

Effective Usage of Large Language Models in Investment with an Example of Asset Allocation Analysis

Kailan Shang

The recent breakthrough of large language models (LLMs) showed us a promising technology that can provide intelligent and creative solutions with little instructions. Although the technology is still at its early stage with many improvements and real-world applications to follow, there is no doubt that LLMs can act as a helpful adviser and significantly improve our productivity. Successful stories of using LLMs are not rare, including writing a book or building a chatbot within a few days, passing exams, developing computer programs, and so on. However, to effectively utilize LLMs in the investment field, some gaps need to be narrowed from the perspective of investment professionals. First, we need to understand where the intelligence comes from. This will help us evaluate whether the intelligence is what we can rely on and how it can be communicated with stakeholders in investment decision-making. In addition, the popular applications of LLMs such as ChatGPT and Bard were built to be a generalist. They need to be adapted to become an expert in the investment field.

The Source of Intelligence

Large language models can achieve human-like conversational intelligence because of extremely large training datasets and the complexity of the underlying models. Using GPT-4 as an example, Table 1 lists some details of the model size.

Table 1
SIZE AND COMPLEXITY OF LLMS

	Detail	GPT-4 ¹
Training Data	Large volume of text data available electronically, including content on the internet such as Wikipedia, and electronic books.	10 trillion words ² or 20 billion ³ pages
Model Structure	Deep learning models that have many layers of neurons to mimic complex relationships.	1.8 trillion model parameters
Model Training	Supervised learning and reinforcement learning from human feedback.	2.15 ²⁵ floating point operations. The same number of arithmetic operations will take a few million years using an ordinary computer.

Notes:

1. GPT-4 (Generative Pre-trained Transformer 4) is an advanced LLM developed by OpenAI who also developed the famous ChatGPT using GPT-3.5.

2. According to Patel and Wong (2023), 13 trillion tokens are used for training GPT-4. Assuming a token is 0.75 word on average, 10 trillion words were used.
3. Assuming each page contains 500 words, 20 billion pages were used for model training.

Compared to a human being, an LLM model may have similar complexity to a human brain in terms of the number of neurons. But it has the advantage of accessing and learning much more knowledge in a parallelized way. Through training, LLMs learned, generalized, and remembered lexical semantics, syntax, reasoning, and mathematics. The LLMs can be considered as intelligent beings who have learned and have the ability to learn quickly by themselves. But at this time point, we may still not have confidence to ask them to perform investment related tasks such as providing financial advice and preparing earnings reports.

The Development of Investment Expertise

With a good foundation for learning investment specific knowledge and skills, LLMs can be retrained to build investment expertise. Figure 1 lists some examples of potentially using LLMs in investment related tasks.

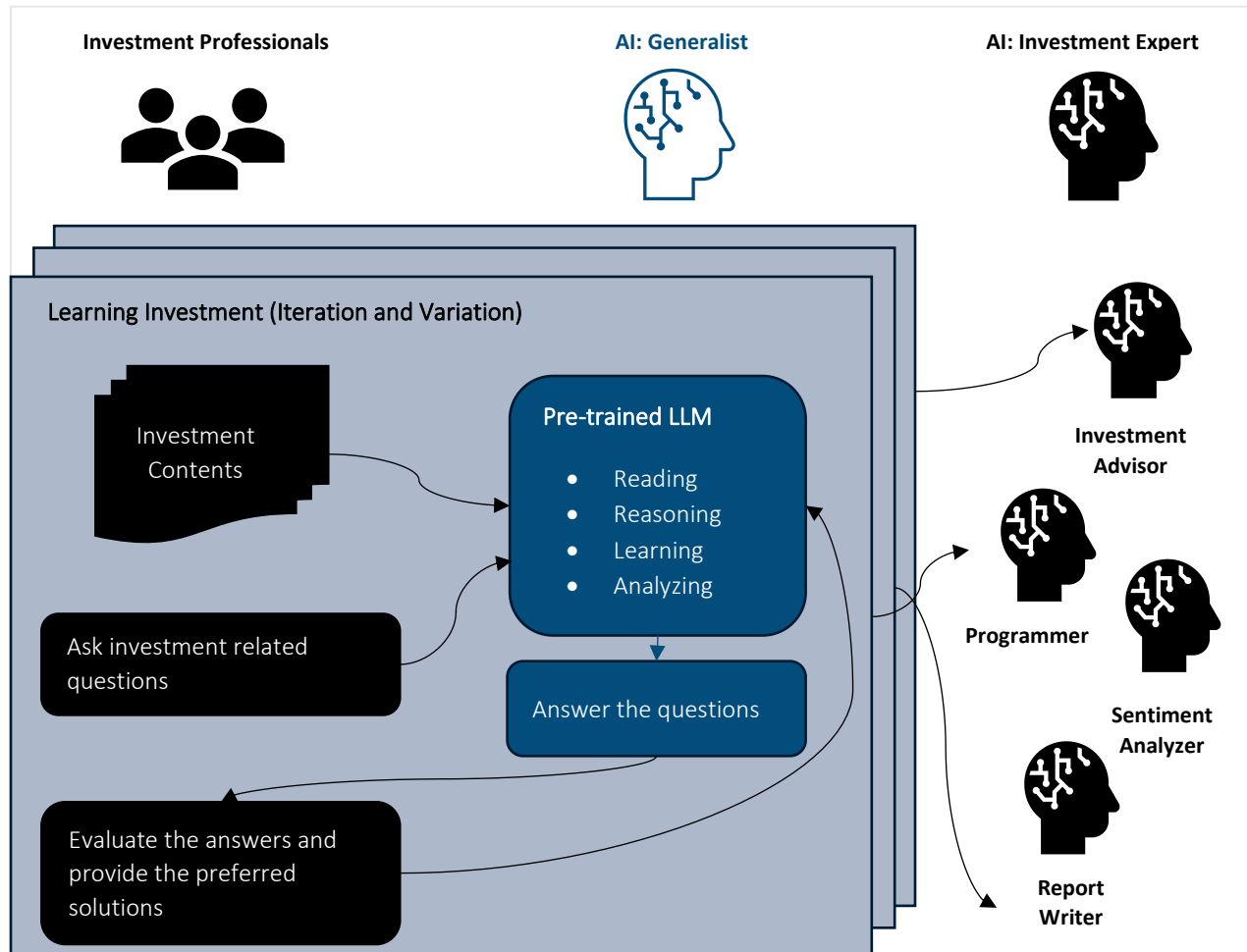
Figure 1

SAMPLE APPLICATIONS OF LLMS TO INVESTMENT

Content Analysis	<ul style="list-style-type: none"> • Summarizing investment literature and financial news • Analyzing social media data and perform sentiment analysis
Knowledge Searching	<ul style="list-style-type: none"> • Search answers to complex questions based on given documents • Evaluate environmental, social, and governance (ESG) investments
Content Generation	<ul style="list-style-type: none"> • Create visualization and presentation materials • Write a sustainability report
Programming	<ul style="list-style-type: none"> • Develop codes for various tasks such as investment analysis and user interface • Debugging computer programs
Custom Services	<ul style="list-style-type: none"> • Chatbots that advise on investment decision-making • Chatbots that can handle customer queries related to their investment.
Risk Management	<ul style="list-style-type: none"> • Assess the compliance of activities against laws, regulations and policies • Fraud detection

Like human beings, LLMs can benefit from specialization. The LLMs are suitable for transfer learning which means that new data can be fed into the pre-trained models to learn the knowledge and skills required for a specific task. This allows a generalist to evolve into different types of experts given targeted training. Figure 2 illustrates the retraining process in which the domain knowledge of investment professionals plays a key role.

Figure 2
SAMPLE PROCESS OF LLM RETRAINING FOR INVESTMENT



Starting from pretrained LLMs, investment professionals can help these generalists develop investment expertise in the following ways:

- Provide the most relevant and high-quality content for the generalist to study and gain knowledge and skills needed to perform specific tasks.
- Help train the model by asking relevant questions, evaluating the answers from the LLMs, and providing feedback to the models. It is also important to know that the generalist is smart. After a certain amount of supervised learning, it can learn by itself through reinforcement learning.
- Help certify if the retrained LLMs are qualified to perform investment related tasks.

Sometimes, the LLMs may have learned the knowledge already but do not know how to answer certain questions in a desired way. Part of the retraining is to adapt them to the problems usually encountered in the investment field. In addition to its importance in model training, asking the right questions can make a significant difference in the outputs generated by existing LLMs.

Asking the Right Questions

Like communicating with people from different cultures or using a different language, the way in which a question is asked matters much when using LLMs. We need to tell the models our objectives in a clear way with specific questions, provide the background information for more targeted answers, interact with the models by back-and-forth dialogue, and provide guidance during the conversation. The practice of designing inputs for LLMs is called prompt engineering. It is more open ended than computer programming, but some rules can be followed to make the best use of LLMs.

- Ask the model to follow certain rules such as breaking down a task into incremental steps when it is asked a question. For example, I want to decide how I should invest my money, bond and/or equity. It is for my retirement. I know it depends on my retirement goal, my age, and my risk appetite. Please ask me a sequence of questions to get the information. Fill in any missing information.
- Ask the model why it gives you this specific answer. For example, why do you suggest a 40/60 bond and equity allocation for my retirement fund?
- Ask the model to suggest a better way to ask the question and ask it again using the suggested way.
- Tell the model to ask you questions one by one to get all the information it needs to achieve a goal, such as designing an investment strategy.
- Tell the model to assume that you have a specific role, such as a child who knows little about investment, or an investment professional.
- Tell the model to answer your question in a specific format, such as completing certain fields in a table.
- Tell the model to provide alternative answers and compare the pros and cons of each answer.
- Tell the model to provide the facts that are contained in the answer.

Example: Asset Allocation

Below is part of a sample conversation on seeking advice to determine a simple asset allocation based on risk return trade-off. ChatGPT-3.5 is used to make the example more accessible.

The LLM already has some basic knowledge of asset allocation in terms of constructing efficient frontier, as shown in Figure 3.

Question: Do you know how to construct an efficient frontier for bond and equity investments?

Figure 3

SAMPLE CHATGPT ANSWER – 1ST EXAMPLE



ChatGPT

Constructing an efficient frontier for a portfolio of bond and equity investments involves determining the optimal mix of these assets to achieve the highest return for a given level of risk or the lowest risk for a desired level of return.

Here are the steps to construct an efficient frontier for bond and equity investments:

1. **Collect Data:** Gather historical returns for various bond and equity assets you're considering in your portfolio.
2. **Calculate Expected Returns and Risks:** Calculate the expected returns and standard deviations (or other measures of risk) for each asset based on historical data. The expected return represents the average return an investor can anticipate, while the standard deviation measures the volatility or risk of an asset's returns.
3. **Create Portfolios:** Construct different portfolios with varying allocations between bonds and equities. For example, you might create portfolios with 100% bonds, 100% equities, and various combinations in between (e.g., 80% bonds and 20% equities, 60% bonds and 40% equities, etc.).
4. **Calculate Portfolio Returns and Risks:** For each portfolio, calculate the expected return and risk. The expected return of a portfolio is the weighted sum of the expected returns of its constituent assets, and the risk is calculated based on the covariance (or correlation) between the assets.
5. **Plot the Efficient Frontier:** Plot these portfolios on a graph, with the x-axis representing the portfolio risk (standard deviation) and the y-axis representing the portfolio return. This graph will illustrate the trade-off between risk and return for different portfolio combinations.
6. **Identify the Optimal Portfolio:** The efficient frontier is the curve that represents portfolios that offer the maximum expected return for a given level of risk or the minimum risk for a desired level of return. The point on this curve that best fits an investor's risk tolerance and return objectives represents the optimal portfolio allocation.
7. **Rebalance and Monitor:** Portfolios need periodic rebalancing to maintain their desired asset allocation due to market fluctuations. Regularly monitor and adjust the portfolio to stay aligned with your investment goals.

Remember, constructing an efficient frontier involves assumptions based on historical data, and it's essential to consider changing market conditions and other factors that might affect future returns and risks.

Tools like mean-variance optimization and Modern Portfolio Theory (MPT) are often used in constructing efficient frontiers to determine the optimal asset allocation for a given level of risk tolerance.

The LLM can also split the task and ask for information one by one, as shown in Figure 4 and Figure 5.

Question: Can you guide me to create an efficient frontier by performing the task step by step? Ask me for inputs if needed.

Figure 4

SAMPLE CHATGPT ANSWER – 2ND EXAMPLE



ChatGPT

Absolutely! Let's walk through the process step by step. To start, we'll need some data on historical returns for bonds and equities you're considering for your portfolio. Do you have any specific historical return data or simulated return data for these assets? If so, please provide the return data or let me know the parameters for simulated returns.

Question/Input: Let's assume the bond investment has an expected return of 3%, a volatility of 6%. The equity investment has an expected return of 7%, a volatility of 20%. Their correlation is 0.2.

Figure 5

SAMPLE CHATGPT ANSWER – 3RD EXAMPLE



ChatGPT

Great, we have the expected returns, volatilities (standard deviations), and the correlation between the bond and equity investments.

Now, let's create a range of portfolios with different allocations between bonds and equities. We can start by considering portfolios with varying percentages allocated to bonds and the remainder to equities (e.g., 100% bonds, 90% bonds - 10% equities, 80% bonds - 20% equities, and so on).

We'll calculate the expected return and volatility of each portfolio using the given data and the relationship between these assets.

Let's create a table for different portfolio allocations:

Portfolio Allocation (Bonds : Equities)	Expected Return (%)	Volatility (%)
100% Bonds		
90% Bonds - 10% Equities		
80% Bonds - 20% Equities		
...		
10% Bonds - 90% Equities		
100% Equities		


Shall we proceed with calculating the expected returns and volatilities for each portfolio based on the given data?

The LLM is also smart enough to make adjustments to the details of each step, as shown in Figure 6.

Question: *Can we have a smaller increment in the portfolio allocation plans? Can we consider an increment of 5%?*

Figure 6

SAMPLE CHATGPT ANSWER – 4TH EXAMPLE

 **ChatGPT**

Absolutely! Let's adjust the allocations by 5% increments to provide an even finer spectrum of portfolio allocations.

Here's the revised table with 5% increments:

Portfolio Allocation (Bonds : Equities)	Expected Return (%)	Volatility (%)
100% Bonds		
95% Bonds - 5% Equities		
90% Bonds - 10% Equities		
85% Bonds - 15% Equities		
80% Bonds - 20% Equities		
75% Bonds - 25% Equities		
70% Bonds - 30% Equities		

The LLM can provide the details of the calculation logic, as shown in Figure 7.

Figure 7

SAMPLE CHATGPT ANSWER – 5TH EXAMPLE



ChatGPT

Great! To calculate the expected return and volatility for each portfolio allocation, we'll use the provided data:

- Bond expected return = 3%
- Bond volatility = 6%
- Equity expected return = 7%
- Equity volatility = 20%
- Correlation between bonds and equities = 0.2

Let's start calculating:

1. Expected Return of Portfolios:

For each portfolio allocation, we'll compute the weighted average of the expected returns of bonds and equities based on their allocations.

2. Volatility (Standard Deviation) of Portfolios:

The volatility of a portfolio is calculated by considering the standard deviation of the portfolio returns, considering the correlations and weights of the assets in the portfolio.

The formula to calculate the expected return of a portfolio is:

$$\text{Expected Return of Portfolio} = \text{Weight of Bonds} \times \text{Expected Return of Bonds} + \text{Weight of Equities} \times \text{Expected Return of Equities}$$

The formula to calculate the volatility of a portfolio is:

$$\text{Volatility of Portfolio} = \sqrt{(\text{Weight of Bonds} \times \text{Volatility of Bonds})^2 + (\text{Weight of Equities} \times \text{Volatility of Equities})^2 + 2 \times (\text{Weight of Bonds} \times \text{Weight of Equities} \times \text{Volatility of Bonds} \times \text{Volatility of Equities} \times \text{Correlation})}$$

Shall we start with the calculations for each portfolio allocation from 100% Bonds to 100% Equities, considering 5% increments?

When something is found wrong, we can give guidance to the LLM to check and rectify, with an example shown in Figure 8 and Figure 9.

Question: Can you show me the volatility calculation of 5% Bonds - 95% Equities?

Figure 8

SAMPLE CHATGPT ANSWER – 6TH EXAMPLE



ChatGPT

Of course! Let's calculate the volatility for the portfolio allocation of 5% Bonds - 95% Equities using the formula for portfolio volatility:

Volatility of Portfolio =

$$\sqrt{(\text{Weight of Bonds} \times \text{Volatility of Bonds})^2 + (\text{Weight of Equities} \times \text{Volatility of Equities})^2 + 2 \times \text{Weight of Bonds} \times \text{Weight of Equities} \times \text{Volatility of Bonds} \times \text{Volatility of Equities} \times \text{Correlation}}$$

Given:

- Weight of Bonds = 5%
- Weight of Equities = 95%
- Volatility of Bonds = 6%
- Volatility of Equities = 20%
- Correlation between bonds and equities = 0.2

Let's compute the volatility for the 5% Bonds - 95% Equities portfolio:

Volatility of Portfolio =

$$\sqrt{(0.05 \times 0.06)^2 + (0.95 \times 0.20)^2 + 2 \times 0.05 \times 0.95 \times 0.2 \times 0.06 \times 0.20}$$

$$\text{Volatility of Portfolio} = \sqrt{(0.003)^2 + (0.171)^2 + 0.000456}$$

$$\text{Volatility of Portfolio} = \sqrt{0.000009 + 0.029241 + 0.000456}$$

$$\text{Volatility of Portfolio} = \sqrt{0.029706}$$


$$\text{Volatility of Portfolio} \approx 0.1723 = 17.23\%$$

Therefore, the calculated volatility for the 5% Bonds - 95% Equities portfolio is approximately 17.23%.

Question: It seems that $0.95 * 0.20 = 0.19$, not 0.171, correct?

Figure 9

SAMPLE CHATGPT ANSWER – 7TH EXAMPLE

 **ChatGPT**

You're absolutely correct. Let's rectify that calculation:

$$\text{Volatility of Portfolio} = \sqrt{(0.05 \times 0.06)^2 + (0.95 \times 0.20)^2 + 2 \times 0.05 \times 0.95 \times 0.2 \times 0.06 \times 0.20}$$

$$\text{Volatility of Portfolio} = \sqrt{(0.003)^2 + (0.19)^2 + 0.000456}$$

$$\text{Volatility of Portfolio} = \sqrt{0.000009 + 0.0361 + 0.000456}$$

$$\text{Volatility of Portfolio} = \sqrt{0.036565}$$

$$\text{Volatility of Portfolio} \approx 0.1913 = 19.13\%$$

Therefore, the corrected volatility for the 5% Bonds - 95% Equities portfolio is approximately 19.13%. Thank you for your patience!

The example illustrates that the LLM has basic knowledge of investment, can correct errors when given instructions, and is smart enough to understand different requests and relationships of questions in a conversation. While the error shown in this example is not present in a more advanced LLM such as GPT-4 with code interpreters, it shows that domain knowledge and retraining is the key to bringing LLMs into practical applications in the investment field.

Managing the Risks for Healthy Usage

Like any model and human being, LLMs can make errors. It is important for us to understand the risks and the mitigation measures that we can use today. In this way, we will not overuse them when they are not ready.

Table 2

RISKS OF USING LLMs

Risk	Detail	Mitigation
Model Risk	Due to the model complexity, without proper knowledge and training, the model may be perceived as a black box without the needed transparency to meet model risk management standards. In addition, the LLMs may generate inappropriate answers for certain areas due to the lack of knowledge and/or training. They are also sensitive to how the questions are asked. With the models evolving with new training through conversations, it is difficult to reproduce the results.	<ul style="list-style-type: none"> • Develop technical expertise needed for using LLMs appropriately. • Ensure that high-quality data is used in the retraining process. This may indicate rigorous data collection and data processing practices. • Perform systematic validation tests to identify weak areas of the models. • Relying on multiple models rather than one single model to evaluate and select the answers. • Detailed documentation for auditing and also providing valuable information to improve the models. • Clear communication of the model risk with stakeholders to establish a reasonable

		expectation, while recognizing the models' full potentials.
Data Bias	Similar to the issue of using imbalanced data in classification model training, LLMs may be trained with biased data and generate unfair outputs.	<ul style="list-style-type: none"> Analyze model outputs regularly to identify any potential biases. Use balanced and diverse training data through simulating or collecting more underweighted data.
Cyber Security and Data Privacy	LLMs are usually cloud-based given their size. Model training and applications require a significant amount of data transfer which can contain sensitive information for investment purposes.	<ul style="list-style-type: none"> Similar to the measures used for other cloud-based services, robust data protection processes including user controls, encryption, and cyber security audits need to be put in place. Sensitive data can be removed or transformed without jeopardizing the model performance. Private data is used only if it is allowed and in a way that complies with regulations.
Ethics	LLMs do not explicitly follow ethical principles which may lead to giving false information, not providing references, impersonating individuals, and causing harm to users and the public.	<ul style="list-style-type: none"> Establish an ethical framework of using LLMs with internal and external stakeholders and a culture of responsibility and accountability. Implement governance and oversight on ethical aspects of all LLM applications. Assess the ethical issues of LLMs regularly and adapt to reduce their impacts.

Many of the risks may be addressed with further improvements of the LLMs but new risks are likely to arise with more applications of the LLMs. For all AI technologies, risk management is an evolving and long-lasting task.

Conclusion

Given the remarkable advancements in large language models (LLMs), a landscape of promising technological opportunities emerges, presenting solutions with minimal instructions. Despite being in its nascent stage, LLMs hold immense potential to revolutionize productivity as adept advisers. Their success stories—writing books in days, passing exams, and developing programs—underscore their capabilities. However, bridging gaps for effective utilization in investment demands additional efforts. Understanding the origins of their intelligence, derived from large datasets and complex models like GPT-4, mirrors human brain complexity but with unprecedented knowledge access. Yet, entrusting them with financial advice or report preparation remains a leap of faith. To evolve LLMs into investment experts requires retraining, guided by investment professionals who can provide domain-specific expertise. This specialized learning hinges on tailored inputs, where the art of asking the right questions becomes pivotal. Prompt engineering requires clear objectives, contextual details, iterative dialogues, and precise guidance for optimal outcomes. At the same time, managing risks inherent in LLMs, from model complexities to biases, cybersecurity, and ethical implications, necessitates stringent measures and ongoing vigilance. Risk management is critical for the long-term success of LLMs.

References

Patel, Dylan and Gerald Wong, "GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE. Demystifying GPT-4: The engineering tradeoffs that led OpenAI to their architecture." July 2023.

<https://www.semianalysis.com/p/gpt-4-architecture-infrastructure>

Kailan Shang, FSA, CFA, PRM, SCJP, is a Director at Aon PathWise. He can be reached at klshang81@gmail.com.