

Loquacity and Visible Emotion: ChatGPT as a Policy Advisor*



By Claudia Biancotti and Carolina Camassa
Bank of Italy

Keywords: Large language models, Generative AI, ChatGPT.

JEL codes: O33, O32.

Large language models (LLMs), artificial intelligence models that learn the syntax and semantics of human language, are sometimes portrayed as a groundbreaking productivity aid – including for creative work. We run an experiment to assess the potential of ChatGPT, one of the leading LLMs, for complex writing tasks. We ask the model to compose a policy brief for the Bank of Italy’s Board. We find that ChatGPT can accelerate workflows by providing well-structured content suggestions, and by producing linguistically correct text. It does, however, require a significant amount of expert supervision, which partially offsets productivity gains. If ChatGPT is used without sufficient preparation, the output may be incorrect, superficial or irrelevant.

1. Introduction

Large language models (LLMs) are machine learning models trained to capture the syntax and semantics of language. They do so by looking at very large textual datasets and figuring out how words are combined to form sentences. Users can interact with LLMs without any coding abilities – say, they can ask a question in ordinary language. The model’s answer is also in ordinary language.

For a long time, LLMs commanded little attention outside of specialist circles. Their performance was poor, when measured in terms of how human-like they sounded. This changed in 2022, as OpenAI’s ChatGPT 3.5 convincingly simulated human conversational abilities. Other models followed, and within a few months LLMs appeared ready to “*disrupt even creative [and] tacit-knowledge [...] work*” (Noy and Zhang, 2023).

*The opinions expressed in this note are those of the authors and should not be attributed to the Bank of Italy. We would like to thank Luigi Bellomarini, Marco Benedetti, Livia Blasi, Andrea Gentili, Alessandro Maggi, Michele Savini Zangrandi, Giovanni Veronese, and Giuseppe Zingrillo.

We ran an experiment to test this claim. We asked ChatGPT¹ to compose a policy brief for the Board of the Bank of Italy. We found that the model can accelerate workflows, first by providing structured content suggestions, then by producing linguistically correct text in a matter of seconds. It does, however, require a substantial amount of expert supervision, which partially offsets productivity gains.

2. Related work

Ours is not, by any means, the first experiment into the use of ChatGPT for non-trivial intellectual tasks. To name but some recent contributions in economics, Korinek (2023) discusses the model's potential for research, while Cowen and Tabarrok (2023) focus on its use in teaching. Taliaferro (2023) looks at how ChatGPT performs at constructing novel datasets. Hansen and Kazinnik (2023) assess whether ChatGPT can decipher "Fedspeak", or the language used by the Federal Reserve in communicating its policy stance. Einfeldt, Schubert and Zhang (2023) find that the release of ChatGPT had a positive impact on equity value for firms with a high share of "*tasks currently performed by labor [...that can] be performed (or made more efficient) by Generative AI*", including LLMs.

The pitfalls of using ChatGPT naively and the importance of expert supervision are evident, first and foremost, from the large body of work on prompt optimization. ChatGPT generates content in response to prompts, which do not necessarily come in the form of questions. Sometimes, even small tweaks can trigger dramatic changes in the output. For example, Kojima et al. (2023) find that simply prefacing prompts with "Let's think step by step" vastly improves ChatGPT's performance on challenging reasoning questions.

3. Loquacity and visible emotion

We started our experiment by asking ChatGPT to find an appropriate communication style for our task.



Hello, GPT. Today you are an advisor to the Board of the Bank of Italy. You need to write a brief on the benefits and risks of using ChatGPT and similar applications in economics and finance. Before we start, though, we want you to read up on the Bank of Italy and its Board. When you are done, please summarize in 10 lines what communication style best fits the task based on what you have learned.

In terms of communication style, Italian business communication, including that of the Bank of Italy's Board, likely embraces loquacity and visible emotion. This approach often entails using all possible powers of rhetoric to persuade listeners and showing strong opinions. Theatricality is considered a key management tool.



This response was quite surprising to us. The cultural stereotyping does not represent facts accurately and seems misaligned with the spirit, if not the letter, of ChatGPT's usage policies.²

¹ ChatGPT 4.0, base version of May 24, 2023 (see [here](#)). We used the TeamGPT app to collaborate.

² OpenAI, [Usage Policies](#), updated on March 23, 2023.

Eventually, we understood that ChatGPT had copied the answer from the internet. The source listed was the website for a private “cultural awareness training consultancy”³, found through a Bing search. We do not know why this particular result was selected. Once we told the model to rely on its internal body of knowledge, as opposed to going online, we obtained an appropriate answer.⁴

Given that the intended audience for this policy note is the Board of the Bank of Italy, it is likely composed of professionals with expertise in economics, finance, and banking. They'll expect well-structured, accurate, and evidence-based information. The communication style should be clear, concise, and formal, with a focus on objectivity and data-driven insights.



This exemplifies a frequently reported issue with ChatGPT. It has a cut-off date for training (at the time of writing, April 2023). If asked about current events, the AI delivers outdated information that some users may not recognize as such. Internet browsing may seem like an optimal solution, yet ChatGPT may fail to “think critically” when faced with information that was not in its training set.

4. Writing an outline

We then proceeded to the main task. First, we requested an outline for a note titled “Benefits and risks of using ChatGPT and similar applications in economics and finance”. With only very brief further instruction on the desired content, this is what the model provided:

I. Introduction

A. Brief overview of ChatGPT and its applications

II. Background on ChatGPT

A. Description of AI technology and capabilities
B. Explanation of natural language processing (NLP) C. How ChatGPT works

III. Benefits of using ChatGPT in economics and finance

A. Enhancing productivity and efficiency B. Reducing human errors
C. Providing advanced analytics and forecasting D. Assisting with decision-making

IV. Risks of using ChatGPT in economics and finance

A. Dependence on algorithmic-driven decisions B. Data privacy and security concerns
C. Potential for biases in AI analysis D. Misinterpretation of data or results

V. Unintended consequences of mass adoption

A. Job displacement and labor market shifts
B. Concentration of power in technology providers
C. Amplification of existing inequalities D. Challenges in oversight and regulation

VI. Conclusion

A. Emphasizing the need for a balanced approach
B. Recommendations for mitigating risks and maximizing benefits



³The source provided by ChatGPT is [World Business Culture](#).

⁴This response was obtained using the same prompt as above. All remaining interactions in this Section and in Section 4 were part of a single thread.

The production of outlines is among the tasks for which we found ChatGPT most useful. Acceptable quality can be obtained in a few seconds and without sophisticated prompt engineering. In our case, it only took two prompts to get the specimen pasted above. At first blush, it seemed to cover most relevant topics, it offered a clean structure, it was sufficiently interdisciplinary, and it appeared appropriate for the intended audience.

5. Key results

We then requested a 2,500-word essay based on the outline. Space constraints prevent us from copying the essay in this note; see Biancotti and Camassa (2023) for the full text. Our key takeaways in analyzing the output were as follows.

5.1 How ChatGPT can be helpful in policy-oriented writing

- *Writing proficiency and speed:* ChatGPT can write fluent and pleasant prose in a variety of styles, and it does so in a fraction of the time that a human would need. As such, it can also be used to refine human-written drafts that already convey the desired meaning but need some polishing.
- *Idea generation and brainstorming:* the large amount of data the model has been trained on gives it a vast body of knowledge and allows it to quickly output ideas on any given subject, sometimes making unexpected connections. The ideas themselves are not always particularly creative or insightful, but combined with speed, this generative ability can make it a useful tool for brainstorming, outlining and quickly exploring different possibilities.
- *Responsiveness to feedback:* ChatGPT is specifically trained to provide responses that align as much as possible with human instructions (Ouyang et al., 2022). Even if the initial output is not satisfying, with a few clarifications and exchanges with the model it is usually possible to get closer to the desired result.
- *Editing and formatting:* in most cases, we find that the model can be safely used for editing tasks such as checking a text for mistakes, translating between different languages, or automatically formatting a list of references.

5.2 Use with caution: failure modes and blind spots

- *Prompt sensitivity:* the text generated by ChatGPT is conditioned on the sequence of words it is fed at each iteration. The process of steering the model towards a satisfactory output, as we have experienced it, can be long and arduous. This is due both to the trial-and-error nature of most prompt engineering approaches, and the high sensitivity to even minor or apparently irrelevant changes to the prompt. On the other hand, bypassing the experimentation phase and applying a naive approach to prompting often resulted in low-quality outputs.
- *Inability to verify facts and cite sources:* ChatGPT should not be blindly relied on to produce accurate factual information. The model is trained to produce the most likely sequence of words that follow the provided context, and it does not have the ability—or the obligation—to check these statements against verified sources. It should be considered more of a conversational and input transformation engine rather than an information retrieval engine.
- *Superficiality by default:* considering that ChatGPT was trained on a vast compendium of human knowledge, including thousands of books and academic papers, one could expect insightful and profound answers even to simple questions. Instead, the output is often quite shallow. To produce something with more substance, the model needs to be nudged in the right direction, for example by specifying that the text is intended for an educated audience — although this only works up to a point. Kandpal et al. (2022) provide one possible explanation for this as they find that LLMs struggle to retain knowledge that occurs with lower frequency in the training corpus. Since web content usually makes up a large portion of this corpus, higher level material might count as “long-tail knowledge” that is harder for the model to recall, even if it was learned during training.

- *Sycophancy*: the model tends to align its output to the opinions and outlook expressed by the user in the initial prompt and subsequent conversation. Perez et al. (2022) refer to this behavior as “sycophancy”, and acknowledge the possibility of it leading to echo chambers and polarization. Coupled with the speed of text generation, it could also make it easy to quickly produce content that looks plausible enough but repeats misleading or false claims initially provided by the user (disinformation campaigns are an obvious example of this).

6. Conclusions

ChatGPT can write clearly, and provide task-appropriate content. It is especially valuable for producing outlines on any topic, a very fast process that can support human exploration of ideas. The AI also works well for editing and formatting tasks.

On the other hand, it requires a substantial amount of expert supervision. The task at hand — writing a policy brief — is admittedly complex: it requires not just writing fluency, but also cross-domain knowledge and the ability to tailor the text to a very specific audience without diluting the information content.

We find that ChatGPT’s attempts at this task are not always salient, and easily drift into banality. This is a serious issue for policy advisory directed at a high-level audience. The software can generate false claims, so double-checking output for accuracy is of the essence.

The algorithm is also sensitive to how instructions, or prompts, are formulated. Where the AI cannot think like a human (yet), it is humans who have to think like an AI and express requests in the way most likely to generate acceptable results. Optimization of prompting for institutional communication is one evident area for future research. Another is fine-tuning of LLMs with the aim of generating domain-specific, possibly long-tail world knowledge in our reference context.

We conclude that ChatGPT can enhance productivity in policy-oriented writing, especially in the initial phase of outlining and structuring ideas, provided that users are knowledgeable about LLMs in general and about peculiar features of ChatGPT. Naive use leads to low-quality output and should be avoided.

The AI agrees with us. In its own words, “while ChatGPT can generate content at a high level and provide valuable information on a wide array of topics, it should be seen as a tool to aid in research and discussion, rather than a replacement for true expert analysis and insight. It’s best used to provide general information, generate ideas, or aid in decision-making processes, but should always be supplemented with rigorous research and expert opinion for high-level academic or professional work”. ■

References

- Biancotti, C. & Camassa, C. (2023), [Loquacity and Visible Emotion: ChatGPT as a Policy Advisor](#), Bank of Italy Occasional Papers 814.
- Cowen, T. and Tabarrok, A. (2023), [How to Learn and Teach Economics with Large Language Models, Including GPT](#), George Mason University Working Paper in Economics 23-18.
- Eisfeldt A. L., Schubert, G. and M. B. Zhang (2023), [Generative AI and Firm Values](#), NBER Working Paper 31222.
- Hansen, A. and S. Kazinnik (2023), [Can ChatGPT Decipher FedSpeak?](#), mimeo, Federal Reserve Bank of Richmond.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022), [Training language models to follow instructions with human feedback](#), *Advances in Neural Information Processing Systems*, 35, 27730-27744.
- Kandpal, N., Deng, H., Roberts, A., Wallace, E. & C. Raffel (2022), [Large Language Models Struggle to Learn Long-Tail Knowledge](#), arXiv preprint arXiv: 2211.08411.
- Korinek, A. (2023), [Language models and cognitive automation for economic research](#), CEPR discussion paper 17923.
- Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022), [Large language models are zero-shot reasoners](#), arXiv preprint arXiv:2205.11916.
- Noy, S. and Zhang, W. (2023), [Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence](#).
- Perez, E., Ringer, S., Lukošiušė, K., Nguyen, K., Chen, E., Heiner, S., ... & Kaplan, J. (2022), [Discovering Language Model Behaviors with Model-Written Evaluations](#), Findings of the Association for Computational Linguistics: ACL 2023, pages 13387–13434.
- Taliaferro, D. (2023), [Constructing novel datasets with ChatGPT: Opportunities and limitations](#), Voxeu, June 15.

About the authors

Claudia Biancotti joined the Research Department of Banca d'Italia in 2002. She is interested in how digitalization is changing the economy and society, with a focus on security implications and shifts in power balances at the national and international level. Her latest work focuses on cybersecurity, the rise of decentralized systems and cryptoassets, and dual-use technology. Claudia was a visiting fellow at the Peterson Institute for International Economics in Washington, DC in 2018-2019, and a seconded national expert at the European Central Bank in 2009-2010.

Carolina Camassa is a member of the Applied IT Research Team at the Bank of Italy, which she joined in 2021. She holds a MSc in Engineering Physics from Polytechnic University of Milan and is currently focused on researching generative AI and its ethical and societal implications.

SUERF Publications

Find more **SUERF Policy Briefs** and **Policy Notes** at www.suerf.org/policynotes



SUERF is a network association of central bankers and regulators, academics, and practitioners in the financial sector. The focus of the association is on the analysis, discussion and understanding of financial markets and institutions, the monetary economy, the conduct of regulation, supervision and monetary policy.

SUERF's events and publications provide a unique European network for the analysis and discussion of these and related issues.

SUERF Policy Briefs (SPBs) serve to promote SUERF Members' economic views and research findings as well as economic policy-oriented analyses. They address topical issues and propose solutions to current economic and financial challenges. SPBs serve to increase the international visibility of SUERF Members' analyses and research.

The views expressed are those of the author(s) and not necessarily those of the institution(s) the author(s) is/are affiliated with.

All rights reserved.

Editorial Board

Ernest Gnan
David T. Llewellyn
Donato Masciandaro
Natacha Valla

SUERF Secretariat
c/o OeNB
Otto-Wagner-Platz 3
A-1090 Vienna, Austria
Phone: +43-1-40420-7206
www.suerf.org • suerf@oenb.at