Can ChatGPT predict future interest rate decisions?

WOODHOUSE, Drew and CHARLESWORTH, Alex

Can ChatGPT Predict Future Interest Rate Decisions?

Drew Woodhouse[1][*] and Alex Charlesworth[*]

First Version: September 2023

Abstract

The public interest in Large Language Models (LLMs) and generative artificial intelligence (AI) has gained significant traction, with academic studies attempting to assess their application to a range of fields. In particular, the rise in Generative Pre-trained Transformer (GPT) models, in the form of ChatGPT has prompted a range of calls to examine beyond its use in writing and language generation; to test its predictive qualities and assess its abilities to process complex textual information. Our paper extends upon the examination of GPT model use in the monetary policy context. We set out to test the hypothesis that LLMs and GPT models can offer predictive qualities of future interest rate decisions through their textual processing and sentiment capabilities. Specifically, we use GPT-3.5 to evaluate and label the speech of every Bank of England Monetary Policy Committee (MPC) member based on linguistic expectations. We then model a preferred policy rate vote equation for each speech giving committee member (i) in each of their future rate decision windows ($t + n$). We find that ChatGPT can predict future interest rate decisions. Our results provide evidence for the potential of LLMs to help us better process latent human beliefs, make out-of-sample predictions and navigate possible models of rational expectations.

Keywords: Generative Pre-trained Transformer (GPT); Bank of England; Monetary Policy; Large Language Models

JEL Classification: G11, E52, E58, G81

[*]Sheffield Business School, Sheffield Hallam University, City Campus, Howard Street, Sheffield, South Yorkshire, S1 1WB, UK.
[1]Corresponding author E-mail: Drew.Woodhouse@shu.ac.uk

1

1 Introduction

The public interest in Large Language Models (LLMs) and generative artificial intelligence (AI) has gained significant traction, with academic studies attempting to assess their application to a range of fields. In particular, the rise in Generative Pre-trained Transformer (GPT) models, in the form of ChatGPT has prompted a range of calls to examine beyond its use in writing and language generation; to test its predictive qualities and assess its abilities to process complex textual information. Despite an overwhelming public and academic interest in broader generative AI, we are still unaware of their array of capabilities, applications and predictive qualities, with applications variously spanning Law (Choi et al., 2023), Public Health (Biswas, 2023) and Mathematics (Frieder et al., 2023) to accelerate our understanding of their uses.

These calls have extended to the context of Finance and Economics, generating a growing volume of papers. For example, one paper has examined LLMs stock return prediction, demonstrating through an econometric approach, that sentiment analysis preformed by ChatGPT is a determinant factor for a range of return windows (Lopez-Lira & Tang, 2023). Cowen and Tabarrok (2023) demonstrate how ChatGPT can be used within the context of economic teaching pedagogy, whilst assisting with research, idea generation and assessments. Ko and Lee (2023) examine the efficacy of ChatGPT in aiding portfolio management with regard to asset allocation and diversification. They find that ChatGPT's selections are statistically significantly better in terms of diversity index than randomly selected assets, suggesting that LLMs can mimic the decisions of portfolio managers. Simiarly, (Yang & Menczer, 2023) assess whether ChatGPT, can evaluate the credibility of news outlets. They find that ChatGPT can provide ratings for a diverse set of news outlets, including those in non-English languages and satirical sources, along with contextual explanations.

Within the context of monetary policy, Hansen and Kazinnik (2023) contribute to an established body of literature that studies the content and sentiment of central bank communication (e.g. Curti and Kazinnik, 2023; Ehrmann and Fratzscher, 2007). They investigate the ability of GPT models to decipher Fedspeak, the technical term used by the Federal Reserve to communicate monetary policy decisions. By evaluating the ability of GPT model to classify the policy stance of Federal Open Mark Committee (FOMC) announcements, they conclude that GPT models demonstrate a strong performance in classification against benchmarks.

Our paper extends upon the examination of GPT model use in the monetary policy context.

We set out to test the hypothesis that LLMs and GPT models can offer predictive qualities of future interest rate decisions through their textual processing and sentiment capabilities. Specifically, we use GPT-3.5 to evaluate and label the speech of every Bank of England Monetary Policy Committee (MPC) member based on linguistic expectations. We then extend this to an econometric setup, modelling a preferred policy rate vote equation for each speech of committee member (i) in each of their future rate decision windows ($t + n$).

We find that ChatGPT can predict future interest rate decisions. Our results provide evidence for the potential of LLMs to help us better process latent human beliefs, make out-of-sample predictions and navigate possible models of rational expectations.

The layout of this paper is as follows. Section 2 outlines the data processing strategy. Section 3 discusses the methods and model. Section 4 presents the results, with Section 5 concluding.

2 Data

2.1 Bank of England & the Monetary Policy Committee (MPC): Rate Decisions and Speeches

We restrict our analysis to decisions made by the Bank of England (BoE) regarding UK monetary policy. Monetary policy decisions are contingent to each central bank. We focus our analysis on the BoE for three reasons. Firstly, the committee structure of the Bank of England Monetary Policy Committee (MPC) offers 'individualistic' vote based decision-making. This individualistic system builds in a diversity of views, reducing scope for group think. On the other hand, when voting is proposed to be split, forward guidance via communications becomes a key transmission channel. We argue that communication via speeches therefore offers a strong insight into future rate decisions of individuals. Secondly, the MPC uses minutes to communicate decisions reached and points of disagreement retrospective to a rate decision. However, as argued by Paul Tucker, a former MPC member, that "it is more difficult for us than for some of our peers to release an informative statement immediately after the policy meeting: if you don't know what you're going to decide, it is pretty hard to prepare a draft in advance" (Tucker, 2008). Autonomy in individual level communication offers prospective insight. Thirdly, the Bank can be described as operating within a framework of 'constrained discretion'. The Bank's monetary policy objective is to deliver price stability–low and stable inflation and, subject to that, to support the Government's economic objectives including those for growth and employment. In a framework of constrained

3

discretion communication is likely to play an important role in resolving uncertainty about the course of monetary policy. As demonstrated by Reeves and Sawicki (2007), we therefore expect communication to unveil short-term interest rate expectations.

The MPC meets eight times a year to discuss the economic outlook and set the direction for monetary policy. These meetings precede and proceed with official speeches given by BoE staff members, including MPC members, alongside the dissemination of MPC meeting minutes. Official speeches involve a range of topic discussion, across varied audiences and through a multitude of formats. Our empirical design and sample focuses on all MPC member speeches and rate decisions between the start of the Bank's independence, June 1997, to our most recent decision in June 2023. We gather both the speeches given by MPC members during their tenure and the historical MPC member voting decisions from publicly available data provided by the Bank of England[1]. We then allocate each MPC member speech to a voting window based on their date of speech delivery. We define a voting window of t as the duration after an MPC decision, ending the day before the next MPC vote. By matching the dates of individual speeches with individual voting decisions, we can evaluate the short-term policy sentiment that can be derived from speeches that occur leading and lagging voting decisions.

2.2 Measuring Generative AI's Expectations of *BoE-speak*

To test for predictive qualities, we apply the textual and sentiment capabilities of GPT models to individual speeches. The GPT (Generative Pretrained Transformer) family of models comprises transformer-based models that undergo training on extensive text data. By employing a self-attention architecture, GPT models effectively incorporate sentence and paragraph context, enabling them to demonstrate proficiency in diverse natural language processing tasks and comprehensively grasp nuanced language understanding. Typically, models such as GPT-3 do not necessitate explicit examples for additional training in order to achieve high performance and can be used 'off-the-shelf', affording of public consumption. This has notable implications in relation to empowering the public within the realms of household finance and economics.

After pre-processing the textual data, we provide predefined prompts to a ChatGPT-3.5 API model, applying it to each individual speech. We do not draw random sentences from speeches

---

[1]BoE speeches are available from the website: https://www.bankofengland.co.uk/news/speeches. Historical MPC Voting Data can be sourced via: https://www.bankofengland.co.uk/-/media/boe/files/monetary-policy-summary-and-minutes/mpcvoting.xlsx

4

but rather read in the processed full speech for contextual sentiment[2]. Prompts are critical in guiding ChatGPT's responses to tasks. The prompt serves an invaluable mechanism to measure the expectation and sentiment, allowing us to decode what we coin *BoE-speak*[3]. Keeping with commonly accepted variations of central bank communication (e.g. Chen, 2016; Jegadeesh and Wu, 2017), we apply the following prompt to each speech:

> You have a PhD in Economics and are working for a central bank. Your task is to classify this speech into one of three categories (dovish, neutral, hawkish).

We decide to employ 3 categories over more as we discern no notable nor subtle differences between closely related labels. We set a temperature of 0 to maximise the reproducibility of the results and limit computational rumination. We assign numerical values to the categories allocated by GPT-3 on a scale of -1 to 1, where -1 denoting Dovish, 0 representing Neutral and 1 defining Hawkish. In Table 1, we present examples of GPT outputs across a distribution of speech classifications.

---

[2]Speeches were textually pre-processed to remove charts, footnotes and references where required

[3]We define 'BoE-speak' as the broader technical language used by the Bank of England (and central banks alike) to communicate monetary policy decisions. An applied Fedspeak term.

Table 1: ChatGPT Speech Outputs

MPC Member	Speech Date	Classification	ChatGPT Output
Andrew Bailey	28-08-2020	Dovish	Based on the speech, I would classify it as dovish. The speaker emphasizes the need for central banks to have flexibility and a range of tools, including expanded purchases of private sector assets, to respond to crises and ensure liquidity reaches where it is needed. They also discuss the state contingent nature of quantitative easing (QE) and the effectiveness of QE in crisis states. The speaker suggests that central bank balance sheets may have a counter-cyclical role and function, and that there may be a need for more potent expansion in the future. Overall, the speech suggests a cautious and accommodative stance towards monetary policy.
Sushil Wadhwani	22-02-2001	Dovish	Based on the speech, I would classify it as dovish. The speaker acknowledges the uncertainty in the international economy and the potential impact of the US economic slowdown on the UK economy. However, they also highlight that indicators of business and consumer confidence in the UK have not changed consistently and that the direct impact of the US slowdown on the UK might be relatively modest. The speaker also mentions that simulations suggest a small impact on UK GDP growth. Overall, the speech suggests a cautious and slightly pessimistic outlook, but not overly hawkish.
Paul Fisher	30-03-2011	Neutral	The speech can be classified as neutral. The speaker discusses various aspects of monetary policy implementation and the mechanics of setting policy, without expressing a strong bias towards either a dovish or hawkish stance. The speech focuses on the Bank of England's objectives and how it seeks to meet them using its balance sheet, as well as the Sterling Monetary Framework and the provision of liquidity insurance. Overall, the tone of the speech is informative and objective, without taking a strong position on the direction of monetary policy.
Stephen Nickell	16-05-2001	Neutral	The speech can be classified as neutral. It provides a detailed analysis of various factors affecting the UK labor market, such as unemployment, inflation, inactivity rates, trade union membership, and changes in the benefit system. The speech does not strongly advocate for any particular policy direction or express a strong bias towards either a dovish or hawkish stance.
Martin Weale	13-06-2011	Hawkish	Based on the speech, I would classify it as hawkish. The speaker acknowledges the weak economic growth and high inflation rate, and argues for an early increase in interest rates to address inflationary pressures. They also mention the risks of delaying the tightening of monetary policy and the need to meet the inflation target.
Dave Ramsden	07-10-2022	Hawkish	Based on the speech, I would classify it as hawkish. The speaker acknowledges the shocks hitting the UK economy and the impacts they are having, particularly in terms of inflation. They highlight the rise in energy prices as a significant contributor to inflation and discuss the tightness in the labor market and its impact on wages. The speaker also mentions the recent turbulence in financial markets and its potential effect on the real economy. Overall, the speech suggests a concern about inflationary pressures and the need for forceful monetary policy action to ensure inflation returns sustainably to the 2% target in the medium term.

3 Model

To test the predictive qualities of ChatGPT within the context of monetary policy decision-making, we define a preferred policy rate model. Specifying a Generalised Least Squares (GLM) panel model, we define a benchmark preferred policy rate as follows:

$$\Delta r_{i[t+1,t+n]} = \theta_j + \beta \cdot ChatGPT_{it} + \gamma \cdot \Delta r_{it} + \lambda \cdot Tenure_{it} + \rho \cdot Outsider_i + \nu_{it} \qquad (1)$$

Where $\Delta r_{i[t+1,t+n]}$ is MPC member i's preferred interest rate change (Δr) in their upcoming MPC meeting ($t + n$). This specification offers a persistence test up to n future MPC meeting, with the cumulative sum of individual i rate decisions from $t + 1$ to $t + n$. Individual dummy intercepts θ_j control for individual fixed effects, reflecting a member i's fundamental limit on what the rate should be. Our regressor of interest $ChatGPT_{it}$, defines the numerical conversion of the ChatGPT read response of a member (i) speech in the current MPC voted window (t). Then, Δr_{it} specifies the last known preferred policy rate change decision at MPC vote (t) of MPC member (i), acting as an autoregressive capture of a preferred interest rate path and control for multi-directionality of speeches. Next, $Tenure_{it}$ accounts for a dynamic (t) member (i) characteristic that captures the influence of committee composition factors on interest rate decisions, specifically in terms of MPC member i's at time t duration of tenure in years (i.e., Meade and Sheets, 2005; Eichler and Lähner, 2017). Furthermore, $Outsider_i$ denotes a member characteristic that remains constant over time and indicates whether member i is an 'outsider' (an external committee member) or not (i.e., Gerlach-Kristen, 2009). Composite errors (ν_{it}) at time t in member i take the form of $c_i + u_{it}$, the sum of the unobserved individual heterogeneity (c_i) and idiosyncratic disturbances (u_{it}).

We argue that $ChatGPT_{it}$, given its analysis of member speeches, attains a significant proportion of implied information pertaining to macroeconomic conditions, thus offering a combined reactionary function. As such, should this parameter of interest be positive and significant, would suggest that there is significant predictive qualities of ChatGPT.

4 Results: ChatGPT can predict future interest rate decisions

To evaluate the performance of LLMs within the context of monetary policy, we specify varied iterations of Equation 1. We present the results for +1 and +2 MPC meetings ahead in Tables 2

and 3, respectively. Our regressions include individual fixed effects with standard errors clustered at the member level.

We find that the ChatGPT score exhibits a statistically and economically significant relation to future interest rate decisions. Across a range of models and time horizons, we see that ChatGPT evaluation of speeches provides predictive use. This result highlights the potential of ChatGPT to act as a valuable tool in predicting preferred interest rate decisions, processing latent human beliefs. Generative AI has significant use in predicting future interest rate decisions.

This paper demonstrates the broader application of sentiment analysis within the context of preferred policy rate analysis. The 'power of words' and communication acts as a powerful signal for future decisions, which can be objectified by publicly available tools such as Generative AI. This emphasises the benefits of incorporating LLMs into policy analysis and forecasting.

We continue to enhance the examination of LLMs in various fields. ChatGPT's excellence in forecasting policies can be attributed to its advanced linguistic comprehension, enabling it to grasp intricacies that conventional methods cannot. Conversely, the realm of qualitative analysis, traditionally the domain of human expertise, holds promise for improvement through generative AI. Most interestingly, we demonstrate that LLM capabilities attain economic rationality, in that policy stance classifications earn sensible economic significance. This extends the work of Hansen and Kazinnik (2023) to the Bank of England context.

Table 2: Preferred Policy Rate: Δr_{it+1}

DV: Δr_{it+1}						
	(1)	(2)	(3)	(4)	(5)	(6)
Main Effects:						
$ChatGPT_{it}$	0.082***	0.057***	0.057***	0.049***	0.019**	0.021**
	(4.093)	(5.063)	(5.202)	(5.267)	(2.929)	(2.370)
Δr_{it}		0.466***	0.466***	0.553***	0.535***	0.530***
		(8.206)	(8.109)	(7.759)	(9.469)	(9.770)
$Outsider_i$					0.024**	
					(2.280)	
$Tenure_{it}$			0.000			0.008**
			(0.225)			(2.789)
Interaction Effects:						
$ChatGPT_{it}$ X Δr_{it}				0.109*		
				(1.663)		
$ChatGPT_{it}$ X $Outsider_i$					0.027	
					(1.563)	
$ChatGPT_{it}$ X $Tenure_{it}$						0.007**
						(2.428)
Estimator	FE	FE	FE	FE	RE	RE
Individual Fixed Effects	Included	Included	Included	Included	Included	Included
R^2	0.05	0.25	0.25	0.26	0.31	0.32
AIC	-218.481	-362.684	-360.722	-363.482	-386.878	-396.525
BIC	-37.262	-177.045	-170.662	-173.422	-369.197	-378.845
RMSE	0.206	0.186	0.186	0.186	0.175	0.174
Std. Errors	by: individ.	by: individ.	by: individ.	by: individ.	by: individ.	by: individ.
Observations	614	614	614	614	614	614

Table shows the results of estimation Equation (1) for predication of Δr_{it+1}. Our estimator is selected based on Hausman Test outcomes on the same variable specification. Our T-scores are in parentheses. Clustered-robust standard errors at individual level adjusted for heteroskedasticity and AR(1) [HC3 Type]. ***, ** and * represent statistical significance at the 1%, 5%, and 10% level respectively. Unit fixed effects, where included, are not reported.

Table 3: Preferred Policy Rate: Δr_{it+2}

DV: Δr_{it+2}						
	(1)	(2)	(3)	(4)	(5)	(6)
Main Effects:						
$ChatGPT_{it}$	0.168***	0.129***	0.129***	0.113***	0.039***	0.046**
	(4.043)	(4.632)	(4.711)	(4.723)	(2.620)	(2.272)
Δr_{it}		0.726***	0.727***	0.924***	0.855***	0.850***
		(6.810)	(6.745)	(6.107)	(8.688)	(9.108)
$Outsider_i$					0.058**	
					(2.352)	
$Tenure_{it}$			0.000			0.018**
			(0.136)			(2.875)
Interaction Effects:						
$ChatGPT_{it}$ X Δr_{it}				0.242*		
				(1.791)		
$ChatGPT_{it}$ X $Outsider_i$					0.074*	
					(1.710)	
$ChatGPT_{it}$ X $Tenure_{it}$						0.017**
						(2.514)
Estimator	FE	FE	FE	FE	RE	RE
Individual Fixed Effects	Included	Included	Included	Included	Included	Included
R^2	0.06	0.20	0.20	0.20	0.24	0.25
AIC	530.745	439.605	441.590	438.218	412.447	402.883
BIC	711.019	624.276	630.658	627.286	430.034	420.471
RMSE	0.378	0.353	0.354	0.352	0.339	0.336
Std. Errors	by: individ.	by: individ.	by: individ.	by: individ.	by: individ.	by: individ.
Observations	600	600	600	600	600	600

Table shows the results of estimation Equation (1) for prediction of Δr_{it+2}. Our estimator is selected based on Hausman Test outcomes on the same variable specification. Our T-scores are in parentheses. Cluster-robust standard errors at individual level adjusted for heteroskedasticity and AR(1) [HC3 Type]. ***, ** and * represent statistical significance at the 1%, 5%, and 10% level respectively. Unit fixed effects, where included, are not reported.

5 Concluding Remarks

The analysis presented in this paper shows that ChatGPT models demonstrate strong predictive qualities in relation to future UK interest rate decisions. Our findings indicate that LLMs offers a crucial instrument to qualify central bank communication and extend as a predictive component. By showcasing the utility of LLMs in the field of economics, we make a meaningful contribution to the expanding literature on the utilisation of AI and natural language processing in this domain.

Our research has several implications for future studies. First, it highlights the importance of continuing the expansion and exploration of how LLMs can provide predictive uses across a range of economic and financial applications. As the analysis of the generative AI capabilities expand, they have the potential to improve the accuracy and efficiency of decision-making processes. This paves the way for a potential new era in research and knowledge exploration, blending the profound insights of human expertise with the expansive and rapid capabilities of AI. As we persist in harnessing and refining these tools, we anticipate a more profound influence on our approaches to analysing and comprehending intricate subjects, spanning from economic policy to various domains necessitating comprehensive textual examination.

Second, we are the first, to our knowledge, to extend LLM analysis to the context of UK monetary policy. Specifically, we find that generative AI offers predictive uses when integrated within a model setup. Future research can explore the role of LLMs in capturing components of 'forward guidance' approaches of central banks.

Third, we find some support for the view that ChatGPT attains economic rationality. By prompting for an assessment of 'BoE-Speak', we contribute with Hansen and Kazinnik (2023) to extend the assessment of economic uses of LLMs. This offers an avenue of research which assesses how generative AI can be used to empower citizens with publicly available tools of economic and financial significance.

In short, our study demonstrates the value of ChatGPT in predicting future interest rate decisions, promoting future research on the application of LLMs in policy analysis.

References

Biswas, S. S. (2023). Role of Chat GPT in Public Health. *Annals of Biomedical Engineering*, *51*(5), 868–869. https://doi.org/10.1007/s10439-023-03172-7

Chen, K. (2016). Interpreting the Fedspeak: Text Analysis of FOMC Statements. https://www.bbvaresearch.com/wp-content/uploads/2016/09/160927_US_Fedspeak.pdf

Choi, J. H., Hickman, K. E., Monahan, A., & Schwarcz, D. B. (2023). ChatGPT Goes to Law School. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4335905

Cowen, T., & Tabarrok, A. T. (2023). How to Learn and Teach Economics with Large Language Models, Including GPT. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4391863

Curti, F., & Kazinnik, S. (2023). Let's face it: Quantifying the impact of nonverbal communication in FOMC press conferences. *Journal of Monetary Economics*. https://doi.org/10.1016/j.jmoneco.2023.06.007

Ehrmann, M., & Fratzscher, M. (2007). Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness? *Journal of Money, Credit and Banking*, *39*(2-3), 509–541. https://doi.org/10.1111/j.0022-2879.2007.00034.x

Eichler, S., & Lähner, T. (2017). Regional, individual and political determinants of FOMC members' key macroeconomic forecasts. *Journal of Forecasting*, *37*(1), 119–132. https://doi.org/10.1002/for.2481

Frieder, S., Pinchetti, L., Chevalier, A., Griffiths, R.-R., Salvatori, T., Lukasiewicz, T., Petersen, P. C., & Berner, J. (2023). Mathematical Capabilities of ChatGPT. https://doi.org/10.48550/ARXIV.2301.13867

Gerlach-Kristen, P. (2009). Outsiders at the Bank of England's MPC. *Journal of Money, Credit and Banking*, *41*(6), 1099–1115. https://doi.org/10.1111/j.1538-4616.2009.00248.x

Hansen, A. L., & Kazinnik, S. (2023). Can ChatGPT Decipher Fedspeak? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4399406

Jegadeesh, N., & Wu, D. (2017). Deciphering Fedspeak: The Information Content of FOMC Meetings. *Monetary Economics: Central Banks - Policies & Impacts eJournal*. https://api.semanticscholar.org/CorpusID:30415687

Ko, H., & Lee, J. (2023). Can Chatgpt Improve Investment Decision? From a Portfolio Management Perspective. https://doi.org/10.2139/ssrn.4390529

Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *SSRN Electronic Journal.* https://doi.org/ 10.2139/ssrn.4412788

Meade, E. E., & Sheets, D. N. (2005). Regional Influences on FOMC Voting Patterns. *Journal of Money, Credit and Banking, 37*(4), 661–677. http://www.jstor.org/stable/3839057

Reeves, R., & Sawicki, M. (2007). Do financial markets react to Bank of England communication? *European Journal of Political Economy, 23*(1), 207–227. https://doi.org/10.1016/j. ejpoleco.2006.09.018

Tucker, P. (2008). Remarks on "Making Monetary Policy by Committee", Quarterly Bulletin 48 (No. 3). https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2008/ quarterly-bulletin-2008-q3.pdf

Yang, K.-C., & Menczer, F. (2023). Large language models can rate news outlet credibility. https://doi.org/10.48550/ARXIV.2304.00228