ChatGPT: Augmenting or Replacing Intelligence in Economic Surveys?

Sungho Park*

New York University

August 30, 2025

Abstract

This study evaluates whether large language models can substitute for human survey respondents. I replicate analyses from a representative households survey (the Italian Survey of Consumer Expectations, ISCE) across three domains: behavioral reactions to information treatments, the formation of economic expectations, and the prediction of persistent household traits. Using gpt-4o-mini with post-training data to mitigate contamination bias, I find that the model reproduces certain aggregate patterns but systematically diverges from observed human behavior. It fails to respond appropriately to information treatments, does not capture demographic heterogeneity in risk perceptions, and does not exhibit prudence. Incorporating demographic embeddings further reduces alignment, indicating that the model struggles to simulate human decision processes. However, the model attains 74% accuracy in predicting income categories and 72% in predicting consumption levels, suggesting potential as an auxiliary tool for imputing persistent traits rather than as a replacement for human respondents.

*Email: sp8783@stern.nyu.edu; This paper is based on my MSc thesis. I am indebted to my MSc advisor, Luigi Guiso, for his invaluable guidance and support. I am particularly grateful to Emiliano Calvano, Francesco Lippi, Claudio Michelacci, and Franco Peracchi, as well as to the workshop participants at the Einaudi Institute for Economics and Finance for suggestions and feedback. I am grateful to John Horton and Robin Horton for technical advice on the implementation of the surveys. I acknowledge drafting assistance from Claude and ChatGPT. All errors are my own.

1 Introduction

ChatGPT's release on 30 November, 2022 sparked a wave of work on the economic and financial impact of generative AI. Recent studies examine how these systems influence firm and worker productivity (Babina et al., 2024; Bertomeu et al., 2023; Noy and Zhang, 2023), corporate valuations (Eisfeldt et al., 2023), asset management (Sheng et al., 2024), and the macroeconomy (Acemoglu et al., 2022; Acemoglu, 2021, 2024; Furman and Seamans, 2019). Researchers are also turning to ChatGPT itself to replicate experiments and surveys (Korinek, 2023). Understanding how these models "think" is essential, because delegating choices to AI can propagate human biases and introduce new statistical distortions. Furthermore, few studies rigorously defend using ChatGPT as a proxy for human decision makers.[1] It remains unclear whether the model resembles an average individual or reflects specific segments of heterogeneity.[2]

Firstly, the framing of questions impacts ChatGPT's responses, which may mirror psychological phenomena in humans. It is unclear if this bias originates from the training data – notably, the contamination of training data with the test data, also known as look-ahead bias – or specific features of its architecture. Secondly, it is unknown if ChatGPT simply reproduces training data, or genuinely mimics human biases. Thirdly, the way one interacts with ChatGPT (e.g., user profiles, framing, or fine-tuning) alters response distributions and behavior. Lastly, ChatGPT has been noted to excel at identifying correlations, but struggle with causal relationships (Manning et al., 2024). In light of the proposed usages of ChatGPT in economics and finance research, this makes it imperative to understand *how ChatGPT reasons and responds to novel information, how it forms expectations, and furthermore if it can detect persistent traits among humans*.

Although large language models (LLMs) such as ChatGPT are increasingly used in economics and finance research that often demand strong numerical reasoning—there remains substantial disagreement over their quantitative capabilities. Levy (2024), for example, shows that even minor perturbations to financial data, such as shuffling the last few digits in accounting statements, can produce large changes in predicted corporate performance. Such look-ahead bias suggests that these models may rely more on surface-level pattern recognition than genuine numerical reasoning,

[1] See, for example, Bybee (2023), Charness et al. (2023), Chen et al. (2023), Duraj et al. (2024), Eisfeldt and Schubert (2024), Horton (2023), Immorlica et al. (2024), Korinek (2023), Lopez-Lira and Tang (2023), Lo and Ross (2024), Manning et al. (2024), Mei et al. (2024), Michelacci and Wu (2024), Zarifhonarvar (2024).

[2] Individual utility maximization does not always aggregate neatly; a choice that raises average welfare may conflict with personal preferences (Kirman, 1992).

raising concerns about their reliability in contexts requiring precise economic judgment. While some studies find that LLMs perform reasonably well on certain economic reasoning tasks (Bybee, 2023; Horton, 2023), others identify significant limitations in how they process information relative to human decision-makers (Fedyk et al., 2024; Chen et al., 2024). The debate extends beyond numerical accuracy to the broader question of whether LLMs can serve as substitutes for human survey respondents or function as "digital twins" in economic experiments. Proponents argue that LLMs offer scalability and cost efficiency, enabling more frequent data collection, whereas critics stress the models' limited ability to capture human heterogeneity, adapt to new information, and engage in causal reasoning rather than merely detecting statistical correlations (Manning et al., 2024).

An ideal setting to test the suitability of ChatGPT as a research tool in light of these concerns is through surveys commonly used in economics and finance. First, if ChatGPT can replicate the responses of an average research participant, including reactions to information treatments, this could offer insights into the sources of bias in its outputs and its potential value in generating high-frequency survey data that vary across time and topics. If it can effectively synthesize and respond to novel information, ChatGPT may serve as a powerful tool for researchers seeking scalable and adaptable inputs. Second, given ChatGPT's widespread use, it is important to understand what kind of agent it represents. Its outputs can have economically meaningful consequences by influencing users, who in turn affect ChatGPT. As users interact with the model, their responses feed back into its training process, both in the short run by conditioning ongoing interactions and in the long run by shaping the pretraining data used in future model versions. This feedback loop highlights the need to evaluate the extent to which ChatGPT's survey responses align with human responses, since any systematic differences could lead to shifts in both user behavior and model behavior over time. Finally, consistent patterns in how ChatGPT responds to specific survey instruments may reveal not only its value as an imputation device but also its capacity to identify correlations or even reason about causal relationships.

In light of this, I replicate elements from two survey-based papers in economics and finance, and examine the responses to those surveys, as well as examine if ChatGPT can accurately guess more persistent features of survey respondents. I focus on three key dimensions of survey responses which the surveys examine: reaction to new information from Guiso and Jappelli (2024b), elicitation of expectations from Guiso and Jappelli (2024a), and prediction of respondents' persistent traits from the same survey. The papers are based on the Italian Survey of Consumer Expectations

(ISCE), which conveniently was administered post-training of the ChatGPT model used in the paper, ruling out any contamination from the survey responses in the training data for ChatGPT. I first examine how respondents react to new information, as these reactions depend fundamentally on their previously elicited expectations. Expectations, in turn, are shaped by persistent traits of respondents, which I analyze as elements influencing the response process.

My findings reveal substantial limitations in using ChatGPT as a proxy for human survey respondents. While the model can sometimes capture the aggregate statistical properties (first and second moments) of response distributions for expectations about idiosyncratic and aggregate factors, the actual response patterns often differ fundamentally from human data, and these differences are starker when demographic traits are injected, in contrast to Fedyk et al. (2024). Moreover, the model fails to reproduce how demographic characteristics influence economic risk perceptions, sometimes predicting effects directly opposite to those observed in human data. Particularly concerning is the model's inability to demonstrate economic prudence - the precautionary saving motive observed in human responses. Most alarmingly, ChatGPT processes new information in ways that contradict human behavior - where humans respond positively to certain information treatments. Notably, ChatGPT consistently demonstrates negative responses for information treatments related to additional information on a disaster (i.e. the economic damages, and the mortality + economic costs) when humans demonstrate otherwise. Furthermore, ChatGPT's outputs for survey responses are notably lower variance than that of humans, consistent with prior literature which observes responses to experiements with ChatGPT result in outcomes of far lower variance (del Rio-Chanona et al., 2025). While ChatGPT excels at predicting static traits like current income and consumption based on demographics (with accuracy rates of approximately 74%), it struggles to form expectations or forecast future economic variables, suggesting limited abilities for economic reasoning that extends beyond recognition of persistent patterns. Specifically, ChatGPT appears adept at inferring static characteristics of synthetic individuals by aggregating information from its training data, yet it struggles to replicate forward-looking expectations and conditional decision-making processes that are not anchored in previously documented choices.

Contribution to the Literature

Recent research in economics and finance has begun exploring the properties and potential applications of ChatGPT, examining both how ChatGPT processes information and how it can be harnessed as a research tool.[3] However, little is known about *how* ChatGPT would respond if it were presented with traditional, unstructured surveys or questionnaires originally designed for humans, nor its similarity or divergence with human responses – especially in the context of economics and finance. Indeed, most prior work even in the fields of statistics and computer sciences, focuses on bespoke questionnaires to probe LLM responses, thus making direct comparisons to naturally generated human survey data more difficult.[4]

A notable exception is Fedyk et al. (2024), who design a survey to elicit human investment preferences and then pose the same questions to ChatGPT after providing demographic cues. Their results suggest that ChatGPT can successfully approximate demographic heterogeneity in investment behavior. Yet, such an approach – where surveys for humans are administered *after* ChatGPT has already been released – presents a risk of "preference contamination."[5] Another notable exception, is Zarifhonarvar (2024), which replications the elicitation of expectations from the New York Fed Consumer Expectations using ChatGPT.

In this paper, I contribute to the literature by administering pre-existing economics and finance surveys to ChatGPT that were originally conducted after ChatGPT's training data cutoff in October 2023. This timing reduces concerns about test set contamination from the training set (look-forward bias).[6] By comparing ChatGPT's responses with human responses collected entirely after the cutoff date, I provide new evidence on whether ChatGPT's answers influence – or are influenced by – human behavior. Assessing whether ChatGPT's response distribution aligns or diverges from human responses, particularly across demographic and cultural groups, is critical for understanding its potential impact on economic and financial decision-making. A key advantage of using pre-existing post-cutoff surveys is that they avoid input contamination. Moreover, for surveys with information treatments, ChatGPT's outputs can be interpreted as the result of

[3] See, for example, Bybee (2023), Charness et al. (2023), Chen et al. (2023), Duraj et al. (2024), Eisfeldt and Schubert (2024), Horton (2023), Immorlica et al. (2024), Korinek (2023), Lopez-Lira and Tang (2023), Lo and Ross (2024), Mei et al. (2024), Michelacci and Wu (2024).

[4] Scherrer et al. (2023) offers a notable contribution from statistics/computer science in examining how ChatGPT responds to moral or social dilemmas, but their bespoke survey questions complicate direct comparison with human responses on pre-existing surveys.

[5] In other words, ChatGPT's responses evolve based on human feedback, and human responses can also be shaped by exposure to ChatGPT's outputs.

[6] In contrast, Zarifhonarvar (2024) uses the NY Fed Survey of Consumer Expectations, which was conducted well before ChatGPT's training cutoff.

an "interaction" with AI. Within this framework, survey questions serve as uncontaminated inputs, while user-specific variations (such as stating a demographic identity) allow for exploration of how ChatGPT's responses to information treatments may shift.

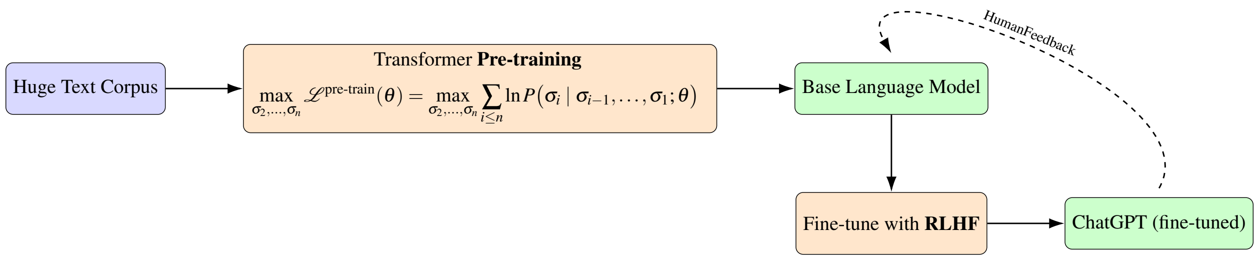
This study also informs the debate on using ChatGPT as a research tool. Although prior work has explored its ability to replicate economic experiments and models, there is limited justification for treating ChatGPT as a credible "representative agent." Existing evidence on its alignment with human responses is mixed.[7] In finance, Bybee (2023) shows that ChatGPT's binary macroeconomic expectations are somewhat predictive of survey-based expectations, but does not test whether ChatGPT mirrors human survey responses or reacts to specific information treatments.

By administering established, structured economics and finance surveys to ChatGPT — and comparing its answers to human responses collected before ChatGPT's release — I evaluate the feasibility and limits of using large language models as "survey respondents." This includes examining how its answers change with demographic cues, language settings, and information treatments, shedding light on both the potential benefits and the risks of aggregating preferences through ChatGPT. The approach adds to the survey methods literature in economics and finance, particularly on subjective beliefs and decision-making, which despite increasing attention to heterogeneous beliefs, remains underexplored.

Additionally, by examining survey responses, my research contributes to the literature on preference aggregation in economics and finance research, particularly related to the assumptions behind representative agents vs. heterogeneous agents models (Kirman, 1992). Given the inherent nature of ChatGPT as a preference aggregator, and a possible nature as an individual preference *dis-aggregator*, it is necessary to understand these features of ChatGPT to understand the effects of its interactions with users of different demographics, influencing economic behavior.

[7]For example, Mei et al. (2024) and Horton (2023) find ChatGPT's answers often resemble human responses, whereas Fedyk et al. (2024), Chen et al. (2024), Ouyang et al. (2024), Kim et al. (2024), and Ross et al. (2024) do not.

2 What is ChatGPT?



A Simple Illustration of ChatGPT's Training Process

ChatGPT is a *large language model*. A large language model aims to *approximate* the *text generating process*, $\mu : \Sigma^* \to \Sigma^*$ where $\mu(\sigma) = \mathbb{P}[\sigma_n|\sigma_{n-1},....,\sigma_1]$, where Σ^* is the space of strings.[8] The text generating function is a function $m(\sigma;\theta) : \Sigma^* \to \Sigma^*$ for $\theta \in \Theta$, the space of parameters of the model. I define a trained LLM, such as ChatGPT, given a training set $\mathscr{T} \subset \Sigma^*$, is defined as $\hat{m}(\sigma,\mathscr{T}) \equiv m(\sigma;\hat{\theta}(\mathscr{T}))$ (Ludwig et al., 2025).

The steps to recover $m(\sigma;\mathscr{T})$ ChatGPT involves three steps. The first step, pre-training, occurs via a process called self-supervised learning, which induces the model to represent the conditional probability distribution of preceding words based on its training data and some provided parameter $\theta \in \Theta$:

$$\mathscr{L}^{pre-train}(\theta) = \sum \ln P[\sigma_i|\sigma_{i-1},\cdots,\sigma_1;\theta]$$

During pre-training, language models inadvertently absorb biases present in their training datasets. These can include stereotypical associations, such as linking doctors predominantly with men and nurses with women. Following pre-training, instruction fine-tuning enhances the model's ability to respond appropriately to human directives through supervised learning. This process exposes the model to millions of examples across thousands of different instructional scenarios. The final phase employs reinforcement learning from human feedback (RLHF), where human evaluators' assessments guide the model in distinguishing between more and less desirable responses.

[8] This set Σ^* is extremely high dimensional. For instance, Italian is said to have around 2000000, which would make $\infty > |\Sigma^*| \geq 2^{2000000}$ at minimum.

3 Data

3.1 Human Survey Data

My primary source of human data is the Italian Survey of Consumer Expectations (ISCE), a quarterly rotating panel that collects demographic, income, wealth, consumption, expectations, and belief data from a representative sample of the Italian population. Conducted quarterly since October 2023, the ISCE is the main dataset used in the studies summarized in Table 1. The discussion below focuses on three waves: October 2023 (wave 1), January 2024 (wave 2), and April 2024 (wave 3).

The survey covers demographic characteristics; household resources (including income and wealth components); consumption; individual expectations (such as anticipated changes in consumption, income, energy costs, and health expenditures); and macroeconomic expectations (inflation, nominal interest rates, GDP growth). The target population is Italian residents aged 18–75. A pilot with 100 interviews was conducted in September 2023. Regular waves take place in the first 7–15 days of each reference month. Wave 1 included 5,006 interviews, wave 2 included 5,001, and wave 3 included 5,005. The retention rate between consecutive waves was 84% (wave 1 to wave 2) and 87% (wave 2 to wave 3).

Sampling follows the Bank of Italy's Survey of Household Income and Wealth (SHIW) design, stratifying by geography (North-East, North-West, Centre, South), age group (18–34, 35–44, 45–54, 55–64, 65+), gender, education (college degree, high school diploma, less than high school), and employment status (employed vs. not employed).

In the empirical analysis, I select the ISCE wave appropriate for each component of the strategy. Summary statistics and variable definitions for each analysis appear in the Appendix.

Table 1: Papers to Test Similarity to Humans

Topic	**Paper**
Information Treatments	Guiso and Jappelli (2024b)
Expectations	Guiso and Jappelli (2024a)

3.2 Simulated Data

I use the "gpt-4o-mini" model available, via the OpenAI API in Python. This model is the most advanced model which users of the ChatGPT web version have access without subscription. While acknowledging that LLMs are not written in stone, the focus on gpt-4o-mini is particularly relevant because it represents the version of ChatGPT available to users for free, and ChatGPT is arguably the most popular LLM model available to general users. Given that humans would most likely delegate their decision-making to the most conventionally available model, understanding how this conventional model processes information and makes decisions represents an important first step in understanding the consequences of novel AI tools about which we actually know very little.

The model's training data extends through October 1, 2023. To balance the reduction of hallucinations and other uninformative output variation with the need to preserve the diversity of responses observed in human data, I follow Fedyk et al. (2024) and set the temperature parameter to 0.8. A temperature of 0 yields fully deterministic outputs, while higher values increase randomness, with 2 producing maximally creative and unpredictable results.[9] The data generation process is outlined in Algorithm 1. For robustness, I repeat the analysis using a temperature of 1.0.

Algorithm 1 Data Simulation without Demographics

Input: $\mathscr{D} := \{(\mathbf{y}_i)\}_{i=1}^N$ and $(q_1, \ldots, q_l)$. $\mathbf{y}_i$ is the target vector of outputs. $(q_1, \ldots, q_l)$ corresponds to the tuple of words constituting a question.

Output: A simulated dataset of observations $\hat{\mathscr{D}} := \{(\hat{\mathbf{y}}_i)\}_{i=1}^N$.

1: Generate a generic system prompt $(s_1, \ldots, s_k)$, a tuple of words, via a standard prompt template function
2: **for** $i = 1$ to N **do**
3: Input $(s_1, \ldots, s_k)$ as the system prompt for `ChatGPT`
4: Input question $(q_1, \ldots, q_l)$ into `ChatGPT` which then outputs $\hat{\mathbf{y}}_i$, the simulated output
5: **end for**
6: **return** $\hat{\mathscr{D}} := \{(\hat{\mathbf{y}}_i)\}_{i=1}^N$

In the baseline simulation, no demographic characteristics are embedded in the prompt apart from the system instruction: "ChatGPT is an Italian survey respondent." For each exercise, the most relevant inputs are selected based on the demographic variables identified as important in the

[9] Temperatures above 1 are generally considered unreliable as proxies for human responses.

corresponding paper. The full set of prompts used to generate outputs, along with summary statistics for the simulated datasets, is provided in the Appendix. All primary exercises are conducted in Italian, except for tasks that involve predicting demographics from other fixed traits, including responses to expectation questions (i.e., inferring the policy function). When demographic information is incorporated into the prompt, the data generation process follows the structure outlined in Algorithm 2.

Algorithm 2 Data Simulation with Demographics

Input: $\mathscr{D} := \{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^N$ and $(q_1, \ldots, q_l)$. $\mathbf{x}_i$ is a non-empty vector of demographic features corresponding to the target vector of outputs, $\mathbf{y}_i$. $(q_1, \ldots, q_l)$ corresponds to the tuple of words constituting a question.

Output: A simulated dataset of observations $\hat{\mathscr{D}} := \{(\hat{\mathbf{y}}_i, \mathbf{x}_i)\}_{i=1}^N$.

1: **for** $i = 1$ to N **do**
2: Generate a demographic-informed system prompt $(s_1, \ldots, s_k)$, a tuple of words, based on $\mathbf{x}_i$ via a prompt template function
3: Input $(s_1, \ldots, s_k)$ as the system prompt for `ChatGPT`
4: Input question $(q_1, \ldots, q_l)$ into `ChatGPT` which then outputs $\hat{\mathbf{y}}_i$, the simulated output conditioned on demographics $\mathbf{x}_i$
5: **end for**
6: **return** $\hat{\mathscr{D}} := \{(\hat{\mathbf{y}}_i, \mathbf{x}_i)\}_{i=1}^N$

4 Empirical Strategy

This section outlines the empirical strategy for comparing ChatGPT-generated survey responses with human responses from the ISCE. The analysis focuses on three key dimensions: (i) reactions to new information, (ii) elicitation of expectations, and (iii) prediction of respondents' persistent traits. I begin by examining how respondents react to new information, as these reactions are inherently shaped by their previously stated expectations. Expectations themselves are influenced by respondents' persistent traits, which I analyze as a foundational factor underlying the entire response process.

4.1 How Does ChatGPT Respond to New Information vs Humans?: Guiso and Jappelli (2024b)

To compare ChatGPT's reaction to new information with that of human survey participants, I replicate Table 7 of Guiso and Jappelli (2024b), shown in Appendix A.3.[10] The original survey employs a two-stage randomization: participants are first assigned to one of two question types, and then, within each question type, to one of three different information treatments. This design produces six experimental groups: T1G1, T1G2, T2G1, T2G2, T3G1, and T3G2 (Table 2).

Table 2: The structure of information treatments

First stage randomization: Describe flood consequence		
T1	**T2**	**T3**
Control group	Treatment: N of deaths	Treatment: N of deaths plus damages
Second stage randomization: Evoke free riding		
G1: No treatment		
G2: Treatment: Fund success depends on how many contribute		
Willingness to pay asked to all		

I seed each survey with this system prompt, in Italian:

> *Answer the questions in the most truthful and accurate way possible. You are participating in a survey in Italy. Current date: January 1, 2024.*

I then elicit ChatGPT's responses 840 times for each treatment group, matching the approximate number of human observations per group in the survey. For each simulation, I provide the corresponding information and question appropriate to that treatment group. The full set of questions, along with summary statistics for both human and ChatGPT-generated data, is provided in Appendix A.

The primary specification Table 7 of Guiso and Jappelli (2024b) involves the following tobit specification:

$$y_i = \mathbf{1}\{y_i^* > 0\} \cdot y_i^* \tag{1}$$

[10] I do not replicate Table 5, which examines willingness to pay, because ChatGPT invariably responds "yes" when asked whether it would contribute to a fund. I investigate the reasons for this behavior later in the same section.

where

$$y_i^* = d + \beta_1 T_{2,i} + \beta_2 T_{3,i} + \beta_3 G_{2,i} + \beta_4 T_{2,i} G_{2,i} + \beta_5 T_{3,i} G_{2,i} + u_i \tag{2}$$

In the regression specification, G_2 is an indicator for whether the prompt invoked free-riding, T_2 indicates treatment with a description of the number of deaths, and T_3 indicates treatment with both the number of deaths and the associated financial damages. Standard errors (u_i) are bootstrapped 10,000 times within each treatment group, using the treatment group as the sampling stratum. The same procedure is applied to human responses from Wave 2 of the ISCE to enable a direct comparison with ChatGPT-generated responses.

The bootstrapped confidence intervals capture sampling uncertainty from the ChatGPT model, which is inherently different from the sampling uncertainty arising from surveying the Italian population. While each ChatGPT response is treated as a single draw from the population, this assumption is unlikely to hold exactly. Nonetheless, the sign and magnitude of estimated coefficients in the simulated data provide a benchmark for assessing whether ChatGPT can generate responses that appear, at least superficially, similar to human responses. When demographic covariates are included, these are held fixed under the assumption that they are independent of unobserved characteristics affecting survey responses – an assumption that may not fully capture the interdependencies present in actual survey data.

4.1.1 Embedding Demographics

To evaluate the performance of ChatGPT when demographics are embedded, I instead seed each survey with this system prompt.

> *Answer the questions in the most truthful and accurate way possible. You are participating in a survey in Italy. You are a [male/female], are [employed/unemployed/a student/a homemaker/retired], with a monthly family income [greater than 2500 euros/between 500 2500 euros]. Current date: January 1, 2024.*

Given computational constraints, I construct a matched sample by drawing the same number of observations for each demographic group as in the human data from Guiso and Jappelli (2024b). Demographic groups are defined by the intersection of college education status, employment status, gender, and monthly family income category, separately for each combination of information treatment group and question group. In the Wave 2 ISCE human data, I drop observations with missing household income, merge employed and self-employed respondents into a single category, and discretize household income into a binary indicator for exceeding the median value (2,500 euros). Employment status is classified into five categories: "Employed" (including self-employed), "Unemployed," "Retired," "Homemaker," and "Student." One observation that did not fit any category is removed from the sample.

I then re-estimate Equation 3, bootstrapping within each of the six treatment groups for both human and simulated data. For the simulated data, I further restrict the sample to demographic group–treatment group cells with more than 50 observations.

4.2 How Does ChatGPT Form Expectations?: Guiso and Jappelli (2024a)

To assess the similarity between human and ChatGPT-generated responses, I perform two complementary analyses examining both distributional properties and demographic heterogeneity in expectation formation. These analyses enable a systematic comparison of how a large language model such as ChatGPT forms expectations relative to humans across demographic groups and risk categories. The distributions are elicited by first seeding the following prompt:

> *Answer the questions in the most truthful and accurate way possible. You are participating in a survey in Italy. Current date: October 1, 2023.*

4.2.1 Distributional Similarity Analysis

The first analysis examines the overall distributional similarity of responses. I compute three correlation coefficients between the distributions of ChatGPT-generated and human responses: Pearson (linear association), Spearman rank (monotonic association), and Kendall's tau (rank-based association, robust to outliers). Coefficients are estimated using a bootstrap procedure with 10,000 iterations. In each iteration, I draw the same number of observations as in the original dataset, compute the mean and variance, and then calculate the correlations between human and ChatGPT data. The resulting bootstrapped confidence intervals provide stable and precise estimates, which is particularly important for smaller demographic subgroups.

To provide a more granular assessment of distributional similarity, I conduct t-tests comparing the mean values of human and simulated responses for each question within each response bin.

This complements the correlation analysis by directly testing whether the distribution of responses differs significantly between the two datasets.

4.2.2 Embedding Demographics

The second analysis incorporates demographic information directly into the prompts provided to ChatGPT, allowing for an examination of how the model captures demographic heterogeneity in expectation formation. This analysis extends beyond simple distributional comparisons to investigate whether ChatGPT can replicate demographic patterns observed in human responses. I begin by restricting the sample to first-wave respondents to ensure temporal consistency and eliminate potential confounds from repeated survey participation. Using this sample, I categorize respondents according to several demographic characteristics determined by their explanatory power in underlying risk-factors per Guiso and Jappelli (2024a): When prompting ChatGPT, I explicitly

Demographic Variable	Categorization
Age	Above/below the age 49 years
Household size	Above/below the household size of 3 members
Geographic region	”South” versus ”North or Centre” of Italy
Education	College-educated versus non-college-educated
Housing status	Homeowners versus non-homeowners

include these demographic characteristics, enabling the model to potentially tailor its responses to different demographic profiles. I set the system prompt to the following, in Italian:

> *You are taking part in a survey in Italy. You are [college-educated/non-college educated], a [homeowner/non-homeowner], living with [above 3 household members/below 3 household members], are [above/below] the age of 49 years, and you come from the [North or Centre/South] of Italy. Current date: 1 October 2023.*

For the demographic-embedded analysis, correlation coefficients are computed using a stratified bootstrap that preserves the original demographic composition of the data. Within each demographic stratum, samples are drawn with replacement in proportion to the stratum's share of the total sample, ensuring that no group is over- or under-represented in the bootstrap iterations.

Pearson, Spearman, and Kendall correlation coefficients are calculated separately for each question–demographic group pair and pooled across all questions. This allows for identification of both question-specific patterns and broader trends in how demographic characteristics shape expectation formation. To maintain statistical reliability, the analysis is restricted to demographic groups with more than 50 observations, thereby avoiding unstable estimates from very small subgroups.

In addition to the correlation analysis, I replicate the regression specifications from Guiso and Jappelli (2024a), focusing on the results reported in Tables 2, 3, and 9.[11] The first specification examines how demographic characteristics influence risk perceptions across different domains:

$$y_i = d + \beta_1 \mathbf{1}_{\{\text{HH Members}>3\}_i} + \beta_2 \mathbf{1}_{\{\text{Age}>49\}_i} + \beta_3 \mathbf{1}_{\{\text{College}\}_i} + \beta_4 \mathbf{1}_{\{\text{Homeowner}\}_i} + \beta_5 \mathbf{1}_{\{\text{North or Centre}\}_i} + u_i \quad (3)$$

In this specification, y_i denotes the risk perception measure for individual i, defined as the variance of the elicited distribution for changes in idiosyncratic or aggregate risk factors across nine categories: consumption, income, health, energy, GDP, unemployment, inflation, interest rates, and house prices. The coefficients β_1 through β_5 capture the marginal effect of each demographic characteristic on risk perception, holding other variables constant. Estimating this model separately for human and ChatGPT-generated responses allows for a direct comparison of how demographic factors shape risk perceptions in humans versus the model.

The second specification examines the relationship between expected consumption growth and the second moment of its elicited distribution:

$$\mathbb{E}_i\left[\frac{c_{t+1}-c_t}{c_t}\right] = d + \beta \mathbb{E}_i\left[\left(\frac{c_{t+1}-c_t}{c_t}\right)^2\right] + u_i \quad (4)$$

This specification is derived from the following relationship:

$$\mathbb{E}\left[\frac{c_{t+1}-c_t}{c_t}\right] \simeq -\frac{u'(c_t)}{u''(c_t)c_t}\frac{r-\delta}{1+r} - \underbrace{\frac{u'''(c_t)c_t}{u''(c_t)}}_{=\text{Prudence}}\frac{1}{2}\mathbb{E}\left[\left(\frac{c_{t+1}-c_t}{c_t}\right)^2\right]$$

The second specification tests whether respondents who expect higher consumption growth also perceive greater variance in their consumption prospects, with β capturing the strength of this relationship. This parameter has a direct economic interpretation related to prudence, the idea that greater future income risk induces precautionary savings. A negative β would be consistent with

[11] The original tables are reproduced in Appendix B.4. The only modification is that age is treated as a discrete variable, and the North and Centre regions are combined into a single binary indicator.

prudent behavior, where higher expected consumption uncertainty reduces current consumption, thereby increasing expected consumption growth.

For both specifications, standard errors (u_i) are computed via 10,000 bootstrap replications within each demographic group. This stratified bootstrap accounts for potential heteroskedasticity and within-group correlation by resampling with replacement within each demographic stratum, re-estimating the model, and calculating the standard deviation of the resulting parameter estimates. Demographic stratification ensures that the standard errors reflect the uncertainty associated with each subgroup, enabling precise comparisons between human and ChatGPT-generated responses.

As in earlier regressions with simulated data, a key limitation is that the bootstrapped confidence intervals capture sampling uncertainty from the ChatGPT model, which differs from the sampling uncertainty of the actual Italian population. Moreover, I hold demographic traits fixed and assume they are independent of unobserved characteristics that might influence survey responses, an assumption unlikely to hold in real data. Nonetheless, this exercise provides a face-value test of whether ChatGPT can replicate patterns observed in human responses.

4.3 Demographic Predictions

Following Fedyk et al. (2024), I examine three sociodemographic factors: gender, age, and employment status, as predictors of monthly household income and consumption. These outcomes are chosen for their economic significance and persistence over time.

To ensure computational tractability, I focus primarily on Wave 1 of the ISCE, which is the closest wave to the end of ChatGPT's training data. Age is binarized at the median value of 49 years, and employment status is consolidated into two categories: employed (including self-employed) and unemployed. All other statuses, such as retired, are excluded. Demographic subgroups with fewer than 50 observations are dropped to ensure statistical power. For each analysis, I sample demographic combinations in the same proportions as in the human reference data to preserve demographic representativeness. I further restrict the sample to individuals who report both household and individual income, as well as homeownership versus rental status. Future income and consumption are taken from ISCE waves in January 2024 and October 2024, while current values come from October 2023 (Wave 1).

For statistical comparison between human and ChatGPT-generated data, I implement a bootstrap procedure to estimate both my accuracy metric and its standard errors. For each demographic combination with at least 50 observations, I draw 10,000 bootstrap samples of equivalent size to the original group. Within each sample, I compute average income or consumption values using the midpoints of ISCE-reported bins, separately for human and ChatGPT data.

These averages are then reclassified into their original income or consumption bins. A "match" is recorded when the binned averages for human and ChatGPT data coincide for a demographic group, and the share of matches across all groups constitutes the primary accuracy metric. As a complementary measure, I perform a binary classification analysis distinguishing values above or below the median thresholds (2,500 euros for income; 1,250 euros for consumption). This captures fundamental patterns of income and consumption persistence across groups.

Finally, I compute Pearson, Kendall, and Spearman correlations between human and ChatGPT averages to assess the strength and direction of association. The full analysis is repeated for current income and consumption in October 2024, as well as for future values four months ahead (January 2024) and twelve months ahead (October 2024), to test for predictive power.

Variable	**Categorization**
Gender	Male/Female
Age	Above/below 49 years (median age)
Employment status	Employed/Unemployed

I implement a stratified paired bootstrap procedure with 10,000 iterations to generate robust statistical inferences. Each stratum is defined by a unique combination of gender, age, and employment status.[12] In each iteration, I draw paired samples with replacement within each demographic subgroup, compute mean monthly household income and consumption, assign these values to the same bins used in the original survey, and create binary indicators for whether the values exceed median thresholds.

Predictive performance is evaluated using two metrics: (1) classification accuracy, defined as the percentage of demographic subgroups in which the predicted income or consumption category matches the actual category in the human data, and (2) numeric correlation, defined as the Pearson correlation coefficient between predicted and observed values across all subgroups.

[12] $2^3 = 8$ categories. While stratified sampling in a real survey would also include geographic regions, model fit is poor when including geography, so I exclude it from the analysis.

To assess ChatGPT's ability to predict future economic outcomes, I elicit expected annual household income and consumption one year ahead. These predictions are compared against two benchmarks: actual reported values in Wave 2 of the ISCE and expected monthly values implied by participants' reported annual growth rates.

This analysis evaluates how accurately ChatGPT predicts average current and future income and consumption for defined demographic groups. It provides evidence on whether large language models have implicitly learned relationships underlying income and consumption patterns across demographic groups, and whether they can generate plausible forecasts based on those relationships.

Algorithm 3 Evaluation of ChatGPT Predictions for Household Income and Consumption

Input: ISCE Wave 1 (Oct 2023), Wave 2 (Jan 2024), Wave 3 (Oct 2024)

```
Select sociodemographic factors: gender, age, employment status
Binarize age at 49 years; recode employment into {employed, unemployed}; drop other categories
Drop subgroups with n < 50; preserve human demographic proportions
Restrict to respondents with household & individual income and housing status
Define outcomes: current (Oct 2023), future (Jan 2024, Oct 2024)
Set thresholds: income = 2500€, consumption = 1250€
Bootstrap procedure:
for each demographic group with n ≥ 50 do
    for b = 1 to 10^4 do
        Sample n obs. with replacement (human and ChatGPT separately)
        Compute mean income & consumption (bin midpoints)
        Re-bin to ISCE categories
        Record:
          • Match indicator: binned ChatGPT = binned human
          • Binary indicators above/below median thresholds
    end for
end for
Metrics:
  • Classification accuracy = % of groups with exact bin match
  • Pearson, Kendall, Spearman correlations between human & ChatGPT means
Repeat for: Oct 2024 current values, Jan 2024 and Oct 2024 future values
Compare future predictions against:
  • Actual Wave 2 values
  • Expected values from reported annual growth rates
```

5 Results and Discussion

5.1 How Does ChatGPT Respond to New Information vs Humans?: Guiso and Jappelli (2024b)

The baseline regression results in Table A5 reveal fundamental disparities between simulated and human responses to information treatments. While the human data shows significant positive effects from both treatments (T3 and T2), the simulated data exhibits consistently negative and highly significant responses. I also run the same regression on the human data after splitting the sample into high AI users, defined as survey participants who reported using AI more than once a month in the ISCE, and low AI users, defined as all others. This allows me to test whether ChatGPT's behavior could reflect that of its users, who it may have influenced to respond similarly, or an over-representation of such users in its training data. Because the survey was administered well after the cutoff date for the gpt-4o-mini training data, I can also rule out contamination from the ISCE responses entering ChatGPT's training set.[13] The contrast in coefficients persists across all specifications.

These contradictions extend to question group effects (G2), where the simulated data shows significant positive effects in some specifications while the human data shows negative or insignificant effects. The simulated data suggests that the AI reacts negatively to free-riding scenarios.

Unlike humans, ChatGPT consistently contributes to disaster funds, consistent with prior work documenting greater altruism and cooperation in AI-generated responses (Mei et al., 2024). This reveals a fundamental inconsistency between the AI's stated preferences and its revealed actions, where it is altruistic when prompted explicitly about ethics or cooperation, but different when these values are only implicit. To explore this further, I elicit responses to trust-related questions from Wave 5 of the ISCE.[14] Table A7 shows that ChatGPT reports systematically higher trust levels than humans. A t-test comparing high AI users and low AI users in Table A8 finds no statistical differences between the two groups, suggesting that the gap originates from model bias rather than the behavior of AI users in the population.

Table A6 shows further differences when the analysis is restricted to demographic and treat-

[13] Although the AI usage question was asked in Wave 3 of the ISCE and the information treatment in Wave 2, it is unlikely that frequent ChatGPT users did not use the model more intensively in the year preceding Wave 3 compared to the complement group.

[14] The questions are reported in Appendix A.2.

ment group cells with at least 50 observations. While the human data shows large positive treatment effects for T3 and T2, the full-sample simulated data produces opposite-signed effects for T3 and smaller positive effects for T2, and the restricted sample produces negative coefficients for both. The G2 effect is broadly positive in the full-sample simulated data but negative in the human data. Interaction terms also diverge, with T3G2 significantly negative in the simulated data but near zero in the human data. Even when demographic effects are directionally aligned, their magnitudes differ substantially. These results indicate that while ChatGPT can reproduce some demographic patterns, it fundamentally diverges in modeling social dilemmas, particularly public goods contributions and responses to free-riding, even when matched on demographics.

5.2 How Does ChatGPT Form Expectations?: Guiso and Jappelli (2024a)

Having established that ChatGPT struggles to replicate human responses to new information, I next test whether it can aggregate historical information to form human-like expectations. The baseline correlation analysis compares the means and standard deviations of distributions generated by humans and ChatGPT across multiple questions. As shown in Table B3, the distributions generated by ChatGPT are strongly correlated with the human data, suggesting that the model captures the first and second moments of human expectations. This is consistent with Fedyk et al. (2024), which finds that ChatGPT's answers to investment product questions correlate closely with human responses on average.

However, moving beyond the moments, I find that the allocation of points into bins differs significantly between human and simulated data (Table B9), indicating distributional divergence despite correlated moments. Figure 5 shows that ChatGPT's distributions are generally more pessimistic and right-skewed than human distributions, except for expected interest rate and interest rate changes. The pattern holds when comparing low versus high AI users in the human sample (Figures 6 and Tables B5, B6), suggesting that the difference arises from the model rather than underlying user composition.

When demographic information is embedded (Table B4), ChatGPT's generated standard deviations become negatively and significantly correlated with the human data, and correlations for means also decrease. This contrasts with Fedyk et al. (2024), which reports reduced heterogeneity after demographic prompting. Further analysis by question (Table B7) shows that only a few topics, such as household income, consumption, household labor income, and mortgage interest rate, display even one correlation measure that is both correctly signed and statistically significant. Others, such as gas and energy bills, show significant negative correlations.

By demographic trait (Table B8), mean values remain positively and significantly correlated, but standard deviations are significantly negatively correlated, implying that ChatGPT may incorporate demographics in ways that introduce systematic bias.

Regression results in Table B10 highlight further misalignment. For household size, humans exhibit consistently positive and significant coefficients across all nine risk categories, while ChatGPT produces insignificant or even negative significant coefficients for several categories, with only income and house price risks showing the correct sign. Age effects are similarly misaligned: humans display strong negative coefficients for all risks, whereas ChatGPT captures this for only a subset and produces positive coefficients for GDP and unemployment risks. Homeownership effects also diverge sharply, with ChatGPT predicting positive coefficients for health and energy risks, opposite to the human data.

This misalignment persists across low and high AI user subsamples in the human data (Table B11). For example, low AI users show strongly positive household size effects across all risks, while ChatGPT responses are mostly insignificant or negative. Age effects and homeownership effects also differ, and regional effects are less consistent for high AI users.

The starkest divergence appears in Table B12, where the regression in Equation 4 shows prudence in the human data but significant negative coefficients in the ChatGPT data. Table B13 confirms that both low and high AI human users display zero or positive coefficients, in contrast to ChatGPT's negative results. A limitation is that, because only Wave 1 is replicated, individual fixed effects and temporal variation cannot be exploited, and the results should not be interpreted causally. Nonetheless, the gap between human and simulated responses is substantial.

Overall, these findings suggest that ChatGPT cannot reliably generate expectations or expectation distributions based on historical data from Italian households. With its current training data, it appears unable to synthesize existing information into future expectations in a manner consistent with human behavior.

5.3 Demographic Predictions

Given that ChatGPT cannot reliably generate expectations or distributions of expectations with its underlying data, I next test whether it can predict more persistent traits, specifically consumption and income. As shown in Table C2, ChatGPT performs well at predicting income, achieving 74% accuracy in classifying the correct income bin across demographic group combinations and 90% accuracy in predicting whether income is above the median (2,500 euros). For consumption (Table C3), the corresponding accuracies are 72.5% and 87.3%. Both traits display positive and statistically significant correlations between the averages in ChatGPT-generated data and human data.

I then repeat the analysis for future monthly income and total consumption, comparing ChatGPT's predictions with realized outcomes. For income (Table C4), accuracy is lower for both four- and twelve-month horizons when using ISCE income bins, although the above-median accuracy remains high. A similar pattern holds for future consumption (Table C5). While bin-level accuracy is statistically significant and above 50% for consumption at both horizons, it is statistically indistinguishable from 50% for below/above-median classification in Wave 1.

To examine whether this inability to predict future outcomes is linked to the inability to form expectations, I re-elicit distributions of future income from Section 4.2 using the same demographic traits as in this exercise. I then calculate the correlation between expected growth in consumption and income from ChatGPT responses and the corresponding values from the human data, following the same procedure as for current income and consumption. As shown in Table C6, these correlations are systematically lower than those for current household income and consumption in Tables C2 and C3, supporting the hypothesis.

Overall, these results suggest that ChatGPT may have practical value for projecting persistent traits such as whether monthly household income is above or below the median, but its predictive ability declines sharply when forecasting less persistent or more expectation-driven outcomes.

5.3.1 Extension: Recovering the Policy Function

Furthermore, it could be the case while ChatGPT cannot generate expectations of distributions with the underlying data well. While due to the dimensionality of the outcome variable, a credible bootstrap algorithm is implausible, I feed into ChatGPT the distributions of variables determined to be crucial for consumption risk according to Guiso and Jappelli (2024a), and ask ChatGPT if each individual owns a home, if he/she is college educated, if the individual's age is greater than 49, if the individual lives in the North or Centre of Italy or the South, and if the individual has more than 3 household members. I use the full sample of individuals from Wave 1 of the ISCE. The results, in Table C7, show that age being greater than 49, homeownership status, and residing in the north or centre have accuracy scores of above 50% consistently – that is ChatGPT correctly guessed the applicable feature for above 50% of the Wave 1 ISCE participants – in particular homeownership has consistently high accuracy rates of above 60%. For a household having more than 3 members, it guessed above 50% correctly for 3/5 questions. This shows that for some possibly persistent traits, ChatGPT may accurately guess these traits given the outcome of a question. Hence, while ChatGPT may not be able to accurately replicate the expectations or responses to information treatments of survey participants, it can complement surveys by predicting the appropriate demographics corresponding to a survey's answers – which can augment survey-based studies which face incomplete information collection.

6 Robustness

A potential concern with the results in Section 4.1 and Section 4.2.2 is that the failure to replicate human responses could be driven by limited variation from my choice of the temperature parameter. To address this, I set the temperature to 1, the maximum value recommended for replicating human responses, and re-run the analyses for Section 4.1 and Section 4.2.2, using 100 bootstrap samples rather than 10,000 for the confidence intervals.[15] The results, reported in Section D.1 in Table D8 and Table D9 for Section 4.1, and in Table D12 and Table D.2 for Section 4.2.2, are qualitatively unchanged. In fact, they diverge even further from human responses, particularly in the positive significance of the G2 coefficient in Table D.1 and the magnitude of the consumption risk coefficient in Table D.2. This suggests that the temperature parameter does not explain the lack of alignment with human responses. The low variance of the outcome variable generated by ChatGPT appears to be an inherent feature of the model, consistent with findings in laboratory market experiments reported by del Rio-Chanona et al. (2025).

Another concern with Section 4.1 is that the baseline responses could be disaster-specific, with ChatGPT incorporating ex-ante knowledge about the event and biasing responses through look-ahead bias. To address this, I remove all references to the flood-affected region (Romagna) from the prompt and repeat the baseline exercise from Section 4.1 using a model temperature of 0.8. The results, in Table D11, show that while the direction and significance of the T2 and G2 coef-

[15] This difference is inconsequential for symmetric confidence intervals.

ficients in Equation (1) differ from Table A5, the G2 coefficient remains opposite in sign to the human responses from Guiso and Jappelli (2024b). In Equation (2), the significance and direction of coefficients are unchanged from Table A5. This indicates that the failure to replicate human responses is not driven by disaster-specific wording in the prompt.

Lastly, I repeat the analysis in Section 4.1 with the same total number of observations as in the human data regressions (5,001), and the same number within each treatment group, by drawing a new sample of observations at a temperature of 0.8. The results, in Table D10, show that while the coefficients on T2G2, T3G2, and G2 are positive and significant, the coefficients on T3 and T2 remain negative and significant, consistent with Table D8. This suggests that ChatGPT may not consistently react to free-riding treatments, in contrast to treatments providing additional information on the scale of a disaster (mortality vs mortality plus economic costs). The inconsistency in G2-related coefficients across Table D11 and Table D10, together with the persistent divergence from human responses for T3 and T2, indicates that ChatGPT fails to consistently replicate human reactions to both free-riding and disaster-related information treatments.

7 Conclusion

My findings reveal substantial limitations in using ChatGPT as a proxy for human survey respondents, that is, replacing human participants, while also highlighting potential complementary applications in augmenting survey data.

The results show that although ChatGPT can sometimes match the aggregate statistical properties (first and second moments) of human response distributions, fundamental differences remain in how it processes information. Most notably, ChatGPT exhibits negative responses to information treatments where humans respond positively, fails to accurately model demographic effects on economic risk perceptions, and cannot replicate the prudence observed in human behavior. These shortcomings are even more pronounced when demographic information is embedded in prompts, contradicting earlier findings that such cues improve alignment with human data (Fedyk et al., 2024). The model's inability to respond consistently to disaster-related information or free-riding scenarios further underscores its limitations in simulating human economic decision-making.

Nonetheless, ChatGPT demonstrates notable skill in predicting persistent traits from demographic characteristics, achieving 74% accuracy in classifying current income categories and 72% accuracy for consumption levels. These findings suggest that while ChatGPT cannot reliably replace human participants in economic surveys, it may function as a valuable complementary tool, particularly for imputing missing demographic information. Consistent with prior work (del Rio-Chanona et al., 2025), ChatGPT's survey outputs display markedly lower variance than human responses, indicating that its strengths lie less in replicating human expectations or processing novel information, and more in identifying relationships between demographic profiles and stable economic traits.

Future research should investigate how AI models can be leveraged to enhance survey methodologies, especially in contexts where demographic data are incomplete. The limitations identified here emphasize the continued necessity of human participation in economic research and caution against overreliance on AI simulations for modeling economic behavior. Rather than replacing human respondents, ChatGPT and similar models may best serve as augmentative tools that work alongside the irreplaceable value of human responses.

References

Acemoglu, Daron, "Harms of AI," Working Paper 29247, National Bureau of Economic Research 2021.

—, "The Impact of AI on Productivity and Employment," Working Paper 31456, National Bureau of Economic Research 2024.

—**, David Autor, Jonathon Hazell, and Pascual Restrepo**, "AI and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson, "Artificial Intelligence, Firm Growth, and Industry Concentration," *Journal of Finance*, 2024, *79* (1), 123–165.

Bertomeu, Jeremy, Yupeng Lin, Yibin Liu, and Zhenghui Ni, "Capital Market Consequences of Generative AI: Early Evidence from the Ban of ChatGPT in Italy," 2023. Available at SSRN: `https://ssrn.com/abstract=4452670` or `http://dx.doi.org/10.2139/ssrn.4452670`.

Bybee, Leland, "Surveying Generative AI's Economic Expectations," *arXiv preprint arXiv:2305.02823*, 2023.

Charness, Gary, Brian Jabarian, and John A. List, "Generation Next: Experimentation with AI," Working Paper 31679, National Bureau of Economic Research 2023.

Chen, Shuaiyu, T. Clifton Green, Huseyin Gulen, and Dexin Zhou, "What Does ChatGPT Make of Historical Stock Returns? Extrapolation and Miscalibration in LLM Stock Return Forecasts," August 2024. Available at SSRN: `https://ssrn.com/abstract=4941906` or `http://dx.doi.org/10.2139/ssrn.4941906`.

Chen, Yiting, Tracy Xiao Liu, You Shan, and Songfa Zhong, "The emergence of economic rationality of GPT," *Proceedings of the National Academy of Sciences*, 2023, *120* (51), e2316205120.

del Rio-Chanona, R. Maria, Marco Pangallo, and Cars Hommes, "Can Generative AI agents behave like humans? Evidence from laboratory market experiments," 2025.

Duraj, Kamila, Daniela Grunow, Michael Haliassos, Christine Laudenbach, and Stephan Siegel, "Rethinking the Stock Market Participation Puzzle: A Qualitative Approach," IMFS Working Paper Series 210, Institute for Monetary and Financial Stability, Goethe University Frankfurt 2024. Accessed: 2024-11-21.

Eisfeldt, Andrea L. and Gregor Schubert, "AI and Finance," Working Paper 33076, National Bureau of Economic Research October 2024.

_ , _ , and Miao Ben Zhang, "Generative AI and Firm Values," Working Paper 31222, National Bureau of Economic Research 2023.

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

Furman, Jason and Robert Seamans, "AI and the Economy," *Innovation Policy and the Economy*, 2019, *19*, 161–191.

Guiso, Luigi and Tullio Jappelli, "Anatomy of Consumption Risk," Technical Report, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy 2024.

_ and _ , "Are People Willing to Pay to Prevent Natural Disasters?," CSEF Working Papers 723, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy June 2024.

Horton, John J, "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?," Working Paper 31122, National Bureau of Economic Research April 2023.

Immorlica, Nicole, Jieming Mao, Aleksandrs Slivkins, and Zhiwei Steven Wu, "Incentivizing Exploration with Selective Data Disclosure," *arXiv preprint arXiv:1811.06026*, 2024.

Kim, Jeongbin, Matthew Kovach, Kyu-Min Lee, Euncheol Shin, and Hector Tzavellas, "Learning to be Homo Economicus: Can an LLM Learn Preferences from Choice Data?," 2024. Available at arXiv: `https://arxiv.org/abs/2401.07345`.

Kirman, Alan P., "Whom or What Does the Representative Individual Represent?," *Journal of Economic Perspectives*, June 1992, *6* (2), 117–136.

Korinek, Anton, "Generative AI for Economic Research: Use Cases and Implications for Economists," *Journal of Economic Literature*, 2023, *61* (4), 1281–1317.

Levy, Bradford, "Caution Ahead: Numerical Reasoning and Look-ahead Bias in AI Models," 2024. Fama-Miller Working Paper.

Lo, Andrew W. and Jillian Ross, "Can ChatGPT Plan Your Retirement?: Generative AI and Financial Advice," Working Paper, SSRN February 2024. Available at SSRN: https://ssrn.com/abstract=4722780 or http://dx.doi.org/10.2139/ssrn.4722780.

Lopez-Lira, Alejandro and Yuehua Tang, "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models," Working Paper, University of Florida 2023.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan, "Large Language Models: An Applied Econometric Framework," Working Paper 33344, National Bureau of Economic Research January 2025.

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

Mei, Qiaozhu, Yutong Xie, Walter Yuan, and Matthew O. Jackson, "A Turing test of whether AI chatbots are behaviorally similar to humans," *Proceedings of the National Academy of Sciences*, 2024, *121* (9), e2313925121.

Michelacci, Claudio and Liangjie Wu, "Profiting from Consumer Politicization," 2024. Seminar presentation, EIEF.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

Ouyang, Shumiao, Hayong Yun, and Xingjian Zheng, "How Ethical Should AI Be? How AI Alignment Shapes the Risk Preferences of LLMs," 2024. Available at SSRN: `https://ssrn.com/abstract=4851711` or `http://dx.doi.org/10.2139/ssrn.4851711`.

Ross, Jillian, Yoon Kim, and Andrew W Lo, "LLM economicus? Mapping the Behavioral Biases of LLMs via Utility Theory," July 2024. Available at SSRN: https://ssrn.com/abstract=4926791.

Scherrer, Nino, Claudia Shi, Amir Feder, and David Blei, "Evaluating the Moral Beliefs Encoded in LLMs," in A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, eds., *Advances in Neural Information Processing Systems*, Vol. 36 Curran Associates, Inc. 2023, pp. 51778–51809.

Sheng, Jinfei, Zheng Sun, Baozhong Yang, and Alan L. Zhang, "Generative AI and Asset Management," 2024. Available at SSRN: `https://ssrn.com/abstract=4501234`.

Zarifhonarvar, Ali, "Evidence on Inflation Expectations Formation Using Large Language Models," SSRN Working Paper April 2024. Also available at `http://dx.doi.org/10.2139/ssrn.4825076`.

Appendix

A Novel Information Processing

A.1 Summary Statistics

Table A1: Summary Statistics by Group – Baseline Human Data

	T1G1			T1G2			T2G1			T2G2			T3G1			T3G2		
Variable	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
High AI	735	0.109	0.312	737	0.079	0.269	735	0.120	0.325	706	0.089	0.285	722	0.120	0.326	703	0.108	0.311
Amount contributed	840	33.08	91.79	827	36.81	105.51	830	42.50	125.07	840	45.15	139.53	837	40.09	120.33	827	43.33	144.06

Notes: High AI is a binary indicator (1 stands for AI usage is greater or equal once per week, 0 for otherwise); Amount contributed is the amount that the survey participant agreed to contribute (midpoint of the elicited bin of the amount contributed).

Table A2: Summary Statistics by Group – ChatGPT Simulated Data

	T1G1			T1G2			T2G1			T2G2			T3G1			T3G2		
Variable	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
Amount Contributed	840	34.98	4.83	840	32.40	6.93	840	30.26	9.61	840	26.81	10.94	840	33.88	5.95	840	28.71	10.17

Notes: Amount contributed is the amount that the survey participant agreed to contribute (midpoint of the elicited bin of the amount contributed).

Table A3: Summary Statistics by Treatment Group (ChatGPT Simulated Data with Demographics)

	T1G1			T1G2			T2G1			T2G2			T3G1			T3G2		
Variable	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
HH Income > 2500	763	0.341	0.474	743	0.316	0.465	753	0.329	0.470	759	0.310	0.463	759	0.336	0.473	740	0.311	0.463
Amount Contributed	763	39.44	33.86	743	45.86	42.92	753	42.78	44.98	759	48.74	50.60	759	38.85	36.44	740	37.32	34.57
College	763	0.223	0.416	743	0.248	0.432	753	0.210	0.407	759	0.261	0.439	759	0.228	0.420	740	0.230	0.421
Male	763	0.505	0.500	743	0.495	0.500	753	0.497	0.500	759	0.503	0.500	759	0.510	0.500	740	0.522	0.500
Employed	763	0.523	0.500	743	0.501	0.500	753	0.515	0.500	759	0.509	0.500	759	0.522	0.500	740	0.528	0.500
Homemaker	763	0.122	0.327	743	0.133	0.340	753	0.114	0.318	759	0.108	0.311	759	0.144	0.351	740	0.108	0.311
Retired	763	0.197	0.398	743	0.202	0.402	753	0.203	0.403	759	0.194	0.395	759	0.171	0.377	740	0.178	0.383
Student	763	0.039	0.194	743	0.030	0.170	753	0.032	0.176	759	0.051	0.221	759	0.038	0.192	740	0.043	0.204
Unemployed	763	0.119	0.324	743	0.135	0.342	753	0.135	0.342	759	0.138	0.345	759	0.125	0.331	740	0.142	0.349

Notes: HH Income > 2500 indicates households with a total monthly income of more than 2500 Euros; Amount Contributed is the elicited amount contributed to the natural disaster fund; College indicates college education; Male indicates male gender. Employed, Homemaker, Retired, Student, and Unemployed are indicators for various employment statuses.

Table A4: Summary Statistics by Treatment Group (Human Data)

	T1G1			T1G2			T2G1			T2G2			T3G1			T3G2		
Variable	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.	N	Mean	Std.Dev.
HH Income > 2500	763	0.341	0.474	743	0.316	0.465	753	0.329	0.470	759	0.310	0.463	759	0.336	0.473	740	0.311	0.463
Amount Contributed	763	34.21	94.99	743	39.14	110.21	753	44.43	128.00	759	47.97	145.29	759	41.41	122.87	740	43.26	137.96
College	763	0.223	0.416	743	0.248	0.432	753	0.210	0.407	759	0.261	0.439	759	0.228	0.420	740	0.230	0.421
Male	763	0.505	0.500	743	0.495	0.500	753	0.497	0.500	759	0.503	0.500	759	0.510	0.500	740	0.522	0.500
Employed	763	0.523	0.500	743	0.501	0.500	753	0.515	0.500	759	0.509	0.500	759	0.522	0.500	740	0.528	0.500
Homemaker	763	0.122	0.327	743	0.133	0.340	753	0.114	0.318	759	0.108	0.311	759	0.144	0.351	740	0.108	0.311
Retired	763	0.197	0.398	743	0.202	0.402	753	0.203	0.403	759	0.194	0.395	759	0.171	0.377	740	0.178	0.383
Student	763	0.039	0.194	743	0.030	0.170	753	0.032	0.176	759	0.051	0.221	759	0.038	0.192	740	0.043	0.204
Unemployed	763	0.119	0.324	743	0.135	0.342	753	0.135	0.342	759	0.138	0.345	759	0.125	0.331	740	0.142	0.349

Notes: HH Income > 2500 indicates households with a total monthly income of more than 2500 Euros; Amount Contributed is the elicited amount contributed to the natural disaster fund; College indicates college education; Male indicates male gender. Employed, Homemaker, Retired, Student, and Unemployed are indicators for various employment statuses. Data is from the ISCE Wave 1, restricted to observations with more than 50 observations per employment status-HH income > 2500-college education-gender combination.

A.2 Survey Question Format

The Treatment Groups are as follows:

T1:

(No additional information provided)

T2:

In Romagna, the evening between May 16 and 17, an unprecedented amount of rain in just a few hours raised river levels until they overflowed. Practically all the waterways between Rimini and Bologna, twenty-one in total, breached their banks or overflowed, flooding vast areas of Romagna. Fifteen people died, and approximately 40 thousand were evacuated.

T3:

In Romagna, the evening between May 16 and 17, an unprecedented amount of rain in just a few hours raised river levels until they overflowed. Practically all the waterways between Rimini and Bologna, twenty-one in total, breached their banks or overflowed, flooding vast areas of Romagna. Fifteen people died, and approximately 40 thousand were evacuated. The Region has calculated damages of almost 9 billion euros for roads, schools, embankments and canals, and to repair damages to homes and businesses.

The corresponding questions are as follows, where {info} corresponds to the treatment group prompts above (no info if T1):

Questions asked to Group 2 (G4_1)

Consider the following information: {info}

Containing environmental degradation and securing areas exposed to hydrogeological risk (floods, landslides, etc.) requires a substantial amount of public resources. To finance these investments, would you be in favor of creating a dedicated public fund?

- Yes

- No
- I don't know

Question asked to Group 3 (G4_2)

Consider the following information: {info}

Containing environmental degradation and securing areas exposed to hydrogeological risk (floods, landslides, etc.) requires a substantial amount of public resources. Success depends on the size of the fund. If few contribute or contribute little, the risk containment policy fails. How much are you willing to contribute? To finance these investments, would you be in favor of creating a dedicated public fund?

- Yes
- No
- I don't know

Questions asked to Group 1 (G5_1)

Previously you were asked '[G4_1 question text]' and you answered: '[G4_1 answer]'

Consider the following information: {info}

How much would you be willing to contribute to this fund each year?

- 5˜10 Euro
- 10˜20 Euro
- 20˜50 Euro
- 50˜100 Euro
- 100˜200 Euro
- 200˜300 Euro

- 300˜400 Euro
- 400˜500 Euro
- 500˜1000 Euro
- More than 1000 Euro

Questions asked to Group 2 (G5_2)

Previously you were asked '[G4_2 question text]' and you answered: '[G4_2 answer]'

Consider the following information: {info}

How much would you be willing to contribute to this fund each year?

- 5˜10 Euro
- 10˜20 Euro
- 20˜50 Euro
- 50˜100 Euro
- 100˜200 Euro
- 200˜300 Euro
- 300˜400 Euro
- 400˜500 Euro
- 500˜1000 Euro
- More than 1000 Euro

A.3 Tables from Guiso and Jappelli (2024b)

Table 7. Tobit estimates of the effect of treatments on WTP

Treatment	Tobit	Tobit
T2	28.878	27.481
	(9.724)***	(7.066)***
T3	22.351	24.1888
	(9.734)**	(7.097)**
G2	-7.859	-7.558
	(9.989)	(5.607)
T2G2	-2.832	
	(13.897)	
T3G2	3.744	
	(13.922)	
P-value test : $\beta_1 = \beta_2$	0.497	0.631
P-value test $\beta_4 = \beta_5 = 0$	0.891	
Average of LHS variable	73.48	73.48
N. of observations	5,001	5,001

Note. The first regression reports marginal effects calculated from Tobit regressions for the amount that respondent intend to contribute to the fund. The estimated equation is $y_i = \beta_1 T_2 + \beta_2 T_3 + \beta_3 G_2 + \beta_4 T_2 G_2 + \beta_5 T_3 G_2 + \varepsilon_i$.. The second column restricts to zero the effects of the joint first-stage and second stage treatments. Heteroskedasticity consistent standard errors are reported in parentheses. One star indicates statistical significance at the 10%, two stars at the 5%, three stars at the 1%. The table also reports the *p*-values of a chi-square test of the listed null.

Figure 1: Table 7 of Guiso and Jappelli (2024b)

A.4 Information Treatment Regressions

Table A5: Baseline Tobit Regression Results

	Full Sample		High AI		Low AI		Simulated Data	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T3	22.351**	24.188***	9.609	7.845	29.664***	25.881***	-1.095***	-2.393***
	(8.964)	(7.065)	(21.753)	(21.247)	(9.957)	(7.975)	(0.263)	(0.246)
T2	28.878***	27.481***	11.620	0.961	36.885***	32.014***	-4.714***	-5.155***
	(9.114)	(7.003)	(20.582)	(19.654)	(10.617)	(8.063)	(0.373)	(0.294)
G2	-7.859	-7.558	11.786	1.113	-2.992	-8.849	-2.571***	-3.730***
	(9.343)	(5.578)	(34.551)	(16.384)	(10.164)	(6.276)	(0.295)	(0.236)
T2G2	-2.832		-26.305		-9.812		-0.881	
	(13.666)		(43.698)		(15.323)		(0.582)	
T3G2	3.744		-5.123		-7.585		-2.595***	
	(13.812)		(45.108)		(15.151)		(0.500)	
Mean of D.V.	40.156	40.156	40.708	40.708	39.168	39.168	31.175	31.175
Var of D.V.	14979.253	14979.253	11806.454	11806.454	14393.183	14393.183	78.557	78.557
N	5001	5001	452	452	3886	3886	5040	5040

Notes: Standard errors are computed with a stratified bootstrapped within each of the 6 information treatment groups of 10000 draws, where I draw a sample size identical to the original size of each of the 6 information treatment groups. The outcome variable (D.V.) is the amount contributed to the disaster fund. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Regression Results

	Full Sample				Restricted				Human Data	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
T3	-0.452	-4.250***	-0.596	-4.519***	-2.642**	-2.460***	-3.301*	-2.707**	22.049**	21.727***
	(1.134)	(0.806)	(1.811)	(1.363)	(1.203)	(0.717)	(1.972)	(1.089)	(9.476)	(7.332)
T2	4.529***	3.615***	3.333	3.126**	0.381	-1.710**	-1.432	-2.349*	31.261***	29.017***
	(1.218)	(0.870)	(2.061)	(1.582)	(1.478)	(0.848)	(2.533)	(1.424)	(9.660)	(7.453)
G2	7.231***	4.048***	6.418***	3.627***	3.386***	1.872***	-9.895***	-10.168***	-50.324	-7.126
	(1.181)	(0.690)	(1.983)	(1.222)	(1.039)	(0.390)	(1.630)	(0.979)	(9.869)	(5.863)
T2G2	-1.874		-0.459		-4.702***		-2.094		-40.590	
	(1.742)		(3.176)		(1.690)		(2.652)		(14.471)	
T3G2	-7.695***		-7.948***		0.351		1.222		0.672	
	(1.613)		(2.698)		(1.392)		(2.110)		(14.209)	
Employed	9.752***	9.689***			5.309***	5.297***				
	(0.771)	(0.771)			(0.456)	(0.448)				
Retired	4.259***	4.213***			-1.480*	-0.374				
	(1.038)	(1.042)			(0.879)	(0.806)				
Male	12.402***	12.387***			3.298***	3.323***				
	(0.706)	(0.698)			(0.572)	(0.564)				
HH Income > 2500	54.307***	54.312***			61.550***	61.553***				
	(0.972)	(0.970)			(2.613)	(2.607)				
College	33.263***	33.335***								
	(1.027)	(1.038)								
Mean of D.V.	42.167	42.167	42.167	42.167	26.952	26.952	26.952	26.952	40.156	40.156
Var of D.V.	1697.627	1697.627	1697.627	1697.627	567.234	567.234	567.234	567.234	14979.253	14979.253
N	4517	4517	4517	4517	1898	1898	1898	1898	4517	4517

Note: Restricted refers to ChatGPT generated simulated samples where groups (tuples) of employed-retired-male-HH Income > 2500-college- information treatment group where less than 50 observations exist are dropped. Full sample refers to samples in which this restriction in not applied, and instead simply just groups (tuples) of employed-retired-male-HH Income > 2500-college with less than 50 observations are dropped. Standard errors are bootstrapped 10000 times, within each combination of demographic groups (Employment-Gender-Income> 2500-College). The outcome variable (D.V.) is the amount contributed to the disaster fund. Bootstrap standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.5 Trust

	Human Data	Simulated Data	P-value
People	5.25	6.85	0.00
Govt.	4.23	5.95	0.00
Police	6.06	7.08	0.00
Judiciary	5.07	7.75	0.00
Health System	5.73	8.97	0.00
Civil Protection	6.65	9.89	0.00

Notes: Both datasets contain 5003 observations. I sampled a paired bootstrap of 1000 draws in the human and simulated data with the original number of samples, and conducted a t-test in difference in means.

Table A7: Trust Ratings

	High AI	Low AI	P-value
People	5.27	5.10	0.13
Govt.	4.22	4.22	0.98
Police	6.06	6.01	0.67
Judiciary	5.07	4.99	0.54
Health System	5.72	5.82	0.38
Civil Protection	6.66	6.58	0.53

Notes: The results displayed are from a t-test with paired bootstrap of 1000 draws in the human and simulated data with the original number of samples, and conducted a t-test in difference in means.

Table A8: High AI vs Low AI

B Expectation Formation

B.1 Survey Question Format

All the questions follow a similar format: "In the next 12 months, you expect that (yMy household's income / total consumption / gas and energy bills / health expenditures / house price / GDP / inflation):[16]

Interval		Probability (%)
will decrease by more than 8%	g_1	p_1
will decrease between 6 and 8%	g_2	p_2
will decrease between 4 and 6%	g_3	p_3
will decrease between 2 and 4%	g_4	p_4
will decrease between 0 and 2%	g_5	p_5
will remain constant	g_6	p_6
will increase between 0 and 2%	g_7	p_7
will increase between 2 and 4%	g_8	p_8
will increase between 4 and 6%	g_9	p_9
will increase between 6 and 8%	g_{10}	p_{10}
will increase more than 8%	g_{11}	p_{11}
Total		100

[16]The format of the questions referring to unemployment and interest is different since respondents are presented with only positive intervals ranging from 0 to "over 14%" for unemployment and from 0 to "over 8%" for interest rate.

B.2 Summary Statistics

Table B1: Summary Statistics for ChatGPT Simulated Data with Demographics

Variable	Obs	Mean	Std. dev.	Min	Max
$\mathbb{E}\Delta$ HH Income	4,814	-3.3849	0.5584	-6.5500	0.6000
SDΔ HH Income	4,814	4.0649	0.5075	2.3537	6.3583
$\mathbb{E}\Delta$ HH Labor Income	4,814	-3.2898	0.6628	-7.2000	0.4000
SDΔ HH Labor Income	4,814	4.0383	0.4894	2.4799	6.3710
$\mathbb{E}\Delta$ Consumption	4,814	-3.3427	0.3174	-5.3300	0.6000
SDΔConsumption	4,814	3.9825	0.3099	2.3537	5.9640
$\mathbb{E}\Delta$ Health Expenses	4,814	-0.6533	2.0065	-3.8900	4.5000
SDΔHealth Expenses	4,814	3.3520	0.6160	1.4491	5.1176
$\mathbb{E}\Delta$ Energy Bill	4,814	-2.4200	1.1999	-3.8980	4.0000
SDΔEnergy Bill	4,814	3.8538	0.4426	1.9339	5.3329
$\mathbb{E}\Delta$ House Price	4,814	-2.4183	1.2645	-5.9700	2.1500
SDΔHouse Price	4,814	3.5719	0.6479	1.3077	5.8258
$\mathbb{E}\Delta$ GDP	4,814	-1.4045	1.1383	-4.7400	1.7000
SDΔGDP	4,814	3.0553	0.6726	1.7436	5.5360
$\mathbb{E}\Delta$ Unemployment	4,814	0.4375	0.8267	-3.9900	3.2000
SDΔUnemployment	4,814	2.6724	0.3752	1.5460	4.6555
$\mathbb{E}\Delta$ Inflation	4,814	6.4258	0.9415	4.3000	11.0412
SDΔInflation	4,814	3.1942	0.4797	1.8856	4.7945
$\mathbb{E}\Delta$ Interest Rate	4,814	3.0034	0.3577	1.9000	4.4000
SDΔInterest Rate	4,814	1.9544	0.2614	1.3416	2.6721
$\mathbb{E}\Delta$ Mortgage Interest Rate	4,814	3.9578	0.3908	2.7000	4.9000
SDΔMortgage Interest Rate	4,814	2.0674	0.1445	1.3964	2.6721
College	4,814	0.2048	0.4036	0	1
Homeowner	4,814	0.7605	0.4268	0	1
North or Centre	4,814	0.6651	0.4720	0	1
Age>49	4,814	0.4790	0.4996	0	1
HH Members>3	4,814	0.2518	0.4341	0	1

Notes: $\mathbb{E}\Delta$ represents expected change in variable from the elicited distribution.

SDΔ represents standard deviation of the distribution of the elicited distributions.

North or South indicates if an individual lives in Northern or Central Italy.

HH Members>3 indicates an individual having more than 3 family members in the household.

Age>49 indicates an individual being older than the age of 49.

College indicates college education.

Homeowner indicates if an individual owns his or her own home.

Table B2: Summary Statistics for Human Data (Restricted)

Variable	Obs	Mean	Std. dev.	Min	Max
$\mathbb{E}\Delta$ HH Income	4,814	-1.1938	3.7149	-10	10
SDΔHH Income	4,814	2.1851	2.1644	0	10
$\mathbb{E}\Delta$ HH Labor Income	4,814	-0.7416	3.5409	-10	10
SDΔ HH Labor Income	4,814	1.9585	2.1752	0	10
$\mathbb{E}\Delta$ Consumption	4,814	0.5180	4.1630	-10	10
SDΔConsumption	4,814	2.2139	2.1583	0	10
$\mathbb{E}\Delta$ Health Expenses	4,814	0.9474	3.5589	-10	10
SDΔHealth Expenses	4,814	2.0305	2.1875	0	10
$\mathbb{E}\Delta$ Energy Bill	4,814	2.1326	3.9643	-10	10
SDΔEnergy Bill	4,814	2.0073	2.0782	0	10
$\mathbb{E}\Delta$ House Price	4,814	0.0327	3.5068	-10	10
SDΔHouse Price	4,814	1.7864	2.1525	0	10
$\mathbb{E}\Delta$ GDP	4,814	-1.7700	3.9274	-10	10
SDΔGDP	4,814	1.8904	2.1079	0	10
$\mathbb{E}\Delta$ Unemployment	4,814	1.6160	4.0764	-10	10
SDΔUnemployment	4,814	1.8937	2.0658	0	10
$\mathbb{E}\Delta$ Inflation	4,814	9.2793	3.5037	1	14
SDΔInflation	4,814	1.4927	1.6225	0	6.5
$\mathbb{E}\Delta$ Interest Rate	4,814	11.8662	20.1571	1	80
SDΔInterest Rate	4,814	7.8434	12.9772	0	39.5
$\mathbb{E}\Delta$ Mortgage Interest Rate	4,814	21.4955	26.1248	1	80
SDΔMortgage Interest Rate	4,814	10.7817	14.3706	0	39.5
Homeowner	4,814	0.7605	0.4268	0	1
College	4,814	0.2048	0.4036	0	1
HH Members>3	4,814	0.2518	0.4341	0	1
Age>49	4,814	0.4790	0.4996	0	1
North or Centre	4,814	0.6651	0.4720	0	1

Notes: $\mathbb{E}\Delta$ represents expected change in variable from the elicited distribution.

SD represents standard deviation of the distribution of the elicited distributions.

Human Data restricted to Homeowner-College-HH Members>3-Age>49-North or South groups with more than 50 obs.

North or South indicates if an individual lives in Northern or Central Italy.

HH Members>3 indicates an individual having more than 3 family members in the household.

Age>49 indicates an individual being older than the age of 49.

College indicates college education.

Homeowner indicates if an individual owns his or her own home.

B.3 Correlation Analysis

Table B3: Correlation Coefficients Between Datasets (Bootstrap Analysis)

	Pearson	Spearman	Kendall
ρ_μ	0.830^{***}	0.628^{***}	0.528^{***}
	(0.003)	(0.003)	(0.004)
ρ_σ	0.732^{***}	0.626^{***}	0.431^{***}
	(0.011)	(0.021)	(0.031)

Note: The point estimates and symmetric bootstrap intervals are calculated after random pair bootstrap design, where I drawing the same number of observations per question 10000 times in the human and ChatGPT simulated data independently, calculating the *averages* of the mean and standard deviation of the elicited distributions per each draw and question respectively, and then calculating for each draw the correlation coefficients across the AI generated *averages* of the mean and standard deviation with that of the human data, and then finally constructing symmetric standard errors and point estimates through this procedure. Bootstrap standard errors are in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B4: Correlation Coefficients Between Human and ChatGPT Responses with Demographics

	Pearson	Spearman	Kendall
ρ_μ	0.761***	0.739***	0.534***
	(0.005)	(0.006)	(0.007)
ρ_σ	-0.697***	-0.284***	-0.159***
	(0.008)	(0.013)	(0.011)

Note: Standard errors are in parenthesis. The point estimates and symmetric bootstrap intervals are calculated after random pair bootstrap design, where I drawing the same number of observations per question-demographic group 10000 times in the human and ChatGPT simulated data independently, calculating the *averages* of the mean and standard deviation of the elicited distributions per each draw and question-demographic group respectively, and then calculating for each draw the correlation coefficients across the AI generated *averages* of the mean and standard deviation with that of the human data, and then finally constructing symmetric standard errors and point estimates through this procedure. Bootstrap standard errors are in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B5: Correlation Coefficients Between Datasets (Bootstrap Analysis) – High AI

	Pearson	Spearman	Kendall
ρ_μ	0.905***	0.775***	0.665***
	(0.006)	(0.021)	(0.028)
ρ_σ	0.850***	0.708***	0.576***
	(0.028)	(0.072)	(0.076)

Note: The point estimates and symmetric bootstrap intervals are calculated after random pair bootstrap design, where I drawing the same number of observations per question-demographic group 10000 times in the human and ChatGPT simulated data independently, calculating the *averages* of the mean and standard deviation of the elicited distributions per each draw and question-demographic group respectively, and then calculating for each draw the correlation coefficients across the AI generated *averages* of the mean and standard deviation with that of the human data, and then finally constructing symmetric standard errors and point estimates through this procedure. I restrict my sample to individuals who reported using AI tools once a week or more. Bootstrap standard errors are in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B6: Correlation Coefficients Between Datasets (Bootstrap Analysis) – Low AI

	Pearson	Spearman	Kendall
ρ_μ	0.907***	0.764***	0.652***
	(0.002)	(0.014)	(0.021)
ρ_σ	0.835***	0.718***	0.556***
	(0.010)	(0.031)	(0.035)

Note: The point estimates and symmetric bootstrap intervals are calculated after random pair bootstrap design, where I drawing the same number of observations per question-demographic group 10000 times in the human and ChatGPT simulated data independently, calculating the *averages* of the mean and standard deviation of the elicited distributions per each draw and question-demographic group respectively, and then calculating for each draw the correlation coefficients across the AI generated *averages* of the mean and standard deviation with that of the human data, and then finally constructing symmetric standard errors and point estimates through this procedure. I restrict my sample to individuals who did not report using AI tools once a week or more. Bootstrap standard errors are in parenthesis. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B7: Bootstrap Estimates of Correlation Coefficients by Question

	Pearson		Kendall		Spearman	
Question Topic	μ	σ	μ	σ	μ	σ
HH Income	0.4588***	0.3437***	0.3119***	0.2701***	0.4392***	0.3867***
	(0.146)	(0.088)	(0.108)	(0.064)	(0.141)	(0.087)
HH Labor Income	0.4958***	0.3378***	0.3196***	0.2417***	0.4483***	0.3430***
	(0.116)	(0.094)	(0.100)	(0.074)	(0.129)	(0.100)
Consumption	0.4016***	0.0278	0.3120***	0.0192	0.4471***	0.0223
	(0.124)	(0.174)	(0.101)	(0.126)	(0.134)	(0.179)
Health Expenses	0.0052	-0.1217	0.0274	-0.0794	0.0364	-0.1078
	(0.382)	(0.201)	(0.275)	(0.138)	(0.395)	(0.189)
Energy Bill	-0.2994***	-0.3255***	-0.1509**	-0.2351***	-0.2314**	-0.3287***
	(0.103)	(0.109)	(0.073)	(0.082)	(0.102)	(0.113)
House Price	0.0764	0.5740***	0.0185	0.3975***	0.0278	0.5301***
	(0.155)	(0.086)	(0.111)	(0.081)	(0.154)	(0.100)
GDP	0.1928	0.3889***	0.1495	0.2631***	0.2130	0.3781***
	(0.150)	(0.094)	(0.106)	(0.072)	(0.152)	(0.099)
Unemployment	0.1276	-0.0021	0.0883	0.0026	0.1213	-0.0032
	(0.187)	(0.147)	(0.126)	(0.101)	(0.175)	(0.138)
Inflation	0.5588***	0.5315***	0.3817***	0.3675***	0.5336***	0.5107***
	(0.111)	(0.089)	(0.093)	(0.077)	(0.123)	(0.100)
Interest Rate	0.0635	0.1286	0.0327	0.0652	0.0578	0.1025
	(0.093)	(0.099)	(0.068)	(0.074)	(0.101)	(0.109)
Mortgage Interest Rate	0.2555**	0.4441***	0.1740**	0.3115***	0.2529**	0.4410***
	(0.104)	(0.114)	(0.078)	(0.093)	(0.113)	(0.124)

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. μ and σ represent correlations computed on the mean and standard deviation of the distribution, respectively. Standard errors from bootstrap samples shown in parentheses. The

Table B8: Bootstrap Estimates of Correlation Coefficients by Group

	Pearson		Kendall		Spearman	
Group	μ	σ	μ	σ	μ	σ
Homeowner	0.7736***	-0.7555***	0.5682***	-0.0849***	0.7448***	-0.2016***
	(0.0049)	(0.0044)	(0.0117)	(0.0176)	(0.0074)	(0.0172)
College	0.8099***	-0.7472***	0.5556***	-0.0058*	0.7174***	-0.0917***
	(0.0032)	(0.0037)	(0.0120)	(0.0296)	(0.0088)	(0.0290)
North or Centre	0.7918***	-0.7451***	0.5606***	-0.0494***	0.7444***	-0.1618***
	(0.0039)	(0.0050)	(0.0117)	(0.0185)	(0.0075)	(0.0156)
Age > 49	0.7918***	-0.7311***	0.5749***	-0.0980***	0.7504***	-0.2193***
	(0.0035)	(0.0058)	(0.0124)	(0.0170)	(0.0105)	(0.0180)
HH Members > 3	0.7967***	-0.7507***	0.5764***	-0.0727***	0.7547***	-0.1883***
	(0.0039)	(0.0045)	(0.0131)	(0.0261)	(0.0052)	(0.0278)

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. μ and σ represent correlations computed on the mean and standard deviation of the distribution, respectively. Standard errors shown in parentheses. Results based on 1000 bootstrap samples for each group.

B.4 Tables from Guiso and Jappelli (2024a)

Table 2 – Idiosyncratic risks

	Consumption risk	Income risk	Health risk	Energy risk
Male	0.002	-0.000	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Age 35 to 50	-0.005	-0.012	-0.003	-0.002
	(0.003)	(0.003)***	(0.003)	(0.003)
Age 51 to 65	-0.013	-0.023	-0.013	-0.013
	(0.003)***	(0.003)***	(0.003)***	(0.003)***
Age 66 to 75	-0.031	-0.045	-0.029	-0.026
	(0.004)***	(0.004)***	(0.004)***	(0.003)***
Family size	0.005	0.005	0.004	0.004
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
College degree	-0.010	-0.010	-0.012	-0.012
	(0.003)***	(0.003)***	(0.003)***	(0.002)***
North	-0.021	-0.022	-0.020	-0.019
	(0.003)***	(0.003)***	(0.003)***	(0.003)***
Centre	-0.015	-0.015	-0.012	-0.011
	(0.003)***	(0.003)***	(0.003)***	(0.003)***
Employed	0.000	0.001	-0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Self-employed	0.005	0.010	0.003	0.002
	(0.004)	(0.004)**	(0.004)	(0.004)
Log cash-on-hand	-0.001	-0.002	-0.000	-0.000
	(0.001)	(0.001)**	(0.001)	(0.001)
Homeowner	-0.016	-0.018	-0.017	-0.014
	(0.003)***	(0.003)***	(0.003)***	(0.003)***
Wave 2	-0.042	-0.031	-0.041	-0.041
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Wave 3	-0.047	-0.036	-0.047	-0.045
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Wave 4	-0.051	-0.043	-0.049	-0.047
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
R^2	0.06	0.06	0.06	0.06
N	20,015	20,015	20,015	20,015

Note. OLS regression estimations. Standard errors in parentheses. * significance at 10%, ** significance at 5%, *** significance at 1%.

Figure 2: Table 2 from Guiso and Jappelli (2024a)

Table 3. Aggregate risks

	GDP risk	Unemployment risk	Inflation risk	Interest rate risk	House price risk
Male	-0.002	-0.002	-0.000	0.001	0.002
	(0.002)	(0.001)*	(0.002)	(0.000)	(0.002)
Age 35 to 50	-0.004	-0.003	-0.004	-0.002	-0.003
	(0.003)	(0.002)	(0.003)	(0.001)***	(0.003)
Age 51 to 65	-0.013	-0.008	-0.013	-0.005	-0.013
	(0.003)***	(0.002)***	(0.003)***	(0.001)***	(0.003)***
Age 66 to 75	-0.028	-0.017	-0.026	-0.008	-0.027
	(0.004)***	(0.002)***	(0.003)***	(0.001)***	(0.003)***
Family size	0.005	0.003	0.004	0.001	0.004
	(0.001)***	(0.001)***	(0.001)***	(0.000)***	(0.001)***
College degree	-0.013	-0.009	-0.012	-0.002	-0.012
	(0.003)***	(0.002)***	(0.002)***	(0.001)***	(0.002)***
North	-0.019	-0.012	-0.019	-0.004	-0.018
	(0.003)***	(0.002)***	(0.002)***	(0.001)***	(0.003)***
Centre	-0.013	-0.008	-0.011	-0.002	-0.009
	(0.003)***	(0.002)***	(0.003)***	(0.001)***	(0.003)***
Employed	0.001	0.000	0.001	0.000	0.001
	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)
Self-employed	0.003	0.003	0.001	0.001	0.005
	(0.004)	(0.003)	(0.004)	(0.001)	(0.004)
Log cash-on-hand	-0.002	-0.001	-0.001	0.000	-0.001
	(0.001)**	(0.001)*	(0.001)	(0.000)	(0.001)
Homeowner	-0.018	-0.009	-0.015	-0.003	-0.013
	(0.003)***	(0.002)***	(0.003)***	(0.001)***	(0.003)***
Wave 2	-0.037	-0.025	-0.040	-0.008	-0.040
	(0.002)***	(0.001)***	(0.002)***	(0.000)***	(0.002)***
Wave 3	-0.040	-0.027	-0.042	-0.009	-0.043
	(0.002)***	(0.001)***	(0.002)***	(0.000)***	(0.002)***
Wave 4	-0.040	-0.027	-0.042	-0.009	-0.044
	(0.002)***	(0.001)***	(0.002)***	(0.000)***	(0.002)***
R^2	0.05	0.05	0.06	0.05	0.05
N	20,015	20,015	20,015	20,015	20,015

Note. OLS regression estimates. Standard errors in parentheses. * significance at 10%, ** significance at 5%, *** significance at 1%

Figure 3: Table 3 from Guiso and Jappelli (2024a)

Table 9. Euler equation estimates

	OLS	Fixed effect	Fixed effect	IV Fixed effect	IV Fixed effect
Interest rate	-0.078	0.024	0.014	0.014	0.015
	(0.012)***	(0.017)	(0.017)	(0.017)	(0.017)
pec2	1.165	1.426	1.800	1.539	1.326
	(0.095)***	(0.124)***	(0.122)***	(0.273)***	(0.264)***
Wave 2	0.003	0.003	0.002	0.002	0.002
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Wave 3	0.003	0.002	0.002	0.002	0.001
	(0.001)***	(0.001)***	(0.001)**	(0.001)**	(0.001)*
Wave 4	0.002	0.001	0.001	0.000	0.000
	(0.001)**	(0.001)*	(0.001)	(0.001)	(0.001)
Income growth			0.259	0.256	0.254
			(0.011)***	(0.011)***	(0.011)***
Constant	0.005	0.001	0.004	0.005	0.005
	(0.001)***	(0.001)*	(0.001)***	(0.001)***	(0.001)***
R^2	0.01	0.01	0.06		
N	20,015	18,031	18,031	18,031	18,031

Note. The dependent variable is expected consumption growth. Consumption risk is the 2nd conditional moment of the distribution of expected consumption growth. Column (1) presents OLS estimations; columns (2) and (3) are panel fixed effects estimations; column (4) presents instrumental variable fixed effects panel estimations using micro risks as instruments; column (5) presents instrumental variable fixed effects panel estimations using micro and macro risks as instruments. Standard errors in parentheses. * significance at 10%, ** significance at 5%, *** significance at 1%.

Figure 4: Table 9 from Guiso and Jappelli (2024a)

B.5 Distributional Plots and Tests

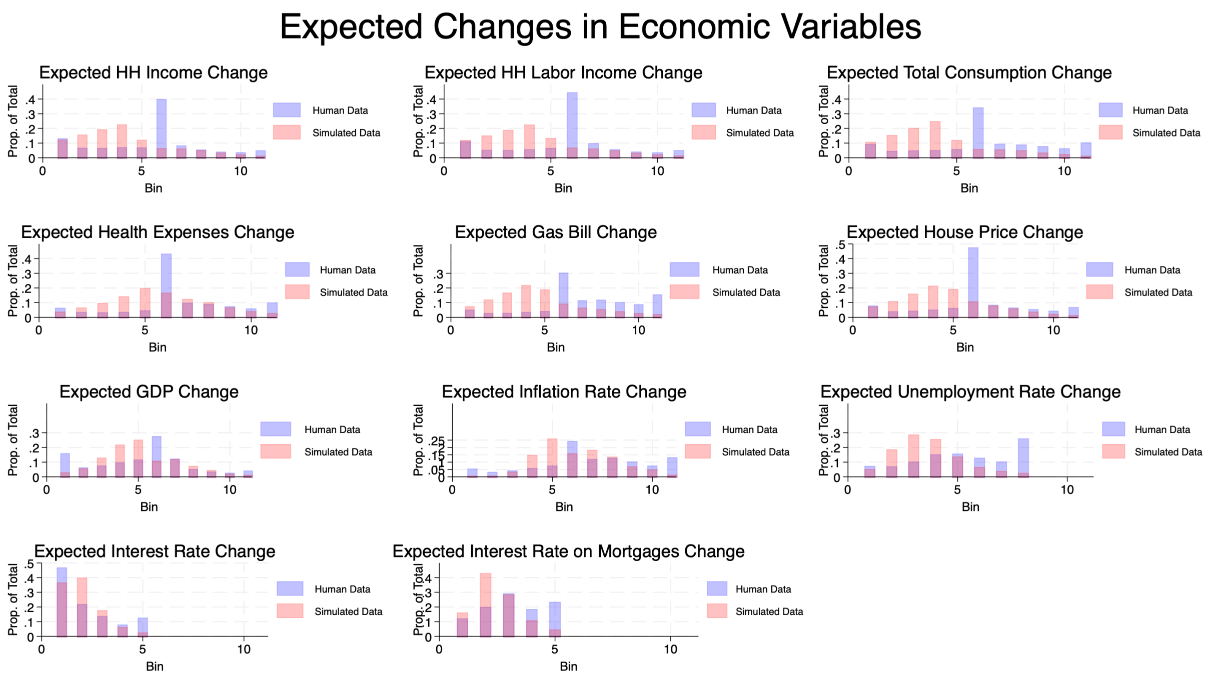


Figure 5: Distribution of Points Allocated: Human vs ChatGPT Simulated Responses

	HH Income				HH Labor Income				Consumption		
	Human	Simulated	P-value		Human	Simulated	P-value		Human	Simulated	P-value
g1	12.5	11.7	.02	g1	10.6	11.4	.02	g1	8.7	10.0	.00
g2	6.1	15.1	.00	g2	4.7	14.5	.00	g2	3.9	14.8	.00
g3	6.0	18.7	.00	g3	4.6	18.2	.00	g3	4.3	19.6	.00
g4	6.5	22.0	.00	g4	5.1	21.9	.00	g4	4.5	24.2	.00
g5	6.3	11.6	.00	g5	6.0	12.8	.00	g5	5.2	11.4	.00
g6	39.3	5.8	.00	g6	43.9	6.2	.00	g6	33.6	5.3	.00
g7	7.7	5.6	.00	g7	9.2	5.6	.00	g7	8.8	5.0	.00
g8	4.9	4.5	.01	g8	5.1	4.4	.00	g8	8.3	4.4	.00
g9	3.4	2.8	.00	g9	3.5	2.7	.00	g9	7.2	2.8	.00
g10	3.0	1.7	.00	g10	3.0	1.6	.00	g10	5.7	1.8	.00
g11	4.4	0.7	.00	g11	4.5	0.8	.00	g11	9.8	0.7	.00

	Health Expenses				Gas Bill				House Price		
	Human	Simulated	P-value		Human	Simulated	P-value		Human	Simulated	P-value
g1	5.8	3.1	.00	g1	4.5	6.8	.00	g1	7.3	6.7	.04
g2	2.9	6.0	.00	g2	2.2	11.4	.00	g2	3.4	10.3	.00
g3	2.7	9.1	.00	g3	2.3	16.1	.00	g3	3.8	15.3	.00
g4	2.9	13.6	.00	g4	2.9	21.2	.00	g4	4.5	20.8	.00
g5	4.0	19.3	.00	g5	3.5	18.3	.00	g5	5.6	18.6	.00
g6	42.7	16.1	.00	g6	29.7	8.6	.00	g6	46.9	10.1	.00
g7	9.2	11.8	.00	g7	10.9	5.9	.00	g7	7.8	7.3	.05
g8	8.4	9.6	.00	g8	11.3	4.8	.00	g8	6.0	5.3	.00
g9	6.8	6.0	.00	g9	9.7	3.4	.00	g9	4.9	3.1	.00
g10	5.3	3.4	.00	g10	8.2	2.1	.00	g10	3.8	1.6	.00
g11	9.4	2.1	.00	g11	14.9	1.5	.00	g11	6.2	0.7	.00

	GDP				Inflation				Unemployment		
	Human	Simulated	P-value		Human	Simulated	P-value		Human	Simulated	P-value
g1	15.4	2.3	.00	g1	4.9	0.0	.00	g1	6.7	4.5	.00
g2	5.7	5.1	.00	g2	2.6	0.1	.00	g2	6.5	18.0	.00
g3	7.0	12.3	.00	g3	3.7	2.8	.00	g3	9.7	28.1	.00
g4	9.1	21.2	.00	g4	5.3	14.3	.00	g4	14.6	25.0	.00
g5	11.1	24.5	.00	g5	6.9	25.4	.00	g5	15.1	13.1	.00
g6	27.0	10.2	.00	g6	23.7	15.4	.00	g6	12.3	5.9	.00
g7	11.7	11.3	.34	g7	11.4	17.7	.00	g7	9.7	3.3	.00
g8	4.6	6.7	.00	g8	12.3	13.0	.02	g8	25.5	1.9	.00
g9	2.9	3.9	.00	g9	9.7	6.4	.00				
g10	2.1	1.4	.00	g10	7.0	4.3	.00				
g11	3.6	0.7	.00	g11	12.6	0.8	.00				

	Interest Rate				Mortgage Interest Rate		
	Human	Simulated	P-value		Human	Simulated	P-value
g1	46.2	36.0	.00	g1	11.5	15.5	.00
g2	21.3	39.3	.00	g2	19.3	42.4	.00
g3	13.1	17.1	.00	g3	28.6	27.9	.20
g4	7.3	5.7	.00	g4	17.9	10.2	.00
g5	12.0	1.9	.00	g5	22.8	3.9	.00

Notes: The data represents a pairwise bootstrapped two-sample t-test comparing the distribution of responses between human participants and ChatGPT simulated data across demographic groups (g1-g11). p-values indicate the statistical significance of differences between human and simulated responses for each variable. Values represent percentages of responses within each group. All variables measured as expected changes relative to baseline period.

Table B9: Balance Table Across Multiple Panels

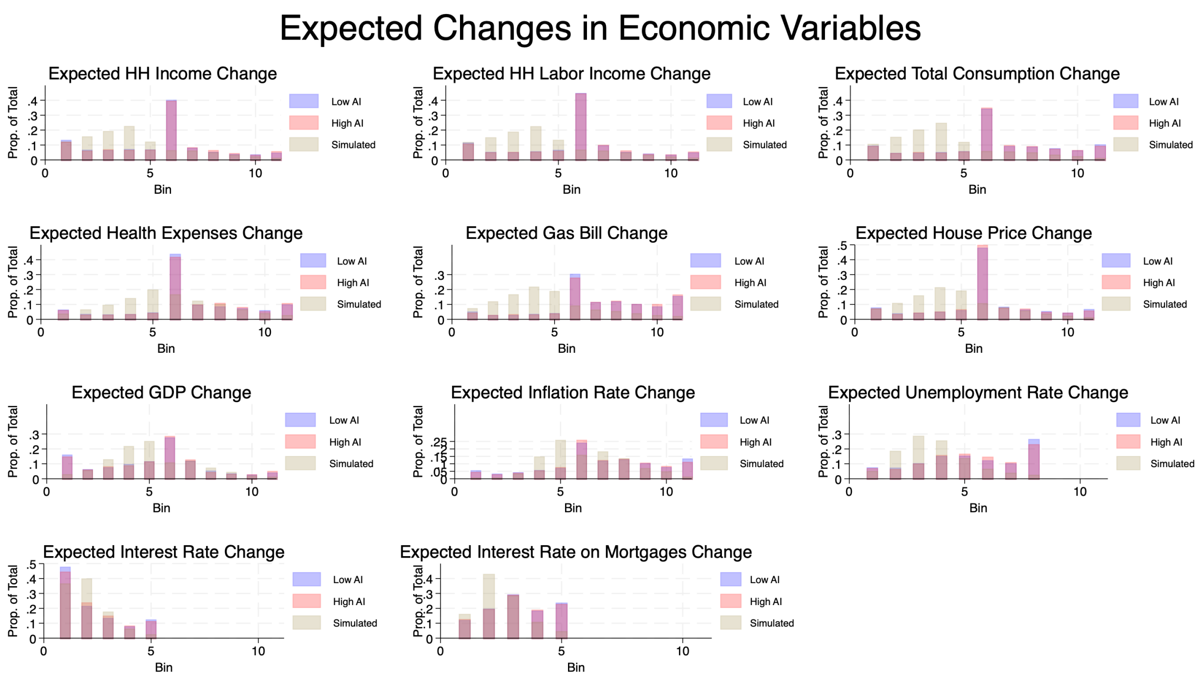


Figure 6: Distribution of Points Allocated: GPT vs Human Data by High or Low AI Use

B.6 Regression Results

Table B10: Regression Results: Human vs ChatGPT Responses

Variable	Consumption Risk	Income Risk	Health Risk	Energy Risk	GDP Risk	Unemp. Risk	Inflation Risk	Interest Rate Risk	House Price Risk
Panel A: Human Responses									
HH Members > 3	0.011**	0.008*	0.014***	0.009**	0.012***	0.008***	0.011***	0.518***	0.015***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.002)	(0.004)	(0.154)	(0.005)
Age > 49	-0.027***	-0.033***	-0.023***	-0.025***	-0.029***	-0.011***	-0.027***	-1.099***	-0.029***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.004)	(0.132)	(0.004)
College	-0.002	-0.005	-0.006	-0.008*	-0.013***	-0.006**	-0.010**	-0.227	-0.010**
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.003)	(0.005)	(0.163)	(0.005)
Homeowner	-0.021***	-0.028***	-0.024***	-0.020***	-0.026***	-0.013***	-0.025***	-0.688***	-0.017***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.003)	(0.005)	(0.157)	(0.005)
North or Centre	-0.025***	-0.021***	-0.023***	-0.021***	-0.024***	-0.013***	-0.022***	-0.630***	-0.017***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.004)	(0.138)	(0.004)
Mean of D.V.	0.096	0.086	0.089	0.083	0.080	0.049	0.079	2.299	0.078
Var of D.V.	0.018	0.016	0.017	0.015	0.016	0.005	0.015	18.872	0.016
N	4,814	4,814	4,814	4,814	4,814	4,814	4,814	4,814	4,814
Panel B: ChatGPT Simulated Responses									
HH Members > 3	0.001	0.003*	-0.003**	-0.005***	-0.003**	0.000	0.001	0.000	0.005***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Age > 49	-0.001	-0.005***	0.001	-0.000	0.009***	0.003***	0.000	-0.004***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
College	0.000	-0.005***	0.011***	0.007***	-0.013***	-0.012***	0.002*	0.005***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)
Homeowner	-0.001	-0.007***	0.018***	0.009***	-0.009***	-0.015***	0.000	0.003***	-0.010***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
North or Centre	-0.002**	-0.013***	-0.004***	0.004***	-0.012***	-0.014***	-0.001	-0.000	-0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Mean of D.V.	0.160	0.165	0.116	0.150	0.098	0.104	0.073	0.039	0.132
Var of D.V.	0.001	0.002	0.002	0.001	0.002	0.001	0.000	0.000	0.002
N	4,814	4,814	4,814	4,814	4,814	4,814	4,814	4,814	4,814

Note: Each column corresponds to a regression for a different perceived risk. Panel A reports estimates using actual human survey responses; Panel B uses ChatGPT-simulated responses (with demographics embedded). Standard errors are in parentheses. The final rows report the mean, variance, and sample size of the dependent variable. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B11: Regression Results High AI vs Low AI

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Low AI									
Variable	Consumption Risk	Income Risk	Health Risk	Energy Risk	GDP Risk	Unemp. Risk	Inflation Risk	Interest Rate Risk	House Price Risk
HH Members > 3	0.275***	0.210**	0.255***	0.211**	0.223**	0.211***	0.226***	1.738***	0.239***
	(0.087)	(0.088)	(0.089)	(0.085)	(0.088)	(0.067)	(0.085)	(0.553)	(0.090)
Age > 49	-0.402***	-0.581***	-0.353***	-0.384***	-0.496***	-0.276***	-0.441***	-3.267***	-0.467***
	(0.080)	(0.079)	(0.081)	(0.076)	(0.076)	(0.061)	(0.076)	(0.476)	(0.078)
North or Centre	-0.315***	-0.253***	-0.312***	-0.241***	-0.323***	-0.176***	-0.284***	-1.682***	-0.179**
	(0.082)	(0.083)	(0.083)	(0.080)	(0.082)	(0.063)	(0.079)	(0.499)	(0.083)
Homeowner	-0.348***	-0.468***	-0.325***	-0.308***	-0.427***	-0.310***	-0.411***	-2.185***	-0.236***
	(0.089)	(0.089)	(0.091)	(0.088)	(0.088)	(0.068)	(0.088)	(0.546)	(0.090)
College	0.071	-0.037	0.002	-0.035	-0.090	0.006	-0.096	-0.312	-0.038
	(0.093)	(0.091)	(0.094)	(0.089)	(0.090)	(0.069)	(0.087)	(0.580)	(0.091)
Mean of D.V.	2.254	1.985	2.057	2.038	1.917	1.516	1.913	7.990	1.815
Var of D.V.	4.693	4.759	4.811	4.318	4.484	2.662	4.317	170.272	4.688
N	3450	3450	3450	3450	3450	3450	3450	3450	3450
Panel B: High AI									
Variable	Consumption Risk	Income Risk	Health Risk	Energy Risk	GDP Risk	Unemp. Risk	Inflation Risk	Interest Rate Risk	House Price Risk
HH Members > 3	0.292	0.203	0.022	-0.012	-0.071	0.030	-0.081	1.611	-0.046
	(0.250)	(0.259)	(0.265)	(0.249)	(0.246)	(0.184)	(0.248)	(1.546)	(0.246)
Age > 49	-0.399*	-0.880***	-0.303	-0.557**	-0.576**	-0.328**	-0.598***	-4.289***	-0.547**
	(0.228)	(0.240)	(0.244)	(0.231)	(0.229)	(0.167)	(0.230)	(1.360)	(0.238)
North or Centre	-0.526**	-0.435*	-0.501**	-0.502**	-0.474**	-0.178	-0.453*	-1.804	-0.454*
	(0.250)	(0.248)	(0.247)	(0.244)	(0.237)	(0.172)	(0.236)	(1.376)	(0.237)
Homeowner	-0.109	-0.296	0.011	0.009	-0.258	-0.087	-0.029	-2.886*	-0.151
	(0.251)	(0.271)	(0.262)	(0.254)	(0.246)	(0.184)	(0.248)	(1.586)	(0.253)
College	0.442	0.220	0.419	0.352	0.139	0.047	0.345	0.539	0.630**
	(0.304)	(0.320)	(0.292)	(0.296)	(0.285)	(0.214)	(0.300)	(1.761)	(0.309)
Mean of D.V.	2.152	1.947	2.109	1.990	1.834	1.374	1.880	7.088	1.798
Var of D.V.	4.498	5.085	4.878	4.655	4.359	2.340	4.304	155.371	4.652
N	391	391	391	391	391	391	391	391	391

Notes: Each column corresponds to a regression for a different perceived risk. Panel A reports estimates using actual human survey responses for Low AI users which I define as survey respondents who use AI tools such as ChatGPT less than once a week; Panel B uses human survey responses for High AI users, which is defined as survey respondents who use AI tools such as ChatGPT once a week or more. Standard errors are in parentheses. The final rows report the mean, variance, and sample size of the dependent variable. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B12: Euler Equation Results

	Simulated Data		Human Data	
	(1)	(2)	(3)	(4)
Consumption Risk	-4.447***	-4.363***	0.168	1.343***
	(0.525)	(0.520)	(0.372)	(0.349)
Expected Labor Income Growth		0.031***		0.349***
		(0.007)		(0.025)
Mean of D.V.	-3.337	-3.337	0.518	0.518
Var of D.V.	0.101	0.101	17.331	17.331
N	4814	4814	4814	4814

Notes: The dependent variable is expected consumption growth. Consumption risk is the 2nd conditional moment of the distribution of expected consumption growth. Standard errors in parentheses, calculated through a bootstrap with with 10000 samples of identical sizes. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table B13: Regression Results: Low vs High AI Users

	Low AI		High AI	
	(1)	(2)	(3)	(4)
Consumption Risk	0.078	1.222***	-0.484	0.569
	(0.432)	(0.412)	(1.205)	(1.133)
Expected Labor Income Growth		0.347***		0.341***
		(0.030)		(0.093)
Mean of D.V.	0.514	0.514	0.340	0.340
Var of D.V.	17.372	17.372	19.015	19.015
N	3450	3450	391	391

Notes: The dependent variable is expected consumption growth. Consumption risk is the 2nd conditional moment of the distribution of expected consumption growth. High AI refers to individuals which responded using AI once a week or more in the ISCE, and Low AI refers otherwise. Standard errors in parentheses, calculated through a bootstrap with with 10000 samples of identical sizes. *** p<0.01, ** p<0.05, * p<0.1.

C Guessing Current Traits

C.1 Question Format

HH Income (October 2023)

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Given the above characteristics, into which bracket does the *average total monthly household income*, net of all taxes, fall for an Italian individual with these characteristics in October 2023? Please consider the entirety of earnings (income, pensions, transfers, income from property and from financial assets) of all household members.

HH Income (expected–January 2024)

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Given the above characteristics, into which bracket does the *average total monthly household income expected for January 2024*, net of all taxes, fall for an Italian individual with these characteristics in October 2023? Consider all forms of earnings for every household member (income, pensions, transfers, income from property and from financial assets).

HH Income (expected–one year ahead)

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Given the above characteristics, into which bracket does the *average total monthly household income expected one year from now*, net of all taxes, fall for an Italian individual with these characteristics in October 2023? Consider all forms of earnings for every household member (income, pensions, transfers, income from property and from financial assets).

Consumption (October 2023)

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Given the above characteristics, into which bracket do the *average total monthly household consumptions* fall for an Italian individual with these characteristics in October 2023? Consider all expenses (food and non-food consumption, rent, loan/mortgage payments, insurance, utilities, etc.) of all household members.

Consumption (expected–4 months ahead 2024)

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Given the above characteristics, into which bracket does the *average total monthly household consumption expected for January 2024* fall for an Italian individual with these characteristics in October 2023? Consider all expenses (food and non-food consumption, rent, loan/mortgage payments, insurance, utilities, etc.) of all household members.

Consumption (expected—one year ahead)

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Given the above characteristics, into which bracket does the *average total monthly household consumption expected one year from now* fall for an Italian individual with these characteristics in October 2023? Consider all expenses (food and non-food consumption, rent, loan/mortgage payments, insurance, utilities, etc.) of all household members.

E1 — HH income-growth distribution

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Distribute **exactly 100 points** across the following scenarios (write just the numbers; they must sum to 100). Over the next year, you expect that yMy family's *total annual income*, net of taxes and state transfers, compared with last year...

- will decrease by more than 8 %: ___________
- will decrease by 6–8 %: ___________
- will decrease by 4–6 %: ___________
- will decrease by 2–4 %: ___________
- will decrease by 0–2 %: ___________
- will remain unchanged: ___________
- will increase by 0–2 %: ___________

- will increase by 2–4 %: __________
- will increase by 4–6 %: __________
- will increase by 6–8 %: __________
- will increase by more than 8 %: __________

E3 — consumption-growth distribution

An Italian [male / female], [aged 50–75 / aged 18–49] who is [employed / unemployed].

Distribute **exactly 100 points** across the following scenarios (numbers only; must sum to 100). Over the next year, you expect that My family's *total consumption* (all expenses)…

- will decrease by more than 8 %: __________
- will decrease by 6–8 %: __________
- will decrease by 4–6 %: __________
- will decrease by 2–4 %: __________
- will decrease by 0–2 %: __________
- will remain unchanged: __________
- will increase by 0–2 %: __________
- will increase by 2–4 %: __________
- will increase by 4–6 %: __________
- will increase by 6–8 %: __________
- will increase by more than 8 %: __________

C.2 Summary Statistics

Table C1: Summary Statistics Comparison: Human vs Simulated Data

	Human Data					ChatGPT Simulated Data				
Variable	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
Demographic Variables										
Male	2,842	0.559	0.497	0	1	2,842	0.559	0.497	0	1
Employed	2,842	0.840	0.366	0	1	2,842	0.840	0.366	0	1
Age $>$ 49	2,842	0.347	0.476	0	1	2,842	0.347	0.476	0	1
Income and Cost Variables										
HH Income October 2023	2,842	2,230.03	1,809.65	750	20,000	2,842	1817.73	331.24	1,250	2,750
Consumption October 2023	2,842	1,510.29	1,568.25	750	20,000	2,842	1,647.26	202.50	1,250	2,250
Future Income Variables										
HH Income January 2024	2,325	2,268.48	1,682.95	750	20,000	2,842	1,813.16	328.68	750	2,250
HH Income October 2024	2,341	2,255.53	1,507.71	750	20,000	2,842	1,813.16	328.68	750	2,250
Expected Consumption Variables										
Consumption January 2024	2,325	1,513.66	1,684.57	750	20,000	2,842	1,637.93	208.54	1,250	1,750
Consumption October 2024	2,341	1,548.06	1,599.22	750	20,000	2,842	1,637.93	208.54	1,250	1,750

Notes: Human data represents the participants in the original Wave 1 ISCE dataset (in waves administered in October 2023, October 2024 and January 2024), while Simulated data represents AI-generated responses. I restrict my sample to individuals who reported individual income and household income, as well as those that explicitly reported home-ownership/rental status.

C.3 Accuracy of Current Traits

Table C2: Bootstrap Results for Current Total Household Income

Parameter	Mean	Std. Dev.	Significance
Accuracy (bin)	0.743	0.099	***
Accuracy (>2,500)	0.901	0.082	***
Pearson's ρ	0.707	0.042	***
Spearman's ρ	0.714	0.023	***
Kendall's τ	0.494	0.052	***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Results based on 10,000 pairs bootstrap iterations of human and simulated data, stratified by employment status, gender, and age>49. For each bootstrap sample, the same number of observations were drawn from each dataset. Accuracy (bin) represents the percentage of observations correctly classified into income bins. Accuracy (>2,500) represents the percentage of observations correctly classified as having income greater than 2,500 Euros.

Table C3: Bootstrap Results for Current Consumption

Parameter	Mean	Std. Dev.	Significance
Accuracy (bin)	0.725	0.135	***
Accuracy (>1,250)	0.873	0.088	***
Pearson's ρ	0.798	0.105	***
Spearman's ρ	0.707	0.124	***
Kendall's τ	0.538	0.115	***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Results based on 10,000 pairs bootstrap iterations of human and simulated data, stratified by employment status, gender, and age>49. For each bootstrap sample, the same number of observations were drawn from each dataset. Accuracy (bin) represents the percentage of observations correctly classified into consumption bins. Accuracy (>1,250) represents the percentage of observations correctly classified as having consumption greater than 1,250 Euros.

Table C4: Bootstrap Results for Future Income Predictions

Parameter	Mean	Std. Dev.	Significance
Income 4 Months in the Future			
Accuracy (bin)	0.248	0.017	***
Accuracy (>€2,500)	0.881	0.049	***
Pearson's ρ	0.187	0.277	
Spearman's ρ	0.147	0.310	
Kendall's τ	0.116	0.242	
Income 12 Months in the Future			
Accuracy (bin)	0.251	0.016	***
Accuracy (>€2,500)	0.965	0.062	***
Pearson's ρ	-0.256	0.178	
Spearman's ρ	-0.133	0.300	
Kendall's τ	-0.096	0.233	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Significance determined by whether 95% confidence intervals (±1.96 SE) include zero. Results based on 10,000 pairs bootstrap iterations of human and simulated data, stratified by employment status, gender, and age>49. For each bootstrap sample, the same number of observations were drawn from each dataset. Accuracy (bin) represents the percentage of observations correctly classified into future income bins. Accuracy (>€2,500) represents the percentage of observations correctly classified as having future income greater than €2,500.

Table C5: Bootstrap Results for Future Consumption Predictions

Parameter	Mean	Std. Dev.	Significance
Consumption 4 Months in the Future			
Accuracy (bin)	0.579	0.146	***
Accuracy (>€1,250)	0.999	0.013	***
Pearson's ρ	0.077	0.315	
Spearman's ρ	0.106	0.279	
Kendall's τ	0.061	0.233	
Consumption 12 Months in the Future			
Accuracy (bin)	0.574	0.116	***
Accuracy (>€1,250)	0.965	0.056	***
Pearson's ρ	-0.146	0.184	
Spearman's ρ	0.152	0.180	
Kendall's τ	0.143	0.159	

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Significance determined by whether 95% confidence intervals (±1.96 SE) include zero. Results based on 10,000 bootstrap iterations. Accuracy (bin) represents the percentage of observations correctly classified into future consumption bins. Accuracy (>€2,500) represents the percentage of observations correctly classified as having future consumption greater than €2,500.

Table C6: Bootstrap Results for Expected Consumption and Income

Parameter	Mean	Std. Dev.	Significance
Mean Expected Consumption			
Pearson's ρ	0.619	0.155	***
Spearman's ρ	0.611	0.158	***
Kendall's τ	0.413	0.149	***
Mean Expected Income			
Pearson's ρ	0.507	0.124	***
Spearman's ρ	0.491	0.133	***
Kendall's τ	0.374	0.116	***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Results based on 10,000 pairs bootstrap iterations of human and simulated data, stratified by employment status, gender, and Age>49.

C.4 Extension: Recovering Policy Function

Table C7: Accuracy Scores by Economic Prediction Model

Variable	**HH Income**	**Consumption**	**Health**	**House**	**GDP**
Homeowner	0.6540	0.7435	0.7309	0.6996	0.6692
College	0.3616	0.2283	0.2557	0.2918	0.3404
Age $>$ 49	0.5178	0.5122	0.5569	0.5344	0.5070
North or Centre	0.5883	0.6568	0.6458	0.6410	0.6217
HH Members $>$ 3	0.4718	0.6872	0.4423	0.7275	0.6488

Note: Accuracy—shown above for five separate prediction targets—is the share of correct guesses of demographic traits given the elicited distribution of the change in a given variable across the full Wave 1 sample. ChatGPT crosses the 50 % threshold consistently for homeownership, age, and region, and in three of five cases for household size, suggesting that even if it cannot reproduce the full expectations distribution, it can still help impute persistent demographic traits when survey data are incomplete.

D Robustness

D.1 Information Treatment

Table D8: Regression Results: Simulated Data Analysis

	Simulated Data (temperature=1.0)	
	(1)	(2)
T3	-2.030***	-1.134***
	(0.491)	(0.285)
T2	-7.452***	-4.702***
	(0.475)	(0.320)
G2	2.452***	4.883***
	(0.402)	(0.282)
T2G2	5.500***	
	(0.682)	
T3G2	1.792***	
	(0.659)	
Mean of D.V.	29.73	29.73
Var of D.V.	107.84	107.84
N	5040	5040

Notes: Standard errors are computed with a stratified bootstrapped within each of the 6 information treatment groups of 100 draws, where I draw a sample size identical to the original size of each of the 6 information treatment groups. The outcome variable (D.V.) is the amount contributed to the disaster fund. As opposed to the baseline case of temperature of 0.8, I use a temperature of 1.0. Bootstrap standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D9: Regression Results: Simulated Data Analysis (Temperature = 1.0)

	Simulated Data (Temperature = 1.0)			
	(1)	(2)	(3)	(4)
T3	0.275	-5.044***	0.128	-5.326***
	(1.168)	(0.907)	(1.584)	(1.422)
T2	5.237***	3.696***	4.064**	3.219*
	(1.142)	(1.148)	(1.875)	(1.751)
G2	10.280***	5.647***	9.517***	5.262***
	(1.296)	(0.657)	(1.949)	(1.152)
T2G2	-3.148*		-1.757	
	(1.871)		(3.059)	
T3G2	-10.778***		-11.050***	
	(1.724)		(2.748)	
Employed	9.194***	9.108***		
	(0.855)	(0.717)		
Retired	4.615***	4.556***		
	(1.170)	(1.148)		
Male	12.187***	12.165***		
	(0.767)	(0.799)		
HH Income > 2500	52.954***	52.960***		
	(1.035)	(1.028)		
College	33.261***	33.356***		
	(1.347)	(1.114)		
Mean of D.V.	42.068	42.068	42.068	42.068
Var of D.V.	1805.989	1805.989	1805.989	1805.989
N	4517	4517	4517	4517

Note: Simulated Data refers to ChatGPT generated simulated samples where groups (tuples) of employed-retired-male-HH Income > 2500-college with less than 50 observations are dropped. Temperature parameter set to 1.0 for all simulations. Standard errors are bootstrapped 100 times, within each combination of demographic groups (Employment-Gender-Income>2500-College). The outcome variable (D.V.) is the amount contributed to the disaster fund. Bootstrap standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D10: Regression Results: Simulated Data with Sample Size Identical to Human Data (N=5001)

	Simulated Data	
	(1)	(2)
T3	-4.026^{***}	-2.425^{***}
	(0.435)	(0.267)
T2	-10.137^{***}	-6.957^{***}
	(0.457)	(0.294)
G2	2.366^{***}	5.566^{***}
	(0.308)	(0.240)
T2G2	6.366^{***}	
	(0.578)	
T3G2	3.228^{***}	
	(0.534)	
Mean of D.V.	29.701	29.701
Var of D.V.	90.396	90.396
N	5001	5001

Notes: Standard errors are computed with a stratified bootstrapped within each of the 6 information treatment groups of 100 draws, where the sample size is identical to the original size of each of the 6 information treatment groups. The outcome variable (D.V.) is the amount contributed to the disaster fund. A baseline temperature of 0.8 is used to simulate the human treatment groups such that the number of observations in each treatment group are identical to that of the human data (summing up to 5001 observations). Bootstrap standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.1.1 New Question Format

The prompts are otherwise identical, except for that on groups T2G1, T2G2, T3G1, T3G2, where they are following this T2 and T3 respectively:

- T2: There was a flood that caused fifteen deaths and about 40,000 displaced people.
- T3: There was a flood that caused fifteen deaths and about 40,000 displaced people. The Region calculated damages of almost 9 billion for roads, schools, embankments and canals, as well as to repair damage to homes and businesses.

Table D11: Baseline Information Treatment Regression with Non-Romagna specific Prompt

	(1)	(2)
T3	0.595	-1.190***
	(0.400)	(0.260)
T2	1.095***	-3.643***
	(0.342)	(0.252)
G2	1.857***	-2.492***
	(0.300)	(0.244)
T2G2	-9.476***	
	(0.406)	
T3G2	-3.571***	
	(0.564)	
Mean of D.V.	30.579	30.579
Var of D.V.	75.552	75.552
N	5040	5040

Notes: Standard errors are computed with a stratified bootstrapped within each of the 6 information treatment groups of 10000 draws, where I draw a sample size identical to the original size of each of the 6 information treatment groups. The outcome variable (D.V.) is the amount contributed to the disaster fund. The sample is sampled using ChatGPT (temperature = 0.8) without demographics embedded, using a prompt which does not specify the location of the flood in Italy. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Expectation Formation

Table D12: Regression Results (Temperature = 1)

Variable	Consumption Risk	Income Risk	Health Risk	Energy Risk	GDP Risk	Unemp. Risk	Inflation Risk	Interest Rate Risk	House Price Risk
HH Members > 3	0.001	0.004***	0.000	-0.004***	-0.001	0.000	0.002***	0.001	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Age > 49	-0.001	-0.002	-0.002	-0.001	0.008***	0.001	0.000	-0.004***	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)
North or Centre	-0.003***	-0.012***	-0.009***	0.002*	-0.012***	-0.014***	-0.001	0.001**	-0.014***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Homeowner	-0.003***	-0.007***	0.020***	0.007***	-0.012***	-0.018***	0.001	0.003***	-0.006***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)
College	-0.001	-0.005***	0.009***	0.008***	-0.011***	-0.014***	0.003***	0.005***	-0.002
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)
Mean of D.V.	0.162	0.166	0.117	0.150	0.101	0.103	0.074	0.040	0.133
Var of D.V.	0.001	0.002	0.002	0.001	0.002	0.001	0.001	0.000	0.002
N	4814	4814	4814	4814	4814	4814	4814	4814	4814

Notes: Each column corresponds to a regression for a different perceived risk. Data is from a ChatGPT simulation, with a temperature of 1 (as opposed to a temperature of 0.8). Standard errors are in parentheses. The final rows report the mean, variance, and sample size of the dependent variable. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D13: Euler Equation (Temperature = 1)

	(1)	(2)
Consumption Risk	-6.453***	-6.379***
	(0.478)	(0.560)
Expected Labor Income Growth		0.030***
		(0.008)
Mean of D.V.	-3.280	-3.280
Var of D.V.	0.226	0.226
N	4814	4814

Notes: The dependent variable is expected consumption growth. Consumption risk is the 2nd conditional moment of the distribution of expected consumption growth. Data is from a ChatGPT simulation, with a temperature of 1 (as opposed to a temperature of 0.8). Standard errors in parentheses, calculated through a bootstrap with with 100 samples of identical sizes. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$