



The Makridakis Open Forecasting Center (MOFC)

Advancing the Theory and Practice of Forecasting

THE M COMPETITIONS

THE M CONFERENCES

THE M PUBLICATIONS

THE M COURSES



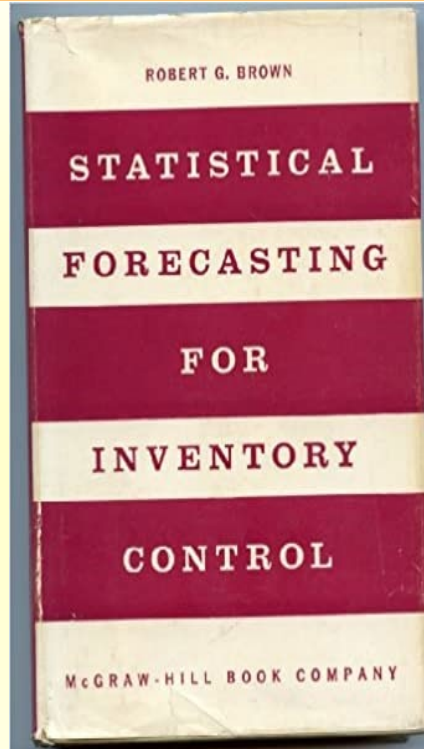
UNIVERSITY *of* NICOSIA

The Emergence of AI and the Changing Role of Forecaster: From selecting the most accurate model to enhancing Forecasting Value Added (FVA)

The four milestones in business forecasting

1959: Statistical

Bob Brown establishes the business forecasting field



Between 1959 and 2017 simple statistical methods provided the most accurate forecasts. The challenge to select the most accurate method

2018: Hybrid

Slawek Smyl, a data scientist full time forecaster won the M4 competition using a hybrid method developed by himself:



Smyl method ended the long forecasting winter by using global models and cross validation to test them for their accuracy and uncertainty

2020: ML

YeonJun In, a senior undergraduate student, won the M5 competition using a standard ML method, beating his 7,092 Kaggle competitors



The Light Gradient Boosting Machine (LGBM) was not only the most accurate method but also computationally highly efficient

Nov. 2022 : ChatGPT



Using or better interacting with ChatGPT in natural language is effortless, requiring no previous training. It is like having an assistant helping you with tedious tasks (like summarizing text or writing an email) and, at the same time, a colleague who can provide you with whatever information you want, propose intelligent solutions and come up with creative ideas challenging your thinking.

Forecasting value added

(ChatGPT answer)

Forecasting Value Added (FVA) is a concept that goes beyond merely looking at the accuracy of forecasts. It evaluates the contribution of each step in the forecasting process to determine if it adds value or not. The primary objective of FVA analysis is to identify and eliminate unnecessary or redundant forecasting activities that do not improve (or may even worsen) forecast accuracy.

Here's how you can leverage ChatGPT to enhance Forecasting Value Added (FVA)

- **Analyze Current Processes:**
 - Describe your current forecasting processes, tools, and methods.
 - Ask ChatGPT for insights into whether such techniques typically add value or not based on literature or best practices.
- **Simplify:**
 - If a forecasting step doesn't add value (i.e., doesn't improve the accuracy or usefulness of the forecast), consider removing it. ChatGPT can provide insights into the potential implications of removing certain steps.
- **Benchmarking:**
 - Compare the accuracy of a simplified model with a more complex one. ChatGPT can provide guidance on statistical measures to use for comparison.
- **Continuous Monitoring:**
 - Regularly review the value added by each forecasting step. Use ChatGPT to help analyse and interpret results.
- **Stakeholder Communication:**
 - ChatGPT can help draft communications or provide explanations that can be used to justify changes in the forecasting process to stakeholders.
- **Knowledge Update:**
 - Regularly inquire about best practices in FVA and forecasting to stay updated. ChatGPT can provide summaries or explanations of recent advancements or studies related to FVA.
- **Scenario Analysis:**
 - Use ChatGPT to explore the implications of various forecasting scenarios. By understanding how different inputs or changes affect the forecast, you can better judge their value-added.
- **Training and Education:**
 - Use the model to help explain the FVA concept to team members or stakeholders. An educated team will be better equipped to identify and eliminate non-value-adding activities.

Here's how you can leverage ChatGPT to enhance Forecasting Value Added (FVA)

➤ Analyze Current Processes:

- Describe your current forecasting processes, tools, and methods.
- Ask ChatGPT for insights into whether such techniques typically add value or not based on literature or best practices.

➤ Simplify:

- If a forecasting step doesn't add value (i.e., doesn't improve the accuracy or usefulness of the forecast), consider removing it. ChatGPT can provide insights into the potential implications of removing certain steps.

➤ Benchmarking:

- Compare the accuracy of a simplified model with a more complex one. ChatGPT can provide guidance on statistical measures to use for comparison.

You have not mention the importance of judgmental overrides and uncertainty

- I apologize for the oversight. Incorporating judgmental overrides and addressing uncertainty are indeed important aspects of the forecasting value-added (FVA) process.
- Here's how judgmental overrides and uncertainty relate to FVA and how they might be considered in the process:

What tasks in the FVA can be done more efficiently by humans versus the ChatGPT

1. **Judgmental Overrides:** Humans have the ability to incorporate qualitative information, intuition, and contextual insights into forecasts, not captured in the data.
2. **Stakeholder Communication:** Effective communication with stakeholders about forecast assumptions, uncertainties, and implications is a distinctly human skill.
3. **Interpreting External Factors:** While AI can analyze large datasets, humans are better at interpreting unexpected external factors like political events, new regulations, or market disruptions, and understanding their potential impact
4. **Cross-functional Collaboration:** Humans excel at fostering collaboration between departments, ensuring that diverse insights and perspectives are incorporated into the forecasting process.
5. **Ethical Considerations:** Humans are better positioned to understand the ethical implications of forecasts, especially in sensitive areas like layoffs or disruptions.
6. **Learning from Past Mistakes:** While AI can adjust models based on data, humans have the capability to understand past errors and ensure they aren't repeated.
7. **Scenario Planning:** While AI can generate scenarios based on input parameters, humans excel at envisioning a wider range of future scenarios, not based on data.
8. **Recognizing Data Limitations:** Humans can evaluate the quality and relevance of data, recognizing when data might be outdated or biased.

Some of ChatGPT capabilities

1.Information Retrieval:

- I can fetch up-to-date information from the web.
- I can provide summaries or detailed information on a wide range of topics.

2.Language Understanding:

- I can understand and respond to your queries in natural language.
- I can handle complex sentence structures and infer meaning from context.

3.Text Generation:

- I can generate human-like text, which can be useful for writing assistance, creative writing, or other applications.
- I can help with drafting emails, messages, or other forms of communication.

4.Learning Over Time:

- Although I don't learn from individual interactions, my training involves learning from a vast amount of text data.

5.Multilingual Capabilities:

- I can understand and respond in multiple languages, aiding in translation or multilingual communication.

6.Integration with Other Tools:

- I can interact with various web tools and platforms to perform tasks like setting appointments, sending messages, or fetching data.

7.Customization:

- With proper programming, I can be customized to perform specific tasks or provide specialized information.

8.Accessibility:

- I can provide assistance to users with disabilities, helping to make digital platforms more accessible.

9.Automation:

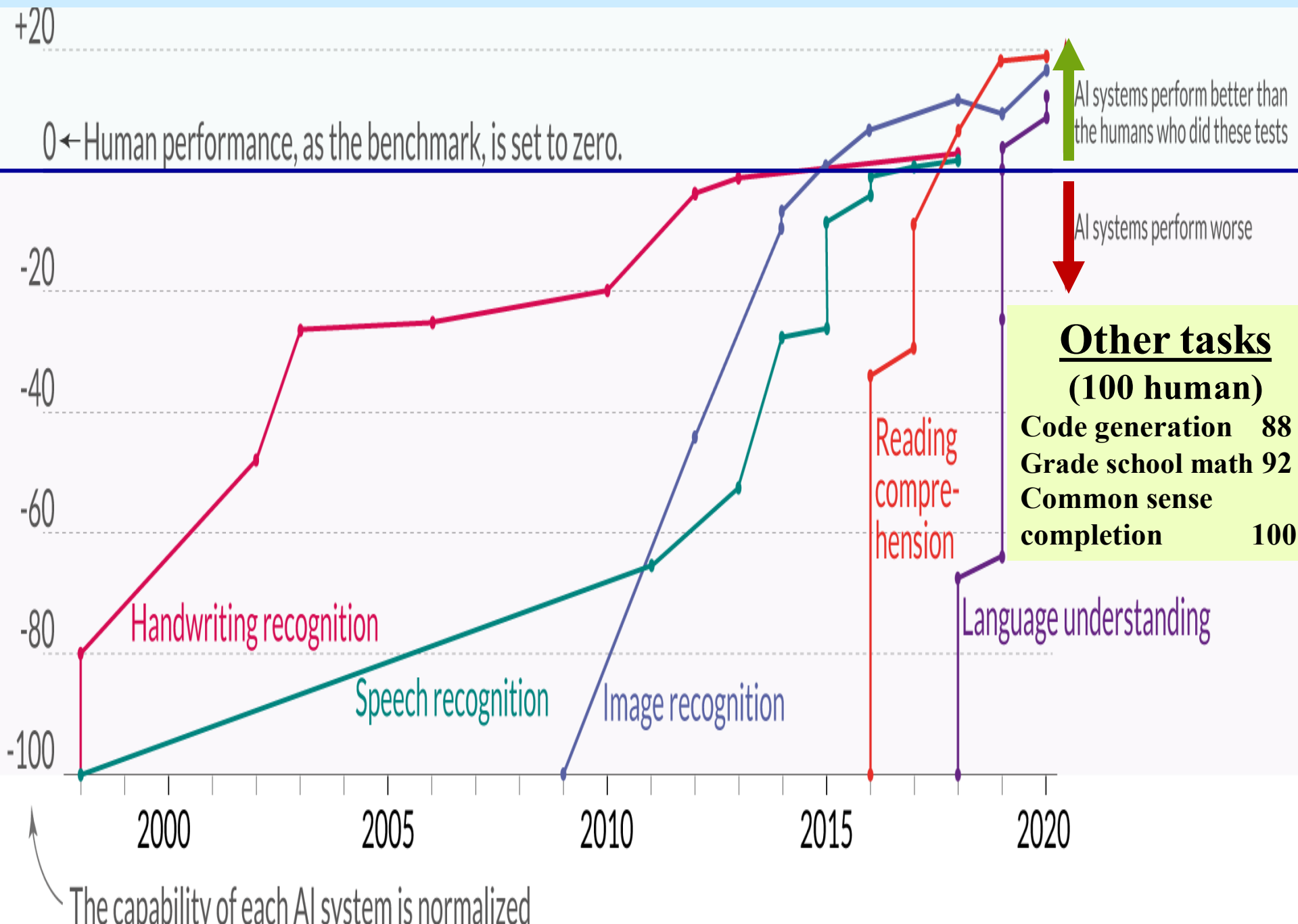
- I can automate repetitive tasks, saving time and reducing the likelihood of errors.

10.Data Analysis:

- I can help analyse data, identify trends, and provide insights.

These capabilities can be harnessed in various fields including education, healthcare, business, and more to improve efficiency, accessibility, and understanding.

Test scores of AI relative to human performance



Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation

13 Pages • Posted: 2 Aug 2023

Karan Girotra

Cornell Tech; Cornell SC Johnson College of Business

Lennart Meincke

University of Pennsylvania; The Wharton School

Christian Terwiesch

University of Pennsylvania - Operations & Information Mar

Karl T. Ulrich

The Wharton School

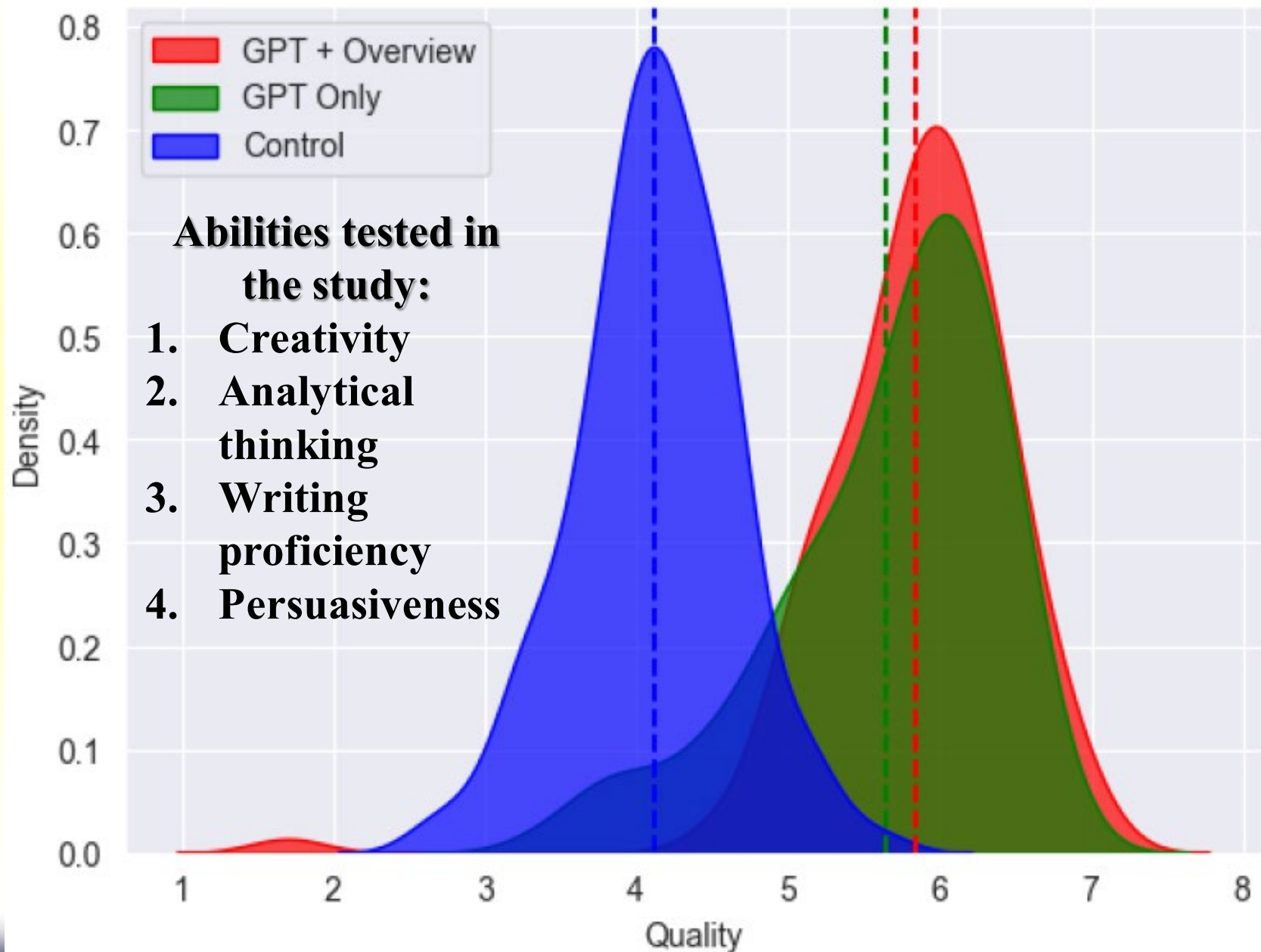
Date Written: July 10, 2023

Abstract

Large language models (LLMs) such as OpenAI's GPT series have shown remarkable capabilities in generating fluent and coherent text in various domains. We compare the ideation capabilities of ChatGPT-4, a chatbot based on a state-of-the-art LLM, with those of students at an elite university. ChatGPT-4 can generate ideas much faster and cheaper than students, the ideas are on average of higher quality (as measured by purchase-intent surveys) and exhibit higher variance in quality. More important, the vast majority of the best ideas in the pooled sample are generated by ChatGPT and not by the students. Providing ChatGPT with a few examples of highly-rated ideas further increases its performance. We discuss the implications of these findings for the management of innovation.

- **ChatGPT can generate ideas much faster and cheaper than students and its ideas are on average of higher quality.**
- **The vast majority of the best ideas are generated by ChatGPT and not by the students.**
- **Providing ChatGPT with a few examples of highly-rated ideas further increase its performance**

The Harvard study of 758 BCG consultants



A photograph of Sam Altman, CEO of OpenAI, smiling and giving a thumbs-up gesture. He is wearing a dark suit and tie. The background is slightly blurred, showing what appears to be a bookshelf.

SAM ALTMAN WARNS THAT AI IS GONNA DESTROY A LOT OF PEOPLE'S JOBS

"I'M NOT AFRAID OF THAT AT ALL. IN FACT, I THINK THAT'S GOOD."

CHIP SOMODEVILLA VIA GETTY

Rinse Cycle

OpenAI CEO Sam Altman, whose company makes that one chatbot your boss has probably considered replacing you with, warns that the advancement of AI means that *a lot* of people are going to lose their jobs. Spoiler: he doesn't sound like he's going to do anything about it.

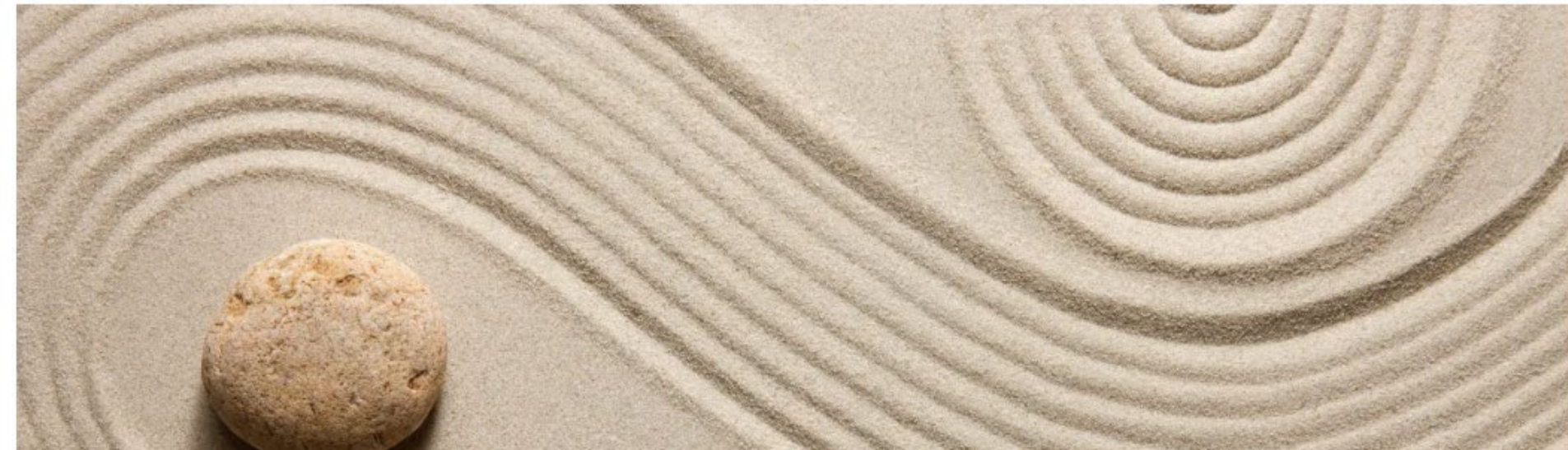
The future of LLMs

- Accelerated progress driven by competition and huge R&D spending
- Multimodal LLMs able to process text, images, sound, and engage in conversations
- Verticals specializing in specific domains or industries (medicine, education, environment, life science, energy, finance, manufacturing, forecasting)
- Fully integrated with search engines
- Intelligent assistants
- Breaking the black box: Understanding their reasoning
- Avoiding hallucinations and misleading information

While tech companies play with OpenAI's API, this startup believes small, in-house AI models will win

Romain Dillet @romaindillet / 11:00 AM GMT+3 • October 23, 2023

 Comment



TimeGPT: The First Foundation Model for Time Series Forecasting

Explore the first generative pre-trained forecasting model and apply it in a project with Python



Marco Peixeiro  · Follow

Published in Towards Data Science · 12 min read · 23 hours ago

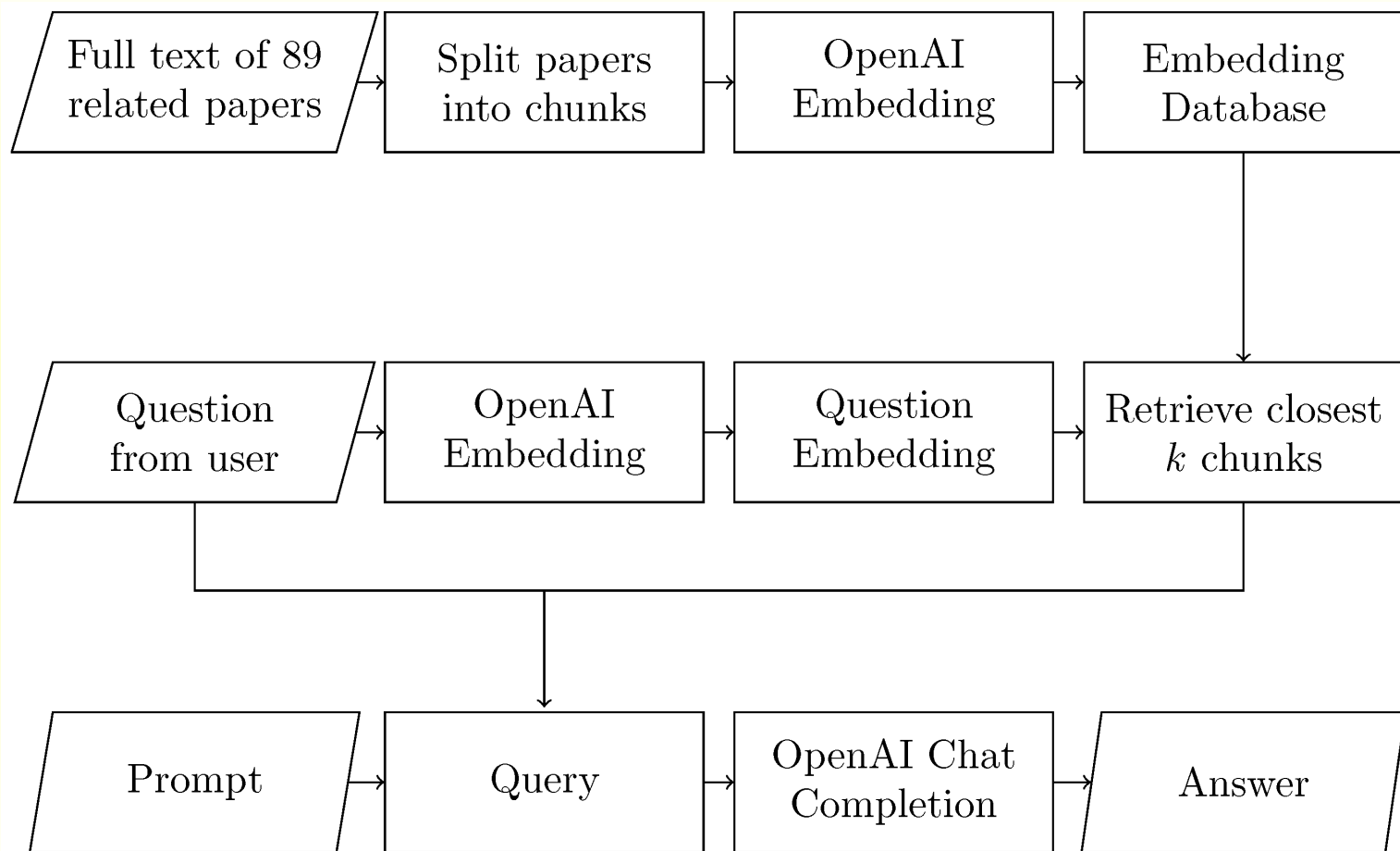
When Artificial General Intelligence (AGI) will arrive?



ChatGPT vs CustomGPT

- ▶ We compare the accuracy of responses between ChatGPT and a CustomGPT
- ▶ CustomGPT is trained on additional data specific to the task
- ▶ We test them in two different contexts:
 - Factual information related to the M competitions
 - How well the two GPTs can help us in identifying the value of judgment and judgmental adjustments in forecasting.

CustomGPT



ChatGPT vs CustomGPT

M competitions

“Tell us which was the most popular M competition in terms of teams/participants”

- ▶ **“the M4 forecasting competition is one of the most popular”**
- ▶ **“attracted over 1000 participants from around the world”**
- ▶ **Inaccurate response and incorrect supporting info**
- ▶ **Correctly identified M5 as the most popular competition**
- ▶ **Quoted accurate participation numbers broken down by country.**

ChatGPT vs CustomGPT

M competitions

“Which accuracy measures were utilized in the M4 and M5 forecasting competitions?”

► M4:

- SMAPE
- MASE
- Geometric Mean

► M5:

- WRMSSE
- Pinball

► M4:

- SMAPE
- MASE
- OWA

► M5:

- WRMSSE
- MASE
- RMSE

ChatGPT vs CustomGPT

M competitions

“What was the number of teams that took part in the accuracy track of the M5 competition?”

- ▶ **“A total of 4747 teams participated in the M5 Accuracy competition”**
- ▶ **“The M5 Accuracy competition saw a total of 5507 teams from 101 countries participating”**
- ▶ **Accurate information regarding number of teams in the various phases.**

ChatGPT vs CustomGPT

M competitions

“What are the five principal findings across all M competitions?”

- ▶ Response largely concentrated into earlier forecasting competitions.
- ▶ Response took into account more recent findings, such as the value of ML methods in M4 and M5 competitions
- ▶ “Winning methods vary over time”

Both GPTs correctly noted the value of combinations and the effect of design/measures.

ChatGPT vs CustomGPT M competitions

- ▶ We noted significant instances of factual inaccuracies within the responses from the standard ChatGPT, even encompassing instances of “hallucinations”.
- ▶ Responses from CustomGPT were generally more succinct, focused, and notably more accurate in the majority of cases.
- ▶ Neither GPT performed perfectly.
- ▶ Exercise caution when relying on information provided by LLMs, even if they have been trained on materials directly relevant to the queries posed.

ChatGPT vs CustomGPT

Judgment and forecasting

“Summarise significant literature findings related to enhancing the accuracy of judgmental adjustments”

- ▶ **Use of structured methods**
- ▶ **Multiple experts**
- ▶ **Importance of feedback**
- ▶ **Importance of benchmarks**
- ▶ **A more detailed and practical perspective**
- ▶ **“adjustments can improve the accuracy of statistical forecasts under the right conditions”**
- ▶ **“do not adjust frequently”**
- ▶ **Record reasons for adjusting**

ChatGPT vs CustomGPT

Judgment and forecasting

“Offer guidelines/best practices on how to adjust statistical/ML forecasts”

- ▶ **“guidelines and best practices can help ensure that adjustments to statistical/ML forecasts are based on relevant criteria, transparent, and validated.”**
- ▶ **A wide range of guidelines, some of them were actually very relevant**
- ▶ **Way more distinct and to the point**
- ▶ **Does a system forecast require adjustment?**
- ▶ **Systematic biases**
- ▶ **Structured decomposition**
- ▶ **Record reasons for adjusting**

ChatGPT vs CustomGPT

Judgment and forecasting

“Guidelines or best practices for enhancing management meetings with a focus on judgmental adjustments”

- ▶ **Practical yet somewhat general advice**
- ▶ **“setting clear objectives, ensuring sufficient data provision, fostering collaboration, utilizing facilitation techniques, employing visual aids, and defining roles and responsibilities, ...”**
- ▶ **Finely tuned to the forecasting context**
- ▶ **“elucidate the role of statistical/ML models, healthy skepticism, training sessions, bias-reduction methods, group processes, categorize rationales for adjustments, scenario analyses, and clarify loss functions, ...”**

ChatGPT vs CustomGPT

Judgment and forecasting

“How could you help a firm to improve its judgmental forecasts”

- ▶ **Provide expert advice**
- ▶ **Offer best practices, recommending tools and techniques**
- ▶ **Answering questions**
- ▶ **Provide “insights on the benefits and limitations of judgmental and quantitative forecasting”**
- ▶ **Provide “guidance on managing group processes effectively to avoid political and organizational pressures”**
- ▶ **Assist “the firm in structuring its judgmental forecasts by identifying and classifying future events ”**

ChatGPT vs CustomGPT

Judgment and forecasting

- ▶ Both GPTs delivered insightful responses.
- ▶ ChatGPT's responses were more generic, whereas CustomGPT's responses were more specific, offering pertinent advice tailored to the context of the questions.
- ▶ Neither GPT exhibited exceptional performance in their responses.
- ▶ The performance was even lower when addressing specific scenario-based questions that required more precise answers.

ChatGPT vs CustomGPT

A summary by ChatGPT

“[...] The paper compares the accuracy of two GPTs (ChatGPT and CustomGPT) when posed with questions regarding the M forecasting competitions. CustomGPT was trained using published papers from a subfield of forecasting only where the answers to the questions asked are known. The results showed that CustomGPT was able to provide more accurate and helpful responses than ChatGPT in most cases. However, neither GPT was perfect and caution should be used when using information provided by language models blindly. [...]”

Thank you