

White paper – April 2023

# Can ChatGPT Help You Trade?

Part II

Misha Malyshev, Alexey Goldin



## Introduction

Since our first article on transformers was published, ChatGPT (in its GPT3.5 and later GPT4 incarnation) created a shockwave-like impact on the AI community and the rest of the world. We have seen reasonable people lose their mind and proclaim the immediate coming of singularity, similar to when peripatetic preachers proclaimed apocalypse in times past.

Indeed, the technology demonstrated by OpenAI is truly amazing. While we do not know all the details of the architecture, it's likely very simple, though the sheer scale of implementation required some brilliant technical solutions.

Despite the apparent simplicity, numerous tasks can optimize a trading workflow, and examining these optimizations proves to be intriguing. In this paper, we will explore the most prevalent financial use cases where ChatGPT can provide assistance, and also assess the distance remaining to achieve genuine singularity.

## Transformers - The Basics and GPTs

There are quite a lot of introductory texts explaining how transformers work, so we will dedicate only a couple of paragraphs to the subject. To those who are new to the field, we would recommend [The Annotated Transformer](#), and [GPT2 paper](#).

The main idea behind transformer architecture is called "attention." It was initially created to improve how recurrent neural networks work. However, in an important paper called "[Attention is all you need](#)", it was demonstrated that it is possible to eliminate recurrence entirely and rely solely on attention mechanisms.

*(Reference: In Natural Language Processing (NLP), attention refers to a technique used in neural network models to selectively focus on relevant parts of the input data. It mimics the way humans pay attention to specific aspects of a given context or situation. By weighting different parts of the input sequence based on their importance, attention mechanisms enable models to better capture long-range dependencies, context, and relationships between words or phrases, ultimately improving their performance in various NLP tasks such as machine translation, sentiment analysis, and question-answering systems.)*

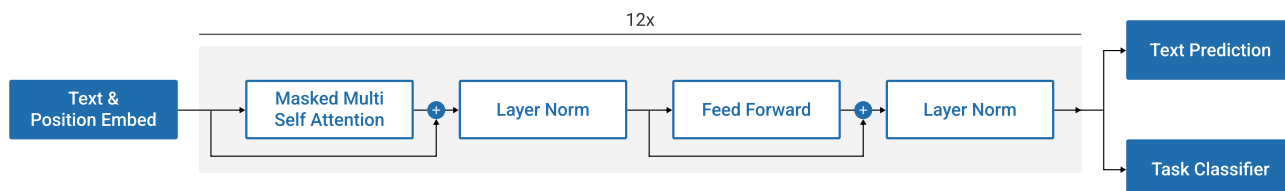
Due to their architecture, transformers can process very diverse classes of data. Inputs always consist of a set of tokens. In Natural Language Processing (NLP), tokens are multidimensional vectors representing words or subwords adjusted to match the transformer's dimensions -  $d$ . For images, tokens represent image patches, and for tabular data, each feature is converted into the same space.

Each token then goes through a nonlinear transformation and is projected into query ( $q$ ), key ( $k$ ), and value ( $v$ ) spaces using matrices  $Q$ ,  $K$ , and  $V$ . To identify the most relevant tokens for predicting at position  $i$ , query  $q_i$  is multiplied by keys  $k_j$ , and a softmax function is applied along the dimension  $j$  to calculate weights  $w_{\{ij\}}$ , such that the sum of all  $w$  along  $j$  is 1. Then, values  $v_j$  are combined with weights  $w_{\{ij\}}$  and processed through another nonlinearity, normalization, and projection into  $q$ ,  $k$ , and  $v$  spaces.

Since all tokens use the same projection matrices and nonlinear functions, their positions can be switched. To ensure the transformer understands the relative positions of tokens (words, image patches, or importance of features in tabular data), position embeddings are added to the tokens after projecting them into dimension  $d$ . In NLP, this can be achieved by multiplying sines and cosines of the position by different dimensions. For images, both horizontal and vertical positions are considered, while for tabular data, separate embedding vectors are used for each variable.

This architecture naturally manages irregular positions (for irregular time series), missing data (for partial images fed into the network), missing features in tabular data (when only available features are input), or mixed modality tokens (all fed into the same transformer, with only the embedding indicating the token's origin).

Moving on to what is inside GPT. The GPT-2 architecture is more straightforward than the one proposed in the original transformers paper, featuring a 12-layer decoder-only network. GPT-3, on the other hand, comprises 96 layers. As for GPT-4, its number of layers can only be speculated upon.



Picture 1: A simplified representation of GPT2 transformer architecture. GPT3 is similar, with 96 layers instead of 12. It is widely believed that GPT4 has similar architecture.

Despite the remarkable computational prowess of these models, their primary objective remains simple: predicting the subsequent token (word or subword) in a text. This is achieved by training these models on an astonishing volume of textual data available on the internet.

The model outputs  $N$  numbers (logits), which correspond to the number of tokens in vocabulary, and the greatest number corresponds to the predicted token. Thus, ChatGPT, and similar models like Alpaca, function primarily as text continuation machines with the added refinement of reinforcement learning to produce more useful responses.

So if all that is achieved by ChatGPT is an improved statistical model of the text - why are so many minds predicting we will witness singularity through the advanced iterations of these models? Why is it more interesting than [SClgen](#) which was built strictly for amusement?

## The power of LARGE and signs of human intelligence?

While NLP models existed in the past and were capable of predicting the next word in a randomly generated sentence, what is truly different about ChatGPT? And what is contributing to this global hype?

In practice, models like GPT3, GPT4, LLaMa (and Alpaca), PaLM, Chinchilla, Claude, etc. are **qualitatively** better than the previous generations and are called LLM ([Large Language Models](#)). It turns out that the sheer size of such models, which have been exposed to extremely different examples of texts, allowed new properties to emerge, which was not possible in smaller linguistic models.

That unpredictable phenomenon is referred to in a paper as “emergent abilities of large language models,” which explains, “*The ability to be emergent if it is not present in smaller models but is present in larger models. Thus, emergent abilities cannot be predicted simply by extrapolating the performance of smaller models. The existence of such emergence raises the question of whether additional scaling could potentially further expand the range of capabilities of language models.*”

The most surprising “emergent ability” of big models is that they can be used for various tasks not even imagined during training! It is this unique ability that makes users believe they are interacting with true intelligence.

Let's consider several examples relevant to the financial industry.

### Example 1. Sentiment.

One of the most common tasks for any trading desk – discretionary or quantitative – is defining the market sentiment based on quarterly corporate filings and reports. No one trained ChatGPT to identify the sentiment of quarterly report data of a stock, and yet, it succeeded in doing so.

At ChatGPT prompt type:

*"Is the tone of the following letter positive, negative, or neutral?"*

*(you can also instruct it to answer in a single word to be friendly for machine usable format) (feed here [https://s201.q4cdn.com/589201576/files/doc\\_financials/2022/q4/2022-Q4-SIVB-CEO-Letter.Final.pdf](https://s201.q4cdn.com/589201576/files/doc_financials/2022/q4/2022-Q4-SIVB-CEO-Letter.Final.pdf))*

ChatGPT answers:

*"The tone of the letter is generally positive. The letter highlights SVB Financial Group's achievements, growth, and strengths, while also addressing challenges and uncertainties in the market. It emphasizes the company's resilience, strategic focus, and confidence in its long-term strategy."*

### Example 2. Text summation.

One of the other common tasks is providing short summaries and market analytics. No one trained ChatGPT to summarize text. Yet, it does it. With the above letter the result is:

*"In the SVB Financial Group Q4 2022 CEO Letter, the company reports strong financial performance with solid earnings, net income, and return on equity. Despite market pressures and slower VC investment, the underlying business remains strong with robust loan growth, record core fees, and healthy investment banking activity. The letter highlights the company's strong balance sheet and expertise to successfully serve clients and manage the current environment. It also mentions a focus on strategic priorities and the long-term growth opportunity. The outlook for 2023 is considered a transitional year with challenges in the market, but the company remains confident in its strategy and commitment to the innovation economy. Several key leadership appointments and hires have been made to strengthen the organization and support growth."*

### Example 3. Data extraction.

Data extraction – especially if reported in a non-normalized format – is yet another time-consuming task if you are building a quantitative model. No one trained the model to extract information in machine readable form. And yet it does with a simple prompt.

*Please read the text after ### and extract relevant data into JSON according to the example:*

```
{'QuarterlyRevenue': 1e9,  
'EarningPerShare': 1.0,}
```

*Use scientific notation for large numbers*

*###*

*CUPERTINO, CALIFORNIA OCTOBER 27, 2022 Apple today announced financial results for its fiscal 2022 fourth quarter ended September 24, 2022. The Company posted a September quarter record revenue of \$90.1 billion, up 8 percent year over year, and quarterly earnings per diluted share of \$1.29, up 4 percent year over year. Annual revenue was \$394.3 billion, up 8 percent year over year, and annual earnings per diluted share were \$6.11, up 9 percent year over year.*

ChatGPT answers:

```
{  
"QuarterlyRevenue": 9.01e10,  
"EarningsPerShare": 1.29  
}
```

### Example 4. Being your personal developer.

**Problem:** Find the sum of primes larger than 10, smaller than 100

Use this format to find the solution:

*Problem: \${Problem}*

*IPython session:*

...

*\${IPython function and commands needed to find solution}*

...

*Answer: \${Answer}*

*The ChatGPT promptly answers:*

***Problem:*** Find the sum of primes larger than 10 and smaller than 100.

IPython session:

scss

Copy code

```
import sympy prime_sum = sum(sympy.primerange(11, 100))
```

```
print(prime_sum)
```

***Answer:*** The sum of primes larger than 10 and smaller than 100 is 1060.

Giving it a proper prompt to steer into the direction of writing a code is crucial. The code is absolutely correct, but the answer (1060) is not. Running the code you can easily get the correct answer: 1043.

This last example shows both power and limitations. The model can readily point us to a correct solution, tell us about libraries one might not have heard about, but will readily and confidently hallucinate a wrong answer, approximately guessing an amplitude. Interestingly, this behavior closely resembles that of humans.

## No Singularity Yet

ChatGPT is undoubtedly a very powerful tool if used correctly. It is evident that such language models can accomplish far more than basic sentiment analysis, proving valuable in extracting structured information from text, and providing insightful research suggestions. The mass adoption of such tools can spark the "AI revolution" akin to how the internet changed our lives. However, it is crucial to remember that, fundamentally, ChatGPT and similar models are still statistical text models designed to continue given prompts, similar to a "statistical parrot." Start a prompt with professional jargon, and you will likely receive a response related to professional literature. Begin with juvenile language, and you might get a children's tale. Use politically biased language, and the model will mirror the bias. The model is simply connecting existing information within its vast memory – which, by the way, ends in September 2021.

In conclusion, we do not perceive these language models as an immediate AI threat exhibiting "sparks of consciousness." However, they do represent powerful tools. While they could potentially draft a "Trading for Dummies" guide, they are unlikely to replace human traders (just ask if you should buy Credit Suisse stock). Yet, in systematic trading processes, these robust models can supplement or enhance various steps leading to investment decisions.

We are (for now) safe from singularity.



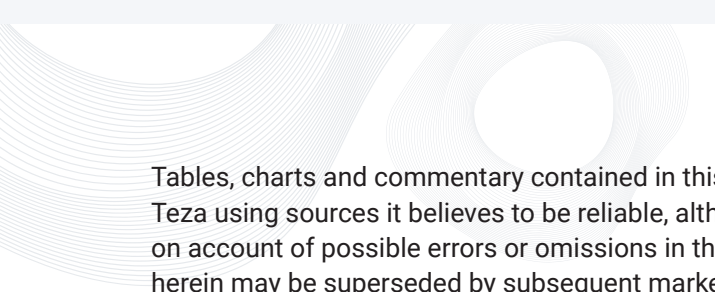
## Disclosures

This document is provided solely for informational and educational purposes, and there is no consideration given to the specific investment needs, objectives, or tolerances of any recipient. This document is not investment research and should not be treated as such, nor does it represent a formal or official view of Teza. Additionally, Teza's investment positions may, and often will, vary from its conclusions discussed herein based on any number of factors, including client investment guidelines and restrictions. No representation is given that any statements made in this document are accurate or that Teza's objectives will be achieved. This document contains Teza's opinions, and such opinions are subject to change without notice.

This document does not constitute an offer to sell or the solicitation of an offer to purchase any security or investment product (each, a "Product") and should not be relied on in making any investment decision. Any such solicitation or offering may only be made by means of delivery of an approved offering document and relevant subscription documents, all of which must be read in their entirety. No offer to purchase shares in a Product will be made or accepted prior to receipt by the offeree of such documents and the completion of all appropriate documentation. No offer to sell (or solicitation of an offer to buy) will be made in any jurisdiction in which such offer or solicitation would be unlawful.

It should not be assumed that investments described herein will be profitable. Nothing described herein is intended to imply that an investment with Teza is safe, conservative, risk free or risk averse. An investment with Teza entails substantial risks, and a prospective investor should carefully consider the summary of risk factors included in Teza's Form ADV Brochure (and the relevant offering document) in determining whether an investment with Teza is suitable. The risk of loss in trading futures is substantial. This document does not consider the specific investment objective, financial situation or particular needs of any investor and an investment with Teza is not suitable for all investors. Prospective investors should not rely upon this document for tax, accounting or legal advice. Prospective investors should consult their own tax, legal, accounting or other advisors about the issues discussed herein. Historic market trends are not reliable indicators of actual future market behavior or future performance of any particular investment which may differ materially, and should not be relied upon as such. Investors are also reminded that past performance should not be seen as indication of future performance and that they may lose the entirety of their investment. No recommendation is made positive or otherwise regarding individual securities, futures, strategies or other investment products mentioned herein. Information provided about positions, if any, and attributable performance is intended to provide a balanced commentary, with examples of both profitable and loss-making positions; however, this cannot be guaranteed. Certain data and analyses contained herein are based on theoretical and/or backtested model portfolios and are not representative of the performance of accounts that Teza currently manages. The information provided herein is not intended to provide a sufficient basis on which to make an investment decision, and investment decisions should not be based on simulated, hypothetical or illustrative information that have inherent limitations. Unlike an actual performance record, simulated or hypothetical results do not represent actual trading or the actual costs of management and may have under or over compensated for the impact of certain market risk factors. Teza makes no representation that any account will or is likely to achieve returns similar to those shown. Gross performance results do not reflect the deduction of investment advisory fees, which would reduce an investor's actual return. There can be no assurance that any Product advised by Teza will implement the strategies or trading signals referred to herein, or that if implemented any such strategies or signals achieve their investment objectives.

Certain information contained in this document constitutes "forward-looking statements," which can be identified by use of forward-looking terminology such as "may," "will," "should," "expect," "anticipate," "project," "target," "estimate," "intend," "continue," or "believe" or the negatives thereof or other variations thereon or other comparable terminology. Such statements are based on the current expectations and certain assumptions of Teza, and are, therefore, subject to certain risks and uncertainties. A variety of factors, many of which are beyond Teza's control, affect the operations, performance, business strategy and results of the accounts that Teza manages and could cause the actual results, performance or achievements of such accounts to be materially different from any future results, performance or achievements that may be expressed or implied by such forward-looking statements or anticipated on the basis of historical trends.



Tables, charts and commentary contained in this document have been prepared on a best efforts basis by Teza using sources it believes to be reliable, although it does not guarantee the accuracy of the information on account of possible errors or omissions in the constituent data or calculations. Further, the information herein may be superseded by subsequent market events or for other reasons. Teza does not assume any duty to, nor does it undertake to, update the information herein. Charts and graphs provided herein are for illustrative purposes only. No part of this document may be divulged to any other person, distributed, resold and/or reproduced without the prior written permission of Teza.

\* \* \*

Teza® is a registered trademark of Teza Technologies LLC.