

# IMPERIAL

## **Exposure and Performance Analysis of Investment Funds using Large Language Models (LLMs)**

An application using investment fund factsheets on the Llama3  
model through LangChain

**Sara Lakatos**  
**10/09/2024**

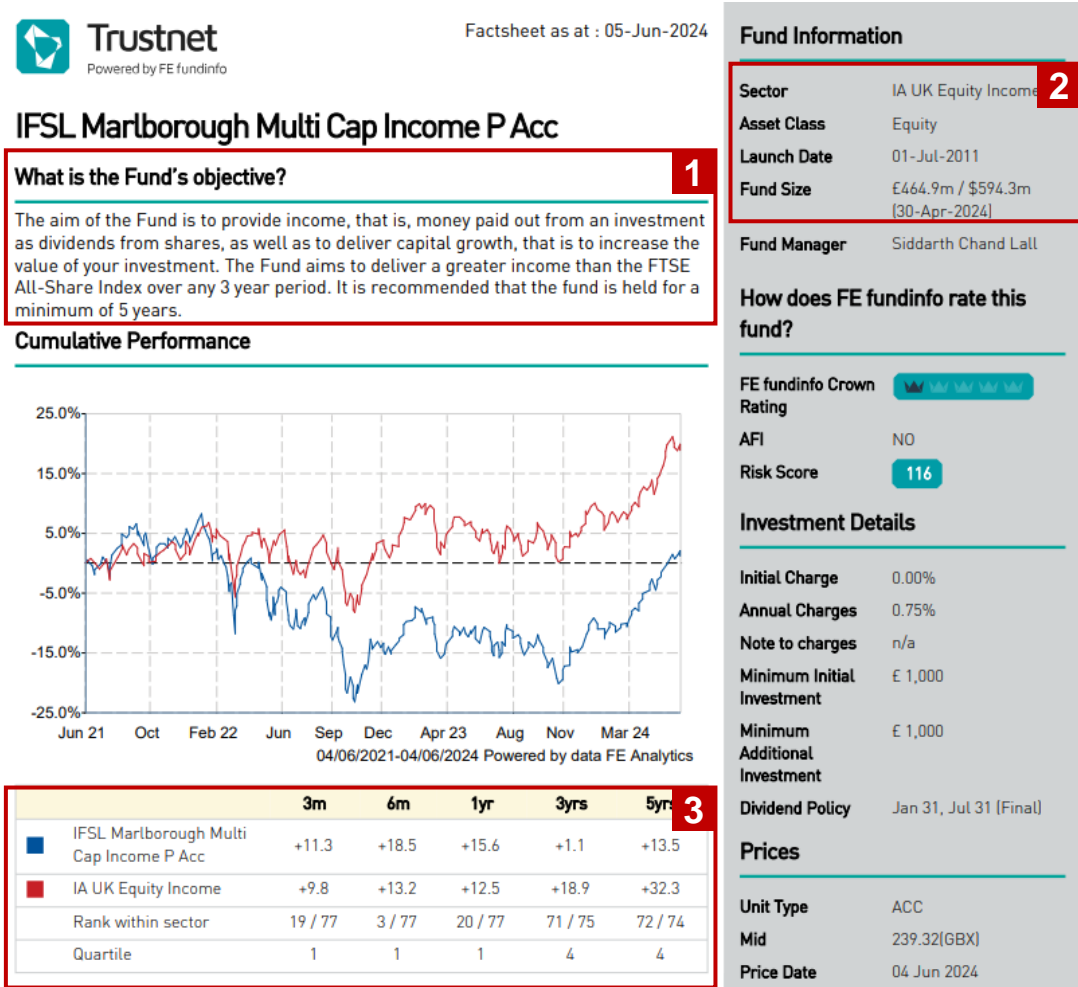
**Supervisors:**  
**Rhodri Nelson**  
**Marijan Beg**

# Dataset and Objective

The dataset consists of investment fund factsheets, a regulatory required report available in portable document format(PDF). Each factsheet must have two tables for historical performance and at least one investment allocation table, exposure across assets, sector, region and holdings.

**Objective:** to develop an investment fund chatbot (finbot), that can not only describe and retrieve information from a domain knowledge, but can **conduct its own investigation, within its domain knowledge** to identify “outlier” funds and describe their performance/exposure.

- 1 **Investment objective** -The fund's primary goal, such as growth, income, or capital preservation.
- 2 **Fund Information:** Basic details about the fund, including its name, type, and management.
- 3 **Historical Performance:** A record of the fund's past returns over specific periods.
- 4 **Investment Allocation Tables:** A breakdown of how the fund's assets are distributed across different investments.



Regional Breakdown (01 May 2024)

4

Rank	Regions	%
1	United Kingdom	95.60
2	Cash	1.50
3	Other	1.20
4	Ireland	0.80
5	Switzerland	0.50
6	United States	0.40

Sector Breakdown (01 May 2024)

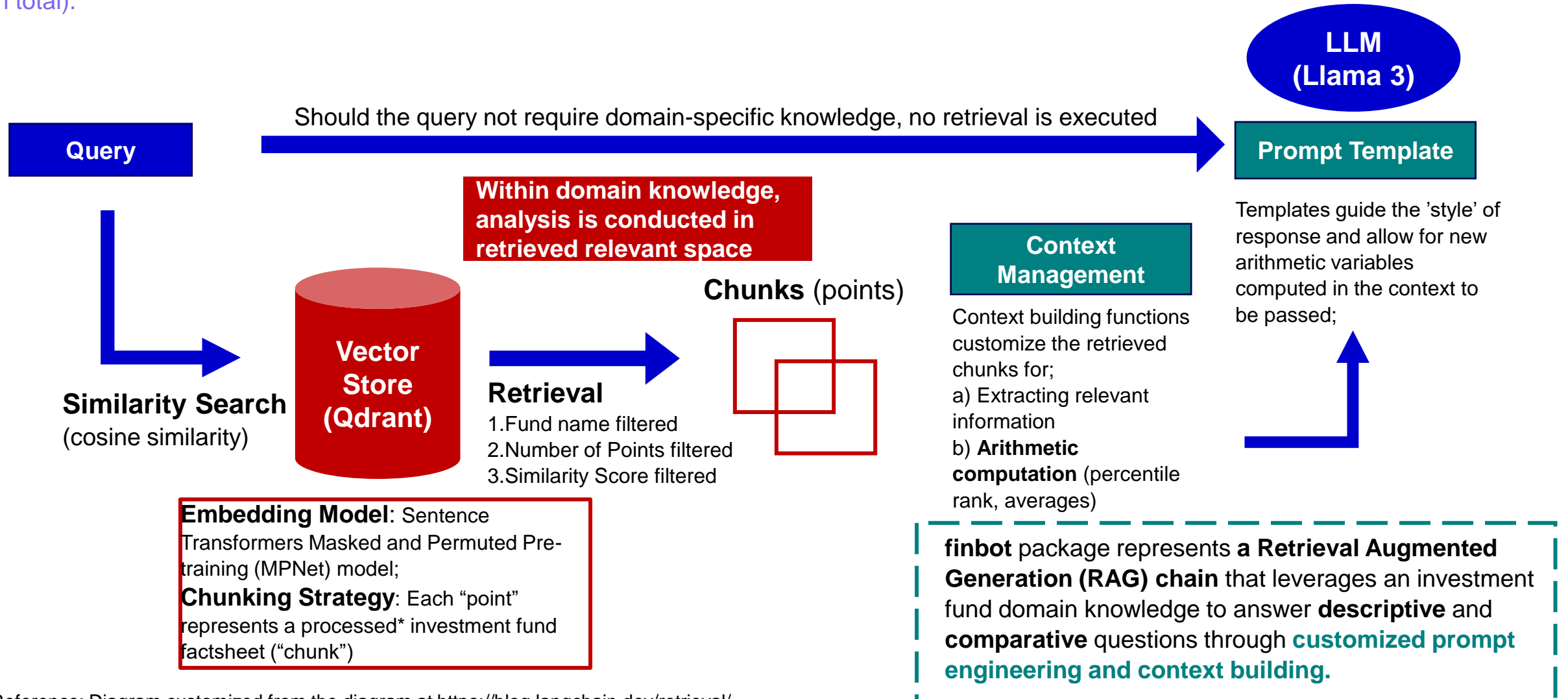
Rank	Sectors	%
1	Financials	27.80
2	Industrials	21.90
3	Consumer Discretionary	17.40
4	Real Estate	7.80
5	Consumer Staples	6.40
6	NonClassified	5.00
7	Basic Materials	3.40
8	Telecommunications	3.10
9	Energy	2.30
10	Others	4.90

# Methodology

## Implementation and Justification

# General Retrieval Augmented Generation (RAG) chain

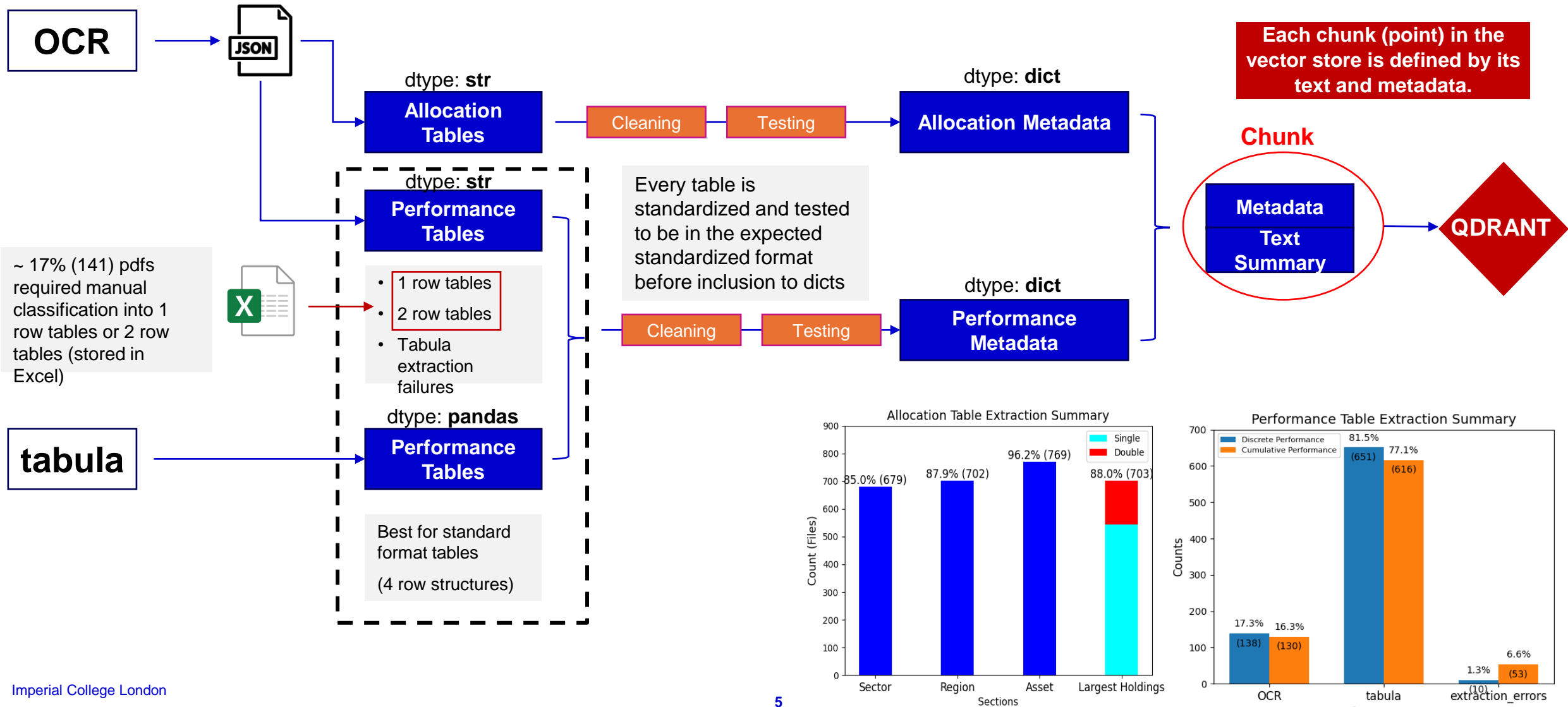
The basic structure of a RAG chain consists of a base LLM that may respond based on retrieved “points” from a vector store (its domain knowledge base) to answer domain-specific questions. The domain knowledge base considered represents investment fund factsheets as of March 2024 (799 in total).



Reference: Diagram customized from the diagram at <https://blog.langchain.dev/retrieval/>

# Data Processing

The fund information and investment objective (unstructured data) were processed separately to the tabular information (structured data). The unstructured data was mainly processed using regex pattern matching, below we present the processing pipeline for the structured data.



# Peer Groups

Peer Groups underpin the ‘relevance’ and ‘insightfulness’ of the comparative analysis, hence motivate the choice of chunking strategy for storing and embedding the factsheets into the Qdrant vector store (knowledge base).

- MorningStar defines a “peer” as a comparable investment fund with similar investment objective and strategy;
- A peer group is defined based on **both investment objective and holdings data** to reflect funds’ actual investment style;
- A minimum requirement for a peer group is a comparable investment universe\*, an indicator is **asset class**;
- The t-SNE plots show **embedding the full document reflects the distribution of asset classes** within the dataset better than just the investment objective, consistent with MorningStar research.

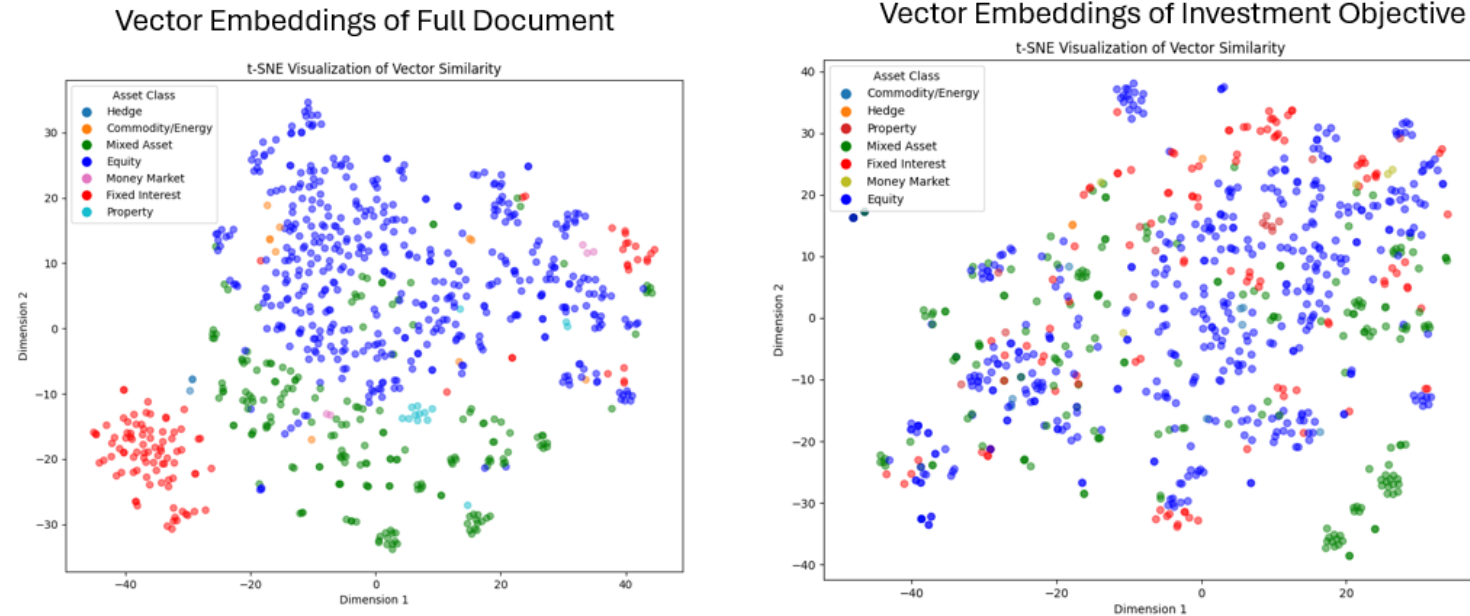
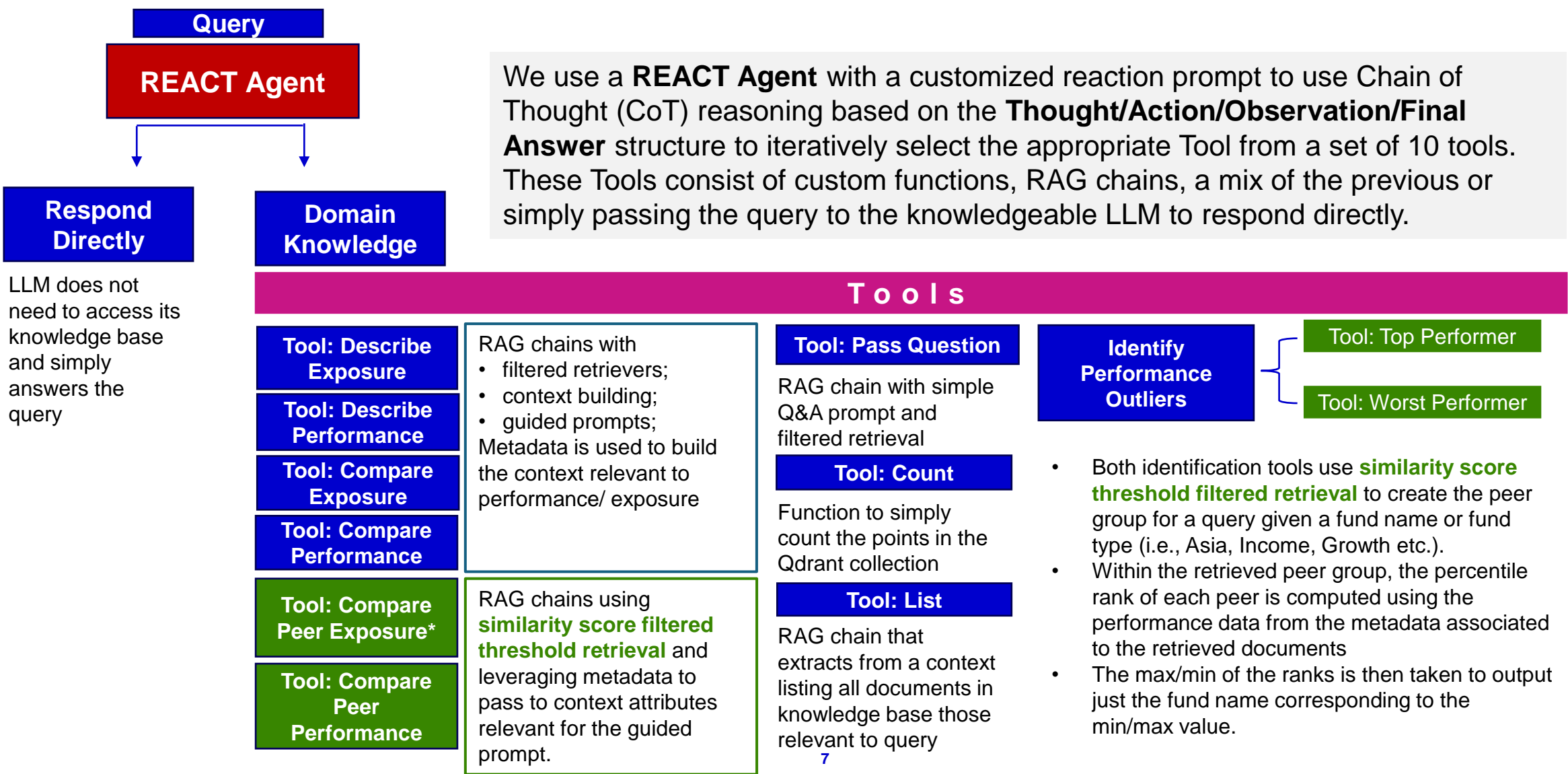


Fig 1. Vector Embedding Similarity: Embedding Model used: Sentence Transformers Masked and Permuted Pre-training (MPNet). The t-SNE (t-distributed Stochastic Neighbor Embedding) plots show the embeddings of the entire factsheet (LHS) for all funds and just the investment objective embeddings (RHS). The scope is to represent high-dimensional data (768 is the dimension of the vectors using the embedding model) in a 2D plane. The x and y axis do not have any interpretable meaning, the numerical values are abstractions such that the proximity of the embeddings in the higher-dimension are represented ‘close’ to each other in a lower dimension.

\*Investment universe: the opportunity set of investments a fund manager is allowed to choose from within the constraints of the fund’s investment objective.

# Finbot Implementation



\* Compare Peer Exposure was ultimately not used in the final implementation

# Results

## LangSmith Evaluations



# LangSmith Evaluations

Each LangSmith evaluation is conducted using the in-built contextual accuracy metric based on a custom built “reference output” defined as the context passed to the LLM to generate the output.

- General Questions – **poor**. The agent chain responds well to specific information requests, but performs very poorly on list queries and some unspecified information requests.
- Description Performance – **excellent**. The agent’s guided performance prompt is meant to: identify key highlights, compute active performance/ percentile rank; review peer metrics.
- Description Exposure – **good**. The agent’s guided exposure prompt is meant to describe the **allocation table in the context of its investment objective**. The evaluator LLM classifies responses as incorrect when the queried allocation table is not in the factsheet, despite the response noting this.
- Comparison – **poor**. The agent has a guided prompt receiving the 1-year average return percentile rank from the peer group. There is a discrepancy between the figures used as reference and those in the response. The peer groups used in the response don’t exactly match those expected in the reference output, impacting the arithmetic computation.

	General Questions (n = 69)	Description Performance (n = 100)	Description Exposure (n = 100)	Comparison (n = 100)
agent (finbot)	58%	99%	78%	57%
route	57%	92%	58%	53%
domain	75%	77%	85%	74%
simple	41%	49%	65%	67%

\*Figures indicate number of questions classified as correct (based on contextual accuracy) by evaluator LLM.

“Benchmark” chains used to compare the agent chain (finbot) to other implementations:

- **simple chain**: investment fund ignorant chain that cannot access the investment fund knowledge base.
- **domain chain**: knowledgeable chain that retrieves from the knowledge base but is not prompt engineered.
- **route chain**: similar chain to the agent chain, however it routes to underlying chains based on two routing functions and two query classification chains.

# Peer Group: Outlier Identification

For successful outlier identification we expect the Agent to call a sequence of tools namely one of the Identification Tools then one of the Description Tools, highlighting that analysis within the knowledge base, rather than extracted from the documents is used to inform the response

**Example Query:** Describe the sector allocation of the fund with the worst performance in the **High Yield** peer group.

Peer Group	Artemis High Income I Inc has rank 75.0
	Invesco High Yield (UK) Z Acc has rank 100.0
	Jupiter Strategic Bond I Acc has rank <b>25.0</b>
	JPM Global Bond Opportunities C Gr Acc has rank 50.0

The **worst** performer in peer group is **Jupiter Strategic Bond I Acc**

## Context: (passed to LLM)

Fund Name: Jupiter Strategic Bond I Acc  
AA: 11.20%  
AAA: 18.88%  
B: 17.43%  
BB: 28.89%  
BBB: 8.30%  
CC: 1.32%  
CCC: 5.01%  
NR: 2.13%  
Other: 6.41%  
Others: 0.43%

## Tool Sequence:

### 1. Identification Tool

The worst/top performer is identified based on the 1-year return of the funds in the peer group



### 2. Descriptive Tool

The identified performance outlier is then passed to the exposure/descriptive tool

## LLM Response:

The **Jupiter Strategic Bond I Acc Fund** has an investment objective to achieve a high income with the prospect of capital growth by seeking out the best opportunities within the fixed interest universe globally.

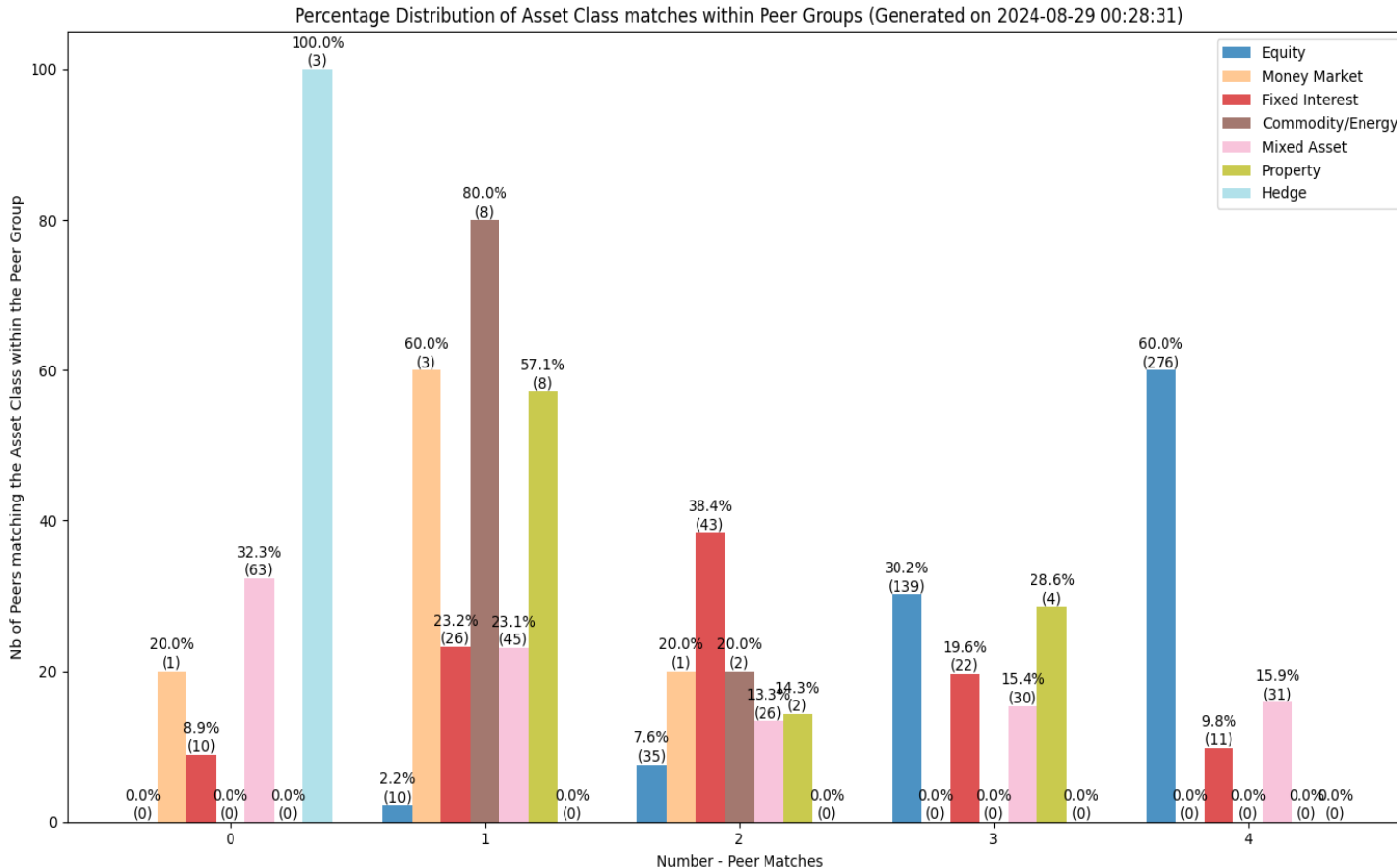
Upon **analyzing the provided allocation table**, we can see that the fund's exposure is **not entirely in line with its investment objective**. The allocation is heavily weighted **towards lower-rated bonds (B: 17.43%, BB: 28.89%, CCC: 5.01%)** which are higher-yielding but also carry more credit risk. While this may generate higher income, it contradicts the objective of achieving capital growth. On the other hand, the allocation to higher-rated bonds (AAA: 18.88%, AA: 11.20%) and government bonds is relatively smaller, which may not provide sufficient stability for capital preservation.

Additionally, the presence of CC-rated bonds (1.32%) and NR-rated bonds (2.13%) suggests that the fund takes on additional credit risk, which may not be aligned with the objective of achieving a high income with capital growth. Overall, while the fund's allocation provides some exposure to higher-yielding assets, it appears to prioritize income generation over capital preservation and growth.

Reference: Example from LangSmith, Project: FinBot; Identification Questions - Exposure

# Peer Group Evaluation

To evaluate the similarity score-based retrieval, we conduct an asset class evaluation of the retrieved peer groups to review the effectiveness of this method for peer group generation. This evaluation method is inspired by Qdrant's method to measure retrieval quality of its Approximate Nearest Neighbors (ANN) algorithms by identifying the relevant documents in the top k results.



Reference: Generated (with ChatGPT) through the extract\_peers.py script in the GitHub repository under peer analysis. For each fund in the knowledge base, we apply filtered similarity score retrieval (80%), store the asset class attributes of the identified peers and match it to the asset class of the given fund. We proceed to count the number of matches within a peer group (with ~4 peers) which are then aggregated across the peer groups. The aggregate numbers represent the number of asset class matches of a given asset class category within a peer group across all the peer groups.

- For **60% of the Equity funds in the knowledge base, all the peers retrieved were Equity funds**, whereas for another 30%, only 3 peers retrieved were Equity funds.
- For **Fixed Interest and Mixed Asset** there is a more even distribution of peer group retrieval, implying that for most of these fund types, **the peer groups contained at least 1 peer of the same asset class**, however for 9% and 32% of funds respectively, none of the peers matched the asset class.
- The **minority asset classes** in the sample, despite being enough to form a peer group, **were never grouped entirely together**, amongst these minorities the Property funds were most often grouped together, while each Hedge fund had peers of completely different asset classes.

# IMPERIAL

# Thank you

Exposure and Performance Analysis of Investment Funds  
using Large Language Models (LLMs)

10/09/2024