developing robust data analytics capabilities that enable deeper insight and better decisions.

https://calimerepoint.com

© 2024 Calimere Point Risk Advisory Ltd. All Rights Reserved.
Registered Address: 63/66 Hatton Garden, Fifth Floor, Suite 23, London, United Kingdom, EC1N 8LE