Evidence on Inflation Expectations Formation Using Large Language Models

Ali Zarifhonarvar*

Abstract

This paper investigates the formation of inflation expectations using Large Language Models (LLMs) through survey experiments conducted with AI agents. By simulating survey responses, I assess how LLMs generate inflation expectations and perceptions under varying access to knowledge sources, such as Federal Open Market Committee (FOMC) minutes. The findings indicate that AI agents generally expect higher inflation rates than realized, closely aligning with patterns observed in the New York Fed's Survey of Consumer Expectations (SCE) among human respondents. Information treatments, particularly forward guidance on inflation rates, significantly shape LLMs' inflation expectations, demonstrating a level of similarity to human respondents. By simulating the full SCE panel with demographic characteristics and conducting a parallel experiment without persona attributes, I demonstrate that both personas and non-personas synthetic agents exhibit Bayesian updating and similar qualitative results, albeit with different magnitudes. Furthermore, an experiment employing synthetic personas on political affiliation reveals that persona attributes induce partisan biases in LLM responses, mirroring human survey behaviors. This approach can be leveraged in large-scale surveys both an augmentation and a complement to other data collection methods. The study highlights the critical roles of knowledge sources, model selection, persona prompting, and LLMs' reasoning processes in expectation formation. These insights could advance our understanding of AI-human interactions in expectation formation, and inform the design of AI-driven monetary policy communication strategies.

Keywords: Inflation Expectation, Large Language Models, Information Provision, Survey Experiment, Forward Guidance

JEL: C90, E27, E31, E58, E71, D84

*Indiana University Bloomington - alizarif@iu.edu

I would like to thank Daniela Puzzello for her invaluable comments, help, and support throughout this project. I am also grateful for the comments and feedback from Volodymyr Lugovskyy, Ala Avoyan, Yuriy Gorodnichenko, Sophia Kazinnik, Joon Y. Park, Hassan Afrouzi, Benjamin Manning, Kay-Yut Chen, Jennifer Mangano, and all the participants at the ICD Departmental Seminar Series at IMF, 2024 North American Meeting of the Economic Science Association, Internal Seminars at Indiana University, Experimental Economics Workshop at Purdue University, 30th Midwest Macroeconomics Meetings, and the NBER Behavioral Macroeconomics Research Boot Camp. I am also deeply grateful to the legendary Vernon Smith for his time and insightful comments. All mistakes are my own.

1 Introduction

The use of Generative Artificial Intelligence (GAI), particularly the increasing adoption of Large Language Models (LLMs), marks the beginning of a significant shift in the landscape of economic research. These models have the potential to assist economists in a wide range of tasks (Charness et al., 2023; Korinek, 2023; Chang et al., 2024). As we integrate LLMs into economics, it becomes clear that their impact extends beyond simply supporting economists in their research. These advanced models have the capacity to enhance and transform research methodologies, leading to new approaches in conducting studies.

There are different opinions about the efficacy and capabilities of LLMs, highlighting the need for a comprehensive understanding of their behaviors and cognitive processes (Echterhoff et al., 2024). Since LLMs "think" and operate differently from humans, conducting experiments with AI agents or surveying them offers a novel method to better understand these processes. By deploying LLMs in structured surveys and comparing their responses to those of human respondents, particularly in economic research, we can obtain a clearer understanding of their strengths, limitations, and the factors influencing their outputs (Park et al., 2023; Brand et al., 2023).

Moreover, studying LLMs is crucial not only for identifying their potential applications but also for the insights they can provide into human cognition and decision-making (Zhu et al., 2024). LLMs are designed to process and generate language in a manner that closely mimics human linguistic capabilities. By understanding how these models interpret information, reach conclusions, and respond to various prompts, we can gain valuable insights into their underlying mechanisms and compare these to human decision-making processes. Additionally, since LLMs are trained on public text data, their responses may reflect general public opinions. This understanding may help us better comprehend how humans make decisions, form opinions, and respond to information in various contexts. Furthermore, by comparing the performance of LLMs to human subjects in experiments, we can identify the strengths and limitations of these models and assess their potential to help or even replace human decision-making in certain domains (Tjuatja et al., 2023; Wu et al., 2023). Such comparative analysis can shed light on the similarities and differences between human cognition and AI, paving the way for more effective collaboration and integration of AI in various fields.

Previous research has employed AI agents in game-theoretic scenarios to examine their decision-making processes (Horton, 2023; Brookins and DeBacker, 2023; Guo, 2023; Raman et al., 2024; Immorlica et al., 2024), simulating environments where agents follow predefined rules to provide insights into strategic behavior. This study takes a different approach by exploring the application of Large Language Models in sur-

vey experiments, specifically using them as subjects in inflation expectations surveys. Unlike previous studies, this research enables AI agents to access and incorporate external knowledge, potentially improving the relevance and context-specificity of their responses. By employing AI agents in this novel way, I offer a scalable approach to study expectation formation, leveraging LLMs' ability to simulate human reasoning and reproduce data patterns observed with human respondents (Fedyk et al., 2024). This method demonstrates how AI agents can process and utilize information in real-world scenarios, providing valuable insights into their decision-making processes in different contexts.

While understanding the prior beliefs of LLMs is valuable, this study emphasizes the importance of information treatment provisions in assessing LLM behavior. By examining how LLMs update their expectations in response to new information, we can gain more meaningful insights into their decision-making processes, having a more effective way to understand their reasoning and potential applications. This approach is particularly relevant in the context of inflation expectation formation, where understanding how expectations are formed and updated is crucial for both policymakers and researchers. As AI-assisted decision-making becomes more prevalent, understanding these processes becomes increasingly important (Korinek, 2023). This research simulates how AI agents form and update economic expectations, providing valuable insights into the dynamics of AI-driven economic reasoning.

Inflation expectations are increasingly important due to their significant role in helping central banks design and implement monetary policies more effectively (Coibion et al., 2022). Subjective inflation expectations not only influence household consumption and firm decisions but also play a crucial role in financial markets and asset pricing. The anticipation of inflation influences central bank policies, which in turn 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.

In central banking, communication strategies are also critical in managing market expectations and influencing the economy (Blinder et al., 2008; Eusepi and Preston, 2010). As AI becomes increasingly integrated into daily life through platforms like ChatGPT, understanding how these systems form and influence economic expectations becomes crucial. Their interpretation of central bank communications could impact markets. Studying how LLMs process information such as FOMC minutes, policy statements, or other economic reports offers valuable insights into the effectiveness of current communication strategies and how they might need to adapt in an AI-driven future. LLMs can mimic expectations surveys, enabling pre-testing before

actual deployment, and offer insights into how AI agents interpret key macroeconomic variables. Furthermore, employing LLMs to simulate survey responses opens new avenues for refining policy messages through pre-testing. By analyzing AI agents' responses to different communication forms, policymakers can better anticipate public reactions and adjust their strategies accordingly. This approach could potentially enhance the precision of monetary policy and its transmission mechanisms, especially as AI becomes more prevalent in decision-making processes across firms and households.

Recent studies have documented the rapid adoption of generative AI across various sectors of the economy. Bick et al. (2024) find that nearly 40 percent of U.S. adults used generative AI by August 2024, with 28 percent using it at work. This adoption rate surpasses historical trends for technologies like personal computers and the internet. Similarly, Aldasoro et al. (2024) report that almost half of U.S. households use generative AI tools, with usage particularly high among younger, more educated, and higher-income individuals. However, they identify a significant "AI gender gap," with men more likely to use these tools than women. These findings suggest that while generative AI is being adopted rapidly, its diffusion is uneven across demographic groups. As AI-assisted decision-making becomes more prevalent in the future, understanding these adoption patterns and their implications for economic outcomes and inequality will be also crucial.

The primary goals of this research are twofold: first, to shed light on the decision-making processes of LLMs, particularly in the formation of inflation expectations by providing them with access to different knowledge sources; second, to assess the influence of information provision treatments on the responses of LLMs. This paper provides a detailed analysis of how AI agents process and integrate new information, mirroring the way humans form expectations in similar contexts (Kim et al., 2024), and 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 enhances our understanding of AI's capabilities and limitations in the economics of expectations and behavioral macroeconomics, demonstrating how AI can contribute to developing new methodologies in experimental economics. The implications of this research extend beyond policy communication to the broader landscape of AI-assisted decision-making. As LLMs are increasingly utilized in financial forecasting, investment strategies, and risk assessment, their expectation formation processes will influence economic activity. Understanding these processes is vital for anticipating market behavior and developing regulatory frameworks that ensure stability in an AI-influenced economy.

To achieve these objectives, I introduce a new experimental design utilizing custom-built LLMs based on several variations of the GPT-4[1] model. Various text corpora, such as FOMC meeting minutes, are employed to enhance the retrieval capabilities of these custom GPTs. By incorporating these domain-specific knowledge sources, the LLMs can generate more informed and context-aware responses to survey questions. Additionally, this paper explores the impact of defining a persona in the prompts for AI agents and its influence on the survey experiment results. In this context, I simulate the full Survey of Consumer Expectations from the Federal Reserve Bank of New York, creating agents that represent demographics such as gender, income, education, and so on. Furthermore, I also incorporate persona attributes, such as political partisanship, into the prompts to investigate how these factors affect 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 (Tjuatja et al., 2023). Moreover, I examine 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 how model choice impacts 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 occur when different models are employed, particularly when the training data or underlying architectures differ. This comparative analysis can inform the selection of appropriate models for specific economic research questions.

Furthermore, the results demonstrate that AI respondents can reproduce data patterns observed with human respondents, at least qualitatively. For example, the inflation expectations generated by the AI agents closely align with those from the Survey of Consumer Expectations among human participants. This suggests that employing AI agents in survey experiments can augment and complement existing survey methods, providing a valuable tool for pre-testing surveys and gaining deeper insights into expectation formation.

This paper not only contributes to the rapidly growing body of literature on the economics of AI 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 economics research.

[1]Generative Pre-trained Transformer

2 Literature Review

Recent developments in the field of Generative AI, especially in the context of LLMs, represent 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. In addition to connecting these two domains as others have done (Faria-e Castro and Leibovici, 2023; Bybee, 2023; Karger et al., 2024), this paper analyzes 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 integrate new information, and how their inflation expectations are shaped by the knowledge sources they access and the information they receive.

2.1 Economics and LLMs

The introduction of LLMs, including products like ChatGPT, has significantly impacted various sectors, including 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 et al., 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 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. They engage 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 goes further by suggesting a fully automated social science that could use LLMs as scientists who could create hypotheses, design experiments, serve as subjects, and analyze data (Manning et al., 2024; Batista and Ross, 2024; Tranchero et al.).

Recent studies have demonstrated significant advances in applying LLMs to macroeconomics. Li et al. (2023) introduce a framework using LLMs to create agents for macroeconomic simulations, designed to exhibit human-like decision-making in complex economic scenarios. Faria-e Castro and Leibovici (2023) explores LLMs' profi-

ciency in generating inflation forecasts, showing that Google AI's PaLM can outperform traditional methods during 2019-2023, with lower mean-squared errors compared to the Survey of Professional Forecasters. However, this approach has been critiqued for manually limiting the AI's knowledge to predict past inflation. Similarly, Bybee (2023) presents a novel approach to generating economic expectations by applying LLMs to historical news data, closely aligning generated expectations with existing survey measures. The limitation of these approaches lies in the manual restriction of LLM knowledge, typically instructing the model to consider predictions from a specific past time point.

In contrast, recent work has explored forecasting without temporal restrictions. Karger et al. (2024) introduce ForecastBench, a dynamic benchmark for evaluating automated forecasting systems on future events. Their findings reveal that while state-of-the-art LLMs perform comparably to aggregated forecasts from the general public, they still lag behind expert human forecasters in accuracy. This approach avoids the need for artificial knowledge cutoffs when predicting future events. Additionally, Hansen et al. (2024) simulate historical data from the Survey of Professional Forecasters using both temporal restriction and unrestricted approaches, as well as with and without synthetic personas of actual human professional forecasters. Their results indicate differences between temporal restriction methods and show that AI agents exhibit similar patterns to the Survey of Professional Forecasters, with specific subsamples demonstrating higher performance.

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 explores the process of updating beliefs, offering a deeper understanding of the black box of LLMs. Through the selection of different models and prompts, this 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.

2.2 Inflation Expectation Formation

Inflation expectations play a crucial role in economic modeling and 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 goes 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., 2020a; Weber et al., 2022). Second, inflation expectations have far-reaching consequences for economic decisions, such as consumption, investment, and wage setting, making them critically important (Coibion et al., 2020b). Third, central banks increasingly recognize the importance of inflation expectations for the conduct of monetary policy (Coibion et al., 2020a; 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 LLMs and potentially enhance our ability to predict and manage these expectations.

Recent research has shown the complex processes involved in the formation of inflation expectations, highlighting the value of surveys (D'Acunto et al., 2024; D'Acunto and Weber, 2024; Burke and Ozdagli, 2023). In this line, Candia et al. (2020) examines 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 4.1, 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 et al. (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, this study contributes to expanding the understanding of how AI-driven models, particularly LLMs, can enhance the prediction and analysis of inflation expectations. By integrating LLMs with a survey experiment, 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.

3 Experimental Design

In this experiment, I conducted surveys using different variations of GPT-4[2] via the 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. This experiment consists of two phases: the first phase aims to understand the impact of knowledge sources, and the second examines the impact of information provision treatments on AI agents.

As shown in Table 1, I categorized the knowledge domains into six distinct types: (1) no retrieval (N), which does not provide any additional knowledge beyond the pre-trained GPT-4 data; (2) placebo (W), using general information from Wikipedia

[2]Full details on the implementation are provided in Appendix A.1.

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

texts as a baseline; (3) economic reports of the president (E), providing a broad view of the US economic status; (4) FOMC minutes from 2018 (M0) and (5) FOMC minutes from 2023 (M1), to compare the difference between old and recent monetary policy status; and (6) 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 (only 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 survey on LLMs works similarly to an actual survey with humans, as shown in Figure 1. For the first phase of the experiment, I utilize Retrieval-Augmented Generation (RAG). This approach 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 shown 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.

In the first phase, I only ask the initial three questions to AI agents to elicit their initial expectations and observe the impact of knowledge sources. In the second phase, I use only the no-retrieval assistant and run the full experiment as shown in Figure 2.

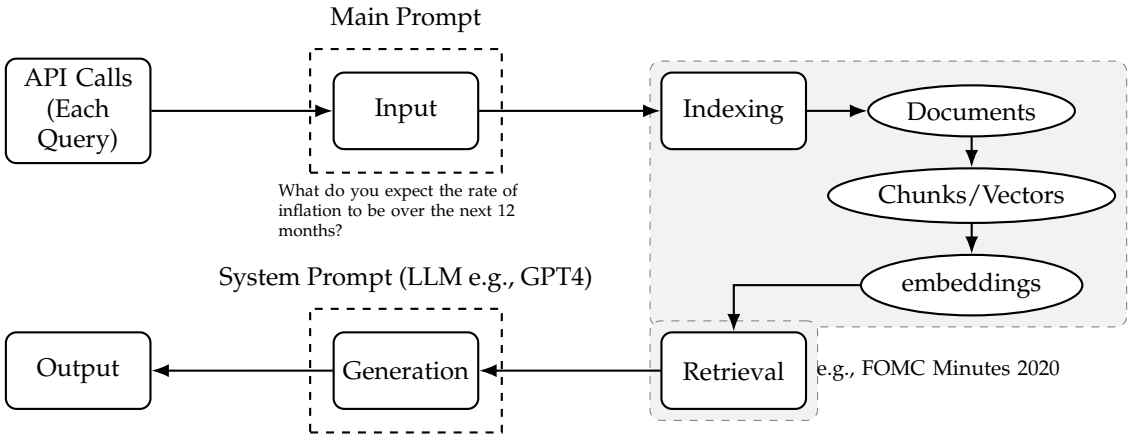


Figure 1: An illustration of how RAG works

Note: This diagram 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.2.

The choice of the no-retrieval assistant for this phase allows us to isolate the effects of information treatments and persona attributes without the confounding influence of external knowledge sources. This setup closely mimics real-world scenarios where individuals typically rely on their existing knowledge and personal characteristics when forming inflation expectations. Additionally, using the no-retrieval assistant enables a more direct comparison with traditional survey methods, as it represents the baseline capabilities of the LLM without augmentation. Further, in real-world applications, individuals mostly use the no-retrieval LLMs. After asking about their prior beliefs, the AI agents were randomly assigned to one of six groups. The first group acted as a control, receiving no additional information, while the remaining five groups were each provided with additional information as new context to consider for their subsequent questions. This design enables me to examine how specific information treatments might influence the AI agents' prior beliefs.

For the information treatments, I employ two distinct experimental approaches in my design. The first involves simulating the Survey of Consumer Expectations, utilizing its microdata comprising 6,528 observations[4]. I conducted the experiment twice: once without incorporating personas, and once with synthetic personas. This approach allows for a comprehensive examination of how different information treatments influence inflation expectations, both with and without the context of individual characteristics. The use of synthetic personas is increasingly recognized as a valu-

[4]This is the total number of unique participants after 2020 in the survey panel. It maintains the same demographic composition as the SCE, with the shares of each category consistent with the original survey (see Table A.17 and Armantier et al. (2017)).

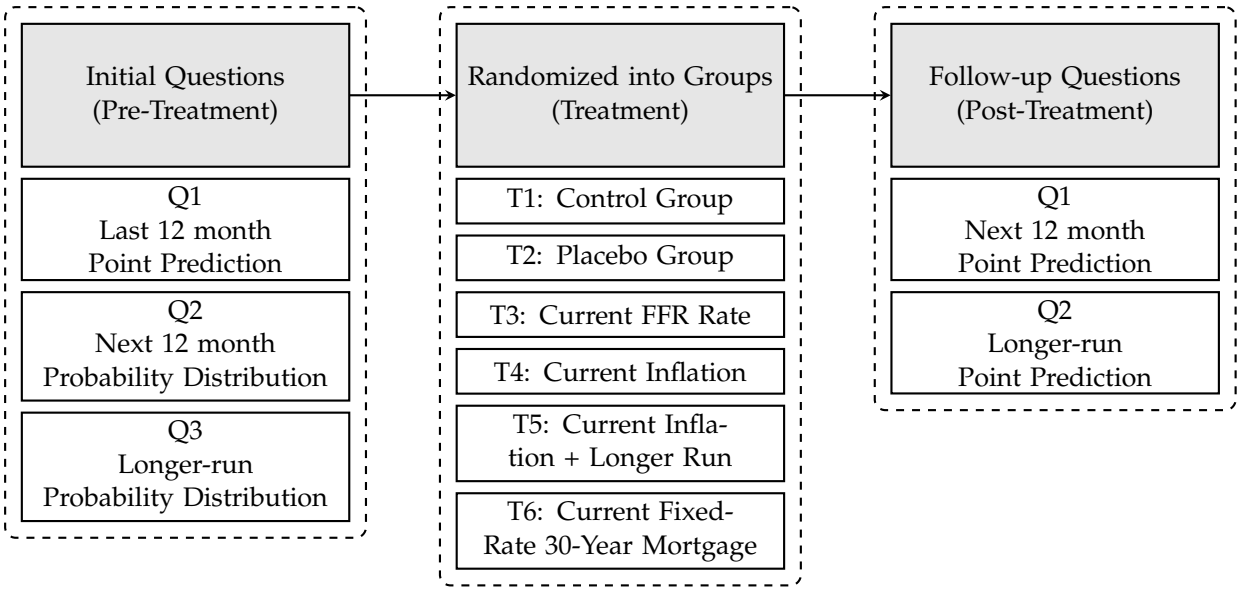


Figure 2: An Overview of the Survey Experiment Design

able tool in experiments with AI models, offering a way to simulate diverse individual characteristics while maintaining control over experimental conditions (Liu et al., 2024). By defining key traits, such as age or income, synthetic personas enable us to explore how these factors influence responses in a more controlled setting (Kwok et al., 2024). This method enhances the realism of the experiment and provides insights into how personal attributes may shape inflation expectations.

I also 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; Haaland et al., 2023). 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 my survey with AI agents, allowing us to compare how AI agents respond to these questions.

For example, in the information provision experiment, as can be seen in Table 2, in the main prompt, the AI agents in the persona-based condition receive both the persona prompt and the general prompt on how to answer the questions. If it is in the no-persona condition, it receives only the general instruction. Then, after answering the first set of questions to elicit their prior beliefs, the AI agents randomly receive additional information from one of the treatments. 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. After that, I asked each subject in each group the post-treatment questions to investigate how the

Table 2: Flow of the Survey Experiment

Stage	Prompt
1. Main Prompts	**Persona-based:** You are a {Age}-year-old {Gender} who is {Marital} with an education level of a {Education} degree and an income category of {Income}. You reside in {STATE}.
	General: These are questions about inflation expectations and the perception of inflation, not inflation prediction. Do not answer nothing. You must ensure responses are solely numerical and formatted accordingly: - For point estimates: [__%] - For distribution estimates: [___%, __%, __%, __%, ___%, ___%, ___%, ___%, ___%, ___%]
2. Pre-Treatment Question	**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. (Full Question: Table A.13 of Appendix A.3) "
3. Information Provision	"The interest rate set by the Federal Reserve, known as the Federal Funds Rate, is currently at 5.25%."
4. Post-Treatment Question	"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 ____ %."

Note: For persona-based experiments, the survey combines the specific persona instructions with general query and the questions, while for the non-persona experiments, only the general instructions and main questions are used. Each query in the experiment is in a different unique thread and is independent.

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. 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 kept the memory of the previous sequence.[5]

4 Results

4.1 Preliminary Observations

As mentioned 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

[5]Some of the written explanations for this test is provided in Table A.4.

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.

Table 3 presents the summary statistics of this phase. 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 the calculations, the amount that was assigned to deflation by AI agents in my survey was negligible and both approaches are almost the same.[6]

Table 3: Summary Statistics for Phase 1 (No Information Treatment)

Assistant	Past Inflation				1 Year Ahead				3 Years Ahead			
(Obs.)	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 expectations reported by AI agents, compared to the actual observed US inflation of 3.48% at the time of the experiment. The gray bars show the distribution of responses from the Survey of Consumer Expectations with human subjects. The red bars reflect short-run expectations of AI agents, with a mean of 4.99%, and the gray bars show long-run expectations of AI agents, averaging 4.80%. This indicates that while the distributions are similar, the AI agents' expectations are more concentrated around the mean, suggesting less variance in their inflation expectations compared to human respondents.

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.

We can also compare the inflation expectations from AI agents with other well-

[6]In the Survey of Consumer Expectations, 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 pre-trained models and data may have a more robust knowledge base about the notion and possibility of deflation.

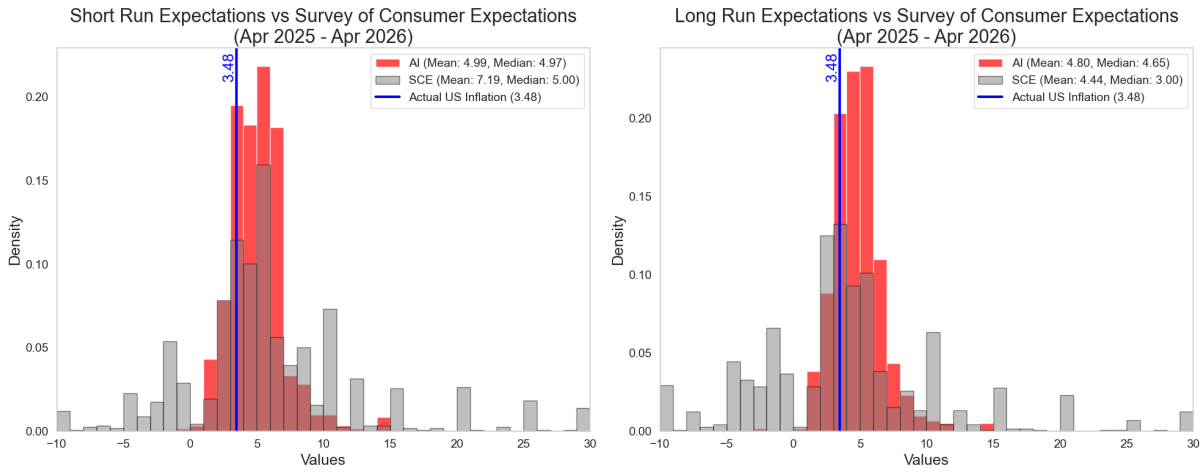


Figure 3: Distribution of responses for human and AI respondents

Note: Data from SCE is trimmed from -10 to 30 as there are many outliers outside this range.

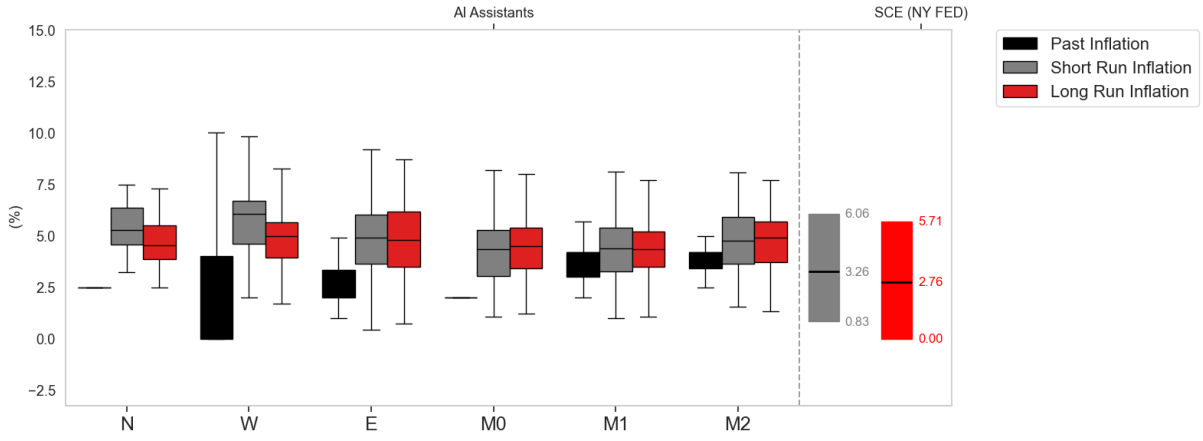


Figure 4: Inflation Expectations in Different Horizon by Assistant

Note: The right side displays inflation expectations from the SCE, showing the 25th, median, and 75th percentiles for comparison. Assistant M2 shows greater comparability, potentially due to broader data access, while Assistant N, without additional retrieval capability, exhibits less variation.

known economic surveys such as the Survey of Professional Forecasters (SPF). Table 4 presents both the mean and median inflation expectations over various time periods, allowing for a comparison between AI and human forecasts.

Table 4: Comparison with SCE and SPF

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

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

Note: SURVEY OF PROFESSIONAL FORECASTERS (SPF) IS FROM THE SECOND QUARTER 2024 REPORT AND SURVEY OF CONSUMER EXPECTATIONS (SCE) IS FROM APRIL 2024.
* 5 Year Forecast

4.2 Using Personas in Prompt

Recent studies has increasingly highlighted the significance of personas in simulating human-like behavior with large language models (Horton, 2023; Chen et al., 2024; Hu and Collier, 2024). Many studies have employed persona prompting to create AI agents with specific personality traits, backgrounds, and characteristics, generating more realistic and nuanced behaviors. While concerns exist about whether inflation expectations or other survey responses generated by LLMs accurately reflect real-world phenomena, recent studies by Fedyk et al. (2024) and Kazinnik (2023) suggest that LLMs can effectively mimic complex human behaviors in financial contexts. These findings support the view that, with appropriate setup, LLMs can provide valuable insights into human decision-making processes.

To explore the potential of using personas in this study, before conducting phase 2 and the full information treatment experiments on all agents, I compared inflation expectations across different agent types. This preliminary analysis included agents with no persona, agents living in Texas (a predominantly "red" state) or California (a predominantly "blue" state), and agents identified as Republican or Democrat. The aim was twofold: to justify the potential use of personas and demographic characteristics in prompts, and to test the hypothesis that LLMs have a robust understanding of underlying beliefs associated with different characteristics.

This approach is based on the idea that assigning specific attributes to an AI agent can yield different responses. For instance, studies have shown that when tackling complex questions, prompting an AI agent to consider itself as an expert can potentially lead to more accurate and insightful responses compared to using no such attribute (Samuel et al., 2024). Applying this concept here, I compared inflation expectations across these varied agent types. This configuration allowed for an examination of how state residency and political affiliation might influence inflation expectations,

with each added aspect facilitating comparison to the baseline no-persona condition.

As shown in Figure 5, the Republican persona consistently demonstrates higher inflation predictions than other groups, particularly in mean values for both short-run and long-run point predictions. This illustrates how political affiliation influences the AI's responses. The Democrat persona, while slightly different, often aligns more closely with the no-persona expectations, especially in short-run predictions. Interestingly, the Texas and California personas show distinct patterns, with Texas generally aligning closer to Republican expectations and California closer to Democrat expectations, reflecting the political leanings of these states. The distributions also reveal that political partisanship had a much stronger effect on agents' inflation expectations compared to their state of residence, especially in point predictions. For probability distribution questions, we observe that agents assigned probabilities across bins more uniformly, with less variation by political persona. However, for point prediction questions, the Republican agent expressed significantly higher inflation expectations than others, particularly in the long run.

These findings closely mirror the inflation expectation differences by political affiliation found in recent survey work with human subjects by Binetti et al. (2024), Stantcheva (2024), and Binder et al. (2024). The similarities between LLM and human responses may be the result of the models' training on diverse human-generated content, including economic discussions and news. However, differences could arise from the LLMs' lack of personal economic experiences or biases present in their training data. This demonstrates the potential for using persona attributes derived from real-world human data to generate realistic variation in LLM experiments. Moreover, the influence of political affiliation on inflation expectations also depends on which party is in power; subjects tend to expect higher inflation when the opposing party holds the presidency, a pattern that seems to be captured in our AI-generated responses as well.

To further validate these results, I do a comparison with data from the Survey of Consumer Expectations. The data in Table 5 shows a big difference in inflation expectations between California (a proxy for being Democrat and residing in a "blue" state) and Texas (a proxy for being Republican and residing in a "red" state), with Texas consistently showing higher expectations for both short-term (1 year ahead) and long-term (3 years ahead) horizons. This aligns with my results, where the Texas/Republican persona generally exhibited higher inflation expectations compared to the California/Democrat persona.

These comparisons between AI experiment results and the Survey of Consumer Expectations data provide compelling evidence that LLM agents with carefully crafted personas can generate inflation expectations that reflect real-world patterns. While there are certainly differences and limitations, the overall trends and relative positions

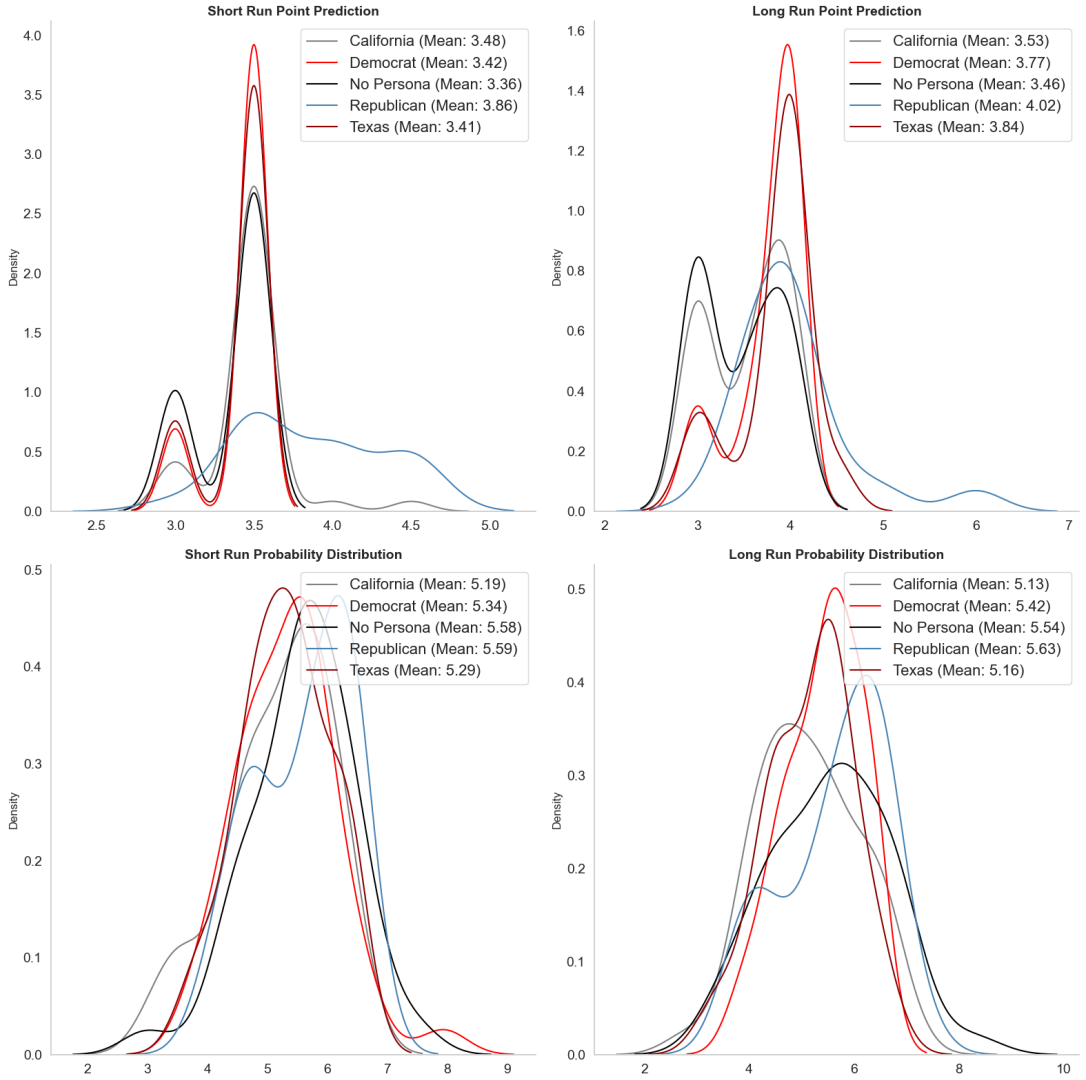


Figure 5: Inflation expectations by persona attributes

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 is 200 (40 for each category).

of different groups (e.g., Republican vs. Democrat, Texas vs. California) are largely consistent between the simulated data and actual survey responses. This opens up new possibilities for using LLM-based experiments to explore economic expectations and decision-making processes.

Table 5: Statistics Summary of Subjects in SCE (Point Prediction)

State	Metric	Mean	Std	Median
CA	1 Year Ahead	5.39	13.25	4.50
CA	3 Year Ahead	2.12	13.37	3.00
TX	1 Year Ahead	9.54	15.42	5.50
TX	3 Year Ahead	6.06	16.77	4.00

Note: Based on the available microdata, this table represents the average of the aggregated data for Texas and California from January 2023 to September 2023 of the Survey of Consumer Expectations.

5 Information Provision Treatments

In the second phase of the experiment, I investigate how specific information treatments influence the inflation expectations of AI agents, both with and without assigned personas. This phase focuses solely on the no-retrieval assistant (Assistant N) and involves a sample of 6,528 AI agents for each condition—one with personas based on the Survey of Consumer Expectations demographic data, and one without any persona attributes. As already mentioned, this design allows for a comprehensive analysis of how personal characteristics and information treatments interact to affect the expectation formation of AI agents. By comparing the impact of information provision treatments on AI and human respondents, we can shed light on whether LLMs can complement experiments involving human participants.

The results show that across almost all groups, there is a decrease in expected inflation following the provision of information treatments. This trend suggests that AI agents, much like human subjects in studies conducted by Coibion et al. (2023) and Coibion et al. (2020b), effectively integrate new information, leading to a revision of their initial expectations.[7] A particularly interesting observation is that, due to the nature of LLMs closely following the text input they receive, they demonstrate an exceptional ability to follow forward guidance. This consistency indicates that AI agents exhibit an almost perfect adherence to forward guidance and long-term predictions when provided with such information, highlighting a key characteristic of how LLMs process and respond to explicit numerical forecasts.[8]

Table 6 presents the changes in short-run (ΔSR) and long-run (ΔLR) inflation expectations across different treatments for both persona-based and non-persona experiments as well as the average treatment effects. In both setups, the prior expectations are relatively consistent across treatments. However, the posterior expectations exhibit greater variation, indicating that the treatments have a significant impact on how expectations are updated. The changes appear more pronounced in the persona-based experiments, suggesting that the inclusion of personal characteristics may lead to more

[7] This could also be because of point prediction vs probability distribution questions.
[8] Full summary statistics of this wave are provided in Appendix A.6.

nuanced processing of new information. This indicates that AI agents with assigned personas might be more sensitive to information treatments, possibly reflecting the heterogeneity seen in human subjects due to demographic factors.

Table 6: Impact of Information Treatments on Inflation Expectations

	With Persona		No Persona	
Treatment	Δ SR	ATE SR	Δ SR	ATE SR
T1: Control Group	0.22		0.19	
T2: Placebo Group	-0.72	-0.94	-0.39	-0.58
T3: Current FFR Rate	-0.84	-1.06	-0.49	-0.68
T4: Current Inflation	-0.77	-0.98	-0.16	-0.35
T5: Current Inflation + Longer Run	-0.79	-1.01	-1.09	-1.27
T6: Current Fixed-Rate 30-Year Mortgage	-0.11	-0.32	0.01	-0.18
	With Persona		No Persona	
Treatment	Δ LR	ATE LR	Δ LR	ATE LR
T1: Control Group	-1.49		-1.66	
T2: Placebo Group	-2.03	-0.54	-1.71	-0.05
T3: Current FFR Rate	-2.17	-0.69	-1.83	-0.17
T4: Current Inflation	-1.72	-0.24	-1.59	0.06
T5: Current Inflation + Longer Run	-2.66	-1.17	-2.68	-1.03
T6: Current Fixed-Rate 30-Year Mortgage	-1.13	0.35	-1.41	0.25

Note: This table represents mean changes in Short Run (SR) and Long Run (LR) inflation expectations. Δ indicates change from prior to posterior; ATE is the Average Treatment Effect calculated as ATE $= E[(\Delta Y_{\text{treatment}} - \Delta Y_{\text{control}})]$. The ATE for all variables is statistically significant at the 1% level.

5.1 Empirical Model

To empirically assess the impact of information treatments on AI agents' inflation expectations, I follow the specification used by Coibion et al. (2018, 2023), which captures how new information influences expectation formation. This model allows us to measure the influence of information interventions, as shown in experiments conducted by Armantier et al. (2016). The empirical model is formulated as follows:

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

where $E_j\pi^{\text{post}}$ and $E_j\pi^{\text{pre}}$ represent the AI agent j's inflation expectations after and before receiving the treatment, respectively. $\text{Treatment}_j^{(k)}$ is a dummy variable indicating whether AI agent j was subjected to treatment k (with $k = 2, \ldots, 6$), where treatment 1 serves as the control group. The coefficients β_k and γ_k measure the level and interaction effects of the treatments on the expectations, respectively, relative to the control group. The vector $\boldsymbol{W}_j$ includes control variables relevant to each agent, and $\boldsymbol{\Psi}$ denotes the parameters associated with these controls.

In line with Bayesian updating, agents revise their expectations by combining their prior beliefs with new information, as highlighted by Baley and Veldkamp (2023) and Coibion et al. (2018). The updated (posterior) expectation can be expressed as:

$$E_j\pi^{\text{post}} = (1 - \kappa_k)E_j\pi^{\text{pre}} + \kappa_k S^{(k)}, \tag{3}$$

where $\kappa_k \in [0, 1]$ represents the weight placed on the new information from treatment k, and $S^{(k)}$ is the signal provided by treatment k.

By rearranging Equation (1) for agents in the control group ($k = 1$) and those who receive treatment ($k > 1$), and assuming $\boldsymbol{W}_j = 0$ for simplicity, we have:

For the control group ($k = 1$):

$$E_j\pi^{\text{post}} = \alpha + \theta E_j\pi^{\text{pre}} + \epsilon_j. \tag{4}$$

For treatment groups ($k > 1$):

$$E_j\pi^{\text{post}} = (\alpha + \beta_k) + (\theta + \gamma_k)E_j\pi^{\text{pre}} + \epsilon_j. \tag{5}$$

Comparing these with the Bayesian updating formula, we can interpret:

For the control group:

$$\theta = (1 - \kappa_1), \tag{6}$$
$$\alpha = \kappa_1 S^{(1)}. \tag{7}$$

For treatment groups:

$$\theta + \gamma_k = (1 - \kappa_k), \tag{8}$$
$$\alpha + \beta_k = \kappa_k S^{(k)}. \tag{9}$$

The coefficient γ_k thus captures how the weight on the prior expectation changes in response to treatment k relative to the control group. A negative γ_k implies that agents in treatment group k place less weight on their prior beliefs and more weight on the new information compared to the control group. This is consistent with Bayesian updating, where agents adjust their expectations more significantly when the new information diverges from their prior beliefs.

Specifically, a negative γ_k indicates that for agents in treatment group k, the higher the prior expectation $E_j\pi^{\text{pre}}$, the larger the adjustment in the posterior expectation $E_j\pi^{\text{post}}$ in response to the treatment, compared to the control group. This reflects the notion that agents with more extreme prior beliefs revise their expectations more upon receiving informative signals (Armantier et al., 2016).

Therefore, in the context of Equation (2), negative γ_k coefficients suggest that the information treatments effectively lead agents to revise their expectations away from their priors, with more informative treatments resulting in larger absolute values of γ_k. This aligns with Bayesian updating principles and indicates that the treatments are impactful in shaping agents' inflation expectations (Coibion et al., 2018, 2023). The β_k coefficients capture the level effect of the treatments, representing the average shift in expectations in response to the information provided in each treatment group relative to the control group.

This empirical model helps us understand how AI agents integrate new information with their pre-existing beliefs, quantifying the adjustments made in their inflation expectations in response to different types of information treatments compared to a control group. The model is applied to both persona-based and non-persona experiments, allowing for a comprehensive analysis of how personal characteristics and information treatments interact to affect expectation formation in AI agents.

5.2 Analysis

To evaluate the effects of the information treatments, I estimate Equation (2) separately for the persona-based and non-persona experiments. Table 7 presents the regression results for both short-run (1 year ahead) and long-run (3 years ahead) inflation expectations. The coefficients θ_{SR} and θ_{LR} represent the weight placed on prior expectations for the control group. The significant and positive values of these coefficients in both experiments indicate substantial persistence in expectations, consistent with adaptive expectations models.

The consistently negative and significant γ coefficients suggest that the information treatments lead AI agents to adjust their expectations away from their priors, with more informative treatments having a larger impact. This pattern is consistent with a form of Bayesian updating, where new information causes agents to revise their prior beliefs, similar to findings by Armantier et al. (2016); Huber et al. (2023), who demonstrate that different information treatments impact inflation expectations across demographics. For instance, in the short-run expectations, Treatment 5 (Current Inflation + Longer Run) has the largest negative β coefficient (-0.5122 with persona and -0.8507 without persona), indicating that providing forward-looking information significantly influences expectation formation. This finding aligns with some studies on human subjects which show that forward guidance on interest rates can influence inflation expectations (Coibion et al., 2022). However, it's important to note that the literature has mixed results on the effectiveness of forward guidance, with its impact often depending on the type of information provided and how it is communicated (D'Acunto et al., 2022).

Comparing the results between the persona-based and non-persona experiments reveals some differences in treatment impacts. The β coefficients are generally larger in absolute terms for the no-persona condition, particularly for short-run expectations. However, the γ coefficients are often more negative in the persona-based experiment, especially for short-run expectations. This suggests that incorporating personas may lead to more sophisticated processing of information, resulting in a greater adjustment of expectations when new information interacts with prior beliefs. This observation supports recent findings that personal characteristics and demographic contexts shape how individuals interpret and respond to economic information (Weber et al., 2022).

Table 7: Regression Results for Persona and No Persona

	$E[\pi]$ **1 Year Ahead (SR)**		$E[\pi]$ **3 Years Ahead (LR)**	
	With Persona	**No Persona**	**With Persona**	**No Persona**
Intercept (Control Group)	3.9104***	3.3349***	3.0439***	2.5756***
	(0.067)	(0.042)	(0.063)	(0.055)
θ_{SR} (Control Group)	0.2112***	0.1599***	–	–
	(0.012)	(0.013)		
θ_{LR} (Control Group)	–	–	0.1539***	0.1254***
			(0.010)	(0.012)
Relative to Control:				
Population Growth (β_2)	-0.2056***	-0.1743***	0.2048***	0.3408***
	(0.076)	(0.059)	(0.071)	(0.067)
Current FFR (β_3)	-0.3382***	-0.1817***	-0.0444	0.1277*
	(0.069)	(0.054)	(0.068)	(0.075)
Current Inflation (β_4)	-0.1236*	-0.0156	0.4381***	0.5804***
	(0.066)	(0.058)	(0.083)	(0.073)
Current + Inflation Prediction (β_5)	-0.5122***	-0.8507***	-0.8385***	-0.4707***
	(0.069)	(0.061)	(0.069)	(0.061)
Mortgage Rate (β_6)	-0.0009	0.2428***	0.5451***	0.7183***
	(0.081)	(0.053)	(0.088)	(0.080)
γ_2	-0.1411***	-0.1337***	-0.1263***	-0.1022***
	(0.015)	(0.016)	(0.013)	(0.014)
γ_3	-0.1661***	-0.1494***	-0.1347***	-0.0864***
	(0.014)	(0.015)	(0.012)	(0.016)
γ_4	-0.1974***	-0.1177***	-0.1412***	-0.1328***
	(0.013)	(0.016)	(0.015)	(0.015)
γ_5	-0.1620***	-0.1090***	-0.1261***	-0.1245***
	(0.014)	(0.017)	(0.012)	(0.013)
γ_6	-0.1381***	-0.1320***	-0.1058***	-0.1149***
	(0.016)	(0.015)	(0.016)	(0.016)
R-squared	0.368	0.368	0.414	0.420
Observations	6,528	6,402	6,528	6,402

Note: I also estimated the Huber regression. As there are not many outliers the results are almost the same. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 6 visualizes the estimated coefficients from the regression models, comparing AI agent responses with human survey data. This figure illustrates the distinct impacts of each information treatment on short-run (1 year ahead) and long-run (3 years ahead) inflation expectations across both persona-based and non-persona AI agents,

as well as human respondents. It can be seen while both persona and non-persona AI agents exhibit similar patterns of response to the treatments, the magnitude of these effects often differs. For instance, Treatment 5 (Current + Inflation Prediction) exerts the most substantial negative effect on AI agents' expectations, especially in the short run, underscoring the influence of forward guidance on inflation expectations—a finding consistent with some of the existing literature on human responses to similar guidance (D'Acunto et al., 2021).

Interestingly, the mortgage rate information (Treatment 6) has a positive effect on inflation expectations among AI agents, contrasting with a negative effect observed in human survey data (Coibion et al., 2023). This difference could reflect distinct interpretative processes between AI agents and human respondents when assessing inflation-related information. Furthermore, while human survey responses often show larger changes in response to treatments—particularly with current Federal Funds Rate (FFR) and inflation information—the AI control group displays a baseline positive effect (θ), whereas human respondents in the control group tend toward negative expectations.

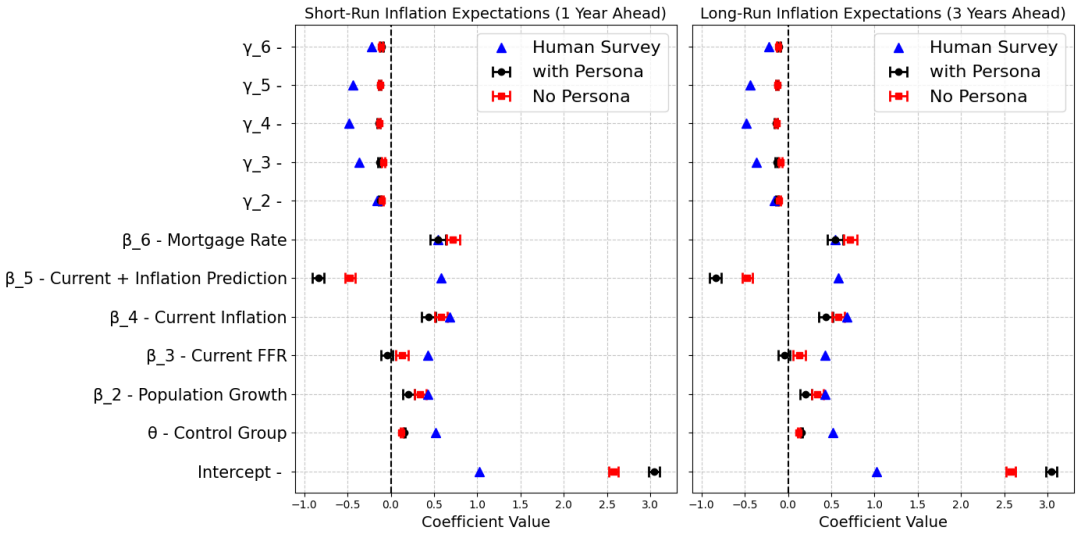


Figure 6: Estimated Coefficients from Regression Models

The differences between AI and human responses highlight important considerations for using LLMs in economic research. AI agents show more consistent and predictable reactions to information treatments, potentially due to their lack of personal biases or experiences. However, this consistency may overlook the nuanced and sometimes irrational ways humans process economic information (Coibion et al., 2022; Weber et al., 2022). It's crucial to remember that LLMs, as transformer-based models, are fundamentally designed to predict the next most likely word or sequence based on patterns in their training data, rather than truly reasoning about economic concepts (Brown et al., 2020).

The integration of reinforcement learning from human feedback (RLHF) in LLM training represents a significant advancement in aligning AI behavior with human preferences and decision-making patterns (Ouyang et al., 2022). This approach could potentially bridge the gap between AI consistency and human nuance in economic reasoning, allowing LLMs to better capture the complexities of human economic decision-making (Bai et al., 2022). However, it also raises questions about whose feedback is incorporated and how this might influence the model's economic perspectives.

The primary motivation for this paper's experiments lies in the growing importance of AI-assisted decision making across various economic domains (Korinek, 2023; Chang et al., 2024). As AI systems increasingly influence financial markets, policy analysis, and individual economic choices, it becomes critical to understand their underlying mechanisms and how they interpret economic information. This understanding is essential for predicting how AI might shape market dynamics, influence policy effectiveness, and interact with human decision-makers in mixed human-AI economic environments (Fedyk et al., 2024; Zhu et al., 2024).

While LLMs' strong responsiveness to clear signals makes them highly adaptable, it may lead to overestimating the effectiveness of certain policy communications in real-world scenarios. The incorporation of RLHF could potentially mitigate this by introducing more human-like variability and context-sensitivity in responses (Christiano et al., 2017). However, these differences still underscore the need for careful calibration when using AI models to predict or simulate human economic behavior, especially as we move towards a future where AI plays a more significant role in economic decision-making processes (Korinek, 2024; Acemoglu, 2024).

5.3 Demographics and AI Agents' Expectations

The demographic variables in the persona-based experiments (Table 8) show some interesting patterns. Higher education levels are associated with lower inflation expectations, while older age groups tend to have higher expectations. These patterns mirror some findings in human survey data (Huber et al., 2023; D'Acunto et al., 2022), suggesting that the AI agents are capturing realistic demographic influences on expectation formation.

These findings have important implications for both the development of AI models in forecasting and the design of effective communication strategies for central banks and policymakers. They suggest that AI agents, particularly those with assigned personas, can serve as valuable tools for simulating and predicting the impacts of various information dissemination strategies on public inflation expectations. However, the differences between AI and human responses highlight the need for careful calibration and interpretation when using AI models to predict human economic behavior.

	$E[\pi]$ 1 Year Ahead		$E[\pi]$ 3 Years Ahead	
	With Persona	**No Persona**	**With Persona**	**No Persona**
Intercept (Control Group)	3.9104***	–	3.0439***	–
	(0.067)		(0.063)	
Demographic Variables:				
Age Group: 18-30	–		–	
Age Group: 31-40	0.0051		0.0409	
	(0.030)		(0.029)	
Age Group: 41-50	-0.0252		0.0517*	
	(0.030)		(0.029)	
Age Group: 51-60	0.0213		0.0424	
	(0.031)		(0.029)	
Age Group: 60+	0.0681**		0.1302***	
	(0.029)		(0.028)	
Gender: Male	0.0140		0.0103	
	(0.016)		(0.016)	
Education: College	–		–	
Education: High School	0.2144***		0.2113***	
	(0.025)		(0.028)	
Education: Some College	0.1709***		0.0966***	
	(0.018)		(0.019)	
Income: Under 50k	–		–	
Income: 50k to 100k	-0.0699***		-0.0360	
	(0.021)		(0.023)	
Income: Over 100k	-0.1701***		-0.0981***	
	(0.022)		(0.025)	
R-squared	0.368	–	0.414	–
Observations	6,528	–	6,528	–

Table 8: Regression Results for Demographic Variables

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Age Group base is 18-30, Gender base is Female, Education base is College, and Income base is Under 50k. The 'No Persona' columns are left blank as the regression did not include demographic variables for this condition.

5.4 Topic Modeling

To gain deeper insights into the reasoning behind the AI agents' responses, I conducted a topic modeling analysis on the open-ended explanations provided in the final question of the survey. Using a zero-shot classification approach[9], I categorized the responses into ten predefined topics based on common themes in inflation discourse. This method allowed for a comparative analysis between persona-based and non-persona AI agents' reasoning patterns.

Figure 7 illustrates the distribution of topics across both experimental conditions. Notably, "Inflation Trends" emerged as the dominant topic for both groups. This shows that the AI agents focus on historical and current inflation patterns in forming their expectations. Interestingly, persona-based agents showed a slightly higher tendency to reference inflation trends, suggesting that demographic characteristics might influence the emphasis placed on historical data. Other significant topics included "Interest Rates" and "Federal Reserve" actions. The similarity in topic distributions

[9]Initially, I used Latent Dirichlet Allocation (LDA) to explore different topics and select the topic titles. Subsequently, I employed a zero-shot classification approach using the BART-large-mnli model. Other models yielded similar results, highlighting the effectiveness of transformer-based models in context-aware topic modeling.

between persona and no-persona conditions indicates a consistent underlying reasoning process across AI agents, regardless of assigned demographic characteristics.

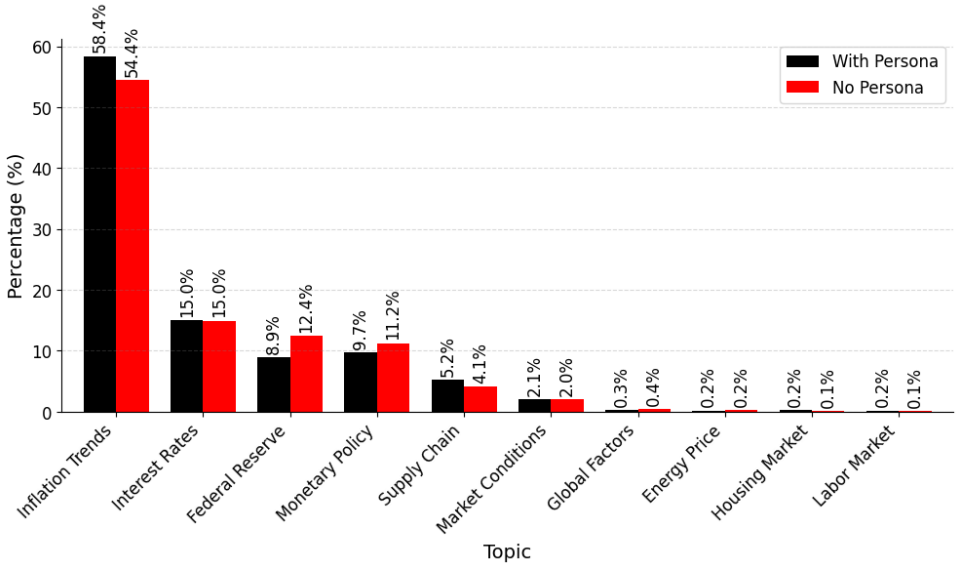


Figure 7: Distribution of Topics in AI Agents' Reasoning

This topic modeling analysis provides valuable insights into the cognitive processes of AI agents in forming inflation expectations. The consistent emphasis on key economic factors across both conditions demonstrates the AI's ability to identify and prioritize relevant information. However, the nuanced differences between persona and no-persona agents suggest that demographic characteristics can subtly influence the weighting of various factors in expectation formation, mirroring patterns observed in human survey data.

6 Extension

6.1 Reasoning Behind LLMs

OpenAI's recent release of the o1-preview and o1-mini models[10] represents a substantial development in Generative AI's ability to engage in complex reasoning processes. The o1 models, as explained in OpenAI's release notes, are designed to spend more time "thinking" before they respond, achieved through a specific version of the chain-of-thought (CoT) prompting pattern. This method, introduced by Wei et al. (2023), has been refined through a large-scale reinforcement learning algorithm that teaches the model to "think" productively using its chain of thought. Unlike previous versions, these models spend more time processing problems before responding,

[10]Released on September 12, 2024 (known as "strawberry") in two variations.

much like a person would. Through training, they learn to refine their thinking process, try different strategies, and recognize their mistakes.

These new models are particularly relevant to my study of inflation expectations. While previous experiments with GPT-4 variations provided valuable insights into how AI models form and update inflation expectations, the o1 models offer the potential to go deeper into the reasoning processes behind these expectations and address some criticisms regarding LLMs as "black boxes." For example, when asked about past and future inflation expectations, an o1 model might reveal a chain of thought comprising the stages shown in Table 9. Due to limited access to this model, I ran the experiment with five AI agents, one for each treatment group, in addition to the three pre-treatment questions.

Table 9: Different Stages Mentioned by AI Agents

Category	**Number of Topics**
Inflation Analysis	13
Probability Assessment	10
Economic Analysis	10
Data Interpretation	20
Guideline Compliance	10

This level of detail in the reasoning process provides economists with valuable insights into how different factors are weighted and integrated in forming inflation expectations. However, the o1 models also introduce new challenges. Depending on the amount of reasoning required, requests can take from a few seconds to several minutes. In this experiment, instead of only collecting numerical responses, I captured all the reasoning and chain-of-thought processes of the model. The numerical answers are shown in Table 10, and the full clustering of reasoning patterns is provided in Appendix A.8.

The o1 model's responses exhibited a higher degree of nuance and consideration of multiple economic factors compared to previous models. For instance, when asked about short-term inflation expectations after receiving the current Federal Funds Rate information (T3), the model provided a detailed explanation with several stages of thinking:

Table 10: Forecasts of GPT o1 and Time Taken per Question

Question	Participant	Forecast	Seconds
Pre-Treatment	Perceived Inflation	3.7%	10
Pre-Treatment	Probability Prediction 1 Year Ahead	2.7%	22
Pre-Treatment	Probability Prediction 3 Years Ahead	3.4%	28
Post-Treatment 1 Year Ahead	Population Growth Group	3%	29
	Current FFR	2.5%	6
	Current Inflation	2.8%	16
	Longer-Run Inflation	3.0%	6
	30-Year Mortgage Rate	3%	9
Post-Treatment 3 Years Ahead	Population Growth	3.4%	44
	Current Rate, FFR	2.5%	20
	Current Inflation	2.5%	56
	Longer-Run Inflation	2.8%	27
	30-Year Mortgage Rate	2.5%	22

Note: Probability predictions for 1 year ahead are [0.1%, 0.2%, 0.5%, 1.0%, 3.2%, 30%, 50%, 12%, 2%, 1%], and for 3 years ahead are [0.1%, 0.2%, 0.5%, 1%, 3%, 20%, 50%, 20%, 5%, 0.2%].

- **Formulating Expectations:** I'm thinking about the question regarding inflation expectations over the next 12 months, aiming to provide a numerical answer without mentioning policies or guidelines.
- **Assessing Inflation Expectations:** OK, let's think it through: predicting the U.S. inflation rate for the next 12 months involves considering the Federal Reserve's target of 2% and the latest CPI of 3.7% from August/September 2023.
- **Examining Inflation Potential:** I'm piecing together the current Federal Funds Rate and economic forecasts. Inflation over the next 12 months might be around 2–3%, with a projected 2.5% as a reasonable estimate.

> Over the next 12 months, I expect the rate of inflation to be 2.5%. (6 seconds)

This level of reasoning aligns closely with the complex decision-making processes we expect from expert economic forecasters, offering a more realistic simulation of human expectation formation. Furthermore, the o1 model's responses to information treatments showed more nuanced belief updating compared to earlier experiments. These results suggest that the o1 model's chain-of-thought capabilities could simulate more accurately how economic agents process and integrate new information into their inflation expectations. While promising, it is important to note that the use of these models in this context is still in its early stages. As we continue to refine methodologies for running experiments on LLMs and AI agents and leverage these new Generative AI capabilities, we may gain deeper insights into the formation and updating of economic expectations. This could potentially lead to more accurate economic mod-

els and inform more effective monetary policy communication strategies in an era of AI-assisted decision making.

6.2 Consistency in Responses

A common concern in experiments involving AI and language models is the consistency of their responses. In a second wave of the first phase, I used two assistants: one without retrieval capabilities (N) and another with four years of FOMC minutes data (M2). Figure 8 illustrates the distribution of responses across these two waves, showing a strong consistency between them.

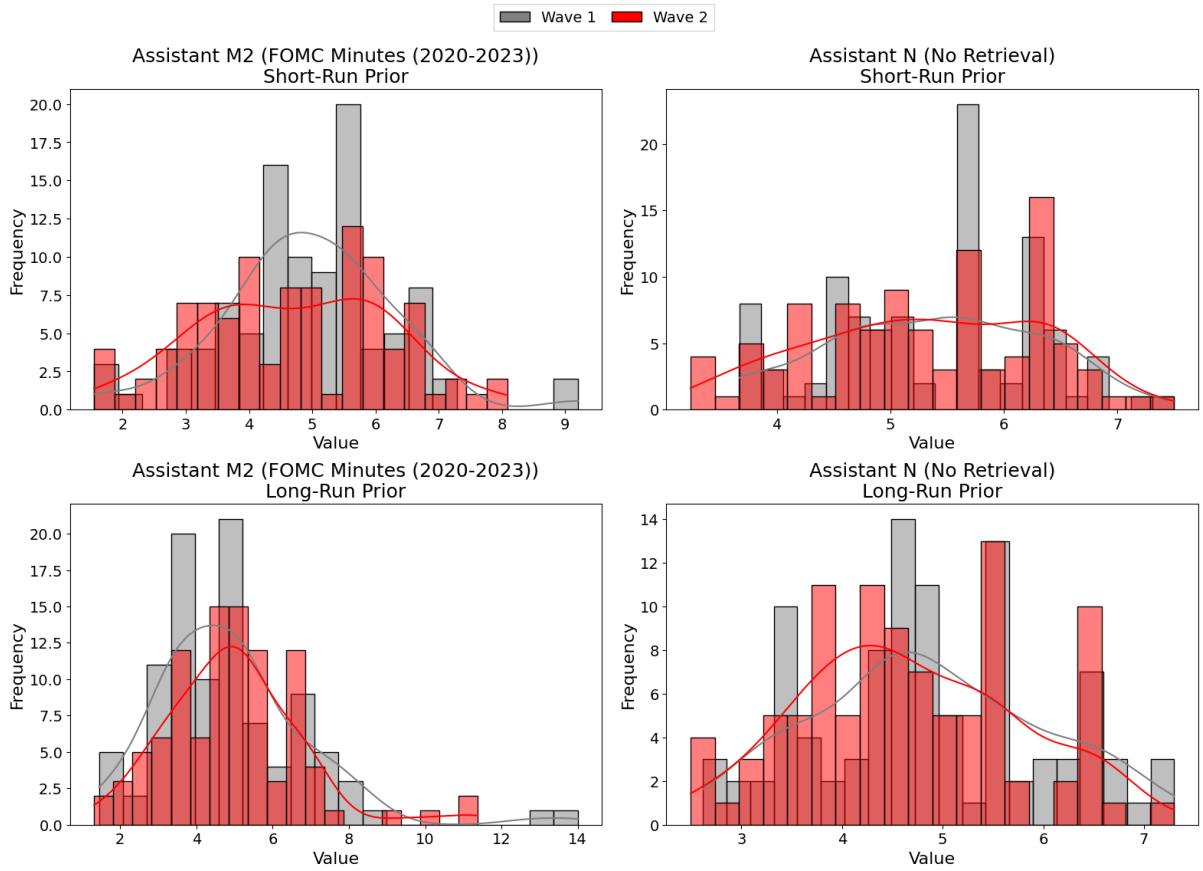


Figure 8: Consistency of Responses in two Waves

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

The strong consistency observed between the two waves has significant implications for the reliability of LLMs in economic experiments. This consistency suggests that LLMs can produce stable and reproducible results across multiple runs, a crucial factor for reproducibility. Also, this consistency demonstrates that LLM responses are time-invariant, suggesting that their economic reasoning remains stable over short periods. However, as we observe in Section 6.3, while LLMs exhibit temporal consistency, they are model-variant, with different architectures or versions producing

varying results. This reliability within models enhances the potential of LLMs as tools for economic research, allowing for more confident extrapolation of findings and enabling comparative studies across different economic scenarios.

6.3 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. This is also important in exploring the formation of inflation expectations using LLMs. GPT models are adept at generating coherent and contextually relevant text, capturing long-range dependencies, and producing extended sequences, making them ideal for tasks requiring a nuanced understanding of language context (Bommasani et al., 2021).

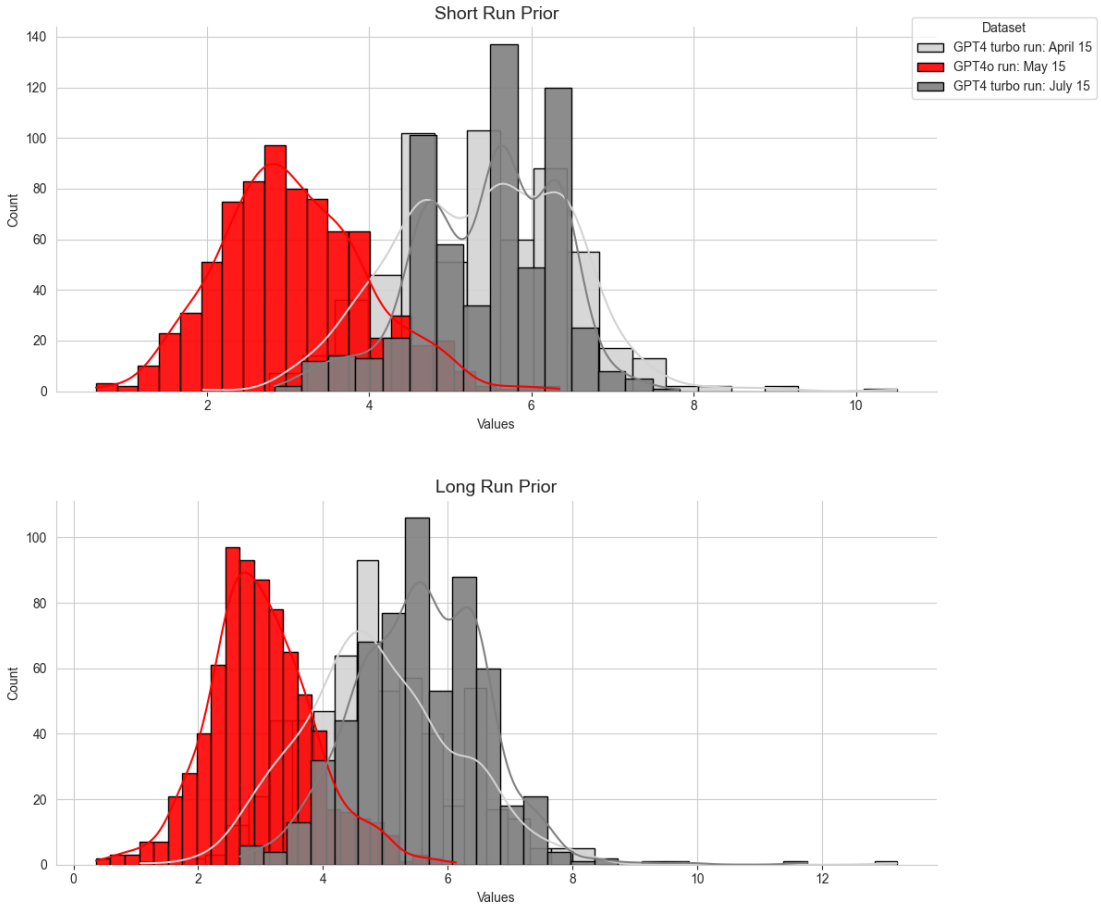


Figure 9: Short-Run and Long-Run Inflation Expectations for Different Models

Note: The figure compares responses to short-run (SR) and long-run (LR) inflation expectation questions for different GPT-4 models. Results are presented for three runs: GPT-4 turbo on April 15, GPT-4o on May 15, and GPT-4 turbo on July 15. The GPT-4o model has lower inflation expectations. However, the overall shape of the distributions remains similar across models and runs, suggesting consistent response patterns despite updates and architectural variations.

The training data used to develop LLMs plays a crucial role in shaping their capabilities and potential biases. Models trained on diverse and extensive datasets develop a broad understanding of language and its complexities, enhancing their generalizability. GPT-4, in its various versions, satisfies this requirement. Furthermore, the use of retrieval-augmented generation (RAG) necessitates a model that can effectively access external knowledge sources to generate informed and accurate responses.

Recently, OpenAI introduced GPT-4o, their latest model with an updated knowledge cutoff. This model also improved in speed and computation time, and claimed to be their best model to date. For robustness, I conducted another experiment to assess the impact of model changes. Figure 9 shows the different distributions for our two main questions based on the model. The mean has shifted to the left; however, the shape of the distribution remains the same. To determine whether this change resulted from the model or a temporal effect, I conducted an additional experiment with the GPT-4 model. The findings suggest that the change is more likely due to the model itself.

Another important aspect of the model is randomness, which can be adjusted by changing the temperature parameter. In the original run, the Assistants API did not allow temperature adjustments[11]. In a recent update, the temperature can be set between 0 and 2[12]. Adjusting the temperature affects the randomness of the model's responses, which is crucial for ensuring consistency and reliability in the experiment.

7 Discussion

Based on the findings presented in the previous sections, this study offers valuable insights into both AI agent modeling and the potential for AI-assisted decision-making in the context of inflation expectations. LLMs, such as GPT-4, trained on vast textual datasets, show an ability to generate coherent and contextually relevant responses. By processing extensive information, LLMs can model how economic agents might form inflation expectations based on available data (Bybee, 2023; Fedyk et al., 2024). The experiments in this paper show that sometimes LLMs exhibit patterns in expectation formation that resemble those of human agents.

While LLMs have significant processing power, they may still display limitations similar to humans due to biases in training data or information cutoffs (Bender et al., 2021). This suggests that while LLMs can serve as useful models for understanding expectation formation, they are not immune to the challenges faced by human agents.

[11]OpenAI announced that the default temperature was $1 \in [0, 2]$ for Assistants API.

[12]This parameter typically ranges between 0 and 1 for most models, but OpenAI sets it between 0 and 2 for the Assistants API. Testing different settings revealed consistent results between 0 and 1; however, increasing the temperature above 1 led to poor and sometimes nonsensical responses.

Integrating AI assistance into human decision-making offers a potential solution to these constraints. Rational inattention models posit that agents face costs when acquiring and processing information, leading them to optimally choose how much attention to allocate to different signals (Sims, 2003; Mackowiak and Wiederholt, 2009). Limited attention models similarly emphasize that agents have bounded cognitive capacities (Simon, 1955; Gabaix, 2014). These constraints can result in agents forming suboptimal expectations due to incomplete information processing.

By leveraging AI tools like LLMs, humans can process complex information more effectively, reducing cognitive burdens and information-processing costs. To provide a general framework for understanding AI-assisted decision-making in the context of inflation expectations, this paper proposes the following model:

Consider a human agent whose inflation expectation $\mathbb{E}_t^H[\pi_{t+1}]$ is updated with the assistance of an AI tool. The combined expectation can be modeled as:

$$\mathbb{E}_t^A[\pi_{t+1}] = (1 - \phi_t)\mathbb{E}_t^H[\pi_{t+1}] + \phi_t \mathbb{E}_t^{\text{AI}}[\pi_{t+1}] \tag{10}$$

$$\phi_t = \frac{\lambda_t \cdot \tau_t \cdot \alpha_t}{1 + c_t} \tag{11}$$

where:

- $\mathbb{E}_t^A[\pi_{t+1}]$ is the AI-assisted expectation
- $\mathbb{E}_t^H[\pi_{t+1}]$ is the human agent's initial expectation
- $\mathbb{E}_t^{\text{AI}}[\pi_{t+1}]$ is the expectation provided by the AI tool
- $\phi_t \in [0, 1]$ is a weighting factor
- $\lambda_t \in [0, 1]$ represents the agent's personal perception of the AI tool
- $\tau_t \in [0, 1]$ is the trust in the AI tool
- $\alpha_t \in [0, 1]$ is the perceived accuracy of the AI tool
- $c_t \geq 0$ is the cost of using the AI tool

This framework acknowledges that the effectiveness of AI assistance depends on both the AI's capabilities and the human agent's perception and willingness to use the tool. Higher personal perception, trust, and accuracy increase ϕ_t, leading the agent to rely more on the AI's input, while higher cost decreases ϕ_t.

By incorporating personal perception, trust, cost, and accuracy into the expectation formation process, this framework acknowledges that the effectiveness of AI assistance depends not only on the AI's capabilities but also on the human agent's perception and willingness to use the tool. This general framework provides a foundation for understanding how AI-assisted decision-making could mitigate rational inattention and limited attention in the context of inflation expectations.

Moreover, as shown in the Extension section, the integration of more advanced AI models with enhanced reasoning capabilities further improves the alignment between AI-generated expectations. This indicates that ongoing advancements in AI technology can potentially enhance the effectiveness of AI-assisted decision-making in economics. However, while AI assistance can mitigate some cognitive constraints, it is important to acknowledge potential limitations. AI tools may introduce biases if the underlying models are trained on skewed data (Bender et al., 2021). Additionally, overreliance on AI could lead to herding behavior or reduce diversity in expectations, which has implications for market dynamics (Afrouzi et al., 2023).

8 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 that the knowledge domain significantly impacts the variation in responses. LLMs with access to relevant economic data, such as Federal Open Market Committee minutes, exhibit 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 previous findings from human subject studies. Specifically, the information treatments effectively influenced the LLMs' expectations, demonstrating their sensitivity to new information in a manner similar to human respondents. The findings reveal that LLMs generally predict higher inflation than actual rates, closely mirroring the responses observed in human surveys like the Survey of Consumer Expectations. This suggests that LLMs can effectively simulate human-like reasoning in economic contexts, providing a valuable tool for pre-testing surveys and understanding expectation formation.

Furthermore, introducing a persona in the prompt results in partisan expectations, reflecting patterns observed in human surveys. For instance, LLMs assigned a Republican persona exhibited higher inflation expectations compared to those with a Democrat persona or no specified persona. This demonstrates the ability of LLMs to capture heterogeneity among economic agents, making them useful tools for modeling diverse perspectives. Conversely, as AI adoption increases, we may observe a reduction in polarization, since a unified AI source might mitigate the disparate impacts of varied news outlets on individuals' expectations. The study also underscores the significance of model selection and the chain-of-thought processes within LLMs.

Variations in underlying architectures, training data, and reasoning capabilities can significantly influence the responses generated. For example, the introduction of the o1 models with enhanced chain-of-thought capabilities showed changes in inflation expectations and provided deeper insights into the reasoning behind those expectations. This emphasizes the need for careful consideration when employing LLMs in economic research, as different models may yield different outcomes.

This experimental design demonstrates strong consistency in responses across different sessions using the same model, suggesting reliability in the LLMs' outputs. However, differences emerge when different models are used, particularly if their training data or underlying architectures differ. This finding is crucial for researchers and policymakers who plan to use LLMs for economic forecasting or policy analysis, as it highlights the importance of model validation and robustness checks. The implications of this research are multifaceted. First, it contributes to the growing body of literature on the intersection of AI and economics, showcasing how LLMs can enhance methodologies used in economic research and provide new insights into expectation formation. Second, it offers practical insights for central banks and policymakers. By using LLMs to simulate how different information treatments influence inflation expectations, policymakers can refine their communication strategies to more effectively manage public expectations.

Moreover, as AI continues to advance and become more integrated into economic and financial decision-making, understanding how AI agents form and update their expectations about key economic variables becomes increasingly important. LLMs can serve as valuable tools for policymakers to anticipate market reactions, design effective communication strategies, and enhance the precision of monetary policy.

Finally, this study opens avenues for future research. Incorporating a wider range of persona attributes—such as varying levels of financial literacy, risk preferences, or socio-economic backgrounds—could simulate a more diverse population of agents and provide deeper insights into expectation formation. Exploring the use of other AI models or hybrid approaches could further enhance our understanding of the role of AI in economic analysis.

In conclusion, this paper demonstrates the potential of integrating generative AI with survey experiments to explore and understand inflation expectations. By leveraging the capabilities of LLMs, we can gain valuable insights into the cognitive processes underlying expectation formation, improve economic forecasting, and inform the development of more effective monetary policies in an increasingly AI-driven economy, fostering the future of AI-assisted decision making.

References

D. Acemoglu. The simple macroeconomics of ai. 2024.

H. Afrouzi, S. Y. Kwon, A. Landier, Y. Ma, and D. Thesmar. Overreaction in expectations: Evidence and theory. *The Quarterly Journal of Economics*, 138(3):1713–1764, 2023.

H. Afrouzi, A. Dietrich, K. Myrseth, R. Priftis, and R. Schoenle. Inflation preferences. 2024.

E. Akata, L. Schulz, J. Coda-Forno, S. J. Oh, M. Bethge, and E. Schulz. Playing repeated games with large language models. *arXiv preprint arXiv:2305.16867*, 2023.

I. Aldasoro, O. Armantier, S. Doerr, L. Gambacorta, and T. Oliviero. Survey evidence on gen ai and households: job prospects amid trust concerns. 2024.

O. Armantier, S. Nelson, G. Topa, W. van der Klaauw, and B. Zafar. The price is right: Updating inflation expectations in a randomized price information experiment. *The Review of Economics and Statistics*, 98(3):503–523, 2016.

O. Armantier, G. Topa, W. Van der Klaauw, and B. Zafar. An overview of the survey of consumer expectations. *Economic Policy Review*, (23-2):51–72, 2017.

Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

I. Baley and L. Veldkamp. Bayesian learning. In *Handbook of economic expectations*, pages 717–748. Elsevier, 2023.

R. Batista and J. Ross. Words that work: Using language to generate hypotheses. 2024.

E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623. ACM, 2021.

B. S. Bernanke and K. N. Kuttner. What explains the stock market's reaction to federal reserve policy? *The Journal of finance*, 60(3):1221–1257, 2005.

A. Bick, A. Blandin, and D. J. Deming. The rapid adoption of generative ai. 2024.

C. C. Binder, R. Kamdar, and J. M. Ryngaert. Partisan expectations and covid-era inflation. *Journal of Monetary Economics*, page 103649, 2024.

A. Binetti, F. Nuzzi, and S. Stantcheva. People's understanding of inflation. 2024.

A. S. Blinder, M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen. Central bank communication and monetary policy: A survey of theory and evidence. *Journal of economic literature*, 46(4):910–945, 2008.

V. R. Boctor, O. Coibion, Y. Gorodnichenko, and M. Weber. On eliciting subjective probability distributions of expectations. 2024.

R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

J. Brand, A. Israeli, and D. Ngwe. Using gpt for market research. *Available at SSRN 4395751*, 2023.

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

T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners, 2020.

M. A. Burke and A. Ozdagli. Household inflation expectations and consumer spending: evidence from panel data. *Review of Economics and Statistics*, 105(4):948–961, 2023.

L. Bybee. The ghost in the machine: Generating beliefs with large language models. *arXiv preprint arXiv:2305.02823*, 2023.

B. Candia, O. Coibion, and Y. Gorodnichenko. Communication and the beliefs of economic agents. 2020.

S. Chang, A. Kennedy, A. Leonard, and J. A. List. 12 best practices for leveraging generative ai in experimental research. 2024.

G. Charness, B. Jabarian, and J. A. List. Generation next: Experimentation with ai. 2023.

C. Chen, B. Yao, Y. Ye, D. Wang, and T. J.-J. Li. Evaluating the llm agents for simulating humanoid behavior. 2024.

P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei. Deep reinforcement learning from human preferences. In *Advances in neural information processing systems*, pages 4299–4307, 2017.

O. Coibion, Y. Gorodnichenko, and S. Kumar. How do firms form their expectations? new survey evidence. *American Economic Review*, 108(9):2671–2713, 2018.

O. Coibion, Y. Gorodnichenko, S. Kumar, and M. Pedemonte. Inflation expectations as a policy tool? *Journal of International Economics*, 124:103297–103297, 2020a.

O. Coibion, Y. Gorodnichenko, and T. Ropele. Inflation expectations and firm decisions: New causal evidence. *The Quarterly Journal of Economics*, 135(1):165–219, 2020b.

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

O. Coibion, D. Georgarakos, Y. Gorodnichenko, and M. Weber. Forward guidance and household expectations. *Journal of the European Economic Association*, pages 2131–2171, 2023.

F. D'Acunto, U. Malmendier, J. Ospina, and M. Weber. Exposure to grocery prices and inflation expectations. *Journal of Political Economy*, 129(5):1615–1639, 2021.

F. D'Acunto and M. Weber. Why survey-based subjective expectations are meaningful and important. *Annual Review of Economics*, 16, 2024.

F. D'Acunto, D. Hoang, and M. Weber. Managing households' expectations with unconventional policies. *The Review of Financial Studies*, 35(4):1597–1642, 2022.

F. D'Acunto, E. Charalambakis, D. Georgarakos, G. Kenny, J. Meyer, and M. Weber. Household inflation expectations: An overview of recent insights for monetary policy. 2024.

J. Echterhoff, Y. Liu, A. Alessa, J. McAuley, and Z. He. Cognitive bias in high-stakes decision-making with llms. *arXiv preprint arXiv:2403.00811*, 2024.

S. Eusepi and B. Preston. Central bank communication and expectations stabilization. *American Economic Journal: Macroeconomics*, 2(3):235–271, 2010.

M. Faria-e Castro and F. Leibovici. Artificial intelligence and inflation forecasts. 2023.

A. Fedyk, A. Kakhbod, P. Li, and U. Malmendier. Chatgpt and perception biases in investments: An experimental study. *Available at SSRN 4787249*, 2024.

X. Gabaix. A behavioral new keynesian model. 2014.

Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, and H. Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023.

F. Guo. Gpt in game theory experiments. *arXiv preprint arXiv:2305.05516*, 2023.

I. Haaland, C. Roth, and J. Wohlfart. Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40, 2023.

A. L. Hansen, J. J. Horton, S. Kazinnik, D. Puzzello, and A. Zarifhonarvar. Simulating the survey of professional forecasters. Working Paper, 2024.

B. Heydari and N. Lorè. Strategic behavior of large language models: Game structure vs. contextual framing. *preprint*, 2023.

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

T. Hu and N. Collier. Quantifying the persona effect in llm simulations. *arXiv preprint arXiv:2402.10811*, 2024.

S. J. Huber, D. Minina, and T. Schmidt. *The pass-through from inflation perceptions to inflation expectations*. Number 17/2023. Deutsche Bundesbank Discussion Paper, 2023.

N. Immorlica, B. Lucier, and A. Slivkins. Generative ai as economic agents, 2024.

J. H. Jiang, R. Kamdar, K. Lu, and D. Puzzello. How do households respond to expected inflation? an investigation of transmission mechanisms. 2024.

E. Karger, H. Bastani, C. Yueh-Han, Z. Jacobs, D. Halawi, F. Zhang, and P. E. Tetlock. Forecastbench: A dynamic benchmark of ai forecasting capabilities. *arXiv preprint arXiv:2409.19839*, 2024.

S. Kazinnik. Bank run, interrupted: Modeling deposit withdrawals with generative ai. *SSRN Electronic Journal*, October 2023. doi: 10.2139/ssrn.4656722.

J. Kim, M. Kovach, K.-M. Lee, E. Shin, and H. Tzavellas. Learning to be homo economicus: Can an llm learn preferences from choice. *arXiv preprint arXiv:2401.07345*, 2024.

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

A. Korinek. Economic policy challenges for the age of ai. 2024.

L. Kwok, M. Bravansky, and L. D. Griffin. Evaluating cultural adaptability of a large language model via simulation of synthetic personas. *arXiv preprint arXiv:2408.06929*, 2024.

N. Li, C. Gao, Y. Li, and Q. Liao. Large language Model-Empowered agents for simulating macroeconomic activities. *arXiv preprint arXiv:2310.10436*, 2023.

R. Liu, J. Wei, F. Liu, C. Si, Y. Zhang, J. Rao, S. Zheng, D. Peng, D. Yang, D. Zhou, et al. Best practices and lessons learned on synthetic data for language models. *arXiv preprint arXiv:2404.07503*, 2024.

B. Mackowiak and M. Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803, 2009.

B. S. Manning, K. Zhu, and J. J. Horton. Automated social science: Language models as scientist and subjects. 2024.

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

L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023.

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

N. Raman, T. Lundy, S. Amouyal, Y. Levine, K. Leyton-Brown, and M. Tennenholtz. Steer: Assessing the economic rationality of large language models, 2024.

V. Samuel, H. P. Zou, Y. Zhou, S. Chaudhari, A. Kalyan, T. Rajpurohit, A. Deshpande, K. Narasimhan, and V. Murahari. Personagym: Evaluating persona agents and llms, 2024.

H. A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118, 1955.

C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3): 665–690, 2003.

S. Stantcheva. Why do we dislike inflation? 2024.

L. Tjuatja, V. Chen, S. T. Wu, A. Talwalkar, and G. Neubig. Do llms exhibit human-like response biases? a case study in survey design. *arXiv preprint arXiv:2311.04076*, 2023.

M. Tranchero, C.-F. Brenninkmeijer, A. Murugan, and A. Nagaraj. Theorizing with large language models.

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

M. Weber, F. D'Acunto, Y. Gorodnichenko, and O. Coibion. The subjective inflation expectations of households and firms: Measurement, determinants, and implications. *Journal of Economic Perspectives*, 36(3):157–184, 2022.

J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023.

J. X. Wu, Y. Wu, K.-Y. Chen, and L. Hua. Building socially intelligent ai systems: Evidence from the trust game using artificial agents with deep learning. *Management Science*, 69(12):7236–7252, 2023.

J.-Q. Zhu, H. Yan, and T. L. Griffiths. Language models trained to do arithmetic predict human risky and intertemporal choice, 2024.

A Appendix

A.1 Notes on Implementations

In my study, I used different versions of the GPT-4 model to see how they would answer survey questions. Each part of the experiment, or "thread," treated a single survey filled out by an AI agent independently. However, the same AI agent answered all the questions in that survey, just like how a person would answer all the questions in a human survey. I noted the details of the models used, the tokens processed, and the dates of the experiments in Table A.11.

ID	Model	Experiment	Token	Date	Obs	Knowledge Cutoff
1	GPT 4 Turbo	Knowledge Source (Phase 1)	7,078,126	Apr 14, 2024	600	2023-01-12
2	GPT 4 Turbo	Pilot (Phase 2)	14,291,337	Apr 15, 2024	2400	2023-01-12
3	GPT 4 Turbo	Model Selection	10,791,991	Jul 15, 2024	600	2023-01-12
4	GPT 4o	Model Selection	5,447,033	May 15, 2024	600	2024-05-01
5	GPT 4o	Basic Persona	322,308	May 22, 2024	600	2024-05-01
6	GPT 4omini	Main Survey Experiments	88,320,112	Sep 18-22, 2024	5628 × 2	2024-07-18
7	o1 preview	Reasoning	—	Sep 14, 2024	30	2023-01-10

Table A.11: Details on Models Used in the Experiment

The implementation process involved several key considerations and strategies:

1. **Model Selection:** I utilized various GPT-4 versions, as they represent state-of-the-art LLMs.

2. **Temperature:** For all experiments in this paper, I used the default temperature of 1, which lies within the range [0, 2] and is commonly employed in many real-world applications. I also conducted experiments with temperatures lower and higher than 1. As the temperature decreases below 1, the distribution of responses becomes narrower, with the mean and median remaining almost unchanged but exhibiting less variation. Conversely, as the temperature increases above 1, responses become much less reliable, sometimes generating non meaningful text. This is why many real-world applications, following best practices, use the default temperature.

3. **Prompts:** I crafted system prompts with neutral language to guide the AI's responses. To ensure robust results, I experimented with prompt variations. I found that imprecise or overly broad prompts often led to the AI stating it couldn't predict the future or offering similar evasive responses. This sensitivity to context is characteristic of transformer models. Consequently, I designed prompts for both numerical answers and open-ended explanations, iterating multiple times to achieve the right balance of clarity and specificity.

4. **Knowledge Access:** For the first phase, I implemented Retrieval-Augmented Generation (RAG) using API Assistants. This allowed models to access relevant economic data during queries. RAG offered a cost-effective method to combine the models' general capabilities with specific economic information. The process involved feeding PDF files downloaded from FOMC and other sources to the API assistant, as illustrated in Figure 1. OpenAI's proprietary embedding and indexing system then identified relevant information for context-aware responses from the AI agent.

5. **Structured Responses:** I established a system for structured numerical responses to inflation expectation questions, as shown in Section A.2. This approach ensured AI outputs aligned with traditional survey data formats, facilitating easier analysis and comparison with human surveys. Additionally, I incorporated open-ended questions to gain deeper insights into the AI's reasoning processes.

6. **Adjustments:** For probability questions, less than 0.5 percent of the observations did not sum to 100 percent. In these cases, I applied a weighted adjustment to normalize the totals to 100 percent.

A.2 Prompts

To ensure clear and consistent answers, I designed straightforward prompts guiding AI agents on response types and formats. These included instructions for numerical percentage estimates and specific formats like single numbers or lists of percentages summing to 100%. By providing clear examples and using simple, neutral language, I ensured accurate instruction following. This approach helped the collection of consistent and comparable data across different AI agents and treatments.

System Prompt

The GPT is designed to answer users' questions about the expectations for future inflation. It responds with numerical percentage estimates, reflecting its opinions and anticipations about inflation while acknowledging the inherent uncertainty in such forecasts. 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 (Only for Retrieval Mode)

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 ____ %."

A.3 More on the Survey

Table A.12 lists all the survey questions used in the experiment. The questions are designed to elicit both point predictions and probability distributions of inflation expectations, both in the short run and long run.

Table A.12: Survey Questions

Q_id	Question
Q1_I	**Last 12 month Point Prediction (Pre-Treatment)** 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	**Next 12 month Probability Distribution (Pre-Treatment)** 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%] • the rate of deflation (opposite of inflation) will be 12% or more _____ • the rate of deflation (opposite of inflation) will be between 8% and 12% _____ • the rate of deflation (opposite of inflation) will be between 4% and 8% _____ • the rate of deflation (opposite of inflation) will be between 2% and 4% _____ • the rate of deflation (opposite of inflation) will be between 0% and 2% _____ • the rate of inflation will be between 0% and 2% _____ • the rate of inflation will be between 2% and 4% _____ • the rate of inflation will be between 4% and 8% _____ • the rate of inflation will be between 8% and 12% _____ • the rate of inflation will be 12% or more_____ Give your answer as a list like this: [...%,...%,...%,...%,...%,...%,...%,...%,...%,...%]

Continued on next page

Table A.12 – continued from previous page

Q_id	Question
Q3_I	**Longer-run Probability Distribution (Pre-Treatment)** 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%] • the rate of deflation (opposite of inflation) will be 12% or more _____ • the rate of deflation (opposite of inflation) will be between 8% and 12% _____ • the rate of deflation (opposite of inflation) will be between 4% and 8% _____ • the rate of deflation (opposite of inflation) will be between 2% and 4% _____ • the rate of deflation (opposite of inflation) will be between 0% and 2% _____ • the rate of inflation will be between 0% and 2% _____ • the rate of inflation will be between 2% and 4% _____ • the rate of inflation will be between 4% and 8% _____ • the rate of inflation will be between 8% and 12% _____ • the rate of inflation will be 12% or more_____ Give your answer as a list like this: [...%,...%,...%,...%,...%,...%,...%,...%,...%,...%]
	Randomly assigned to one of the group in Table A.13.
Q1_F	**Next 12 month Point Prediction (Post-Treatment)** 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	**Longer-run Point Prediction (Post-Treatment)** 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 __ %.
Q3_F	**Open-ended Question** Please provide a brief explanation for your answers in 1-2 sentences.

Note: "I" indicates initial questions and "F" indicates follow-up questions.

Table A.13 outlines the control and treatment groups used in the experiment. Each AI agent after the first sets of initial questions was randomly assigned to one of these groups. The data used in this part is from FOMC's Summary of Economic Projections.

Table A.13: Control and Treatment Groups

id	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	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.

A.4 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 remain 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

A.5 Additional Observations

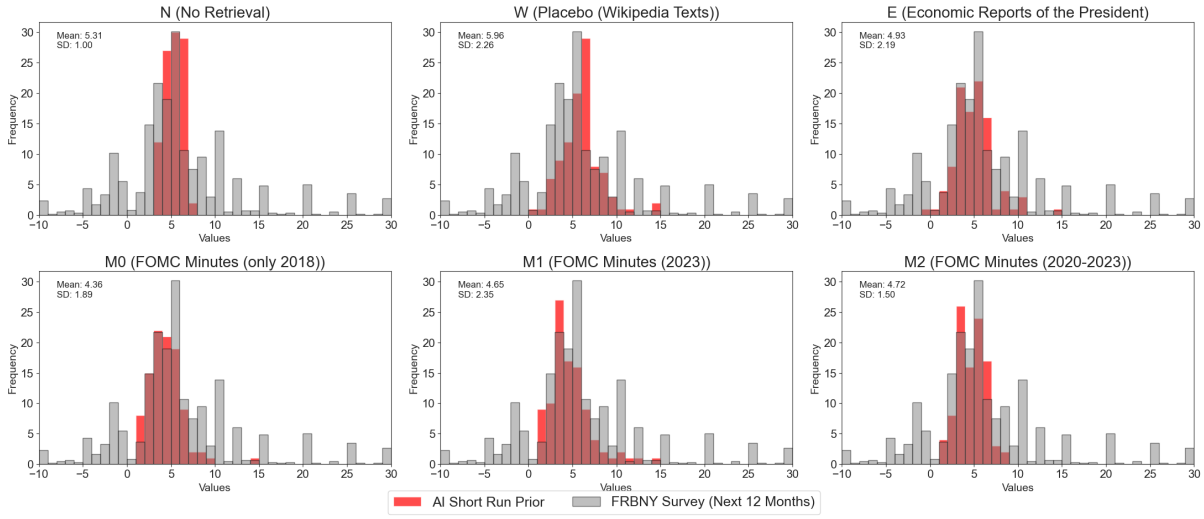


Figure A.10: Distribution of AI responses per assistant type (1 year ahead)

Note: The gray area is the NY FED Consumer Expectation Survey (5-2023), scaled down for comparison and also trimmed for the range between -10 and 30.

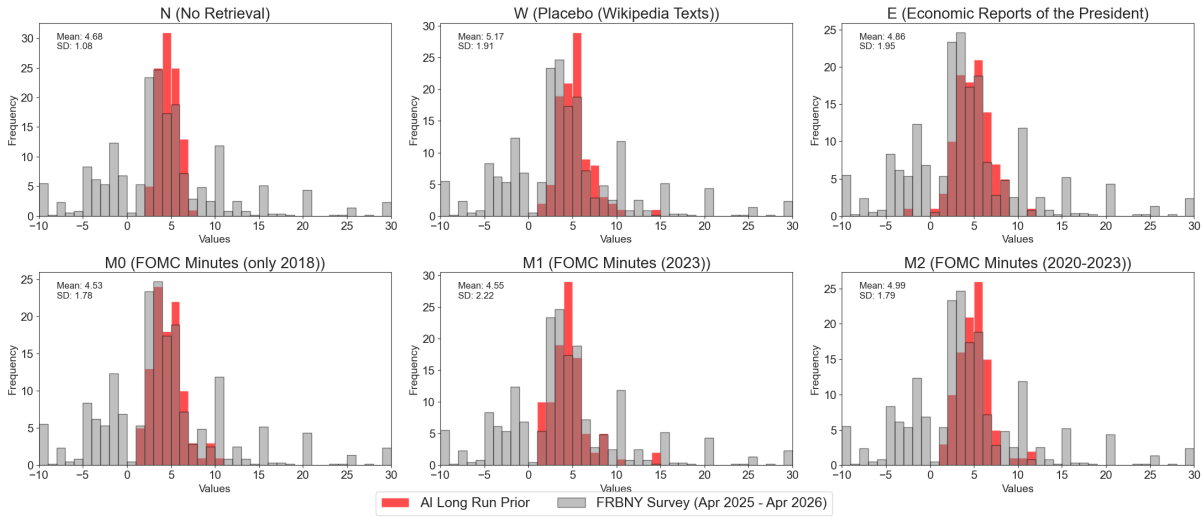


Figure A.11: Distribution of AI responses per assistant type (3 year ahead)

Note: The gray area is the NY FED Consumer Expectation Survey (5-2023), scaled down for comparison and also trimmed for the range between -10 and 30.

A.6 Additional Result of Persona Prompting

A.7 Additional Result of Persona Prompting

In the experiment, I wanted to see how giving different personas to the AI agents would affect their inflation expectations. I assigned various characteristics to the AI agents, such as being from California or Texas, or identifying as Democrat or Republican. The tables above show the summary statistics of the inflation expectations made by these AI agents with different personas.

Table A.15 presents the average inflation expectations for each persona group. For instance, AI agents with a Republican persona tend to predict higher inflation rates compared to those with a Democrat persona or no persona at all. This suggests that the assigned persona influences the AI's responses, mirroring how personal beliefs or regional differences might affect human expectations.

Table A.15: Summary Statistics of Basic Persona Experiment

Category	Variable	Mean	Median	Std Dev	Min	Max
California	Short Run Point Prediction	3.48	3.50	0.25	3.00	4.50
	Long Run Point Prediction	3.53	3.75	0.43	3.00	4.00
	Short Run Probability Distribution	5.19	5.48	0.86	3.19	6.35
	Long Run Probability Distribution	5.13	5.06	0.96	2.84	6.95
Democrat	Short Run Point Prediction	3.43	3.50	0.18	3.00	3.50
	Long Run Point Prediction	3.77	4.00	0.36	3.00	4.00
	Short Run Probability Distribution	5.34	5.49	0.82	3.84	7.94
	Long Run Probability Distribution	5.42	5.60	0.71	3.84	6.35
No Persona	Short Run Point Prediction	3.36	3.50	0.23	3.00	3.50
	Long Run Point Prediction	3.46	3.50	0.43	3.00	4.00
	Short Run Probability Distribution	5.58	5.63	0.86	2.99	7.50
	Long Run Probability Distribution	5.54	5.60	1.11	3.41	8.28
Republican	Short Run Point Prediction	3.86	4.00	0.45	3.00	4.50
	Long Run Point Prediction	4.02	4.00	0.61	3.00	6.00
	Short Run Probability Distribution	5.59	5.65	0.80	3.99	6.70
	Long Run Probability Distribution	5.63	5.75	0.99	3.76	7.30
Texas	Short Run Point Prediction	3.41	3.50	0.19	3.00	3.50
	Long Run Point Prediction	3.84	4.00	0.41	3.00	4.50
	Short Run Probability Distribution	5.29	5.35	0.73	3.70	6.35
	Long Run Probability Distribution	5.16	5.39	0.81	3.24	6.70

Table A.16 compares the inflation expectations between AI agents with personas and those without for the simulation of SCE panel. The results indicate that incorporating personas leads to higher average inflation expectations in both the short run and the long run. This means that the AI agents with assigned personas anticipate more inflation than those without any specific persona.

Table A.16: Summary Statistics of Inflation Expectations

Stat	Dataset	P_Inflation	Short-run Prior	Long-run Prior	Short-run Post.	Long-run Post.
Mean	With Persona	5.047	4.591	5.297	4.091	3.431
	No Persona	3.901	3.691	4.757	3.368	2.945
SD	With Persona	0.748	2.244	2.090	0.783	0.838
	No Persona	0.787	1.781	1.565	0.691	0.631
Min	With Persona	3.000	-1.300	-0.400	2.000	2.000
	No Persona	2.000	-0.100	-0.220	2.000	2.000
Median	With Persona	5.000	4.400	5.000	4.000	3.000
	No Persona	3.500	2.727	4.300	3.500	3.000
Max	With Persona	8.000	10.560	10.640	7.000	6.000
	No Persona	8.500	10.840	10.780	6.000	5.000
N	With Persona	6,528	6,528	6,528	6,528	6,528
	No Persona	6,402	6,528	6,528	6,402	6,402

Table A.17 shows the demographic distribution of the personas I used, based on the microdata from the Survey of Consumer Expectations.

Table A.17: Demographic Distribution of Persona Used from the SCE

Variable	Category	Distribution
Gender	Male	3,297 (50.5%)
	Female	3,233 (49.5%)
Marital Status	Married	4,117 (63.0%)
	Not Married	2,413 (37.0%)
Age	Under 40	2,071 (31.7%)
	40 to 60	2,602 (39.8%)
	Over 60	1,857 (28.5%)
Education	High School	786 (12.0%)
	Some College	2,068 (31.7%)
	College	3,676 (56.3%)
Income	Under $50k	2,026 (31.0%)
	$50k to $100k	2,236 (34.2%)
	Over $100k	2,268 (34.8%)
Total	(Unique Participants from 01/2020 to 09/2023)	6,528

Figures A.12 and A.13 illustrate the distribution of posterior inflation expectations for both short-run (1 year ahead) and long-run (3 years ahead) horizons, comparing cases with and without persona attributes. Each one shows the density of posterior expectations across different treatment groups. In the short-run, we observe more varied distributions, particularly for the persona-based scenario, indicating that different information treatments lead to diverse expectation updates. The long-run expectations show more convergence, especially in the non-persona case, suggesting that AI agents tend to align their long-term inflation expectations more closely regardless of the information provided. Notably, the "Current Inflation + Longer Run" treatment (T5) consistently produces a more concentrated distribution in both time horizons and groups, implying that this type of forward guidance has a strong anchoring effect on AI agents' expectations.

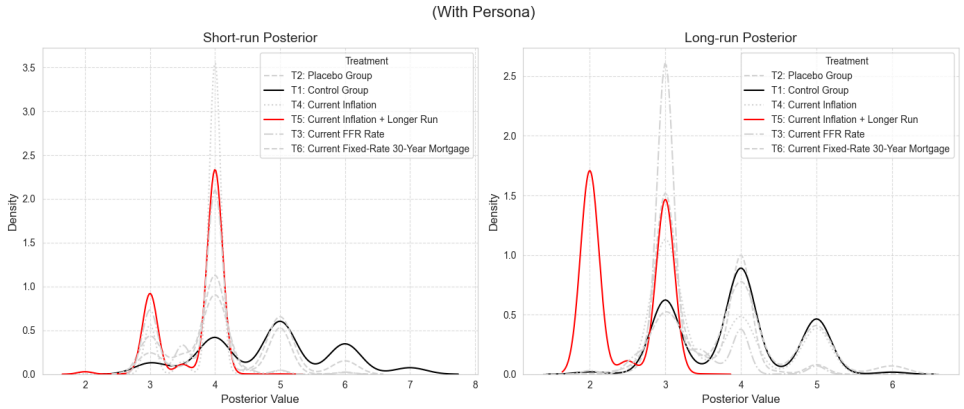


Figure A.12: Distribution of Posterior for Short-run(1 Year Ahead)

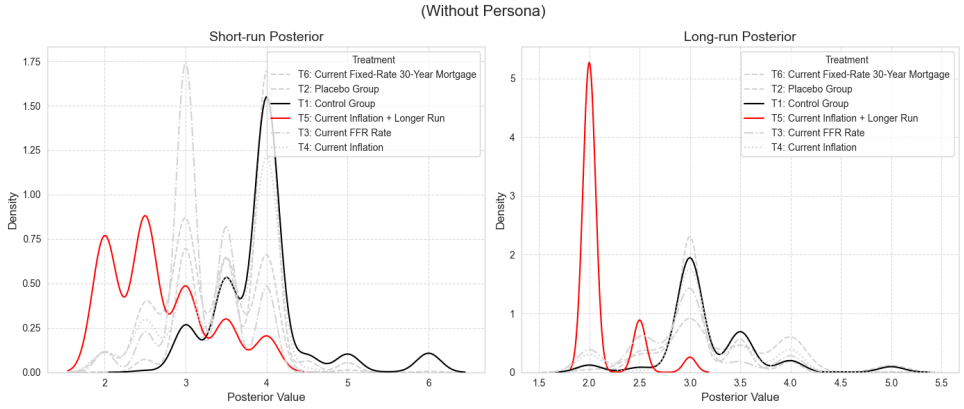


Figure A.13: Distribution of Posterior for Long-run(3 Year Ahead)

Figure A.14 is showing how AI agents with different demographic personas, simulated on SCE participants, form varying inflation expectations. The posterior expectations are more concentrated compared to the prior expectations, indicating a great convergence in the AI agents' expectations after receiving additional information.

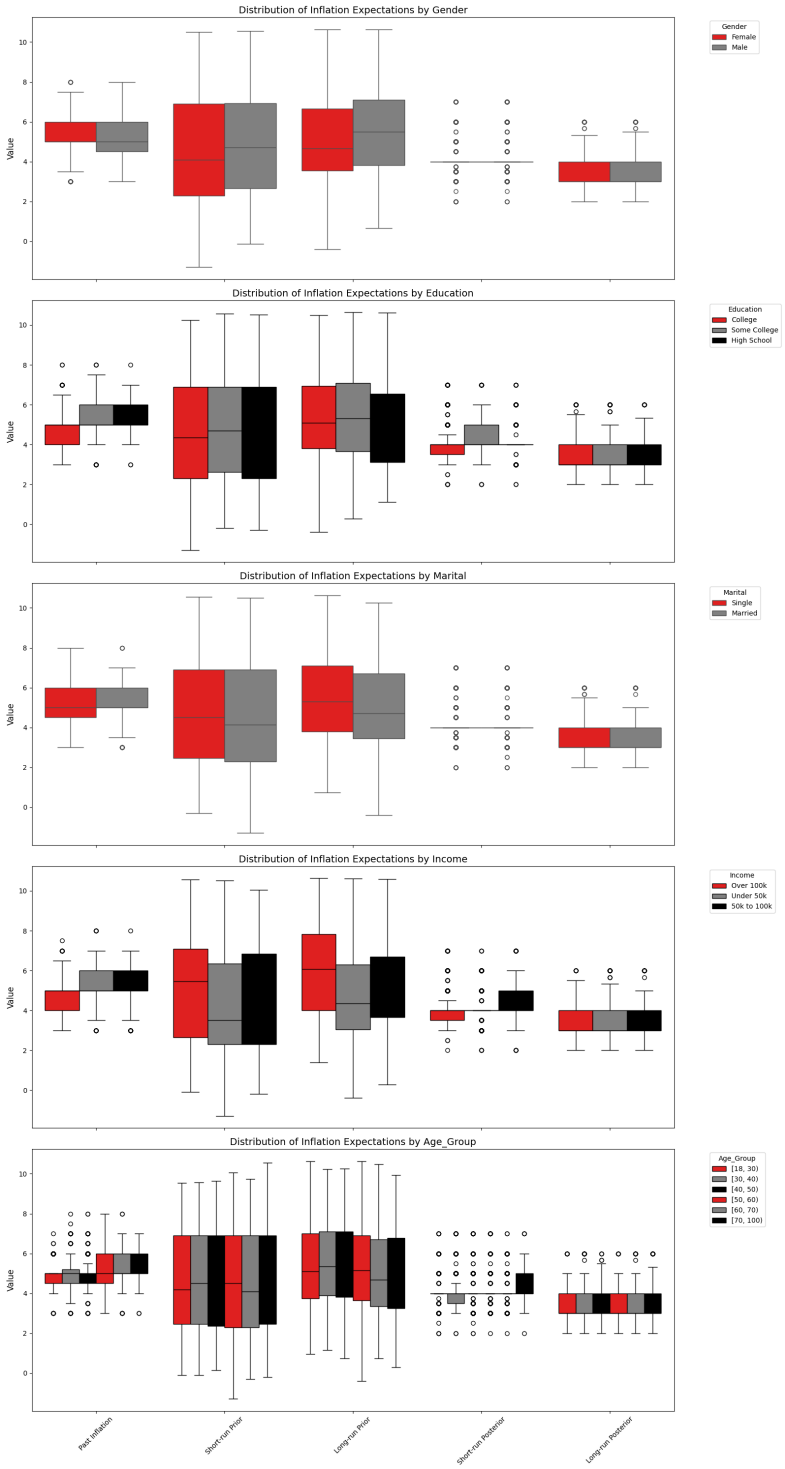


Figure A.14: Expectations Per Category of Demographic Characteristics

Table A.18: Regression for the Determinants of Inflation Perceptions/Expectations

	Dependent Variable		
	Past Inflation	Short-Run Expectation	Long-Run Expectation
Constant	5.133***	3.777***	4.859***
	(0.109)	(0.330)	(0.308)
Age	-0.008**	-0.023**	-0.024**
	(0.004)	(0.012)	(0.011)
Age Squared	0.0001***	0.0002*	0.0002
	(3.72e-05)	(0.0001)	(0.0001)
Gender (Male)	-0.015	0.347***	0.245***
	(0.019)	(0.056)	(0.052)
Marital Status (Married)	-0.051**	0.185***	0.208***
	(0.021)	(0.063)	(0.058)
Education: High School	0.288***	0.185**	0.102
	(0.029)	(0.092)	(0.085)
Education: Some College	0.208***	0.479***	0.369***
	(0.021)	(0.063)	(0.059)
Income: Over $100k	-0.099***	0.742***	0.715***
	(0.023)	(0.070)	(0.065)
Income: Under $50k	-0.020	-0.334***	-0.505***
	(0.024)	(0.071)	(0.066)
Census Region:			
East North Central	0.037	0.142	0.091
	(0.048)	(0.137)	(0.125)
East South Central	0.102*	0.596***	0.490***
	(0.057)	(0.166)	(0.150)
Mid-Atlantic	0.053	0.156	0.142
	(0.048)	(0.141)	(0.128)
Mountain	0.031	0.443***	0.377***
	(0.053)	(0.154)	(0.140)
Pacific	-0.016	0.370***	0.326**
	(0.049)	(0.140)	(0.127)
South Atlantic	-0.000	0.486***	0.410***
	(0.047)	(0.135)	(0.123)
West North Central	0.105**	0.073	0.094
	(0.052)	(0.152)	(0.138)
West South Central	0.064	0.449***	0.428***
	(0.052)	(0.149)	(0.135)
Observations	6,528	6,528	6,528
R-squared	0.045	0.049	0.060

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Reference groups: Education - College degree; Income - $50k to $100k; Census Region - New England. [a]

[a] The regression specification is: $Y_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Age}_i^2 + \beta_3 \text{Gender}_i + \beta_4 \text{Marital}_i + \sum_{j=1}^{2} \gamma_j \text{Education}_{ji} + \sum_{k=1}^{2} \delta_k \text{Income}_{ki} + \sum_{m=1}^{8} \theta_m \text{CensusRegion}_{mi} + \epsilon_i$, where Y_i represents the dependent variable (inflation perception or expectation) for individual i.

I estimated the same model for both September 2023 and December 2023, which is the most recent available microdata, to compare the coefficients of determinants for both short-run and long-run expectations. We need to remember that the Survey of Consumer Expectations (SCE) asked for both point predictions and probability distributions (first for short-run and then for long-run expectations). In my survey, I can only use the pre-treatment questions, which are the probability distributions. Also, the timing, based on the lack of similar data from the microdata, is not completely correct, but it could be a starting point to see the determinants of expectations for AI agents versus human participants. (See Figure A.15)

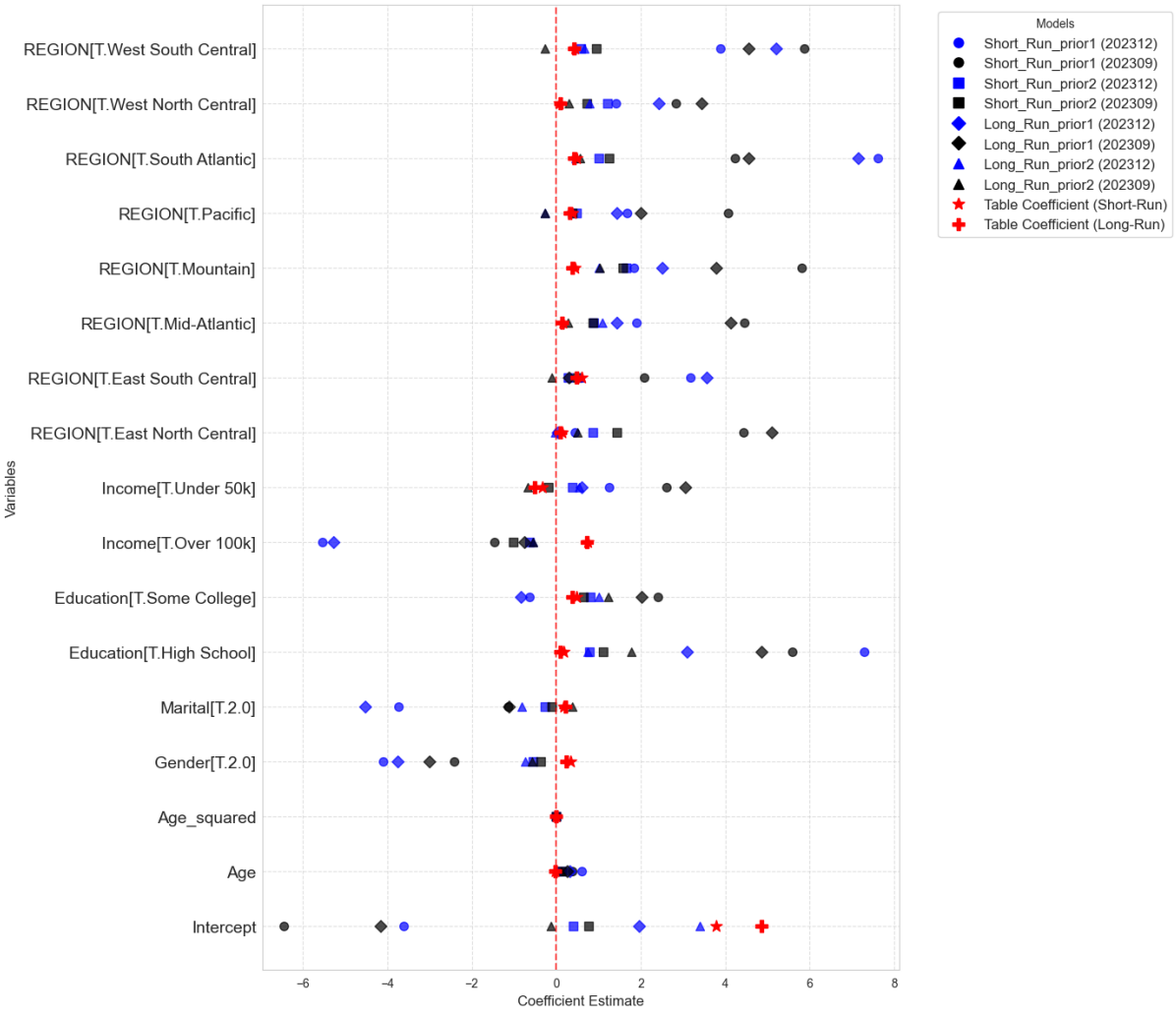


Figure A.15: Estimates for Determinants of Inflation Expectations (SCE vs. AI)

A.8 Reasoning of GPT o1

Table A.19: Reasoning Patterns Across Treatments

Pattern	Population Growth	Federal Funds Rate	Current Inflation	Long-Run Inflation	Mortgage Rate
Data Analysis and Trend Tracking	General economic trends and historical inflation up to October 2023.	Focus on monetary policy's impact and current Federal Funds Rate.	Emphasis on recent inflation data, including monthly and annual rates.	Integration of short-term data with long-term projections for a comprehensive outlook.	Use of broader indicators like fixed-rate mortgage rates to assess inflation dynamics.
Probability Mapping and Scenario Evaluation	Balanced probabilities across various inflation and deflation scenarios.	Probabilities skewed towards policy targets, reflecting Federal Funds Rate influence.	Distribution centered on recent trends with moderate variance.	Bimodal distribution distinguishing short-term and long-term expectations.	Incorporation of market expectations influencing inflation probabilities.
Time Horizon Considerations	Consistent methodology for one-year and three-year forecasts.	Projections align with policy targets over the long term.	Gradual normalization of inflation based on current trends.	Separate expectations for short-term and long-term horizons, considering different factors.	Inflation expectations tied to long-term market rates, aligned with mortgage trends.
Economic Indicators and Policy Considerations	Limited focus on specific policies, emphasizing general economic conditions.	Strong emphasis on Federal Reserve policies affecting inflation paths.	Moderate consideration of policies, balancing data-driven insights with policy impacts.	Long-term inflation expectations strongly anchored to policy targets, incorporating current and future policies.	Indirect policy influence through market-driven indicators like mortgage rates.
External Factors	Inclusion of population growth as an influencing external factor.	Minimal consideration of external factors beyond monetary policy.	Primarily internal economic data focus with limited external factors.	Some inclusion of structural factors affecting long-term inflation, such as demographics.	Consideration of housing market conditions as key external influences on expectations.
Methodological Approach	Utilizes general economic knowledge and historical data for predictions.	Policy-focused analysis prioritizing the Federal Funds Rate's role in inflation forecasting.	Data-driven projections based on recent inflation statistics and analysis.	Integrates short-term and long-term data for comprehensive multi-horizon forecasts.	Employs market-based extrapolation using indicators like fixed-rate mortgage rates to predict.