LLMs and Portfolio Attribution

Bruno de Melo

March 20, 2025

About me

Investment due diligence and risk oversight

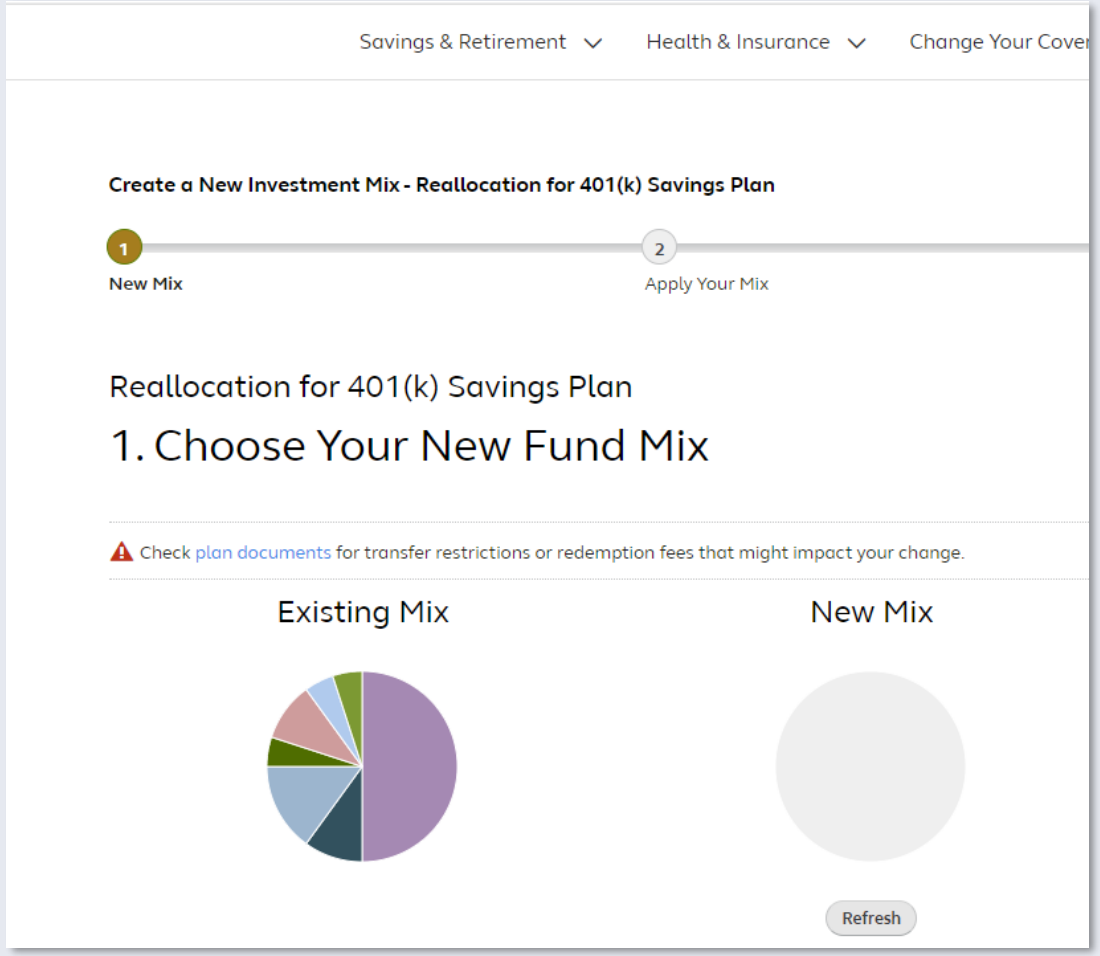
Gen AI applications on portfolio and risk management

Master's in Data Science

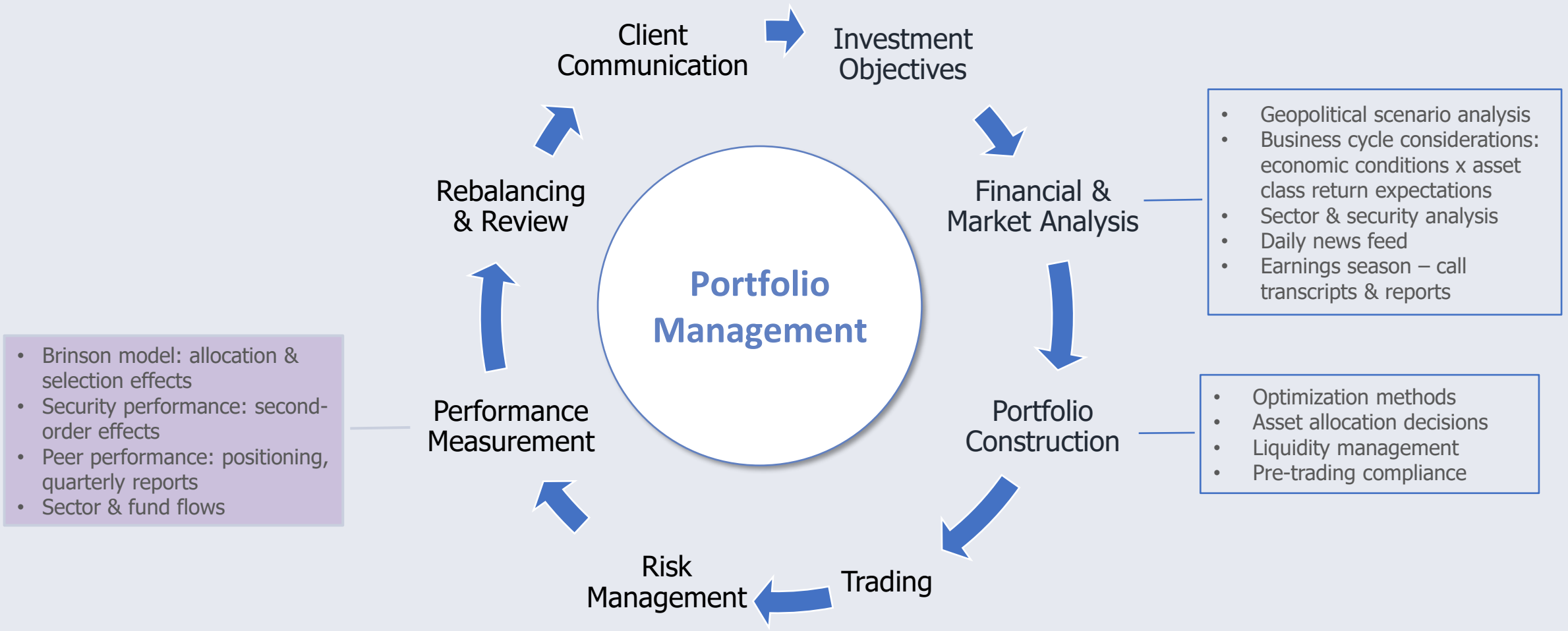
Portfolio Management & investment process

We are all Portfolio Managers!

- Reallocating/Rebalancing of your 401(k)
- That NVDA investment in your stock brokerage account
- Deciding to refinance when interest rates are lower
- Communicating to your partner or parents when you lost money on Crypto!

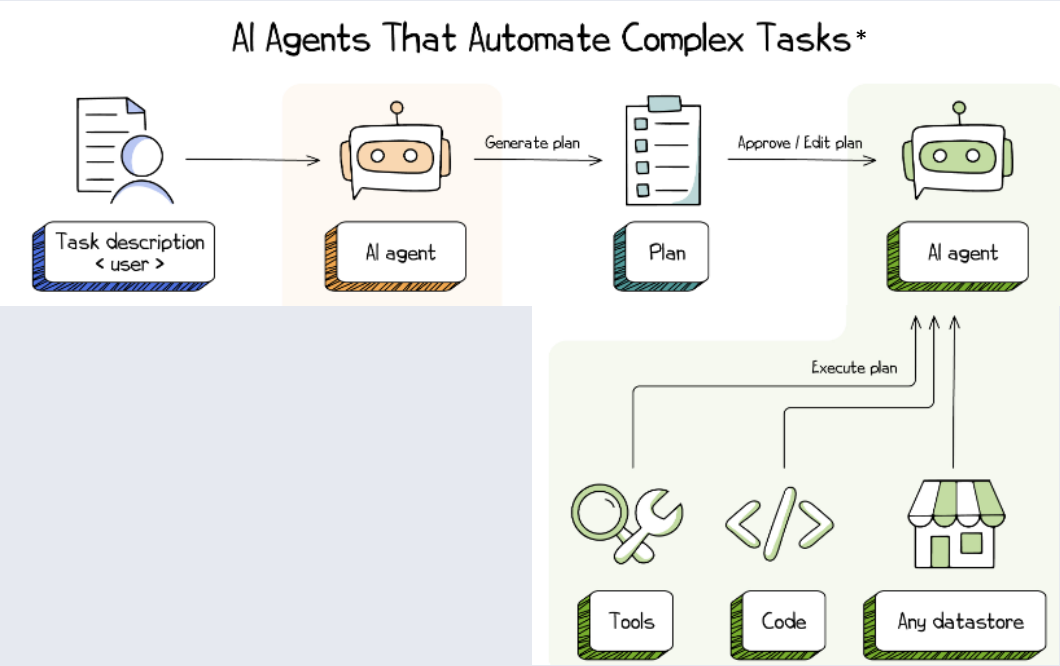


Portfolio Management is a complex endeavor



How can LLMs help with portfolio management?

Agentic Workflows / AI agents



Features

- Enhanced by new reasoning models
- Low barrier to experiment
- Highly customizable
- Incorporate cognitive science: memory layers, pre-frontal cortex planning

*https://twitter.com/bindureddy/status/1778508892382884265

Agent Leaderboard: Geminis offer the best cost/benefit

Rank	Model	Type	Vendor	Cost (I/O)	Avg Category Score (TSQ)
1st	claude-3-7-sonnet-20250219	Private	Anthropic	$3.00/$15.00	0.953
2nd	gemini-2.0-flash-001	Private	Google	$0.15/$0.60	0.938
3rd	gemini-2.0-flash-lite-001	Private	Google	$0.07/$0.30	0.933
4	gpt-4o-2024-11-20	Private	OpenAI	$2.50/$10.00	0.900
5	gpt-4.5-preview-2025-02-27	Private	OpenAI	$75.00/$150.00	0.900
6	gemini-1.5-flash	Private	Google	$0.07/$0.30	0.895
7	gemini-1.5-pro	Private	Google	$1.25/$5.00	0.885
8	o1-2024-12-17	Private	OpenAI	$15.00/$60.00	0.876
9	o3-mini-2025-01-31	Private	OpenAI	$1.10/$4.40	0.847
10	mistral-small-2501	Open source	Mistral	$0.10/$0.30	0.832
11	gpt-4o-mini	Private	OpenAI	$0.15/$0.60	0.832
12	qwen2.5-72b-instruct	Open source	Alibaba	$0.90/$0.90	0.817
13	mistral-large-2411	Private	Mistral	$2.00/$6.00	0.810
14	claude-3-5-sonnet-20241022	Private	Anthropic	$3.00/$15.00	0.801
15	Llama-3.3-70B-Instruct-Turbo	Open source	Meta	$0.90/$0.90	0.774
16	claude-3-5-haiku-20241022	Private	Anthropic	$0.80/$4.00	0.765
17	mistral-small-2409	Private	Mistral	$0.20/$0.60	0.750
18	ministral-8b-2410	Private	Mistral	$0.10/$0.10	0.689
19	Meta-Llama-3.1-8B-Instruct-Turbo	Open source	Meta	$0.20/$0.20	0.678
20	open-mistral-nemo-2407	Open source	Mistral	$0.15/$0.15	0.661

Features

- Based on Tool Selection Quality (TSQ)
- Assess agents' tool call performance, evaluating tool selection accuracy and effectiveness of parameter usage.
- Developed by `Galileo.ai`

A deeper dive into **Performance Attribution**

Question: Could an agentic workflow help me with performance attribution analysis*?

What it is

The measure of portfolio performance or excess return versus a benchmark, which actively informs the investment decision process

GICS Sector	Portfolio Weight (%)	Benchmark Weight (%)	Portfolio Return (%)	Benchmark Return (%)	Investment Decisions
Technology	40%	35%	8.0%	7.0%	**Overweight**
Healthcare	30%	25%	5.0%	6.0%	**Overweight**
Financials	30%	40%	4.5%	5.0%	**Underweight**
					Excess Return
Total	**100%**	**100%**	**6.05%**	**5.95%**	**0.10%**

*: arXiv:2403.10482

Question: Could an agentic workflow help me with performance attribution analysis*?

Why it's important

- Essential part of the investment-process
- Transparent way to assess portfolio managers
- Based on established financial and mathematical foundation

AI agent objectives

- #1: To analyze, reason, and explain a performance attribution report
- #2: Calculation engine (single and multi-level)
- #3: Complete question-answer tasks mimicking a Financial Analyst-level exam
- *Premise: Use of a readily available AI Agent*
- *Model used: gpt-4-0613*

**: arXiv:2403.10482*

Table + theory understanding

- More than just sorting highest/lowest row values
- Involves reasoning across the Weights and Returns columns to arrive at the correct answer
- Requires Intermediate calculations such as Benchmark Total Return

Breakdown | **Portfolio and Benchmark Returns & Weights** | **Effects**

GICS Sector	Portfolio Weight (%)	Benchmark Weight (%)	Portfolio Return (%)	Benchmark Return (%)	Investment Decisions	Allocation Effect	Selection Effect	Total Contribution
Technology	40%	35%	8.0%	7.0%	**Overweight**	0.053%	0.400%	0.453%
Healthcare	30%	25%	5.0%	6.0%	**Overweight**	0.002%	-0.300%	-0.298%
Financials	30%	40%	4.5%	5.0%	**Underweight**	0.095%	-0.150%	-0.055%
					Excess Return			
Total	**100%**	**100%**	**6.05%**	**5.95%**	**0.10%**	**0.15%**	**-0.05%**	**0.10%**

Allocation Effect Explanation

Allocation Effect: Measures whether overweighting or underweighting certain sectors added or detracted value.

$$A_i = (w_i - W_i) \times (b_i - b)$$

Technology: (40%-35%) * (7%-5.95%) = 0.053%

GICS Sector	Portfolio Weight (%)	Benchmark Weight (%)	Portfolio Return (%)	Benchmark Return (%)	Investment Decisions	Allocation Effect	Selection Effect	Total Contribution
Technology	40%	35%	8.0%	7.0%	Overweight	0.053%	0.400%	0.453%
Healthcare	30%	25%	5.0%	6.0%	Overweight	0.002%	-0.300%	-0.298%
Financials	30%	40%	4.5%	5.0%	Underweight	0.095%	-0.150%	-0.055%
					Excess Return			
Total	**100%**	**100%**	**6.05%**	**5.95%**	**0.10%**	**0.15%**	**-0.05%**	**0.10%**

Selection Effect Explanation

Selection Effect: Measures whether security selection within a sector was successful.

$$S_i = w_i \times (r_i - b_i)$$

Technology: (40%) * (8% - 7%) = 0.400%

GICS Sector	Portfolio Weight (%)	Benchmark Weight (%)	Portfolio Return (%)	Benchmark Return (%)	Investment Decisions	Allocation Effect	Selection Effect	Total Contribution
Technology	40%	35%	8.0%	7.0%	Overweight	0.053%	0.400%	0.453%
Healthcare	30%	25%	5.0%	6.0%	Overweight	0.002%	-0.300%	-0.298%
Financials	30%	40%	4.5%	5.0%	Underweight	0.095%	-0.150%	-0.055%
					Excess Return			
Total	**100%**	**100%**	**6.05%**	**5.95%**	**0.10%**	**0.15%**	**-0.05%**	**0.10%**

Allocation Effect: Weights and Returns

Allocation Effect: Sector bets

$$A_i = (w_i - W_i) \times (b_i - b)$$

Weights W	Returns R: $(b_i - b) > 0$	Returns R: $(b_i - b) < 0$	Returns R: $(b_i - b) = 0$
$(w_i - W_i) > 0$	Allocation Positive W: Overweight R: Outperformance	Allocation Negative W: Overweight R: Underperformance	Zero Allocation W: Overweight R: Neutral
$(w_i - W_i) < 0$	Allocation Negative W: Underweight R: Outperformance	Allocation Positive W: Underweight R: Underperformance	Zero Allocation W: Underweight R: Neutral
$(w_i - W_i) = 0$	Zero Allocation W: Neutral R: Outperformance	Zero Allocation W: Neutral R: Underperformance	Zero Allocation W: Neutral R: Neutral

Selection Effect: Weights and Returns

Selection Effect: Security bets

$$S_i = w_i \times (r_i - b_i)$$

Weights W	Returns R: $(r_i - b_i) > 0$	Returns R: $(r_i - b_i) < 0$	Returns R: $(r_i - b_i) = 0$
$w_i > 0$	Selection Positive Outperformance	Selection Negative Underperformance	Zero Selection Neutral
$w_i < 0$	Not applicable	Not applicable	Not applicable
$w_i = 0$	Zero Selection No exposure	Zero Selection No exposure	Zero Selection No exposure

Research Objectives

#1: Analysis Engine

Write sentences with correct explanation (~300 sentences)

Allocation: The Technology sector had a **positive** allocation effect due to the fund being **overweight** in a sector that **outperformed** the benchmark.
Selection: The Technology sector also had a **positive** selection as investments in this sector **outperformed** the sector benchmark.

GICS Sector	Portfolio Weight (%)	Benchmark Weight (%)	Portfolio Return (%)	Benchmark Return (%)	Investment Decisions	Allocation Effect	Selection Effect	Total Contribution
Technology	40%	35%	8.0%	7.0%	**Overweight**	0.053%	0.400%	0.453%
Healthcare	30%	25%	5.0%	6.0%	**Overweight**	0.002%	-0.300%	-0.298%
Financials	30%	40%	4.5%	5.0%	**Underweight**	0.095%	-0.150%	-0.055%
					Excess Return			
Total	**100%**	**100%**	**6.05%**	**5.95%**	**0.10%**	**0.15%**	**-0.05%**	**0.10%**

Research Objectives

#2: Calculation Engine (12 portfolios)
#3: Question-Answering Tasks (140 questions)

The exhibit below shows the weights and returns of a portfolio and its benchmark.

	Portfolio Weight	Benchmark Weight	Portfolio Return	Benchmark Return
Health Care	10%	20%	3%	2%
Utilities	30%	30%	4%	4%
Consumer Goods	60%	50%	7%	8%
Total	**100%**	**100%**	**5.7%**	**5.6%**

According to the Brinson–Fachler attribution model, the allocation effect from the Consumer Goods sector is *closest* to:

A –0.50%.
B 0.24%.
C 0.80%.

Exhibit 17 – QAMC question - Bacon and Wright (2019)

What the AI agent does with the table input

create_csv_agent: This agent calls the *Pandas DataFrame* agent under the hood, which in turn calls the *Python agent*, which executes LLM generated Python code - this can be bad if the LLM generated Python code is harmful. Use cautiously.

Overall Results

- Required **Prompt Engineering** with advanced instructions like **Chain-of-Thought** and **Plan & Solve**
- Encouraging Results!

#1: Analysis Engine: Sentence Generation

- 94-99% of Semantic Similarity
- 93-100% Accuracy in Reasoning

#2: Calculation Engine

- 100% Accuracy

#3: Question-Answering tasks

- Multiple Choice: 89% accuracy
- Direct Answer: 84% accuracy

Using Qwen-QwQ-32B:

- Achieving 100% Accuracy in Reasoning
- No 'agents' used

Takeaways – the results are encouraging!

01

Good Enough at Analysis

Produces a good enough performance attribution analysis

02

Great Calculation Engine

Calculates its main components following specific formulae and step-by-step instructions

03

Winning couple

Use of a pre-canned AI agent coupled with prompt engineering yields very robust results

Next Step: Insights Generation

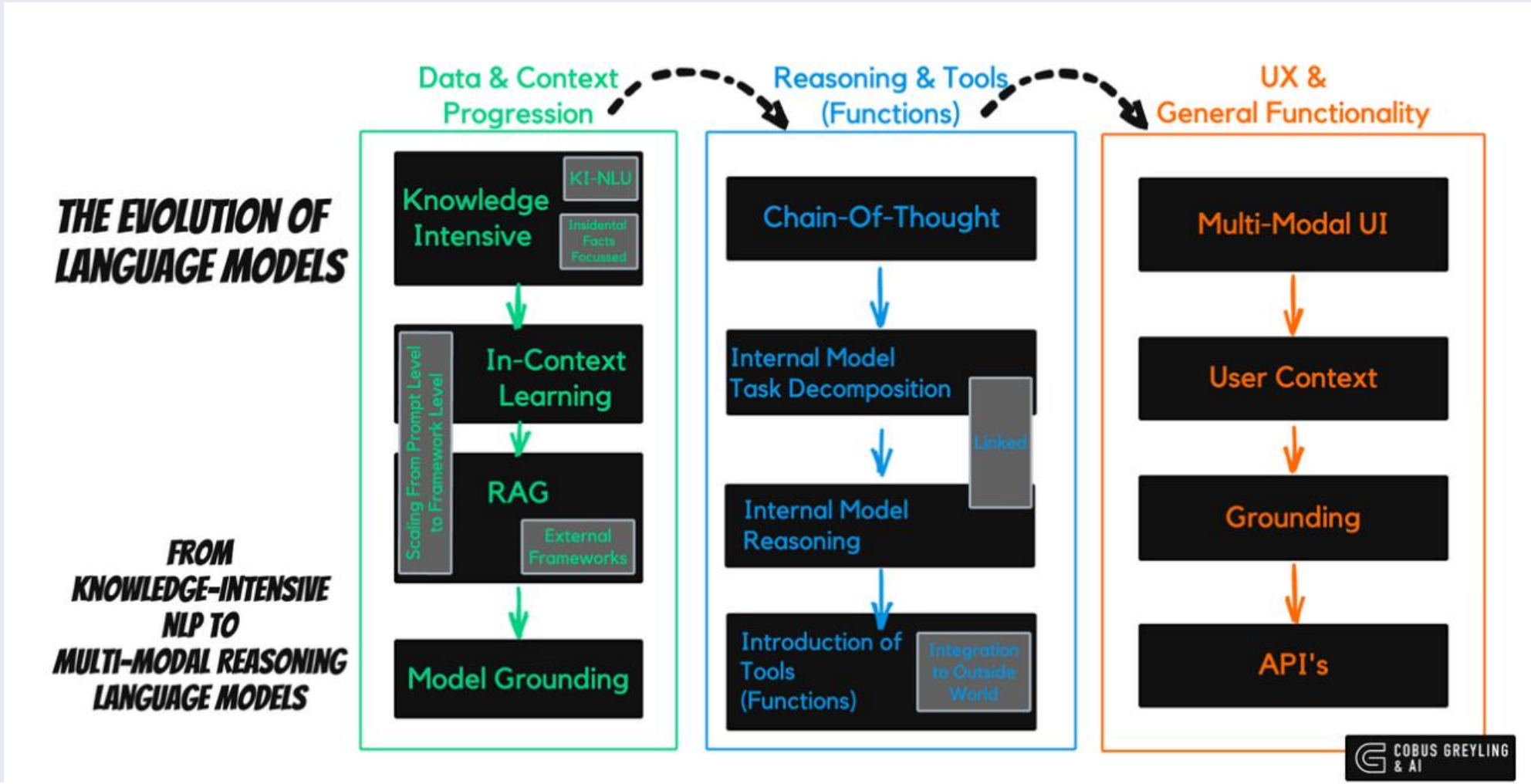
Scenario 1: Investment Due Diligence on a new manager

Scenario 2: Investment thesis validation and monitoring

Scenario 3: Insights across a platform of funds

Scenario 4: Report generation

Appendix


THE EVOLUTION OF LANGUAGE MODELS
FROM KNOWLEDGE-INTENSIVE NLP TO MULTI-MODAL REASONING LANGUAGE MODELS
Data & Context Progression
Reasoning & Tools (Functions)
UX & General Functionality
Knowledge Intensive
KI-NLU
Insidental Facts Focussed
In-Context Learning
Scaling From Prompt Level to Framework Level
RAG
External Frameworks
Model Grounding
Chain-Of-Thought
Internal Model Task Decomposition
Linked
Internal Model Reasoning
Introduction of Tools (Functions)
Integration to Outside World
Multi-Modal UI
User Context
Grounding
API's
COBUS GREYLING & AI