Generative AI in Finance

Student Investment Fund

Amin Izadyar

May 9, 2025

IMPERIAL

What is an LLM?

What *is* a Large Language Model?

- A giant statistical *autocomplete* trained on an internet-scale text library
- Learns patterns of how words follow each other (tokens)
- Internalises those patterns so well it can **mimic human language**—using words to "communicate" or even "think out loud"
- Generates, summarises, translates and answers questions

Turning Words into Numbers (Vectors)

- Each word is turned into a long list of numbers —a *vector* in a high-dimensional space.
- Words with similar meaning sit near each other.
- Famous analogy:

$$\text{vector}(\text{“King”}) - \text{vector}(\text{“Male”}) \approx \text{vector}(\text{“Queen”}) - \text{vector}(\text{“Female”})$$

  semantic relationships appear just from reading lots of text.
- These vectors are the raw ingredients we feed into the model.

Predicting the Next Word — How the Model Learns

- Stack the word vectors into a sequence and pass them through a (transformer) "function":

$$\hat{x}_{t+1} = f_\theta(x_1, \ldots, x_t)$$

  where θ are the model's adjustable weights.
- The function guesses the next word (or the missing word) in the sentence.
- Training loop:
  1. Compare the guess with the **real** next word.
  2. Measure the error (how wrong were we?).
  3. Tiny adjustment to all weights θ to reduce the error (gradient descent).
- Billions of these tiny nudges—over billions of sentences—shape θ so the model speaks fluid, human-like language.

How a Trained LLM Generates an Answer

1. **Prompt in** – User types: *"What causes Gold prices to rise?"*
2. Text is converted into vectors and fed into the trained function $f_\theta(x_1, \ldots, x_t)$.
3. The model returns a probability for every possible next token.
4. Pick the word with the highest probability.
5. Append the chosen token to the prompt, feed the longer sequence back in, and repeat (until some stopping rule).

Key Idea

Generation is just "predict-the-next-word".

- **Token**
  The smallest text chunk the model sees—often a word piece. "Investing" may split into two tokens: ``` ``Invest'' ``` + ``` ``ing'' ```. Model length limits and cost are counted in tokens.

- **Temperature**
  A variable that controls randomness in the generated output.
  - $\tau = 0$ → deterministic "most-likely" answer.
  - $\tau > 1$ → more exploration, creative text.

- **Embedding**
  A high-dimensional vector representing a token's meaning.

Jargon — Fine-Tuning (Domain Adaptation)

- **What?** Continue training a general LLM on a focused corpus so its weights "speak finance."
- **How?** Supply pairs of *(question, answer)* or *(prompt, completion)* for the model to learn.
- **Mini example:**
  - *Base model*: GPT-3.5-like.
  - *Extra data*: 1000 annotated earnings-call transcripts.
  - *Outcome*: answers finance questions more accurately.
- **Sample Q–A pairs used for fine-tuning**
  - **Q:** "Tesla beats Q2 earnings expectations amid record deliveries"
    A: Positive
  - **Q:** "Nike warns of slowing sales in China as consumer demand cools"
    A: Negative
  - **Q:** "IBM revenue matches estimates; company reaffirms full-year guidance"
    A: Neutral

LLM Applications in Finance

Idea 1 — Forecast Future Returns from Vectors

- Feed a business news headline, earnings-call paragraph, 10-K snippet or FED transcripts into an LLM.
- The model outputs a high-dimensional **embedding vector** that captures meaning and tone.

$$r_{i,t+1} = \beta_0 + \boldsymbol{\beta}^\top \mathbf{x}_{i,t} + \varepsilon_{i,t+1}$$

- $\mathbf{x}_{i,t}$ = embedding of latest news for stock i at time t.
- Fit $\boldsymbol{\beta}$ on past data — then plug in today's vector to get tomorrow's expected return.

Idea 1 — Forecast Future Returns from Vectors

- **Estimation ideas:** Apply *PCA* to compress vectors to a few factors *or* let a sparse method such as *Lasso* pick only the most predictive factors.
- Once text is in vector form, it drops into the same toolbox—OLS, PCA, Lasso, Random Forest, you name it.

- **Cash-flow growth** — predict next-quarter operating cash flow from management language.
- **Credit-spread moves** — forecast daily change in CDS spreads from bond-market news.
- **Interest-rate surprises** — map FOMC statement wording to post-meeting yield shifts.

Example use cases: 1- Chen et al 2023 | 2- Sarkar 2025

Why prompts?

- LLMs can act as analysts.
- Works on *any* corpus – a CSV row, a headline, even a PDF page.

User Input: Headline news, inflation reading, ...
Prompt: *“Classify the impact on {TICKER} as Positive, Neutral or Negative. Reply with one word.”*

Example I/O

Input: “Nike warns of slowing sales in China.”

LLM output: Negative

Same pattern scales: ask for probability scores, Valuation uncertainty, ESG tone, supply-chain risk, etc.

Idea 2 — Structured Prompts → Clean Signals

1. **Numerical coding**

$$s_{i,t} = \begin{cases} +1 & \text{Positive} \\ 0 & \text{Neutral} \\ -1 & \text{Negative} \end{cases}$$

   One score per firm i per period t.

2. **Aggregate if needed**: simple moving average could be applied.
3. *End-product:* matrix $\mathbf{X}_t$ ready for your favourite econometrics / ML tool.

$$r_{i,t+1} = \beta_0 + \boldsymbol{\beta}^\top \mathbf{x}_{i,t} + \varepsilon_{i,t+1}$$

- $\mathbf{x}_{i,t}$ – vector of prompt-based variables.

Example use cases: 1- Chen et al 2023 | 2- Lopez-Lira and Tang 2023

Application to the FX market

Time	Cur.	Imp.	Event	Actual	Forecast	Previous
			Monday, February 5, 2024			
03:30	SEK	★☆☆	Services PMI (MoM) (Jan)	51.8		50.3
05:00	GBP	★☆☆	Car Registration (YoY) (Jan)	8.0%		9.8%
05:00	EUR	★★☆	HCOB Eurozone Composite PMI (Jan)	47.9	47.9	47.6
05:00	EUR	★★☆	HCOB Eurozone Services PMI (Jan)	48.4	48.4	48.8
05:30	GBP	★★☆	S&P Global Composite PMI (Jan)	52.9	52.5	52.1
05:30	GBP	★★☆	S&P Global Services PMI (Jan)	54.3	53.8	53.4
05:30	EUR	★☆☆	Sentix Investor Confidence (Feb)	-12.9	-15.0	-15.8
06:00	GBP	★☆☆	United Kingdom New Passenger Cars Registration (Jan)	142,876.0		141,092.0
06:00	NOK	★☆☆	House Price Index (YoY) (Jan)	1.40%		0.90%
06:00	EUR	★☆☆	PPI (MoM) (Dec)	-0.8%	-0.8%	-0.3%
06:00	EUR	★☆☆	PPI (YoY) (Dec)	-10.6%	-10.5%	-8.8%
09:15	CAD	★☆☆	Reserve Assets Total (Jan)	116.3B		118.3B
10:45	USD	★★☆	S&P Global Composite PMI (Jan)	52.0	52.3	50.9
10:45	USD	★★★	S&P Global Services PMI (Jan)	52.5	52.9	51.4

Economic Data Releases:

- Economic calendar data
- January 1996 to September 2024
- G10 currencies.
- Number of observations: 174,820

Return data:

- End-of-day (London time) exchange rates and one-month forward rates from Bloomberg.

Example

Currency: AUD

Input: *News headline:* Trade Balance, *Actual:* 4.502B, *Forecast:* 6.100B, *Previous:* 7.180B

Output: *Analysis*: The trade balance figure came in lower than both the forecast and the previous month, indicating a smaller than expected surplus. This could suggest that exports have decreased or imports have increased more than anticipated, potentially signaling weaker economic activity. A lower trade balance can negatively impact the currency as it implies that fewer Australian dollars are demanded in international markets to purchase Australian goods and services., *Direction*: WEAKEN.

Prompt

Forget all previous instructions. You are a financial analyst. You will be given a news headline. Please analyze the news's effect on {currency} and provide a brief response. Then indicate whether this headline implies {currency} will STRENGTHEN, WEAKEN, or have an INSIGNIFICANT OR UNCERTAIN effect. Generate the output in this format: {(ANALYSIS: short analysis discussing the channel), (DIRECTION: one of STRENGTHEN, WEAKEN, INSIGNIFICANT OR UNCERTAIN)}

- Terminology:
  - Positive news: (Direction: *STRENGTHEN*)
  - Negative news: (Direction: *WEAKEN*)
  - Neutral news: (Direction: *INSIGNIFICANT OR UNCERTAIN*)

Variable Construction

- At time t, consider a lookback period of τ.
- L = $[t - \tau, t]$ is the lookback window.
- For currency c at time t we define:

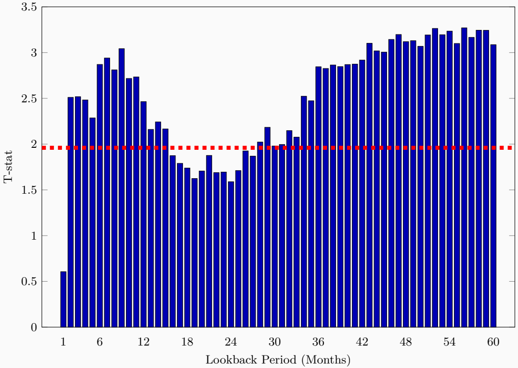
$$\textbf{Strength ratio}_{c,t} = \frac{\text{Number of Positive News Related to c in L}}{\text{Total Number of News Related to c in L}}$$

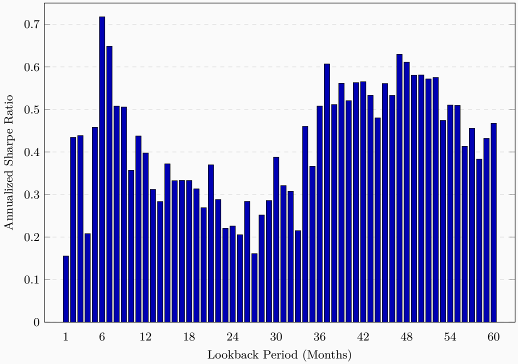
$$\textbf{Weakness ratio}_{c,t} = \frac{\text{Number of Negative News Related to c in L}}{\text{Total Number of News Related to c in L}}$$

$$\textbf{AI-FX ratio}_{c,t} = \text{Strength}_{c,t} - \text{Weakness}_{c,t}$$

$$R_{c,t+1} = \alpha_t + \beta \times \mathrm{AId}_{c,t} + \epsilon_{c,t}$$

- $R_{c,t+1}$ denotes the monthly excess return for currency c at time $t+1$.
- The AI-FX ratio significantly forecasts next month's currency returns across various lookback periods.





- Sort currencies based on AI-FX ratio.
- At the end of each month, go long the top 2 and short the bottom 2.
- The strategy earns an economically significant Sharpe ratio for different choices of lookback period.

More ideas

- Detecting unusuall financial communication (Beckmann et al 2024)
- Measuring information content in earnings calls (Bai et al 2023)
- Thematic Scoring: Quantifying Contextual Narratives (Lopez-Lira et al 2023)

Big Challenge — Look-Ahead Bias (Memorization)

- **Trained on the past → remembers the past.**
  GPT-4o can remember information that sit *inside* its training set. The cutoff date is October 2023.
- **Back-tests inside that window ≠ real forecasting.** High "prediction" accuracy may just be the model recalling memorised details, not derived based on reasoning.
- **Simple Work-arounds often fail.**
  - Telling the model "forget data after 2010" barely helps.
  - Masking company names or release dates might not be enough.
- **Practical safeguard.**
  - Evaulte the performance of the model after the training sets cutoff date.
  - Use chronologically consistent LLMs. (He et al 2023)
  - Desing tests to check whether your results are driven by look-ahead bias.

Key takeaway: if the model could have seen tomorrow's number during training, any "alpha" you measure today might be an illusion.

Thank you for your attention.

Questions?