

Are LLMs the future of API programming?

Vamshi Jandhyala

*The transformatory role of Application Programming Interfaces(APIs) in software development and accelerating new product/service development is well understood. With the advent of Large Language Models(LLMs) like GPT-4, the software development process is undergoing yet another radical shift. The sophisticated code generation capabilities of LLMs are increasing developer productivity and the natural language interface (**aka prompt**) of the LLMs is driving the democratization of the creation of new wave of digital products/services. In this opinion piece, we discuss the implications of domain specific(Financial Services) LLMs, training LLMs to use “external tools/APIs” on developer productivity and enabling rapid experimentation by non-developer personas and conclude with concrete recommendations for API Product Managers.*

1. Ubiquity of APIs

The dynamic nature of the digital landscape in Financial Services necessitates businesses to rapidly innovate and introduce new products and services. One significant factor enabling this accelerated development is the efficient use of APIs. APIs are the backbone of software development, providing a set of protocols and tools for building software applications. They define methods and data formats that a program can utilize to perform specific tasks, interact with other software, and facilitate interoperability.

1.1. Rapid Product/Service Development

The adoption of APIs can expedite the creation of new products and services in several ways. First, APIs allow businesses to leverage existing technologies and services rather than developing solutions from scratch. This reuse of code not only saves time and resources but also reduces the possibility of bugs and errors. APIs can also extend the functionality of products and services, enabling companies to innovate and offer value-added features to their customers.

1.2. Increased Developer Productivity

A compelling advantage of APIs is their significant impact on developer productivity. APIs provide pre-defined, reusable, and modular code pieces that developers can utilize in their applications. This leads to less time spent writing code and more time focusing on core product features. Additionally, well-documented APIs have comprehensive guidelines and examples, further aiding developers in their tasks.

1.3. Faster Experimentation

APIs foster an environment conducive to faster and more effective experimentation. They allow developers to easily plug in and try out different functionalities, enabling a more iterative and agile development approach. This agility is key in the current digital landscape, where the ability to quickly adapt and evolve is essential for survival and success.

2. Advent of LLMs

LLMs, such as OpenAI's GPT-4, are increasingly shaping the trajectory of various industries, with financial services being a notable example. LLMs' capability to generate human-like text, understand nuances in languages, and generate contextual responses are integral to NLP, which has significant applications in finance. LLMs, equipped with advanced NLP capabilities, can understand, interpret, and generate human language in a meaningful and contextually appropriate manner. They are trained on diverse and extensive textual data, enabling them to handle complex language tasks with unprecedented accuracy. As a result, they can facilitate real-time translation, content generation, sentiment analysis, and more.

2.1. Applications in Financial Services

The financial services sector stands to gain substantially from the integration of LLMs. Below are some concrete examples:

- **Sentiment Analysis:** Sentiment analysis is a powerful tool that helps businesses understand public opinion towards their brand, products, or services. In financial services, LLMs can be employed to analyze

investor sentiment from social media, forums, news articles, and more. For example, negative sentiment detected in real-time can signal potential issues with a company's financial health or a change in market dynamics.

- **News Classification:** The financial industry is heavily influenced by news and events happening globally. LLMs can help classify these news events based on their relevance to different sectors, companies, or financial instruments. For instance, news articles can be categorized as relevant to macroeconomic trends, specific industries, or individual corporations. This helps analysts to stay abreast of significant developments impacting their areas of interest.
- **Named Entity Recognition (NER):** NER is a sub-task of information extraction that seeks to locate and classify named entities in text into predefined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc. In the financial sector, LLMs can identify and classify such entities from complex financial documents or news articles, helping to streamline the extraction of key information.
- **Question Answering:** LLMs can be employed in developing sophisticated question-answering systems for the financial services industry. These systems can provide clients with immediate, accurate responses to queries regarding their accounts, transactions, or financial products. Furthermore, they can assist financial advisors by providing instant information on a vast array of financial topics, thereby enhancing their efficiency.
- **Customer Service:** LLMs can power chatbots to provide customer service around the clock. They can handle a wide range of queries, from account balance inquiries to more complex requests like explaining various investment products. Their capacity for understanding and generating natural language responses can significantly enhance the customer experience.
- **Risk Management:** LLMs can analyze textual data, such as news articles or financial reports, to identify potential risks. For instance, they can detect negative sentiment in news articles about a company, suggesting possible financial risks.
- **Financial Analysis:** LLMs can assist in analyzing financial reports, earnings calls transcripts, and more. They can identify key trends, highlight critical data points, and even provide summaries of lengthy

reports. This can help financial analysts in decision-making and forecasting.

2.2. Financial Domain LLMs

General models cover many domains, are able to perform at a high level across a wide variety of tasks and obviate the need for specialization during training time. However, results from existing domain-specific models show that general models cannot replace them. In any financial services company there are large and diverse set of tasks which can be well served by a general model but given that the majority of the applications are within the financial domain, a domain specific model might be more effective. For this reason, Bloomberg has developed **BLOOMBERGGPT [1]** which is a model that achieves best-in-class results on financial benchmarks, while also maintaining competitive performance on general-purpose LLM benchmarks. The key to the model performance is training data which is a combination of general data sources as well as high quality curated domain specific datasets.

3. The New Challenge

With the plethora of APIs and the powerful new capabilities afforded by general models and domain specific LLMs, there is now a new challenge facing the developer - **“How do I combine all these capabilities to create a compelling product/solve a business problem?”**. Unsurprisingly the approaches to deal with this challenge involve eliciting new capabilities from LLMs as LLMs today are fundamentally limited by the information they can store in the fixed set of weights and the things they can compute using a static computation graph and limited context. Furthermore, as the world changes, LLMs require retraining to update their knowledge and reasoning capabilities. By empowering LLMs to use tools, we can grant access to vastly larger and changing knowledge bases and accomplish complex computational tasks. By providing access to search technologies and databases, we can augment LLMs to address a significantly larger and more dynamic knowledge space. Similarly, by providing access to computational tools, LLMs can accomplish complex computational tasks. This transition from a small set of hand-coded tools, to the ability to invoke a vast space of changing cloud APIs could transform LLMs into the primary interface to computing infrastructure and the web. Application developers in the future might interact with an API provider/store using **prompts** e.g. **Please give me a set of APIs to summarize the content of a financial news article**

and convert the summary to speech. In the rest of this section we discuss a couple of innovative new promising tools/approaches to deal with the challenge.

3.1. TaskMatrix.AI

TaskMatrix.AI [2] is an AI ecosystem that can connect foundation models with millions of APIs for task completion. Unlike most previous work that aimed to improve a single AI model, it focuses more on using existing foundation models (as a brain-like central system) and APIs of other AI models and systems (as sub-task solvers) to achieve diversified tasks in both digital and physical domains. Different from any single AI model, TaskMatrix.AI can be seen as a super-AI with abilities to execute both digital and physical tasks, which has the following key advantages:

- TaskMatrix.AI can perform both digital and physical tasks by using the foundation model as a core system to understand different types of inputs (such as text, image, video, audio, and code) first and then generate codes that can call APIs for task completion.
- TaskMatrix.AI has an API platform as a repository of various task experts. **All the APIs on this platform have a consistent documentation format that makes them easy for the foundation model to use and for developers to add new ones.**
- TaskMatrix.AI has a powerful lifelong learning ability, as it can expand its skills to deal with new tasks by adding new APIs with specific functions to the API platform.
- TaskMatrix.AI has better interpretability for its responses, as both the task-solving logic (i.e., action codes) and the outcomes of the APIs are understandable.

3.2. Gorilla

Gorilla [3] is a finetuned LLaMA-7B-based model that surpasses the performance of GPT-4 on writing API calls. It also substantially mitigates the issue of hallucination, commonly encountered when prompting LLMs directly. Furthermore, it is able to adapt to changes in the API documentation, understand and reason about constraints. It is trained on a large corpus of APIs from three major model hubs for dataset construction: TorchHub, TensorHub and HuggingFace. Here is an example of a prompt that Gorilla can deal with - **Help me find an API to convert the spoken language in a recorded audio to text using Torch Hub.**

4. Conclusion

The API ecosystem is going through a fundamental transformation with the advent of LLMs. Companies which can offer APIs that leverage powerful new capabilities of domain-specific LLMs (e.g. BLOOMBERGPT using proprietary data) and enable developers and non-developers alike to effectively use their APIs in conjunction with other popular models/APIs (e.g. HuggingFace) via **prompts** will have a significant competitive advantage. API Product Managers will now have to meet the requirements for a new set of customers (the LLMs). Consistent API design and documentation will be more important than ever for search, discovery and use through LLMs. API documentation will be the new in-demand training data as APIs continue to be increasingly consumed by machines. As LLMs become increasingly well-versed in using APIs, Product Managers need to become adept at “prompt engineering” to be able to quickly generate and test new ideas. Domain-specific API/model aggregators in the future will compete on the quality of the LLM which when given a prompt, is able to automatically identify the right set of domain-specific models and APIs, assemble and orchestrate the APIs and evaluate the results to identify optimal combinations of models/APIs that best meet the requirements of the prompt. Long live the prompt.

References

- [1] S. Wu, O. Irsoy, et al., “Bloomberggpt: a large language model for finance,” 2023.
- [2] Y. Liang, C. Wu, et al., “Taskmatrix.ai: completing tasks by connecting foundation models with millions of apis,” 2023.
- [3] S. G. Patil, T. Zhang, X. Wang, and J. E. Gonzalez, “Gorilla: large language model connected with massive apis,” *Arxiv Preprint Arxiv:2305.15334*, 2023.