# EconSentGPT: A universal economic sentiment engine?

Aref Mahdavi Ardekani[1], Julie Bertz[1], Michael Dowling*[1, 2] and Suwan(Cheng) Long[3,4]

[1]DCU Business School, Dublin City University, Ireland
[2]Rennes School of Business, France
[3]Trinity Business School, Trinity College Dublin, Ireland
[4]Judge Business School, Cambridge University, United Kingdom

**Abstract**

We present EconSentGPT, an economic sentiment prediction model based on a fine-tuned version of the recently-launched artificial intelligence language model, ChatGPT. To assess the model's effectiveness, we analyze a sample of US economic news and a multi-language dataset of European Central Bank Monetary Policy Decisions. Our findings demonstrate that EconSentGPT's sentiment classification ability aligns well with a prominent English-language economic sentiment model, surpasses an established alternative machine learning model, and is capable of predicting sentiment across various languages. Consequently, we offer preliminary evidence that advanced large-language AI models can facilitate flexible and contextual economic sentiment determination, transcending language barriers.

**JEL codes:** C22; F47
**Keywords:** ChatGPT; large language models; economic sentiment; monetary policy

*Michael Dowling: DCU Business School, Dublin City University, Glasnevin, Dublin 9, Ireland. Email: michael.dowling@dcu.ie

# 1 Introduction

Economic sentiment understanding, particularly across languages, is vital yet challenging for economists. Contextual understanding is key for sentiment determination but difficult in practice. Most models also only focus on the English language, leaving other languages underserved. This paper develops an economic sentiment model harnessing the power of a new generation of large-language models, specifically ChatGPT[1], to identify economic sentiment in multiple languages.

Textual analysis has gained traction in studying economic phenomena (Ash and Hansen, 2022), but standard text analysis can encounter difficulties in capturing language nuances in economic contexts. Enter ChatGPT, an artificial intelligence language model developed by OpenAI that has revolutionized the field of natural language models with claims of demonstrating 'sparks' of artificial general intelligence (Bubeck et al., 2023). The model shows strong potential in contextual analysis (Korinek, 2023) and text simulation (Dowling and Lucey, 2023), offering new opportunities for economic research.

Our GPT application addresses economic sentiment, a rapidly growing research area, with recent studies such as Barbaglia et al. (2022) and Picault et al. (2022) emphasizing the importance of contextual understanding. However, existing approaches require cumbersome human labelling and struggle with language context determination and multi-language applicability (Degani and Tokowicz, 2010).

We exploit GPT's contextual understanding abilities (Tur and Traum, 2022) and multi-language capabilities (Hendy et al., 2023). A challenge is GPT's generalization

[1] We use the term ChatGPT for the product, GPT for the underlying model, and EconSentGPT for our developed model.

tendency when domain knowledge is limited (Wang et al., 2020), but fine-tuning with private data mitigates this issue. We, therefore, develop EconSentGPT, a fine-tuned GPT model trained on economic sentiment classification, demonstrating its proficiency in identifying sentiment across languages.

The core contribution of our study is a novel first investigation of large-language AI models for economic sentiment analysis, showing the feasibility of predicting sentiment from structured (central bank policy announcements) and less-structured (economics news) textual data. Our EconSentGPT model performs well contextually and across languages, revealing that fine-tuning a GPT model in English language sentiment examples is sufficient for applying knowledge to other languages' unseen text. This finding has significant implications for creating a universal economic sentiment engine, reducing existing research biases favouring the English language.

## 2 Methodology

We test various EconSentGPT versions to determine GPT performance in predicting economic sentiment. Our first study focuses on the English language, utilizing a dataset of 2,226 New York Times articles during the year 2022 that mention domestic US inflation. With this data, we examine ability to predict sentiment within a single language.

Following the methodology by Barbaglia et al. (2022), we first calculate economic sentiment scores for sentences across all articles containing the term "inflation". These scores become our reference scores to compare our model accuracy. Next, using OpenAI's ADA GPT3 model, we create a fine-tuned GPT model[2] called Econ-

[2]https://platform.openai.com/docs/guides/fine-tuning

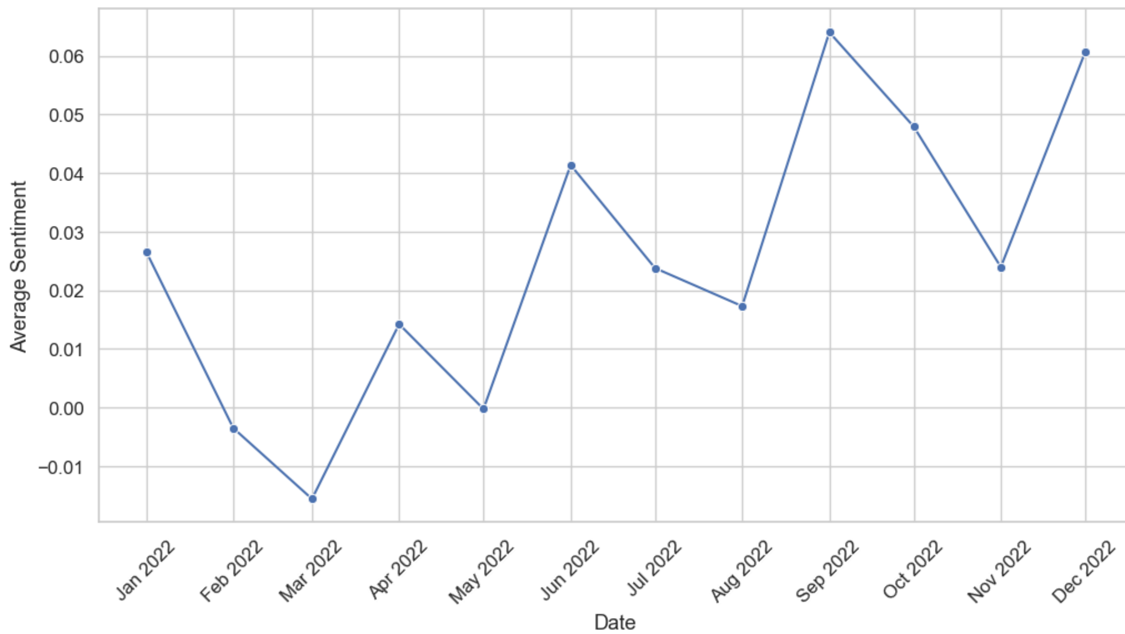
SentGPT_NYT, training it on relevant sentences and sentiment scores from January to June 2022.

To test EconSentGPT_NYT, we input unseen sentences containing "inflation" from July to December 2022 and compare the model's sentiment scores to Barbaglia et al. (2022) scores. In total, 600 sentences were tested, with a random sample of 100 sentences per month. We evaluate accuracy using Spearman's Rank Correlation, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Average Error (MAE). We also train a text-based linear regression machine learning model using the same dataset for comparison of results with the EconSentGPT_NYT model, following the approach by (Hasan et al., 2019). The linear regression model uses a TF-IDF vectoriser to convert word usage to numerical counts and then fits a linear regression model based on this data. This allows us to understand if the EconSentGPT_NYT findings are simply capturing the general benefits of machine learning for text analysis in economics, or if there are additional accuracy benefits to EconSentGPT_NYT.

Our second study develops EconSentGPT_ECB1 and EconSentGPT_ECB2, focusing on GPT's multi-language ability. The core dataset is monthly European Central Bank (ECB) Monetary Policy Decisions statements in English, French, German, Spanish, and Portuguese. We use English-language statements from 2017 to 2021 to train both ECB models, calculating economic sentiment scores following the Barbaglia et al. (2022) method. Using OpenAI's DaVinci GPT model, we create an initial fine-tuned model, EconSentGPT_ECB1, based on this training data.

The next model, EconSentGPT_ECB2, addresses the limited number of relevant ECB statement sentences in our training dataset by also adding in the NYT economic

Figure 1: Economic sentiment towards US domestic inflation for 2022 calculated from New York Times news articles




news training dataset, assessing whether the additional data improves prediction. To test the multi-language models, we predict sentiment scores for unseen 2022 sentences containing the non-English term for "inflation" and compare them to the English-language sentiment scores, following Barbaglia et al. (2022). We rely on the model's ability to interpret different language usage contexts without any fine-tuning for non-English languages.

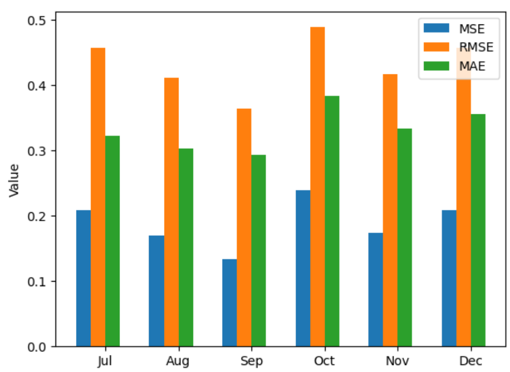
## 3 Results

We start with the economics news study. Figure 1 displays the monthly average economic sentiment towards inflation using the Barbaglia et al. (2022) method. Sentiment is notably negative at the year's beginning and remains subdued throughout.

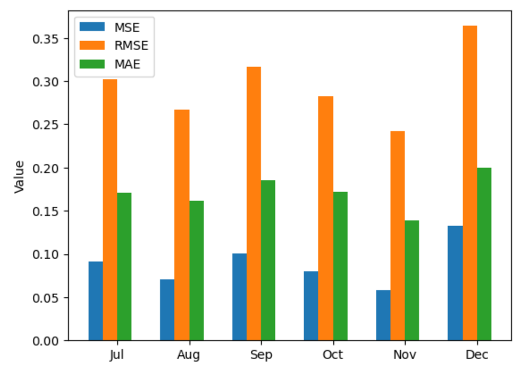
Table 1: US inflation sentiment prediction

| | Overall | Month | | | | | |
|---|---|---|---|---|---|---|---|
| | | Jul | Aug | Sep | Oct | Nov | Dec |
| *Linear regression model* | | | | | | | |
| Correlation | 0.28 | 0.27 | 0.28 | 0.52 | 0.10 | 0.37 | 0.24 |
| MSE | 0.19 | 0.21 | 0.17 | 0.13 | 0.24 | 0.17 | 0.21 |
| RMSE | 0.43 | 0.46 | 0.41 | 0.36 | 0.49 | 0.42 | 0.46 |
| MAE | 0.33 | 0.32 | 0.30 | 0.29 | 0.38 | 0.33 | 0.36 |
| *EconSentGPT_NYT model* | | | | | | | |
| Correlation | 0.61 | 0.62 | 0.58 | 0.63 | 0.59 | 0.75 | 0.53 |
| MSE | 0.09 | 0.09 | 0.07 | 0.10 | 0.08 | 0.06 | 0.13 |
| RMSE | 0.30 | 0.30 | 0.27 | 0.32 | 0.28 | 0.24 | 0.36 |
| MAE | 0.17 | 0.17 | 0.16 | 0.19 | 0.17 | 0.14 | 0.20 |

Table reports US inflation economic sentiment model accuracy findings. Reported are the results of a linear regression model with TF-IDF vectorisation of text (top panel), and a fine-tuned GPT model - EconSentGPT_NYT (bottom panel). Training for both models is on news stories mentioning US inflation in the New York Times from January to June 2022. Testing is on news stories between July 2022 and December 2022. Correlation is Spearman's Rank Correlation Test, MSE is Mean Squared Errors, RMSE is Root Mean Squared Errors. MAE is Mean Average Errors.




Linear Regression accuracy




EconSentGPT accuracy

Figure 2: Comparison of Linear Regression vs GPT accuracy for US economics news sentiment prediction

Table 1 presents the EconSentGPT_NYT model findings for this dataset, with Figure 2 providing a visual representation. The table's top panel shows the text-based linear regression machine learning model results, while the bottom panel displays

EconSentGPT_NYT findings. Overall and monthly accuracy scores for the six tested months are reported, revealing no decrease in accuracy over time.

Focusing on overall results, EconSentGPT_NYT outperforms the linear regression model on all measures. The Spearman Rank Correlation is 0.61, indicating good correlation with the underlying economic sentiment model, and the MAE score demonstrates an average 0.17 difference between SentGPT and the underlying model for average sentiment scores.

Examining the individual datapoints reveals skewness in score accuracy, with over 40% of individual sentiment scores exhibiting less than a 0.05 absolute difference between the Barbaglia et al. (2022) method and EconSentGPT_NYT. Removing the top 10% of score differences would increase the correlation to 0.79. This outlier finding suggests the model could be iterated for closer correlation by examining sentences with large differences and providing additional training examples to rectify underlying errors. GPT fine-tuning allows the addition of new training examples.

Our judgement of the ability of the EconSentGPT_NYT model to contextually understand economic sentiment in sentences relies primarily on prior studies showing that it does contextually understand text (Tur and Traum, 2022). However, our EconSentGPT_NYT model itself only returns sentiment scores. By way of additional investigation, albeit constrained by article space, we ask GPT to explain the reasoning it would use for a selection of economic sentiment use case examples. These are reported in Table 2 and do appear to show contextual understanding. Ideally, future versions of the GPT model would allow a more definitive investigation of these aspects of model generation.

The multi-language study results are reported in Table 3, with Spearman corre-

Table 2: Inferring contextual understanding by SentGPT

| Sentiment | Input sentence | GPT contextual explanation |
|---|---|---|
| Negative (-0.97) | 'shelter inflation picked up sharply in august, with a rent measure climbing at its fastest pace since the 1980s.' | 'Rent prices increasing rapidly suggests a negative outlook for inflation.' |
| Neutral (0.00) | 'the decrease on overall inflation came as some durable goods, like household appliances, televisions and luggage, became cheaper, and as prices for financial services and insurance eased.' | 'The sentiment score of 0 likely indicates that the overall impact of the decrease in inflation was neutral, as the decrease in prices for durable goods and financial services was balanced out.' |
| Positive (+0.82) | 'bank officials have said they are determined to bring inflation down to their 2 percent target and will use higher interest rates to do so.' | 'The score likely reflects the model's positive sentiment towards the bank's determination to bring inflation down to their target.' |

Table reports GPT3 output explaining the context of sample economic sentiment score determination from the EconSentGPT_NYT model. The actual sentiment score for each sentence is in brackets in the Sentiment column. The prompt to GPT includes the sentence and the prompt and the request "what is the likely reason for such a score based on your understanding of GPT3 reasoning?"

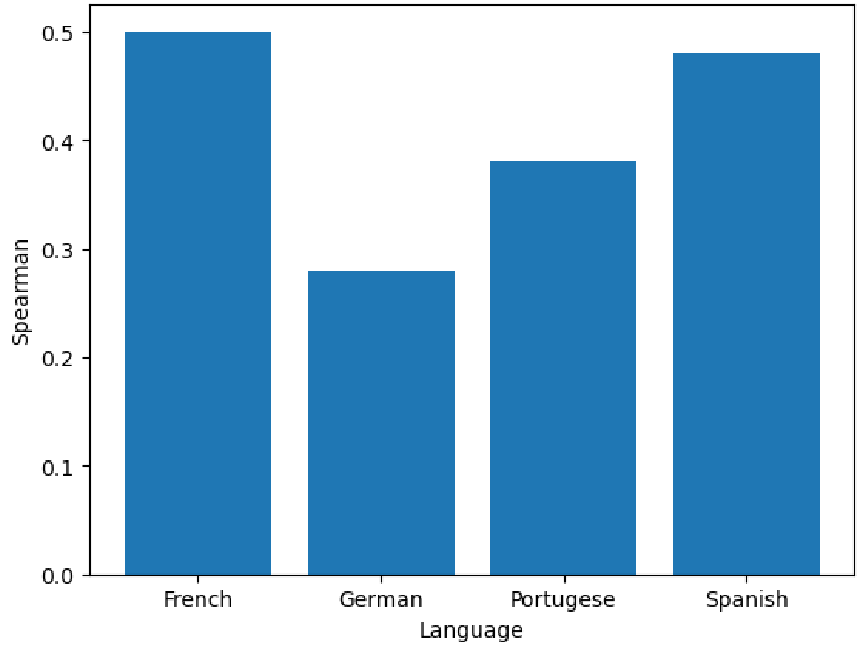
Table 3: ECB multi-language sentiment prediction

| | French | German | Portugese | Spanish |
|---|---|---|---|---|
| *EconSentGPT_ECB1 model* | | | | |
| Correlation | 0.22 | 0.34 | 0.33 | 0.36 |
| MSE | 0.14 | 0.17 | 0.15 | 0.19 |
| RMSE | 0.38 | 0.41 | 0.39 | 0.44 |
| MAE | 0.23 | 0.22 | 0.22 | 0.22 |
| *EconSentGPT_ECB2 model* | | | | |
| Correlation | 0.50 | 0.28 | 0.38 | 0.48 |
| MSE | 0.19 | 0.21 | 0.19 | 0.17 |
| RMSE | 0.43 | 0.46 | 0.44 | 0.41 |
| MAE | 0.31 | 0.36 | 0.36 | 0.32 |

Table reports ECB Monetary Policy Decisions economic sentiment model accuracy findings for unseen non-English ECB sentences containing translated versions of the keyword 'inflation' for 2022. All variables as defined in Table 1. Top panel shows results for EconSentGPT_ECB1 - a fine-tuned GPT model based only on prior ECB English-language sentence sentiment. Bottom Panel shows results for EconSentGPT_ECB2 - a model trained on both prior ECB English-language sentence sentiment and New York Times inflation news sentence sentiment.

lation scores visualized in Figure 3 using the EconSentGPT_ECB2 model. Results are reported for French, German, Spanish, and Portuguese.

Figure 3: Spearman correlations for ECB non-English language economic sentiment prediction




EconSentGPT_ECB2's performance is impressive considering its lack of exposure to non-English training text, suggesting the limited training examples for EconSentGPT_ECB1 hindered performance. However, EconSentGPT_ECB2's accuracy is lower than the English-to-English sentiment prediction in the economic news study. The MAE for EconSentGPT_ECB2 indicates an average 0.34 score variation from the underlying sentiment model.

With EconSentGPT_ECB2, French and Spanish both yield around 0.50 for Spearman's correlation, while German and Portuguese are lower. This discrepancy may result from GPT's higher use of training data in Spanish and French, the 4th and 5th most spoken languages globally, compared to Portuguese (9th) and German (12th)[3]. This observation bodes well for multi-language sentiment prediction from more ad-

[3]Source: https://www.ethnologue.com/insights/ethnologue200/

vanced GPT models, such as GPT4, when they allow fine-tuning.

## 4 Conclusions

This initial exploration of large-language AI models' effectiveness in economic sentiment prediction demonstrates their capability for accurate sentiment understanding, essential for policymakers and industry. Our EconSentGPT models offer insight into how we can derive contextual sentiment within a single language and understand economic sentiment across languages, promising a new era of global sentiment analysis.

We acknowledge limitations in this early-stage research. Our EconSentGPT models are intended only as initial explorations, and we could only fine-tune using the less-powerful GPT3 model since GPT4, which is known to significantly outperform prior GPT models in contextual analysis and non-English languages (OpenAI, 2023), does not currently allow fine-tuning. Additionally, we rely on an underlying sentiment generative model for our training data as GPT cannot yet understand economic sentiment independently. Future studies might also confirm the single-language economic sentiment prediction ability for languages other than English.

This study's broader benefit is moving the conversation beyond speculation about large-language models' potential capabilities, illustrating what they can genuinely accomplish in terms of economics research. We can now, for example, start to envision a universal real-time economic sentiment engine that reduces barriers to global economic understanding, and this is just the beginning.

## References


Ash, E. and Hansen, S. (2022). Text algorithms in economics. Annual Review of Economics.

Barbaglia, L., Consoli, S., and Manzan, S. (2022). Forecasting with economic news. Journal of Business & Economic Statistics, pages 1–12.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712.

Degani, T. and Tokowicz, N. (2010). Semantic ambiguity within and across languages: An integrative review. Quarterly Journal of Experimental Psychology, 63(7):1266–1303.

Dowling, M. and Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. Finance Research Letters, page 103662.

Hasan, M. R., Maliha, M., and Arifuzzaman, M. (2019). Sentiment analysis with NLP on Twitter data. In 2019 international conference on computer, communication, chemical, materials and electronic engineering (IC4ME2), pages 1–4. IEEE.

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., and Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. arXiv preprint arXiv:2302.09210.

Korinek, A. (2023). Language models and cognitive automation for economic research. National Bureau of Economic Research Working Paper 30957.

OpenAI (2023). GPT-4 Technical Report.

Picault, M., Pinter, J., and Renault, T. (2022). Media sentiment on monetary policy: Determinants and relevance for inflation expectations. Journal of International Money and Finance, 124:102626.

Tur, A. and Traum, D. (2022). Comparing approaches to language understanding for human-robot dialogue: An error taxonomy and analysis. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 5813–5820.

Wang, Y., Yao, Q., Kwok, J. T., and Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys, 53(3):1–34.