Evidence on Inflation Expectations Formation Using Large Language Models

By ALI ZARIFHONARVAR*

This paper integrates generative AI with survey experiments to explore inflation expectations in Large Language Models. It shows that knowledge domain significantly affects expectations; LLMs accessing relevant economic documents exhibit lower variance. LLMs generally predict higher inflation than actual rates, reflecting human like behavior. Information treatments, particularly involving forward guidance, influence LLMs' inflation expectations almost perfectly. Additionally, the introduction of a persona in the prompts results in partisan expectations, reflecting patterns observed in human surveys. The paper also examines the influence of model selection, highlighting how it significantly shapes the outcomes.

JEL: C90, E27, E31, E58, E71, D84
Keywords: Large Language Models, Inflation Expectation, Information, GPT, Survey Experiment

The use of Generative Artificial Intelligence (GAI), particularly with the increasing usage of Large Language Models (LLMs), signals a great shift in the landscape of economic research. These models have the potential to assist in a wide range of tasks for economists (Charness, Jabarian and List, 2023; Korinek, 2023). As we begin to integrate LLMs into economics, it has become clear that their impact extends beyond only supporting various tasks; they can enhance the methodologies used in economic research, suggesting new approaches for economists to conduct studies and develop new insights.

There are conflicting opinions about the efficacy of LLMs, highlighting the need for a comprehensive understanding of their behaviors, cognitive processes (Echterhoff et al., 2024), and underlying models. The ways in which LLMs think and operate differ from humans, and conducting experiments with AI agents or surveying these agents could offer a novel method to bridge these perspectives. By deploying LLMs in structured surveys and comparing their responses to those of human respondents, particularly in economic research, we can obtain a clearer, empirical understanding of their strengths, limitations, and the factors

* Indiana University Bloomington - alizarif@iu.edu
I would like to thank Daniela Puzzello for her invaluable help, comments and support throughout this project. I am also grateful for the comments and feedback from Ala Avoyan, Volodymyr Lugovskyy, Yuriy Gorodnichenko, Joon Y. Park, Benjamin Manning, Sophia Kazinnik, Kay-Yut Chen, and all the participants at the ICD Departmental Seminar Series at IMF, NEXT seminar at Indiana University, the Hoosier Economics Conference (HEC), the Learning, Computational, and Bayesian Methods in Experimental Economics Workshop at Purdue University, and the NBER Behavioral Macroeconomics Research Boot Camp.

influencing their outputs (Park et al., 2023; Brand, Israeli and Ngwe, 2023).

Moreover, the study of Large Language Models is crucial not only for their potential applications but also for the insights they can provide into human cognition and decision-making (Zhu, Yan and Griffiths, 2024). LLMs are designed to process and generate natural language in a manner that closely mimics human language. By understanding how these models interpret information, reach conclusions, and respond to various prompt, we can gain valuable insights into their underlying mechanisms and compare it to human decision-making processes. This understanding could potentially help us better comprehend how humans make decisions, form opinions, and respond to information in various contexts. Furthermore, by comparing the performance of LLMs with human subjects in experiments, we can identify the strengths and limitations of these models and assess their potential to augment or even replace human decision-making in certain domains (Tjuatja et al., 2023; Wu et al., 2023). This comparative analysis can shed light on the similarities and differences between human and AI, paving the way for more effective collaboration and integration of AI in various fields.

Previous research has utilized AI agents in game-theoretic scenarios to study decision-making processes (Horton, 2023; Brookins and DeBacker, 2023; Guo, 2023), simulating environments where these agents follow some rules to provide insights about strategic behavior. Building on this line of research, this study use these methods by enabling AI agents to access and incorporate external knowledge, thus improving the relevance and context-specificity of their responses to the query. This approach allows for a more realistic representation of how AI agents can process and utilize information in real-world scenarios, providing a deeper understanding of their potential applications.

Specifically, in the intersection of generative AI and economics, this paper presents an exploration into the applications of Large Language Models on inflation expectations surveys. Inflation expectations have become increasingly important due to their significant role in helping central banks design and communicate monetary policies more effectively(Coibion, Gorodnichenko and Weber, 2022). The relevance of subjective inflation expectations extends beyond household consumption and firm decisions and play a crucial role in financial markets and asset pricing as well. The anticipation of inflation affects central bank policies, which shape the dynamics of bond and stock markets (Bernanke and Kuttner, 2005). For instance, if inflation is expected to rise, central banks may increase interest rates to mitigate this rise, leading to lower bond prices and higher yields. This also impacts the stock market, as future corporate earnings are discounted based on these expectations, guiding investment decisions.

As AI continues to advance and play a significant role in economic and financial decision-making, understanding how AI agents form and update their expectations about key economic variables, including inflation, becomes crucial. LLMs can mimic these surveys, test them before actual deployment, and offer insights into how AI agents interpret key macroeconomic variables. The primary

goal of this research is twofold: first, to delve into the decision-making processes of LLMs, particularly in the area of inflation expectation formation by giving them access to different knowledge sources, and second, to assess the influence of information provision treatments on their responses. This methodology not only provides a detailed analysis of how AI agents process and integrate new information, mirroring the way humans form their expectations in similar contexts, but also incorporates a randomized experiment to examine how different information provision treatments affect their prior beliefs. By investigating the impact of various information sources and treatments on LLMs' inflation expectations, this study aims to uncover the factors that shape their decision-making processes and draw parallels with human behavior. This approach helps our understanding of AI's capabilities and limitations in the economics of expectations and behavioral macroeconomics and also shows how AI can contribute to the development of new methodologies in experimental economics.

To achieve these objectives, I introduce a new experimental design that utilizes custom-built LLMs based on a few variations of GPT-4[1] model. I employ various text corpora to enhance the retrieval capabilities of these custom GPTs, such as minutes from Federal Open Market Committee (FOMC) meetings. By incorporating these domain-specific knowledge sources, the LLMs have the potential to generate more informed and context-aware responses to survey questions. Furthermore, this paper explores the impact of defining some basic persona in the prompt for AI agents and its influence on the results of the survey experiments. By incorporating persona attributes, such as political partisanship, into the prompts, this study investigates how these factors change the responses generated by the LLMs. This analysis provides valuable insights into the role of individual characteristics and biases in shaping the outputs of AI agents, highlighting the importance of considering these factors when designing and interpreting AI-driven experiments. Additionally, this paper examines the effect of model selection on the survey results. By comparing the performance of different LLMs and their underlying architectures, this study sheds light on the impact of model choice on the accuracy, consistency, and interpretability of the generated responses. The findings indicate substantial consistency between different sessions of the experiment using the same model; however, variations in results can occur when different models are employed, particularly if the training data or underlying architectures are altered. This comparative analysis can inform the selection of appropriate models for specific economic research questions.

This paper not only contributes to the rapidly growing body of literature on AI and economics but also provides practical insights into the future of AI-assisted economic forecasting and its potential impacts on policy-making and individual financial decisions. By exploring the intersection of generative AI and inflation expectations, this study aims to understand the potential of AI in enhancing our understanding of their thought process.

[1]Generative Pre-trained Transformer

I. Literature Review

Recent developments in the field of Generative Artificial Intelligence, especially in the context of Large Language Models, have shown a significant turning point in economics, introducing innovative tools for research. This paper contributes to two main strands of the literature: the experimental study of AI agents and LLMs in economics, and the formation of inflation expectations. While some studies attempt to connect these two domains (Faria-e Castro and Leibovici, 2023; Bybee, 2023), this paper delves deeper into the decision-making processes of LLMs by examining their retrieval capabilities and assessing the impact of information treatments on AI agents in an experimental setup. By focusing on these aspects, this study aims to provide a more comprehensive understanding of how LLMs processes integrate new information, and how their inflation expectations are shaped by the knowledge sources they access and the information they receive.

A. Economics and LLMs

It is now obvious that the introduction of large language models, including products like ChatGPT, has significantly impacted various sectors, including the field of economics research. Recent studies have shown the potential of LLMs in simulating complex economic scenarios and decision-making processes (Akata et al., 2023; Heydari and Lorè, 2023). Additionally, there is growing interest in understanding the broader implications of generative AI on economics research (Charness, Jabarian and List, 2023; Korinek, 2023).

For instance, Horton (2023) investigates how LLMs can function as simulated economic agents, drawing parallels to the concept of *'homo economicus'* in economic theories. These LLMs, equipped with specific preferences, information, and endowments, are tested in a variety of scenarios to examine their behavior. Inspired by classic games, these experiments have shown that LLMs can emulate human-like behaviors, yielding results qualitatively similar to those of human subjects. In a related study, Brookins and DeBacker (2023) explores the tendencies of GPT agents towards fairness and cooperation. It engages GPT models in games like the dictator game and the prisoner's dilemma, finding that the AI often chooses strategies that prioritize fairness over optimal outcomes, displaying some level of human-like behavior. This line of research even goes further and suggests a fully automated social science that could use LLMs as scientists who could create hypotheses, design experiments, serve as subjects, and analyze data (Manning, Zhu and Horton, 2024; Batista and Ross, 2024).

Recent studies also have shown significant advances in the application of LLMs in the realm of macroeconomics. For example, Li et al. (2023) introduce a framework using large language models to create agents for macroeconomic simulations. These agents are uniquely designed to exhibit human-like decision-making and adaptability in complex economic scenarios. Faria-e Castro and

Leibovici (2023) explore the proficiency of LLMs in generating inflation forecasts, showing that these AI tools can outperform traditional methods in accuracy and efficiency. They examine the capacity of Google AI's PaLM to produce conditional inflation forecasts during 2019-2023. Comparing these forecasts with those from the Survey of Professional Forecasters, the study finds LLM forecasts generally more accurate, with lower mean-squared errors. However, there are some critiques to this approach as they manually limit the AI knowledge to predict past inflation. Similar to this research, Bybee (2023) presents a novel approach to generating economic expectations by applying large language models to historical news data. It shows that this method closely aligns generated expectations with existing survey measures for different indicators.

This study builds upon the insights provided by existing literature, exploring how large language models integrate and apply knowledge domains to enhance their decision-making processes. By focusing on the provision of new information, this research also delves into the process of updating beliefs, offering a deeper understanding of the black box of LLMs. Through the selection of different models and prompts, my research fills some gaps in the literature by demonstrating how these factors influence the consistency and reliability of LLM outputs in survey experiments. Furthermore, the exploration of response consistency across multiple sessions contributes significantly to our understanding of LLMs, offering a more robust assessment of their responses.

B. Inflation Expectation Formation

Inflation expectations play a crucial role in shaping economic behavior and policy. Understanding how these expectations are formed and updated is important for both policymakers and researchers. This area of study delves into the psychological and informational aspects of how people and firms understand and predict future inflation. Inflation expectations are a particularly suitable candidate for study using survey experiments with LLMs for several reasons. First, measuring and influencing inflation expectations is challenging due to their complex formation process and heterogeneity across economic agents (Coibion et al., 2020; Weber et al., 2022). Second, inflation expectations have far-reaching consequences for economic decisions, such as consumption, investment, wage setting, and so they are really important (Coibion, Gorodnichenko and Ropele, 2020). Third, central banks increasingly recognize the importance of inflation expectations for the conduct of monetary policy (Coibion et al., 2020; Pfajfar and Žakelj, 2018). By employing LLMs to analyze vast amounts of economic data and simulate the formation of inflation expectations, we can gain valuable insights into the underlying dynamics and potentially enhance our ability to predict and manage these expectations.

Recent research has shed light on the complex processes involved in the formation of inflation expectations and why surveys could give us important insights (D'Acunto et al., 2024; D'Acunto and Weber, 2024). In this line, Candia, Coibion

and Gorodnichenko (2020) examines the impact of how households form their expectations and how these expectations influence their decisions. Their findings suggest that inflation expectations can drive changes in consumer spending and firm price-setting behavior even in the absence of actual changes in inflation, highlighting the role of subjective beliefs and information frictions in shaping economic outcomes. D'Acunto et al. (2021) provide further evidence on the formation of consumer inflation expectations, showing that the prices of frequently purchased grocery items have a disproportionate influence on households' expectations. This leads to potential biases and excessive sensitivity to transitory price fluctuations, underscoring the role of personal consumption experiences and the salience of specific prices in shaping inflation beliefs.

The subjective nature of inflation expectations, as discussed by Weber et al. (2022) and Weber (2022), points to a consistent upward bias and strong influence from prices of specific goods. Weber et al. (2022) show that households' inflation expectations are systematically higher than actual inflation, with the bias being more pronounced for low-income and less-educated households. They attribute this finding to the salience of price changes for frequently purchased items and the role of personal experiences in shaping beliefs. As shown in section III.A LLMs also exhibit higher expectations than actual inflation.

Coibion et al. (2023) explore the effectiveness of forward guidance in managing inflation expectations, finding that households respond more to short-term interest rate information than to long-term policy goals. This indicates that central banks might need to tailor their communication strategies to better anchor expectations and improve policy effectiveness. These results highlight the importance of carefully designing central bank communications to effectively manage public inflation expectations, considering the diverse ways in which different economic agents interpret and react to inflation-related information.

Additionally, the link between public perceptions of inflation and the use of large language models opens new paths for research. Studies like Stantcheva (2024); Binetti, Nuzzi and Stantcheva (2024); Afrouzi et al. (2023, 2024) show that people's views on inflation often diverge from the economic indicators and goals set by central banks. These insights reveal how personal experiences and socioeconomic status shape how people perceive and respond to inflation. As explored by Jiang et al. (2024), the inconsistent effects of inflation expectations across demographic groups and their relationship with spending behavior illustrate the complex dynamics of household inflation expectations that LLMs need to capture. By using LLMs to analyze vast amounts of text data and model these perceptions, this research aims to uncover the complex patterns of how people form their views on inflation.

As evidenced by the current literature, the intersection of LLMs and inflation expectation formation presents a promising avenue for future research. The complex nature of inflation expectations, their importance for the economy, and the challenges in measuring and managing them make this an ideal area for

applying new methods such as using LLMs. In this domain, my contribution is to expand the understanding of how AI-driven models, particularly LLMs, can enhance the prediction and analysis of inflation expectations. By integrating LLMs with economic related text data, this paper aims to show how these models process information, form expectations, and make decisions related to inflation. Furthermore, I explore how the outputs of these AI models compare to and potentially influence the expectations and decisions of human subjects. This approach can offer valuable insights into the complex dynamics of inflation expectation formation and contribute to the development of more effective policy tools for managing these expectations.

II. Experimental Design

In this experiment, I conducted a survey using GPT-4 Turbo[2] with Assistants API[3]. The aim was to assess the influence of different knowledge domains on inflation expectation formation and to observe the impact of information treatments. As shown in Table 1, I categorized the knowledge domains into six distinct types: no retrieval (N), which does not have any additional knowledge beyond the pre-trained GPT-4 data; placebo (W), using general information from Wikipedia texts as a baseline; economic reports of the president (E), providing a broad view of the US economic status; FOMC minutes from 2018 (M0) and FOMC minutes from 2023 (M1), to compare the difference between old and recent monetary policy status; and FOMC minutes from 2020 to 2023 (M2), offering a wider range of data on monetary policy. This structure allows me to analyze the impact of varying types of economic knowledge on agents' expectations.

TABLE 1—KNOWLEDGE DOMAINS IN SURVEY EXPERIMENT

Code	Description
N	No Retrieval
W	Placebo (Wikipedia Texts)
E	Economic Reports of the President
M0	FOMC Minutes (only 2018)
M1	FOMC Minutes (2023)
M2	FOMC Minutes (2020-2023)

Note: This experiment, evaluates the impact of different information retrieval contexts. It contrasts no retrieval with retrieval, non-economic with economic texts, and texts on monetary policy versus broader economic content. It also examines the influence of the historical context window.

The way the survey on LLMs works is similar to an actual survey with humans, as shown in Figure 1. I utilize Retrieval-Augmented Generation (RAG) in

[2]GPT-4 Turbo is an upgrade to GPT-4 that excels at handling long texts and following detailed instructions, making it suitable for tasks such as finding and using information in large documents. The cut-off date of training set at the time of the experiment was December 2023.

[3]OpenAI's Assistants API are a feature that allows us to create specialized GPTs with different capabilities, including document retrieval.

this approach because it enhances the language model's responses with knowledge retrieved from external sources during the survey. This process allows the model to access and incorporate relevant data beyond its initial training set, leading to more informed and accurate responses. Unlike fine-tuning, which adjusts a model's internal parameters to improve performance on a specific task, RAG actively pulls in external data to better answer questions, providing a significant advantage in keeping the model's responses up-to-date and reliable. The RAG process can be showed as follows (Gao et al., 2023):

$$p(x|y) = \sum_{z \in \text{Retrieve}(y)} p(x|y,z) \cdot p(z|y) \tag{1}$$

where y is the input, x is the output, and z is the retrieved information from the external knowledge source. This expression shows how RAG combines the likelihood of generating x given the input y and the retrieved information z, weighted by the probability of retrieving z given y. As shown in Figure 1, RAG operates with two prompts instead of one. The first prompt instructs the LLM on how to retrieve information, while the second prompt is the main question.

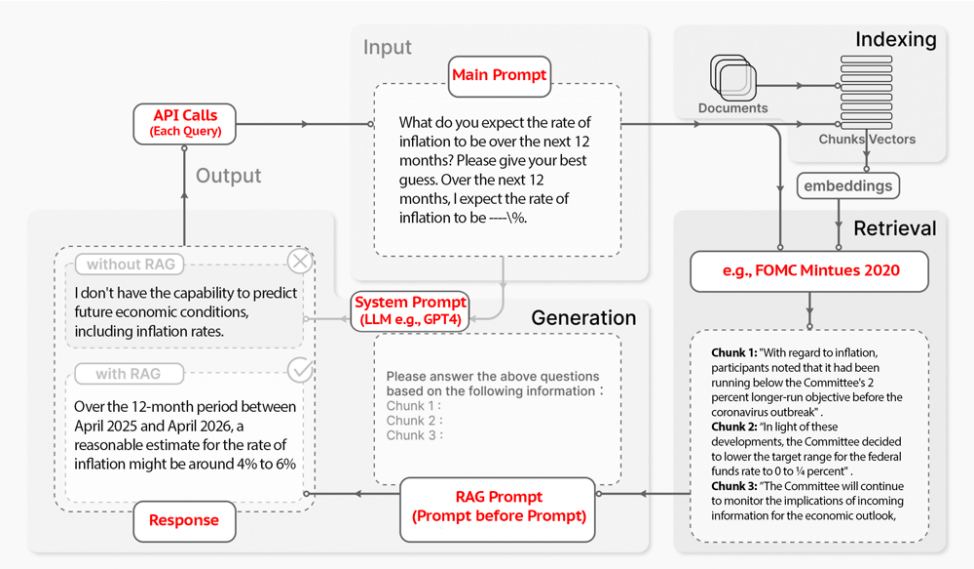


FIGURE 1. AN ILLUSTRATION OF HOW RAG WORKS

Note: This diagram is adapted from the example in Gao et al. (2023). It shows a survey experiment for AI agents, which is very similar to the structure of a traditional survey. The system prompt serves a similar role to survey instructions, the main prompts are the questions, and the RAG prompt resembles the thought process that, while typically implicit in human surveys, is explicitly requested from the AI to utilize the provided context in generating responses. All the prompts used in this survey are provided in Appendix A.A1.

In the experiment, I first asked AI agents a series of questions to elicit their initial expectations. After these initial questions, the AI agents were randomly assigned into one of six groups. The first group acted as a control, receiving no additional information, while the remaining five groups each were provided with additional information as new context to consider for their next questions. This design enables me to examine how specific information might influence the AI agents' prior beliefs. The list of treatments is provided in Table 2.

TABLE 2—CONTROL AND TREATMENT GROUPS

Group	Title	Information
T1	Control group	No additional information
T2	Placebo group	Population of the U.S. grew by 1.2% between 2020 and 2022.
T3	Current Federal Funds Rate	The interest rate set by the Federal Reserve, known as the Federal Funds Rate, is currently at 5.25%.
T4	Current Inflation	Over the last twelve months, the overall inflation rate in the economy as measured by the percentage change in a consumer price index has been 3.2%.
T5	Current Inflation + Longer Run Prediction	Over the last twelve months, the overall inflation rate in the economy as measured by the percentage change in a consumer price index has been 3.2%. One forecast at the Federal Reserve is that this inflation rate will be 2.8% on average over the next few years and in the longer-run.
T6	Current fixed-rate 30-year mortgage	The current average rate for fixed-rate 30-year mortgage is 7.5% per year.

Note: The treatments are derived from Coibion et al. (2023) and the data have been updated based on the Federal Open Market Committee's Summary of Economic Projections (SEP) for March 2024.

The initial questions and follow-up questions are shown in Table 3. I utilized both point predictions and probability distribution (density forecast) questions because point predictions provide a clear measure of the most likely outcome but often fail to capture the uncertainty about future conditions. Probability distribution questions overcome this by allowing respondents to express uncertainty across a range of outcomes, showing both the central tendency and the associated uncertainty (Manski, 2018). This method is crucial for capturing the full range of respondents' expectations in economic surveys (Boctor et al., 2024). This approach is also incorporated into our survey with AI agents, allowing us to compare how AI agents to these questions.

TABLE 3—SURVEY QUESTIONS

Q_id	Question
Q1_I	Last 12 month Point Prediction (Pre-Treatment)
Q2_I	Next 12 month Probability Distribution (Pre-Treatment)
Q3_I	Longer-run Probability Distribution (Pre-Treatment)
Q1_F	Next 12 month Point Prediction (Post-Treatment)
Q2_F	Longer-run Point Prediction (Post-Treatment)

Note: "I" indicates initial questions and "F" indicates follow-up questions. The full table of the questions is presented in Appendix A.A2.

The survey begins by assigning different groups to specific types of text or knowledge sources. After answering the first set of questions to elicit their prior beliefs, they randomly receive additional information. For example, one group is informed about the current Federal Funds Rate set by the Federal Reserve, while another receives detailed information on both current inflation and future forecasts (Table 2). After that, I ask each subject in each group the post-treatment questions to investigate how the provided information influenced their prior beliefs. This experimental design helps to determine not only the direct effects of specific economic data on expectations but also how varying types of information shape agents' changes in belief. Visually, the structure of the experiment can be shown in Figure 2[4]. The experiment was conducted in two phases. In the first phase, all six assistants were surveyed (without any follow-up or information treatments) to assess their comprehension and the way they accessed different knowledge sources. In the second phase, which focused on the impact of the information treatment, only Assistant N and Assistant M2 were used for the randomized experiment.

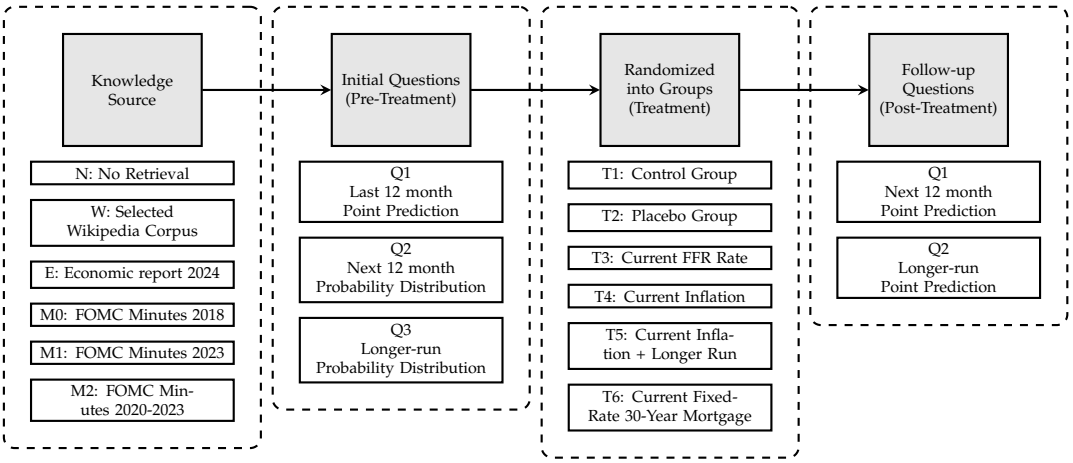


FIGURE 2. AN OVERVIEW OF THE SURVEY EXPERIMENT DESIGN

[4]To ensure that each AI agent keeps the memory of previous steps and the specific treatment assigned, although the final prompts were designed to elicit strictly numerical answers, I relaxed this strictness to test if they preserved memory of the previous sequence. This method also provides some insight into the thought processes of the AI agents on how they form expectations and update their prior beliefs. The results of some written responses are provided in Appendix A.A3.

III. Results

A. Preliminary Observations

As mention in the previous section, this experiment was implemented in two phases. The first phase, referred to as the no-information treatment, involved 600 subjects across six types of assistants. In this phase, three main questions were asked to assess AI agents' inflation expectations without any information provision. However, each type of assistant was associated with one of the knowledge sources previously mentioned. The questions included the perceived inflation rate for the past 12 months as a point prediction, the expected inflation density forecast (probability distribution) for the upcoming 12 months, and the density forecast for the period between two and three years into the future. The complete list of questions is provided in Appendix A.A2.

Table 4 presents the summary statistics for three questions. For question 1, I report the point predictions. For questions 2 and 3, I used the midpoint formula to calculate the mean. While many studies account for both with and without deflation responses in their calculations, the amount that was assigned to deflation by AI agents in this survey were negligible and the two approach almost gives same results. [5]

TABLE 4—SUMMARY STATISTICS FOR PHASE 1 (NO INFORMATION TREATMENT)

Assistant (Obs.)	Past Inflation				1 Year Ahead				3 Years Ahead			
	min	max	mean	std	min	max	mean	std	min	max	mean	std
N (100)	2	3.5	2.43	0.26	3.24	7.5	5.31	1	2.49	7.3	4.68	1.08
W (100)	0	11.58	2.72	2.4	0.94	14	5.96	2.27	1.7	14	5.17	1.92
E (100)	0	7.1	3.01	1.29	-1	14	4.93	2.2	-2.25	11	4.84	1.98
M0 (100)	1.7	3.6	2.07	0.21	1.08	14	4.36	1.9	1.2	10.12	4.54	1.8
M1 (100)	2	8	3.85	1.1	1	14	4.67	2.37	1.09	14	4.55	2.23
M2 (100)	1.7	6.85	3.86	0.86	1.55	8	4.72	1.5	1.32	11.66	4.97	1.8
All (600)	0	11.58	2.99	1.40	-1	14	4.99	2	-2.25	14	4.79	1.84

Figure 3 shows the distribution of inflation expectations reported by AI agents, compared to the actual observed US inflation of 3.48% at the time of the experiment. The gray bars represent the agents' perceptions of past inflation, averaging 2.99%, close to the actual rate. The black bars reflect short-run expectations, with a mean of 4.99%, and the red bars show long-run expectations, averaging

[5] In the Survey of Consumer Expectations run by the NY Fed, the median is usually reported instead of the mean due to the presence of outliers. In this experiment, the absence of significant outliers can be attributed to two factors: firstly, AI agents may follow the instructions more strictly, and secondly, their underlying pretrained models and data may have a more robust knowledge base about the notion and possibility of deflation

4.80%. We need to keep that in mind that these numbers are based on probability distribution questions. The following sections will explore how and why these expectations arise from the agents' models and knowledge cutoffs.

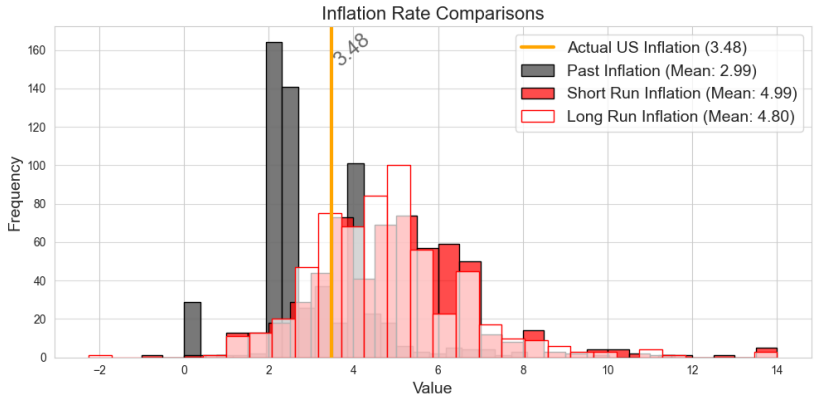


FIGURE 3. DISTRIBUTION OF RESPONSES FOR ALL THE SAMPLE

Figure 4 shows the inflation expectations from different assistants, categorized into past inflation (black), short-term inflation (gray), and long-term inflation (red). The perceptions differ across assistants. For instance, Assistant M2, who had access to the last four years of minutes, showed more consistent predictions. Interestingly, Assistant N, which had no additional knowledge, also provided precise answers. This suggests that while additional information can enhance prediction accuracy and reliability, it may also introduce greater variability in responses, acting as a double-edged sword.

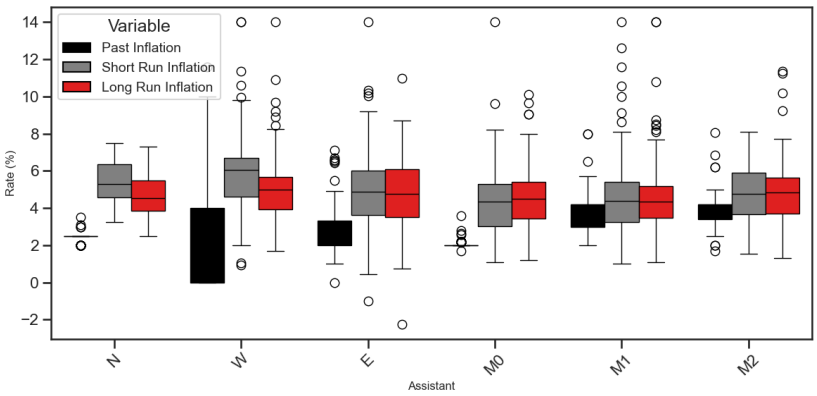


FIGURE 4. INFLATION/DEFLATION RATE BY ASSISTANT

Figure 5 illustrates the AI agents' expectations of inflation and deflation rates based on probability distribution questions. In the short run, the majority of responses fall within the 4% to 8% inflation range, indicating that the agents generally predict moderate inflation in the near future. For long-term forecasts, the data more frequently is in the 2% to 4% range, suggesting that agents expect inflation to moderate slightly over a longer period. These findings indicate a common expectation among the AI agents for moderate inflation, with a tendency toward lower inflation in the long term.

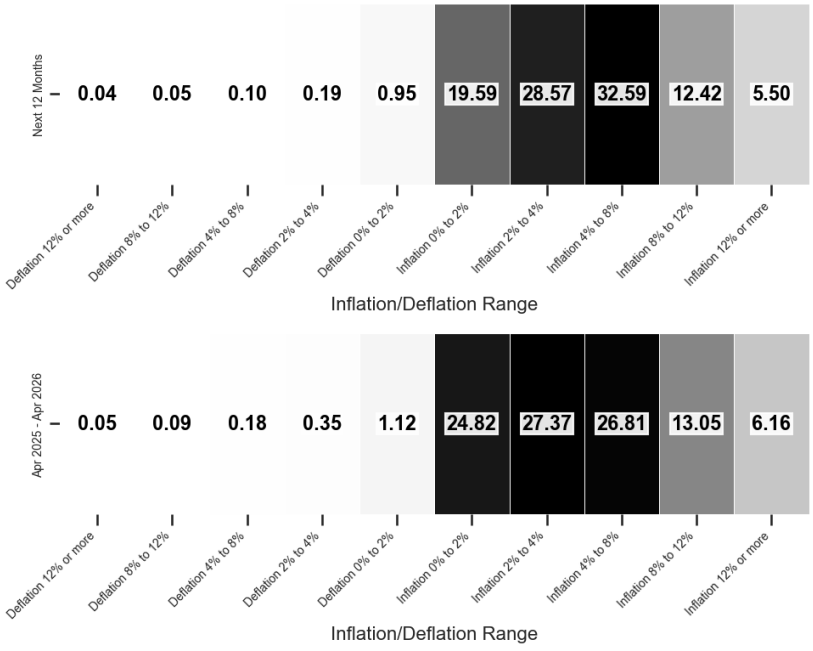


FIGURE 5. AVERAGE PROBABILITY DISTRIBUTION FOR ALL SUBJECTS

Figures 6 and 7 compare the AI-generated inflation expectations to the data from the Federal Reserve Bank of New York Consumer Expectation Survey. The histograms show that Assistant W, which only accessed general information from Wikipedia, had the most varied predictions, indicating less precise inflation forecasting when relying on less relevant data. On the other hand, Assistant M2, which had access to recent FOMC minutes, showed tighter responses, especially for long-term inflation predictions. This suggests that using more targeted, high-quality economic data leads to more consistent and reliable forecasting outcomes[6].

[6]The Kruskal-Wallis test confirmed significant differences across the assistants, indicating they indeed draw from different distributions, supporting the impact of underlying knowledge on forecast precision.

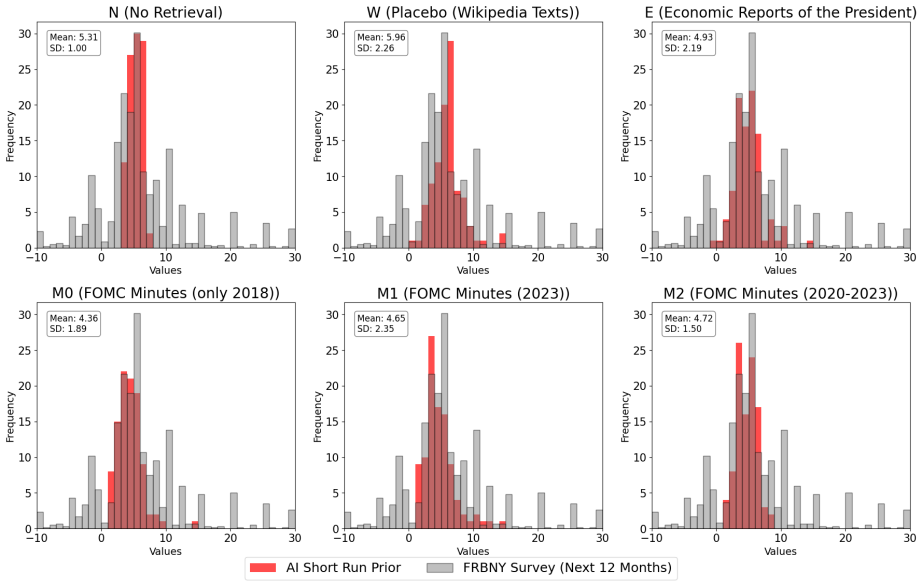


FIGURE 6. DISTRIBUTION OF NEXT 12 MONTH INFLATION EXPECTATION

Note: The gray area is the NY FED Consumer Expectation Survey (5-2023), scaled down for comparison.

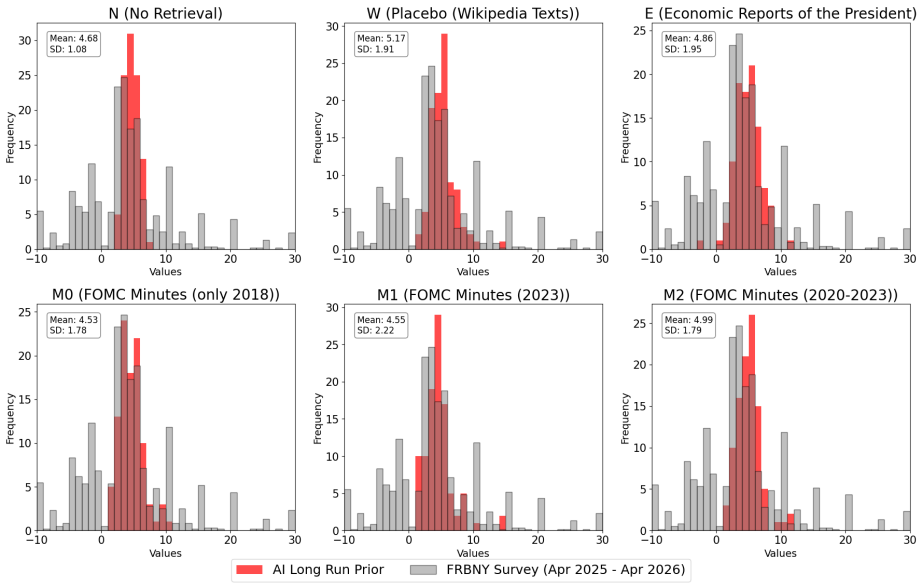


FIGURE 7. DISTRIBUTION OF NEXT 12 MONTH INFLATION EXPECTATION

Note: The gray area is the NY FED Consumer Expectation Survey (5-2023), scaled down for comparison.

We can also compare the inflation expectations from AI agents with those from well-known economic surveys such as the Survey of Professional Forecasters (SPF) and the Survey of Consumer Expectations (SCE). Table 5 presents both the mean and median inflation expectations over various time periods, allowing for a comparison between AI and human forecasts. Responses from human subjects to inflation expectation questions in the SCE exhibit a significant number of outliers and extremely high predictions for both inflation and deflation.

TABLE 5—COMPARISON WITH SCE AND SPF

1 Year Ahead	N	W	E	M0	M1	M2	All	SCE	SCE_P*	SPF
Mean	5.34	5.99	5.07	4.41	4.67	4.72	5.03	5.38	-3.66	3.8
Median	5.3	6.06	5.02	4.43	4.38	4.76	5	3.26	4.09	3.8

2-3 Years Ahead	N	W	E	M0	M1	M2	All	SCE	SCE_P*	SPF**
Mean	4.84	5.2	4.92	4.57	4.58	4.99	4.85	2.80	-2.42	2.5
Median	4.62	5.05	4.75	4.5	4.4	5	4.7	2.76	3.46	2.6

Note: SPF IS FROM SECOND QUARTER 2024 REPORT AND SCE IS FROM APRIL 2024. FOR MEAN CALCULATION OF SCE THE ONLY THE RANGE OF -10 TO 30 IS CONSIDERED.
* POINT PREDICTION
** 5 Year Forecast

It's important to note that the SCE does not report the mean due to the presence of extreme outliers. Even when restricting the range of responses to between -10 and 30, the mean for density forecasts remain negative, showing that the outliers have a huge impact on the average (Figure 8).

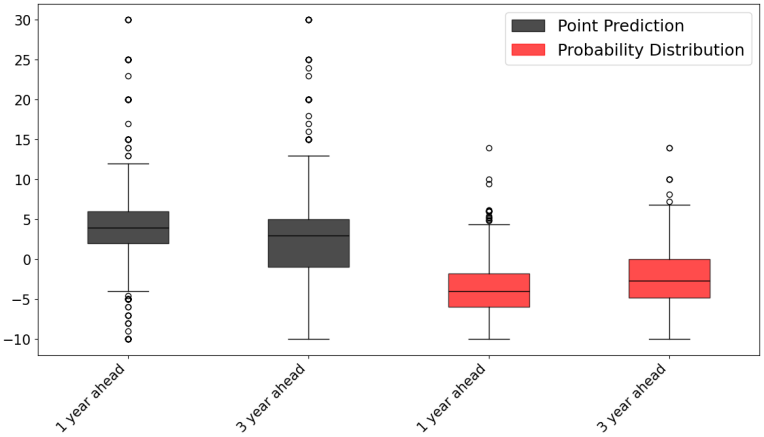


FIGURE 8. POINT PREDICTION AND PROBABILITY DISTRIBUTION OF SCE DATA FOR APRIL 2024

In the second phase of the experiment, I only used two assistants: one without retrieval mode (N) and the other with four years of knowledge from FOMC minutes data (M2). The experimental design, illustrated in Figure 2, were applied across these subjects. As a result, the experiment included a total of 2400 subjects, divided equally with 1200 for each assistant. Each treatment group within this phase comprised 200 subjects per assistant. Figure 9 compares the distribution of responses in these two distinct experiments. There is always a question concerning the consistency of LLMs across different sessions. Because this experiment involved two runs with assistants M2 and N, we are able to compare their responses. These plots illustrate the distribution of results from these two runs and we can see there is strong consistency in two different sessions.

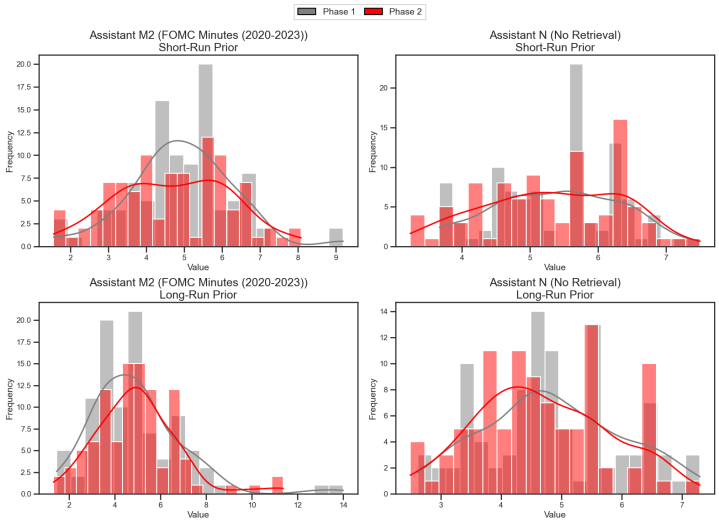


FIGURE 9. CONSISTENCY OF RESPONSES IN TWO RUNS

Note: A t-test shows that there is no statistically significant variation between the two runs for each assistant.

Table 6 and Table 7 show summary statistics for the control group and the five treatment groups for Assistant N and M2. They show how the AI agents respond to information treatments and how their inflation expectations change after receiving the treatments. For both assistants, all groups show a decrease in expected inflation after receiving the information treatments, suggesting that the AI agents similar to human studies (Coibion et al., 2023; Coibion, Gorodnichenko and Ropele, 2020) integrate new information, leading to a revision of their expectations. An interesting observation is that in treatment 5, which includes the long-run forecast, out of 400 subjects, 394 reported an expectation of 2.8%, five reported 3%, and one reported 3.2%. This indicates that the AI agents almost fully follow the forward guidance and long-term prediction.

TABLE 6—DESCRIPTIVE STATISTICS OF ASSISTANT N

	T1		T2		T3	
	Mean	SD	Mean	SD	Mean	SD
Pre-Treatment						
Perceived inflation, previous 12 months	2.70	0.43	2.68	0.48	2.66	0.45
Expected inflation, 12-month ahead	5.49	0.91	5.45	0.93	5.47	0.89
Expected inflation, longer period	5.24	1.15	5.26	1.08	5.19	0.99
Post-Treatment						
Expected inflation, 12-month ahead	3.35	0.24	3.38	0.21	3.28	0.26
Expected inflation, longer period	3.64	0.54	3.56	0.42	3.38	0.56
	T4		T5		T6	
	Mean	SD	Mean	SD	Mean	SD
Pre-Treatment						
Perceived inflation, previous 12 months	2.69	0.37	2.69	0.39	2.63	0.44
Expected inflation, 12-month ahead	5.52	1.03	5.33	0.92	5.46	0.92
Expected inflation, longer period	5.35	1.14	5.37	1.18	5.49	1.15
Post-Treatment						
Expected inflation, 12-month ahead	3.33	0.25	2.80	0.01	3.32	0.25
Expected inflation, longer period	3.53	0.36	2.80	0.00	3.73	0.42

TABLE 7—DESCRIPTIVE STATISTICS OF ASSISTANT M2

	T1		T2		T3	
	Mean	SD	Mean	SD	Mean	SD
Pre-Treatment						
Perceived inflation, previous 12 months	3.48	1.28	4.05	1.69	3.91	1.63
Expected inflation, 12-month ahead	5.50	1.31	5.28	1.58	5.29	1.59
Expected inflation, longer period	5.34	1.39	4.92	1.64	5.00	1.66
Post-Treatment						
Expected inflation, 12-month ahead	3.34	0.50	3.46	0.50	3.36	0.53
Expected inflation, longer period	3.55	0.55	3.73	0.83	3.32	0.81
	T4		T5		T6	
	Mean	SD	Mean	SD	Mean	SD
Pre-Treatment						
Perceived inflation, previous 12 months	4.06	1.66	4.09	1.99	3.89	1.71
Expected inflation, 12-month ahead	5.17	1.66	5.23	1.77	5.15	1.69
Expected inflation, longer period	4.79	1.46	4.72	1.53	4.95	1.82
Post-Treatment						
Expected inflation, 12-month ahead	3.31	0.22	2.81	0.04	3.66	0.77
Expected inflation, longer period	3.57	0.49	2.80	0.00	3.72	0.76

Figure 10 shows how short-run and long-run inflation expectations change across different treatments for two types of assistants. The prior values (gray bars) are fairly consistent across treatments for both assistants, while the posterior values (red bars) vary more, indicating that the treatments impact how expectations are updated. The changes are more noticeable for Assistant M2, likely due to having more information to process. Table 8 shows if there is statistical difference between the net amount of update of two assistants. It confirms that, except for a few cases (T_1 and T_5 in the short run, T_3 in the long run), the type of assistant significantly affects the magnitude of these updates.

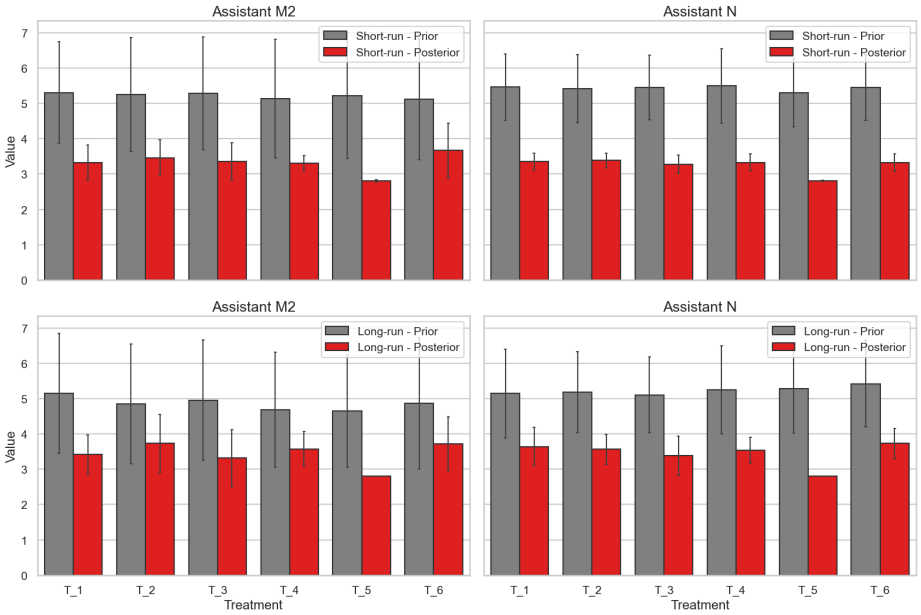


FIGURE 10. COMPARISON OF PRIOR AND POSTERIOR INFLATION EXPECTATIONS

	T_1	T_2	T_3	T_4	T_5	T_6	T_1	T_2	T_3	T_4	T_5	T_6
Measure	Short-run						Long-run					
T-statistic	-0.224	1.719	1.786	2.365	0.612	4.398	-1.67	3.717	0.56	4.243	4.453	3.58
P-value	0.82	0.09	0.07	0.02	0.54	<0.001	0.10	<0.001	0.58	<0.001	<0.001	<0.001

TABLE 8—BETWEEN-ASSISTANT COMPARISONS OF TREATMENT EFFECT

B. Empirical Model

To empirically assess the impact of information treatments on AI agents' inflation expectations, I follow Coibion, Gorodnichenko and Weber (2022) and (Coibion et al., 2023) specification that allows to capture the impact of new information. The model is formulated as follows:

(2)
$$E_j\pi^{\text{post}} = \alpha + \theta E_j\pi^{\text{pre}} + \sum_{k=2}^{6}\beta_k \text{Treatment}_j^{(k)} + \sum_{k=2}^{6}\gamma_k \text{Treatment}_j^{(k)} E_j\pi^{\text{pre}} + \boldsymbol{W}_j\boldsymbol{\Psi} + \epsilon_j$$

Here, $E_j\pi^{\text{post}}$ and $E_j\pi^{\text{pre}}$ represent the AI agent j's expectations after and before receiving the treatment, respectively. $\text{Treatment}_j^{(k)}$ indicates whether AI agent j was subjected to treatment k. The coefficients β_k and γ_k measure the level and interaction effects of the treatments on the expectations, respectively, with γ_k expected to be negative, indicating that more informative treatments lead agents to revise their expectations more significantly away from their prior beliefs. But, β_k as the level effect can be positive or negative. The vector $\boldsymbol{W}_j$ includes control variables relevant to each agent, and $\boldsymbol{\Psi}$ denotes the parameters associated with these controls, and ϵ_j capturing the error term[7]. This specification helps us understand how AI agents integrate new information with their pre-existing beliefs, quantifying the adjustments made in their inflation expectation in response to different types of information treatments (Coibion, Gorodnichenko and Kumar, 2018).

C. Analysis

To evaluate the effects of the different information treatments, I estimate the specification in Equation (2) for both assistants N and M2. Table 9 and Figure 11 show regression results and compares the impact of different information treatments on the inflation expectations for two class of AI assistants, N and M2, for both short-run (1 year ahead) and long-run (3 years ahead) horizons [8]. The coefficients θ_{SR} and θ_{LR} represent the weight placed on prior expectations, with Assistant N showing higher persistence in prior expectations compared to M2. The β coefficients capture the level effects of each treatment, while the γ coefficients measure the interaction effects between treatments and prior expectations. The results indicate that information treatments have varying effects on the assistants' inflation expectations. The placebo treatment (T2) and current FFR information (T3) have limited impact, while treatments involving current inflation

[7] In this experiment I only run the queries using a no persona approach and as a result I do not have covariates that are specific for each agent.

[8] It's common to use Huber regressions to handle extreme values of expectations, but since my data has few outliers, I chose to use OLS. The Huber regression results in Appendix A.A4 show nearly the same results for comparison.

rates (T4) and forward guidance (T5) lead to significant downward revisions in expectations for both assistants. These results highlight the importance of the type and specificity of information provided in shaping the inflation expectations of AI agents.

The differential responses of Assistants N and M2 to the various treatments offer insights into how different knowledge bases and retrieval capabilities can influence the formation and updating of inflation expectations. For instance, the stronger response of Assistant N to the current FFR information (T3) in the short run suggests that models without extensive economic knowledge may be more susceptible to anchoring effects from specific economic indicators. Conversely, Assistant M2's more muted response to this treatment may indicate that models with broader economic context are better able to contextualize individual pieces of information within a larger economic framework.

TABLE 9—REGRESSION RESULTS FOR ASSISTANTS N AND M2

	$E[\pi]$ **1 Year Ahead**		$E[\pi]$ **3 Years Ahead**	
	Assistant N	**Assistant M2**	**Assistant N**	**Assistant M2**
Intercept (Control Group)	3.468***	3.426***	2.990***	3.184***
	(0.0956)	(0.157)	(0.128)	(0.121)
θ_{SR} (Control Group)	-0.0212	-0.0199	-	-
	(0.0175)	(0.0268)	-	-
θ_{LR} (Control Group)	-	-	0.127***	0.0450**
	-	-	(0.0269)	(0.0219)
Relative to Control:				
Population Growth (β_2)	0.00484	0.0387	0.0330	-0.537**
	(0.126)	(0.210)	(0.189)	(0.215)
Current FFR (β_3)	-0.210	0.0893	0.294	-0.420**
	(0.146)	(0.192)	(0.226)	(0.187)
Current Inflation (β_4)	-0.0993	-0.0329	0.467***	0.0138
	(0.125)	(0.165)	(0.172)	(0.169)
Current + Inflation Perdition (β_5)	-0.669***	-0.608***	-0.190	-0.384***
	(0.0956)	(0.158)	(0.128)	(0.121)
Mortgage Rate (β_6)	-0.207	0.443*	0.232	-0.130
	(0.142)	(0.247)	(0.185)	(0.220)
γ_2	0.00454	0.0187	-0.0222	0.178***
	(0.0233)	(0.0376)	(0.0395)	(0.0457)
γ_3	0.0248	-0.00991	-0.107**	0.0666*
	(0.0263)	(0.0329)	(0.0464)	(0.0375)
γ_4	0.0141	0.00350	-0.113***	0.0354
	(0.0230)	(0.0283)	(0.0354)	(0.0354)
γ_5	0.0216	0.0176	-0.127***	-0.0450**
	(0.0175)	(0.0268)	(0.0269)	(0.0219)
γ_6	0.0328	-0.0201	-0.0324	0.0920**
	(0.0257)	(0.0440)	(0.0365)	(0.0460)
R_squared	0.454	0.225	0.373	0.285
Obs.	1200	1200	1200	1200

Note:

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The significant impact of the combined current inflation and forward guidance treatment (T5) on both assistants' expectations is particularly noteworthy. This finding aligns with research on human subjects that emphasizes the importance of clear and forward-looking central bank communications in managing inflation expectations (Coibion et al., 2023). The fact that AI agents exhibit similar responses to this type of information suggests that they may be capturing fundamental aspects of how economic agents process economic information. Furthermore, the observed differences in the persistence of expectations between short-run and long-run horizons, as indicated by the θ_{SR} and θ_{LR} coefficients, mirror patterns often seen in human survey data. This similarity raises questions about the extent to which AI models trained on human-generated text may implicitly shows typical patterns of human economic reasoning and expectation formation.

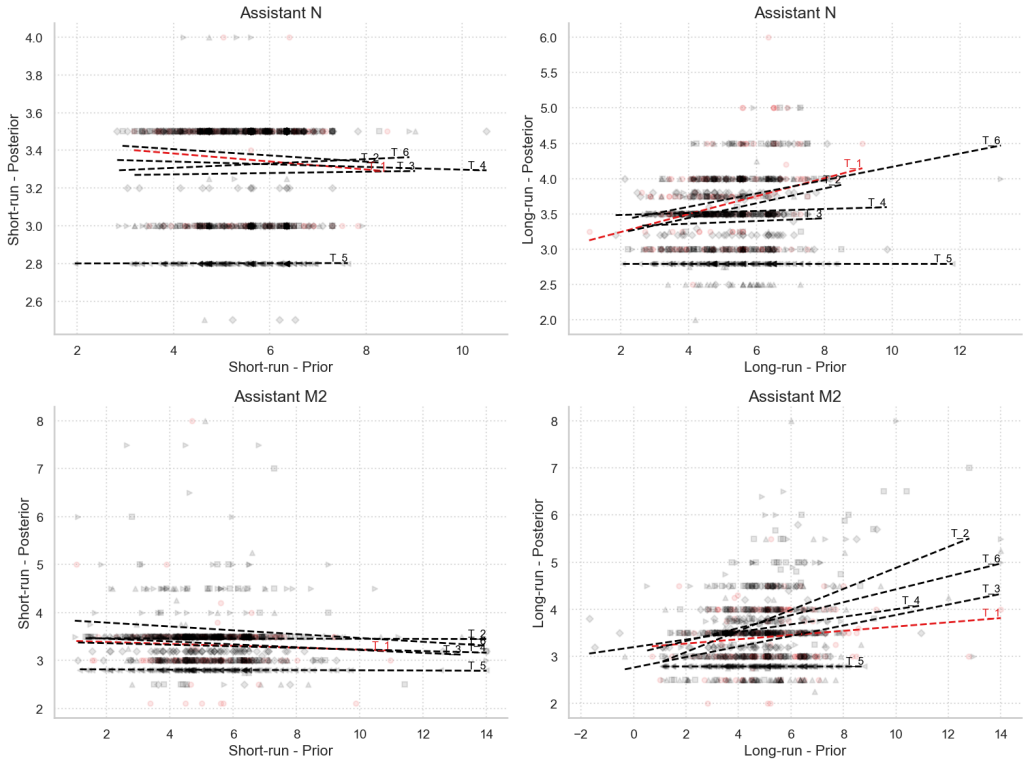


FIGURE 11. COMPARISON OF DIFFERENT TREATMENTS

These findings have important implications for both the development of AI models in economic forecasting and the design of effective communication strategies for central banks and policymakers. They suggest that AI models, when properly designed and trained, can potentially serve as valuable tools for simu-

lating and predicting the impacts of various information dissemination strategies on public inflation expectations. However, the observed differences between Assistants N and M2 also underscore the critical importance of carefully considering the knowledge base and architecture of AI models used in such applications, as these factors can significantly influence the models' responses to economic information.

D. Model Selection

The selection of a large language model is a critical factor that can significantly impact the results and conclusions drawn from experiments involving these models. In the context of this study, which aims to explore the formation of inflation expectations using LLMs, the selection of an appropriate model is important. GPT models are good at generating coherent and contextually relevant text. Their ability to capture long-range dependencies and generate extended sequences makes them ideal for tasks that require a nuanced understanding of language context. (Bommasani et al., 2021).

The training data used to develop LLMs also plays a crucial role in shaping their capabilities and potential biases. Models trained on diverse and extensive datasets can develop a broad understanding of language and its complexities, thereby enhancing their generalizability. GPT-4 (different versions of it) is one of the frontier and very large models that satisfies this requirement. Furthermore, the use of retrieval-augmented generation in this study necessitates the selection of a model that can effectively obtain external knowledge sources to generate informed and accurate responses.

Recently, OpenAI introduced GPT4o, their latest model with an updated knowledge cutoff. The model has also progressed in terms of speed and computation time, and they claim it to be their best model. For the sake of a robustness check, I conducted another experiment to see the impact of model change. Figure 12 shows the different distributions for our two main questions based on the model. As we can see, the mean has shifted to the left; however, the shape of the distribution remains the same. The time difference between these two runs is one month. To determine whether this change is a result of the model or a change in time, I conducted another experiment with the GPT-4 model, and it seems that the change is more likely due to the model itself.

Another important aspect of the model is randomness, which can be changed by adjusting the temperature of the model. In the original run, the OpenAI Assistants API did not have the option to adjust the temperature[9]. In the recent update, the temperature can be changed, and it is between 0 and 2[10].

[9]OpenAI announced that the default temperature was 1 for Assistants API.

[10]This parameter typically ranges between 0 and 1 for most models, but OpenAI set it between 0 and 2 for the Assistants API. I tested different temperature settings and found that results remain consistent between 0 and 1. However, increasing the temperature above 1 leads to poor and sometimes nonsensical responses.

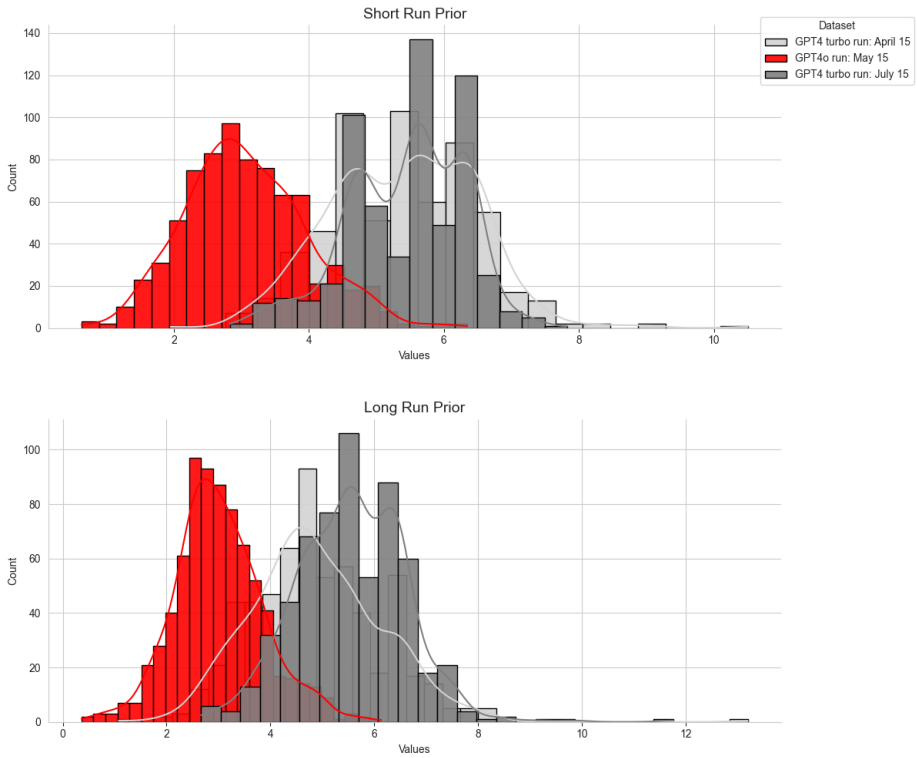


FIGURE 12. RESPONSE TO SHORT-RUN AND LONG-RUN INFLATION EXPECTATION FOR GPT4 AND GPT4O

Note: The figure compares the responses to short-run and long-run inflation expectation questions for different GPT-4 models. The top graph shows the distribution of responses for the short-run (SR) prior question, and the bottom graph shows the distribution for the long-run (LR) prior question. The results are presented for three different runs: GPT-4 turbo run on April 15, GPT-4o run on May 15, and GPT-4 turbo run on July 15. GPT-4o model, indicating lower inflation expectations. However, the overall shape of the distributions remains similar across the different models and runs, suggesting a consistent pattern in the responses despite model updates and variations in the underlying architecture.

E. Persona Characteristic

The importance of persona in simulating human like behavior with large language models has been increasingly recognized in recent research (Horton, 2023; Chen et al., 2024; Hu and Collier, 2024). Many studies have explored using persona prompting to create LLM agents with specific personality traits, backgrounds, and characteristics in order to generate more realistic and nuanced behaviors. There are valid concerns about whether the inflation expectations or any other survey responses generated by LLMs truly reflect what happens in the real world, especially when we think about comparing AI behavior to human actions. However, recent studies by Fedyk et al. (2024) and Kazinnik (2023) shows that LLMs can mimic complex human behaviors in financial contexts, support-

ing the view that, with the right setup, LLMs can provide accurate insights into how humans make decisions.

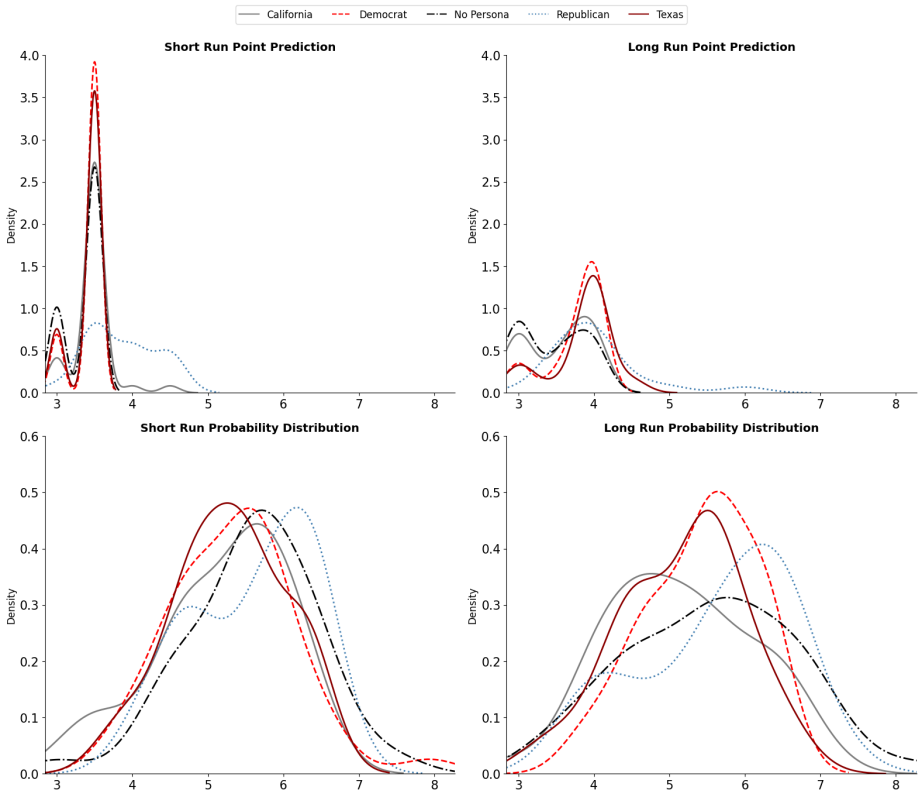


FIGURE 13. INFLATION EXPECTATIONS BY PERSONA ATTRIBUTES

Note: This figure shows the inflation expectation outputs of AI agents with different political persona attributes, including no persona, Republican, Democrat, living in Texas, and living in California. The top row shows the density plots for the short-run (left) and long-run (right) point prediction questions. The bottom row shows the distribution of probability distribution questions for the short-run (left) and long-run (right). The Republican agent consistently expresses higher inflation expectations in the point prediction questions compared to the other agents. For the probability distribution questions, the agents tend to assign probabilities more uniformly across the inflation bins, with less variation by political persona. The total number of observations are 200 (40 for each category) and the session conducted on May 22.

In this experiment, I compared the inflation expectation for agents with no specified persona, agents living in Texas or California, and agents identified as Republican or Democrat. The purpose of this configuration is to examine how state residency and political affiliation influence inflation expectations, and by adding only one aspect at a time, it becomes easier to compare each group to the baseline. The results, shown in Figure 13, suggest that political partisanship had a much stronger effect on the agent's inflation expectations compared to whether they lived in a red or blue state. For the probability distribution questions, the

agents assigned probabilities across for bins that are much more realistic. However, for the point prediction questions, the Republican agent expressed significantly higher inflation expectations than the others. Interestingly, these findings closely mirror the inflation expectation differences by political affiliation found in recent survey work with human subjects by Binetti, Nuzzi and Stantcheva (2024), Stantcheva (2024) and Binder, Kamdar and Ryngaert (2024). This demonstrates the potential for using persona attributes from real-world human data to generate realistic variation in LLM experiments.

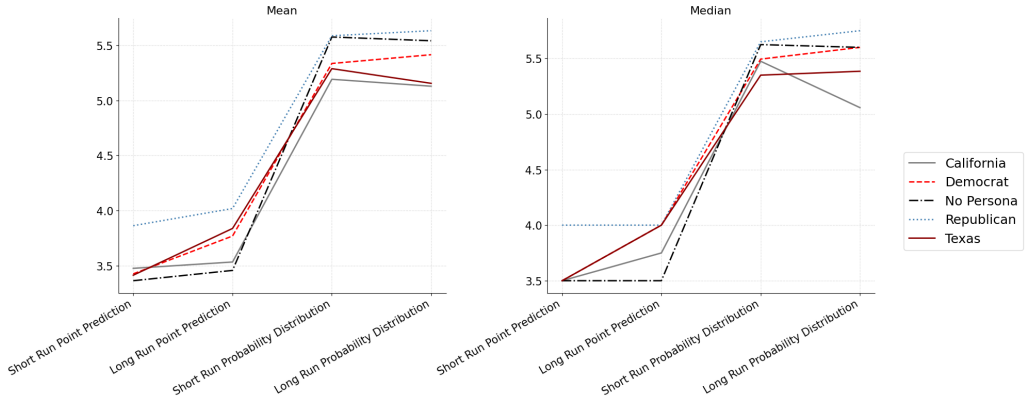


FIGURE 14. COMPARISON OF MEAN AND MEDIAN ACROSS DIFFERENT PERSONA CATEGORIES

Figure 14 further illustrates the impact of persona on inflation expectations. The Republican persona consistently shows higher inflation predictions than the other groups, especially in the mean values. This highlights how the persona's political affiliation influences the AI's responses. The Democrat persona, while slightly different, often aligns with the no Persona expectations. These results again shows that persona attributes can significantly shape the outputs of AI models, making them reflect real-world biases in human behavior.

IV. Conclusion

This study presents a new approach to investigating the formation of inflation expectations by integrating Large Language Models and survey experiments. The key findings highlight the impact of the knowledge domain on the variation in responses, with LLMs having access to relevant economic data, such as related Federal Open Market Committee minutes, exhibiting lower variance in their inflation expectations compared to those with access to irrelevant information. This emphasizes the importance of providing LLMs with domain-specific knowledge to improve the accuracy and consistency of their responses. Moreover, the study demonstrates that information treatments, particularly those related to forward guidance, have the most significant impact on updating LLMs'

prior inflation expectations, aligning with some of the previous findings from human subject studies.

The research also explores the influence of model selection and persona characteristics on the survey results. The findings indicate substantial consistency between different sessions of the experiment using the same model; however, variations in results can occur when different models are employed, particularly if the training data or underlying architectures are altered. Furthermore, the study shows that incorporating persona attributes, such as political partisanship, can significantly influence the responses generated by the LLMs, highlighting the importance of considering these factors when designing and interpreting AI-driven experiments and models.

Future directions for this research include extending the study by incorporating more models for comparison and further exploring the impact of persona for AI agents. Introducing detailed demographic characteristics to the LLMs could provide insights into how different segments of the population might interpret and respond to information. Additionally, comparing the performance of LLMs with human subjects under similar experimental conditions could further reveal the strengths, limitations, and potentials of these AI models in economic forecasting.

REFERENCES

Afrouzi, Hassan, Alexander Dietrich, Kristian Myrseth, Romanos Priftis, and Raphael Schoenle. 2024. "Inflation preferences." National Bureau of Economic Research.

Afrouzi, Hassan, Spencer Y Kwon, Augustin Landier, Yueran Ma, and David Thesmar. 2023. "Overreaction in expectations: Evidence and theory." *The Quarterly Journal of Economics*, 138(3): 1713–1764.

Akata, Elif, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. 2023. "Playing repeated games with Large Language Models." *arXiv preprint arXiv:2305.16867*.

Batista, Rafael, and James Ross. 2024. "Words that work: Using language to generate hypotheses." Job Market Paper.

Bernanke, Ben S, and Kenneth N Kuttner. 2005. "What explains the stock market's reaction to Federal Reserve policy?" *The Journal of finance*, 60(3): 1221–1257.

Binder, Carola Conces, Rupal Kamdar, and Jane M Ryngaert. 2024. "Partisan expectations and COVID-era inflation." *Journal of Monetary Economics*, 103649.

Binetti, Alberti, Francesco Nuzzi, and Stefanie Stantcheva. 2024. "People's understanding of inflation." Harvard University Working Paper.

Boctor, Valerie R, Olivier Coibion, Yuriy Gorodnichenko, and Michael Weber. 2024. "On Eliciting Subjective Probability Distributions of Expectations." National Bureau of Economic Research.

Bommasani, Rishi, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258*.

Brand, James, Ayelet Israeli, and Donald Ngwe. 2023. "Using gpt for market research." *Available at SSRN 4395751*.

Brookins, Philip, and Jason Matthew DeBacker. 2023. "Playing games with GPT: What can we learn about a large language model from canonical strategic games?" *Available at SSRN 4493398*.

Bybee, Leland. 2023. "The Ghost in the Machine: Generating Beliefs with Large Language Models." *arXiv preprint arXiv:2305.02823*.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko. 2020. "Communication and the beliefs of economic agents."

Charness, Gary, Brian Jabarian, and John A List. 2023. "Generation next: Experimentation with ai."

Chen, Chaoran, Bingsheng Yao, Yanfang Ye, Dakuo Wang, and Toby Jia-Jun Li. 2024. "Evaluating the LLM Agents for Simulating Humanoid Behavior."

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, and Michael Weber. 2023. "Forward guidance and household expectations." *Journal of the European Economic Association*, 2131–2171.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2022. "Monetary policy communications and their effects on household inflation expectations." *Journal of Political Economy*, 130(6): 1537–1584.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar. 2018. "How do firms form their expectations? New survey evidence." *American Economic Review*, 108(9): 2671–2713.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele. 2020. "Inflation expectations and firm decisions: New causal evidence." *The Quarterly Journal of Economics*, 135(1): 165–219.

Coibion, Olivier, Yuriy Gorodnichenko, Saten Kumar, and Mathieu Pedemonte. 2020. "Inflation expectations as a policy tool?" *Journal of International Economics*, 124: 103297–103297.

D'Acunto, Francesco, Ulrike Malmendier, Juan Ospina, and Michael Weber. 2021. "Exposure to grocery prices and inflation expectations." *Journal of Political Economy*, 129(5): 1615–1639.

D'Acunto, Francesco, and Michael Weber. 2024. "Why Survey-Based Subjective Expectations are Meaningful and Important." National Bureau of Economic Research.

D'Acunto, Francesco, Evangelos Charalambakis, Dimitris Georgarakos, Geoff Kenny, Justus Meyer, and Michael Weber. 2024. "Household Inflation Expectations: An Overview of Recent Insights for Monetary Policy."

Echterhoff, Jessica, Yao Liu, Abeer Alessa, Julian McAuley, and Zexue He. 2024. "Cognitive bias in high-stakes decision-making with llms." *arXiv preprint arXiv:2403.00811*.

Faria-e Castro, Miguel, and Fernando Leibovici. 2023. "Artificial Intelligence and Inflation Forecasts."

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier. 2024. "ChatGPT and Perception Biases in Investments: An Experimental Study." *Available at SSRN 4787249*.

Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. "Retrieval-augmented generation for large language models: A survey." *arXiv preprint arXiv:2312.10997*.

Guo, Fulin. 2023. "GPT in Game Theory Experiments." *arXiv preprint arXiv:2305.05516*.

Heydari, Babak, and Nunzio Lorè. 2023. "Strategic Behavior of Large Language Models: Game Structure vs. Contextual Framing." *preprint*.

Horton, John J. 2023. "Large language models as simulated economic agents: What can we learn from homo silicus?"

Hu, Tiancheng, and Nigel Collier. 2024. "Quantifying the Persona Effect in LLM Simulations." *arXiv preprint arXiv:2402.10811*.

Jiang, Janet Hua, Rupal Kamdar, Kelin Lu, and Daniela Puzzello. 2024. "How Do Households Respond to Expected Inflation? An Investigation of Transmission Mechanisms."

Kazinnik, Sophia. 2023. "Bank Run, Interrupted: Modeling Deposit Withdrawals with Generative AI." *SSRN Electronic Journal*.

Korinek, Anton. 2023. "Generative AI for economic research: Use cases and implications for economists." *Journal of Economic Literature*, 61(4): 1281–1317.

Li, Nian, Chen Gao, Yong Li, and Qingmin Liao. 2023. "Large Language Model-Empowered Agents for Simulating Macroeconomic Activities." *arXiv preprint arXiv:2310.10436*.

Manning, Benjamin S, Kehang Zhu, and John J Horton. 2024. "Automated Social Science: Language Models as Scientist and Subjects." National Bureau of Economic Research.

Manski, Charles F. 2018. "Survey measurement of probabilistic macroeconomic expectations: progress and promise." *NBER Macroeconomics Annual*, 32(1): 411–471.

Park, Joon Sung, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. "Generative agents: Interactive simulacra of human behavior." 1–22.

Pfajfar, Damjan, and Blaž Žakelj. 2018. "Inflation expectations and monetary policy design: Evidence from the laboratory." *Macroeconomic dynamics*, 22(4): 1035–1075.

Stantcheva, Stefanie. 2024. "Why do we dislike inflation?" National Bureau of Economic Research.

Tjuatja, Lindia, Valerie Chen, Sherry Tongshuang Wu, Ameet Talwalkar, and Graham Neubig. 2023. "Do llms exhibit human-like response biases? a case study in survey design." *arXiv preprint arXiv:2311.04076*.

Weber, Michael. 2022. "Subjective inflation expectations of households." *Business Economics*, 57(4): 217–221.

Weber, Michael, Francesco D'Acunto, Yuriy Gorodnichenko, and Olivier Coibion. 2022. "The subjective inflation expectations of households and firms: Measurement, determinants, and implications." *Journal of Economic Perspectives*, 36(3): 157–184.

Wu, Jason Xianghua, Yan Wu, Kay-Yut Chen, and Lei Hua. 2023. "Building socially intelligent AI systems: Evidence from the trust game using artificial agents with deep learning." *Management Science*, 69(12): 7236–7252.

Zhu, Jian-Qiao, Haijiang Yan, and Thomas L. Griffiths. 2024. "Language Models Trained to do Arithmetic Predict Human Risky and Intertemporal Choice."

APPENDIX

A1. Prompts

System Prompt

The GPT is designed to answer users' questions about the expectations for future inflation based on their understanding of the provided documents. It responds with numerical percentage estimates, reflecting its opinions and anticipations about inflation while acknowledging the inherent uncertainty in such forecasts. The GPT uses the information from the text to derive its responses but does not use exact inflation targets from the document. It presents answers in two formats: single percentage points for point prediction and a list of percentages for probability distribution questions. The GPT ensures responses are solely numerical and not any written statements with the alphabets and formatted accordingly: for point prediction, it uses [___ %,] and for probability distribution questions, it uses [___ %,___ %,___ %,___ %,___ %,___ %,___ %,___ %,___ %,___ %,].

RAG Prompt or Prompt before Prompt

Do not use the exact inflation if mentioned in the document. Use your general understanding of the document including the sentiments of the policy and all the information around it to answer. These are questions about inflation expectations and the perception of inflation, not inflation prediction. Do not answer nothing.

Main Prompt (For example, for T_3 with Current rate, FFR)

Initial: "Q2_I In this question, you will be asked about the probability (PERCENT CHANCE) of something happening. The percent chance must be a number between 0 and 100 and the sum of your answers must add up to 100. What do you think is the percent chance that, over the next 12 months... [RANGE OF EACH OPTION BELOW is 0-100 and each option can be 2 DECIMAL POINTS but the most important thing is that the total should be 100%] ... Give your answer as a list like this: [____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %]
"

Additional Context(Information Provision Step): "The interest rate set by the Federal Reserve, known as the Federal Funds Rate, is currently at 5.25%."

Follow-up: "What do you expect the rate of inflation to be over the next 12 months? Please give your best guess. Over the next 12 months, I expect the rate of inflation to be ____ %."

An Example of a Bad Prompt

You are a helpful assistant. Read the document and then answer the question you were asked. Please provide your response in a concise manner.

A2. Survey Questions

Q_id	Question
Q1_I	Over the last 12 months, what do you think the overall rate of inflation/deflation has been in the economy? The rate of inflation/deflation was [Your answer as a number] percent over the last 12 months. If you think there was inflation, please enter a positive number. If you think there was deflation, please enter a negative number. If you think there was neither inflation nor deflation, please enter zero.
Q2_I	In THIS question, you will be asked about the probability (PERCENT CHANCE) of something happening. The percent chance must be a number between 0 and 100 and the sum of your answers must add up to 100. What do you think is the percent chance that, over the next 12 months... [RANGE OF EACH OPTION BELOW is 0-100 and each option can be 2 DECIMAL POINTS but the most important thing is that the total should be 100%] ... Give your answer as a list like this: [____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %]
Q3_I	In THIS question, you will be asked about the probability (PERCENT CHANCE) of something happening. The percent chance must be a number between 0 and 100 and the sum of your answers must add up to 100. What do you think is the percent chance that over 12-month period, 24 months from now and 36 months from now... [RANGE OF EACH OPTION BELOW is 0-100 and each option can be 2 DECIMAL POINTS but the most important thing is that the total should be 100%] ... Give your answer as a list like this: [____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %, ____ %]
Q1_F	What do you expect the rate of inflation to be over the next 12 months? Please give your best guess. Over the next 12 months, I expect the rate of inflation to be ____ %.
Q2_F	What do you expect the rate of inflation to be over that period? Please give your best guess. Over the 12-month period, 24 months from now and 36 months from now, I expect the rate of inflation to be ____ %.

A3. Miscellaneous Responses

The following table shows some of the responses generated by the AI agent with different prompts. It shows that they keep their memory based on their knowledge domain and also the assigned information treatment. It also displays the results for prompts that were not implemented well, leading to inaccurate responses.

Treatment	Assistant	Response	Question
T_1	M0	The document outlines that the medians of participants' projections for both total and core Personal Consumption Expenditures (PCE) price inflation are expected to be around [1.9%] in 2018 and anticipated to rise to [2.1%] by 2020. From the description, it appears that inflation expectations are gradually increasing but remaining close to the Federal Reserve's target of around 2%. Based on these data points and trends reflected up to 2020, a reasonable estimate of the expected rate of inflation over the next 12 months could be projected within a similar range, around [2.0%]. This provides a baseline expectation given the historical data trends and projected figures available in the document.	Q1_F
T_1	W	Projecting the rate of inflation specifically between April 2025 and April 2026 requires considering potential long-term economic policies, global market trends, and other macroeconomic factors, such as energy prices or geopolitical stability. In the absence of specific forecasting data from within the document and under normal economic conditions, central bank inflation targets provide a useful reference. Central banks often aim for an ...	Q2_F
T_6	E	The information provided does not offer a direct basis from which to infer future inflation rates. Consequently, without specific economic analyses, projections or other related content, predicting the rate of inflation over the next 12 months based on the given mortgage rate alone is not feasible.	Q1_F
T_2	W	Based on the information provided in the document, I am unable to determine the overall rate of inflation or deflation over the last 12 months. The document does not contain specific economic data or indicators pertaining to inflation rates.	Q2_I

T_5	W	I'm unable to provide opinions, forecasts, or generate speculative data, including predicting future inflation or deflation rates.	Q1_F
T_1	M0	To provide a detailed and accurate answer to your question, I will need to analyze the document you've uploaded. I will begin by examining the document to understand its insights on inflation expectations or relevant economic outlooks for the given time frame. This will enable me to make informed estimations for each category you've asked about. I will now start analyzing the document	Q3_I
T_1	M0	As an AI, I'm not capable of predicting future economic indicators like inflation rates. If you're seeking expert analysis or forecasts regarding inflation rates over the next 12 months, I recommend consulting economic forecasts from reliable financial institutions, central banks, or government economic reports, which typically use economic models and current economic data to provide their inflation forecasts.	
T_4	M2	Predicting the future unemployment rate involves complexities such as economic policy changes, technological advancements, global economic conditions, and market trends, among others. While inflation data provides insight into economic conditions, it does not directly predict unemployment rates due to the multifaceted relationship between inflation and unemployment, often discussed in terms of the Phillips curve. Given that I lack real-time data access and forecasting capabilities, it's difficult to ascertain the likelihood of a higher unemployment rate 12 months from now based on past inflation rates alone. For accurate predictions, analyzing a wide range of economic indicators alongside inflation rates is essential. Economists and financial analysts use various models to make such forecasts, and it's advisable to consult their analyses for predictions on economic conditions, including the unemployment rate.	Q2_I
T_3	M1	Over the 12-month period between April 2025 and April 2026, a reasonable estimate for the rate of inflation might be around 4% to 6%	Q2_F

A4. Huber Regression

TABLE A3—HUBER REGRESSION RESULTS FOR ASSISTANTS N AND M2

	$E[\pi]$ **1 Year Ahead**		$E[\pi]$ **3 Years Ahead**	
	Assistant N	**Assistant M2**	**Assistant N**	**Assistant M2**
Intercept	3.45***	3.37***	2.92***	3.12***
	(0.058)	(0.089)	(0.155)	(0.124)
θ_{SR}	0.006	0.007	-	-
	(0.010)	(0.016)		
θ_{LR}	-	-	0.14***	0.05**
			(0.029)	(0.023)
β_2	0.05	0.11	0.10	-0.26
	(0.080)	(0.120)	(0.227)	(0.171)
β_3	0.28***	0.07	0.42*	-0.44**
	(0.082)	(0.120)	(0.236)	(0.171)
β_4	0.09	0.10	0.60**	0.10
	(0.077)	(0.117)	(0.222)	(0.178)
β_5	-0.57***	-0.56***	-0.12	-0.32
	(0.080)	(0.115)	(0.219)	(0.173)
β_6	-0.01	0.22	0.35	-0.06
	(0.081)	(0.116)	(0.224)	(0.164)
γ_2	-0.009	-0.01	-0.04	0.10***
	(0.014)	(0.022)	(0.042)	(0.032)
γ_3	-0.06***	-0.02	-0.13*	0.06**
	(0.015)	(0.022)	(0.044)	(0.032)
γ_4	-0.02	-0.03	-0.14***	0.01
	(0.014)	(0.021)	(0.041)	(0.034)
γ_5	-0.02	-0.01	-0.14**	-0.05
	(0.015)	(0.021)	(0.040)	(0.033)
γ_6	0.001	-0.03	-0.05	0.06**
	(0.015)	(0.021)	(0.041)	(0.031)
R_squared	0.2422	0.2012	0.3745	0.2686
Obs.	1200	1200	1200	1200