

Supplemental Appendix

A Additional tables

Table A1: Overview of studies: Political economy

Paper name	Domain	Measurement	Analysis of text data
Andre (2025)	Explanation of people's distributive choices	"Please explain why you made your choice the way you did."	Hand-coding of responses.
Bursztyn et al. (2023)	Motives for choosing Tweet	"Why did you choose this Tweet rather than the other Tweet?"	Word counts and simple machine learning techniques.
Braghieri, Schwardmann and Tripodi (2024)	Reasons for self-selection into "echo chamber" and positive effects of cross-partisan communication	Conversation topics in a recorded and transcribed video chat, as well as reasons for the willingness to pay for the option to have a video chat.	Hand-coding of responses regarding willingness to pay and topic prediction using ChatGPT 3.5 turbo.
Bursztyn et al. (2020)	Beliefs about motives underlying xenophobic expression	"Why do you think your matched respondent chose to donate to Fund the Wall?" On top of this, the authors employ structured measures of beliefs about the matched respondent's type.	Pre-specified word counting procedure; Support Vector Machine classifier to predict structured belief measures based on text data.
Dechezleprêtre et al. (2025)	Considerations about climate change	"When thinking about climate change, what are your main considerations? What should [country] government do regarding climate change?"	Text analysis.
Galasso et al. (2024)	Anti-populist video ads regarding a populist referendum in Italy	Respondents were invited to share their considerations about a video in an open-ended question to compare how two videos are comparatively perceived.	Hand-coding and supervised text analysis (FEEL-IT)
Gründler and Potrafke (2020)	Attitudes towards fiscal rules	"What are your main considerations about fiscal rules?"; "What should be the goal of fiscal rules?"; "What are the main shortcomings of fiscal rules."	Word cloud and ML methods for sentiment analysis.
Hager et al. (2025)	Voice and political engagement	4 Treatment groups with open-ended questions: "Would you like to tell us more about which issues we should particularly emphasize in the election campaign?(/Would you like to tell us more about which topics are particularly close to your heart?)(/After the completion of the survey, we will send you a summary of the results.)"	Hand-coding of responses.
Hager et al. (2023)	Motives underlying change in effort in response to info	"Why would the results of the survey affect or not affect your decision? Please answer using whole sentences"	Human coding of scripts.
Hüning, Mechtenberg and Wang (2022)	Attitudes towards rent control	Discuss pro and cons of rent control	NLP and text analysis techniques.
Jessen et al. (2024)	Policy demand to reduce socioeconomic inequality in life expectancy	List as many measures as possible that the government could use to improve the life expectancy of the poor.	Word cloud.
König and Schmacker (2022)	Sin taxes, i.e., taxes on sugar-sweetened beverages (SSB)	4 Open-ended questions: 1st the general considerations of SSB taxes; 2nd regarding the goals; 3rd and 4th the benefiting respective losing groups.	Word cloud and keyness analysis.
Lang and Schneider (2023)	Influence of post-WWII German immigrant movement on nationalist sentiment and electoral responses	"What do you think is the significance of the fact that many Germans had experience of expulsion, flight, and immigration?"	Hand-coding of responses.
Liscow and Fox (2022)	Attitudes towards capital tax realization rule	Why preferred to defer taxation until sold respectively why preferred to tax before sold.	Word counting.
Luttmer and Samwick (2018)	Impact of policy uncertainty in social security on individual welfare	"We are interested in better understanding why you chose uncertain benefits around [B]% of the Social Security benefits you are supposed to get under current law over guaranteed benefits equal to [L]% of the Social Security benefits you are supposed to get under current law. Could you tell us the main reason for your choice?"	Hand-coding of responses.
Nathan, Perez-Truglia and Zentner (2025)	Reasoning for filling property tax complain or not.	"If you can, please explain why you will (or will not) protest in 2020."	Handcoding of responses.
Stantcheva (2020)	Understanding, reason and learning about 4 economic policies: i) income, and ii) estate taxation, iii) health insurance, iv) trade	"What are your main considerations about [policy]...?" and more specific sub-questions regarding perceived goals and shortcomings, as well as the anticipated effects (e.g., which group would gain) from the specific policy.	Text analysis techniques (keyness analysis, topic analysis, word clouds).
Stantcheva (2021)	First-order concerns about income and estate tax	"When you think about federal personal income taxation and whether the U.S. should have higher or lower federal personal income taxes (/federal estate tax), what are the main considerations that come to your mind?"	Text analysis techniques (keyness analysis, topic analysis, word clouds).
Stantcheva (2022)	Attitudes towards trade	"When you think about trade policy and whether the U.S. should put some restrictions on trade with other countries, such as tariffs, what are the main considerations that come to your mind?" "What would be the effects on the U.S. economy if barriers to trade, such as tariffs, were increased?" "Which groups of people do you think would gain if trade barriers such as tariffs were increased?"	Text analysis techniques (keyness analysis, topic analysis, word clouds)

Table A2: Overview of studies: Political science

Paper name	Domain	Measurement	Analysis of text data
Breyer, Palmtag and Zollinger (2023)	Perceived status gains for women or minorities.	"E.g., Now think about the people who have tended to gain recognition compared to the past. How would you describe these people? What kind of characteristics, lifestyles, and opinions do these people have?"	Text analysis with a parsimonious dictionary.
Roberts et al. (2014)	Views on immigration; Intuition versus reflection in Public Goods Game; American National Election Survey	"Of the stories you read, what stories do you remember best? (If you don't remember the names, just describe the stories)."; "Please write a paragraph (approximately 8-10 sentences) describing a time your intuition/first instinct(/time carefully reasoning) led you in the wrong(/right) direction and resulted in a bad(/good) outcome."; "What has been the most important issue to you personally in this election?" and "What do you think is the most important political problem facing the United States today?"	Introduce their Structural Topic Model (STM), which relies on machine learning methods, and apply it to the three examples and compare it to hand-coding.
Rothschild et al. (2019)	Stereotypes about the two American parties	"Please write down four words that typically describe people who support the [Democratic/Republican] Party."	Structural Topic modelling.
Zollinger (2022)	Attitudes towards voter-party links	"If you imagine people with a lifestyle and opinions similar to your own, what kind of people would these be? How would you describe them?"; "And someone who is not at all like you? Someone who lives completely differently and who has very different opinions? How would you describe them?"	Text analysis techniques (keyness analysis, latent semantic scaling)

Table A3: Overview of studies: Macroeconomics

Paper name	Domain	Measurement	Analysis of text data
An, Binder and Sheng (2023)	Gas price and inflation expectations	Asked respondents to describe the main considerations that come to their mind regarding the impact of the war on China's economy, overall prices in China, and gas prices in China.	Word cloud and hand-coding.
Andre et al. (2022)	Unemployment and inflation predictions	Ask respondents about their "main considerations in making the prediction" and about how they "come up with [their] prediction". On top of this, the authors employ structured measures of the considerations respondents had on their mind.	Human coding of scripts and simple word counting techniques.
Andre et al. (2025)	Narratives about the rise of inflation	Ask respondents "Which factors caused the rise in inflation?"	Human coding of text into DAGs.
Binetti, Nuzzi and Stantcheva (2024)	People's understanding of inflation	Ask respondents a series of qualitative questions about the relationship between inflation and economic activity, its causes, distributional impacts, and perceived consequences. For example, 'In your opinion, what are the primary causes of inflation?'	Hand-coding of responses.
Hommes, Pinter and Salle (2023)	Prior Knowledge of Public Finance	"Which risk(s) do you have in mind?" or "Which advantage(s) do you have in mind?"	Word cloud and classification.
Leiser and Drori (2005)	Inflation expectations	Ask participants to specify terms, concepts, or short phrases related to inflation.	Human coding of text and classification into broader categories.
Link et al. (2024)	Current economic situation	"What topics come to mind when you think about the economic situation of your company/household?"	Human coding of scripts and word counting.
Stantcheva (2024)	Causes and personal impact of inflation	E.g., "What were the most important impacts of inflation on your life?"; "When inflation gets very high, what do you think is the reason?"	Topic analysis and word clouds.

Table A4: Overview of studies: Labor economics

Paper name	Domain	Measurement	Analysis of text data
Abeler, Huffman and Raymond (2023)	Incentive complexity and effort provision	"If someone were trying to get the most money, total, from [Period 3 and Period 4], what do you think would be the best approach?"	Hand-coding of responses.
Ayyar et al. (2024)	Gender attitudes influence on lifetime earnings.	"Imagine you are now 25 years old. Write about the life you are leading, your interests, your home life and your work at the age of 25. (You have 30 minutes to do so)."	Word-embedding model to identify gender attitudes in essays.
Capozza (2024)	Concerns of women regarding gender gap in salary negotiations	"Which factors do you think caused the gender gap in salary negotiation?"	Word cloud, keyness analysis and hand-coding.
Casarico, Schuetz and Uebelmesser (2024)	Causes of gender gap in earnings and pensions in Germany	"What do you think are the causes of the differences between men and women in gross annual earnings and retirement pensions?"	Word cloud and keyness analysis.
Dube, Naidu and Reich (2022)	Preferences for wages and non-wage amenities	Research Assistants conducted in-depth interviews with 87 employees	Define survey measures of dignity based on the topics mentioned in the interviews
Erkal, Gangadharan and Xiao (2022)	Default selection processes in leadership influence on gender gaps	"In Stage 1 of Experiment II, you chose the following method: (Method X / Method Y / Indifferent). Please explain your decision. "	Incentivized (alignment with modal of other coders) hand-coding of responses by subjects.
Kaur et al. (2025)	Financial worries	"What makes you worry about money issues?"	Word clouds and text-counting.
Miano (2023)	Beliefs about on-the-job search	"Imagine you wanted to look for a new job at a new employer now, while still working at your current employer. Are there any issues that would make looking for a new job difficult for you now? What are the first ones that come to your mind?"	Word cloud.
Oh (2023)	Labor supply decisions related to caste	During the follow-up survey, workers were asked why they turned down specific offers.	Surveyors classified free-form answers based on training.
Rodrik and Stantcheva (2021)	Beliefs about what makes a good job	"What is a good job?"	Text analysis techniques

Table A5: Overview of studies: Finance

Paper name	Domain	Measurement	Analysis of text data
Andre, Schirmer and Wohlfart (2024)	Prediction of Stock Market Returns	Ask respondents for their reasoning for their stock return predictions based on a pair of hypothetical scenarios involving stale news about future company earnings.	Word count and hand-coding of open-ended data.
Ba, Rivera and Whitefield (2023)	Forecasting the stock market impact of racial uprisings	"Please explain your prediction using 2 to 3 sentences."	Hand-coding of responses.
Bailey et al. (2019)	Mortgage leverage choice	Ask respondents why their mortgage leverage choice differs across hypothetical scenarios with different projected home price changes.	Overview of representative text responses
Bauer et al. (2024)	Experts' beliefs about climate risk pricing	"Which factors do you think cause a deviation from correct pricing of climate risks? Please respond in full sentences."	Hand-coding of responses.
Chinco, Hartzmark and Sussman (2022)	Stock investment decisions	Ask respondents what factors are most important to them when deciding what fraction of an endowment to invest in stocks	Self-classification of open-ended responses by survey participants.
Chopra and Haaland (2024)	Stock market non-participation puzzle	Conduct AI-assisted interviews with respondents, exploring their reasons for not investing in the stock market including an "what if" scenario and counterfactual reasoning.	Hand-coding 50 interviews and using these for OpenAI's API to query GPT-4 for code assignment.
Chopra, Roth and Wohlfart (2025)	Home Price Expectations	"How would this change in your expectations about future home prices affect your expectations about your household's future economic situation. Please explain why. Respond in full sentences."	Hand-coding of open-ended data.
Filippini, Leipold and Wekhof (2024)	Financial literacy	"Describe which characteristics you think distinguish sustainable financial products from conventional investments. Please write a short text of about three sentences"	Text analysis techniques (topic specific word counts)
Jiang et al. (2025)	Selective recall of past returns	"First think about the overall stock market movement since you opened an account. Since you started trading, what is the episode of market movement that first comes to mind? Please enter the starting month and ending month of this episode."	Hand-coding of dates.
Laudenbach et al. (2024)	Beliefs about the stock market	Ask respondents to describe in open text which specific historical episodes – if any – they had in mind when estimating the historical autocorrelation of aggregate stock returns.	Human coding of text responses into different historical episodes.
Wekhof (2024)	Intention-Behavior gap sustainable investments	Hypothetical investment scenario: "What criteria would be important to you when choosing a fund? Please write a short text with about 3 sentences."	Semi-manual classification approach from Wekhof and Houde (2023).

Table A6: Overview of studies: Behavioral economics

Paper name	Domain	Measurement	Analysis of text data
Arrieta and Nielsen (2024)	Explanation of people's lottery and charity choices	"Please write a message to another participant describing how you made your last five decisions."	Other respondents replicate choice with or without message, Robustness with GPT-4 which also classified text in procedural categories.
Bordalo et al. (2025)	Explanation of people's solving strategies for statistical problems	"Could you describe to us in your own words how you came up with your answer to the previous question?"	Classification with GPT-3.5 which specific features of the problem were attended to.
Bordalo et al. (2023)	Free Recall of shown words	"Please list up to 15 [category] that you remember seeing in the list of words we showed you. You do not have to fill in all 15 lines."	Hand-coding of responses.
Bursztyjn et al. (n.d.)	Motives for preferring a world without Tiktok or Instagram and feelings about being the sole user to leave the platform	"You mentioned you would prefer to live in a world without [platform]. Why do you still use it?" and "How would you feel if you were the only one who deactivated [platform] and everyone else kept using it?"	Hand-coding of responses and AI based classification using LLM.
Castagnetti and Schmacker (2022)	Motivated information selection and updating	"Please explain, in general, how you decided between feedback modes across the five scenarios. For example, why did you choose one feedback mode over another? What specific characteristics of the feedback modes were you looking at?"	Hand-coding of responses.
Chopra, Haaland and Roth (2024)	Motives for (not) subscribing to newsletter	"Why did you (not) subscribe to the newsletter?"	Word counts and simple machine learning techniques.
Conlon (2024)	Role of rehearsal	Conversation with LLM which focus either on Successes or Failures in past academic experiences.	Conversation incentivized by engagement which was classified by LLM in a scale 0 to 100.
Agranov and Ortleva (2017)	Motives for choices between lotteries	"In Part III of the experiment each question was asked to you three times. If you chose different options, could you please tell us why did you do it? (Please elaborate)."	Hand-coding of responses.
Elias, Lacetera and Macis (2023)	Attitudes towards sudden price increases and price regulation	"Using the slider below, please rate this scenario as: -10 (completely unfair) to +10 (completely fair)", "We now ask you to select, among the two scenarios described above, the one that you would prefer to have in place in your country.", "Please briefly describe in the space provided the main reason(s) for your answers and choice above"	Text analysis.
Graeber, Roth and Zimmermann (2024)	Memories about information provided	"Please tell us anything you remember about this product scenario. Include as much detail as you can. Most importantly, please describe things in the order they come to mind, i.e., the first thought first, then the next one etc."	Hand-coding of responses.
Graeber, Roth and Schesch (2024)	Verbal explanations of financial reasoning choices	"We are interested in how you would give advice in an informal conversation: You should share an explanation behind your response. Your recording will be played to a few other participants who will have to respond to the same question."	Transcribe transcripts by Phonic using Amazon Transcribe and GPT-4 for classification of transcribed text.
Graeber, Noy and Roth (2024)	Oral transmission of information using speech recordings	"Think about the first(/ or second) opinion you listened to about changes in house price growth in a large US city. We will now ask you to record a voice message summarizing this opinion."	Classify by hand-coding and GPT of responses whether level and reliability are transmitted.
Grunewald et al. (2024)	Potential reinforcement of motivated beliefs through communication.	"[Quotes & Chat] To start the conversation and to give you some food for thought, here are two quotes by famous personalities: I think we are living in selfish times. — Javier Bardem, Hollywood actor and Oscar winner I'm just thankful I'm surrounded by good people. — Jon Pardi, singer and songwriter"	Word lists and bigram and trigram analysis.
Houser and Xiao (2011)	Communication influence on coordination game	Messages from Charness and Dufwenberg (2006).	Hand-coding of responses vs. incentivized (alignment with modal of other coders) hand-coding of responses.
Jabarian and Sartori (2024)	How the storytelling surrounding an information impacts the effectiveness of a survey eliciting reasoned preferences	Critical reasoning essays written by the respondents on the topic at hand.	Grading by doctoral-level psychologists determining whether the respondent is a critical or naive thinker.
Kaufmann, Andre and Kőszegi (2024)	Explanation of people's impact prediction on externalities	"Please explain why you chose this response.", "Please explain why you gave the same(/different) answer(/s) in the two situations." and "Please explain why you would be willing to pay money in situation 2 where the total impact would be zero."	Hand-coding of responses.
Martínez-Marquina, Niederle and Vespa (2019)	Provide incentivized advice to another participant for making a guess	"In the box below you can provide advice on what price you think the advisee should submit and a justification for your recommendation."	Hand-coding of responses.
Roth, Schwarzmann and Tripodi (2024a)	Social Stigma and demand for psychotherapy	"Imagine a person with depression. What views about depressed people by others does this person worry about most?"	Hand-coding of responses.
Roth, Schwarzmann and Tripodi (2024b)	Effectiveness and demand for psychotherapy	"What considerations do you have on your mind when choosing how much you would be willing to spend on 4 weeks of online therapy from BetterHelp? Please write 2-3 sentences. You may mention both downsides and benefits of buying therapy (if any were on your mind)."	Hand-coding of responses.
Saccardo and Serra-Garcia (2023)	Enable or limit their capacity to distort beliefs in moral dilemmas	"When you had to decide between learning about your commission Before or After getting information about the quality of Product B [A, if the order was flipped], how did you make this decision?"	Hand-coding of responses.
Tiezzi and Xiao (2016)	Market experiment evaluating the role of delaying externalities on support for taxes	"How did you decide to vote in favor or against the tax?" or "If your second vote was different from your first vote during the experiment, why did you change your mind?"	Incentivized (same coding with other randomly matched respondent) hand-coding by subjects.

Table A7: Overview of studies: Development economics and other areas

Paper name	Domain	Measurement	Analysis of text data
Ashwin et al. (2022)	Parent's aspirations for children in Bangladesh	Analysis of qualitative interview.	Developing supervised classification model trained with hand-coded subsample.
Ashwin et al. (2022)	Parent's aspirations for children in Bangladesh	Analysis of qualitative interview.	Using different LLMs (GPT-3.5 turbo and two Llama 2 versions) to compare to method from Ashwin et al. (2022).
Baird, McIntosh and Özler (2011)	Role of conditionality in cash transfers	Qualitative interview of random subsample.	Hand-coding of responses.
Burgstaller, Dorr and Necker (2023)	Tax credits influence on demand for legally provided services	"What reasons could there be for someone not claiming the government support?"	Hand-coding of topics which are used for keyness analysis.
Dillon et al. (2012)	Child labor	In the pilot phase, qualitative interviews with open-ended questions were conducted to solicit how respondents thought about the survey questions, why they chose the responses they did, and how they thought about concepts such as work, household production, and their primary activities.	Hand-coding of responses.
Wekhof and Houde (2023)	Investment in energy efficiency	"Describe the reasons why you decided (not) to carry out energy efficiency retrofits. Please write a short text of about 4 sentences."	Hand-coding and machine learning methods.
Jayachandran, Biradavolu and Cooper (2023)	Woman's agency	Semi-structured interview with open-ended questions.	Hand-coding of responses to calculate benchmark score.
Parker and Kozel (2007)	Poverty and vulnerability in India	'Semi-structured interview with open-ended questions.	Qualitative analysis methods are used to inform a quantitative survey.
Romero et al. (2022)	Direct vs. indirect management training	DWMS, an adaptation of the World Management Survey, was used for an interview that included 23 open-ended questions, such as "How do you keep track of what teachers are doing in the classrooms?"	Hand-coding of responses.

B API Implementation Details

To implement this workflow using the API of a language model, researchers can follow these steps. First, obtain access to the API, along with a programming interface such as Python. The coding scheme should be formatted for programmatic use, such as in JSON or YAML format. An example of a formatted coding scheme could look like the following:

- FOMO: Respondent mentions fear of missing out, feeling out of the loop, or justifies usage through others' usage.
Examples:
 - "I feel compelled to keep 'in touch' with what I perceive as being the culturally relevant 'thing' at the moment. It breeds a sense of FOMO when you don't use it."
 - "Everyone else uses it so I feel that I will be missing out if I don't."

Next, design the prompt to incorporate this coding scheme. Below is an example prompt:

You will classify text responses based on the coding scheme below. Each response can fall into multiple categories or none. Use the categories' names and definitions to make your decisions.

Coding Scheme:

- FOMO: Respondent mentions fear of missing out, feeling out of the loop, or justifies usage through others' usage.
Examples:
 - "I feel compelled to keep 'in touch' with what I perceive as being the culturally relevant 'thing' at the moment. It breeds a sense of FOMO when you don't use it."
 - "Everyone else uses it so I feel that I will be missing out if I don't."

Response to classify: "{response_text}"

This prompt can be programmatically passed to the latest available OpenAI model via the API. For instance, using Python with the GPT-4o model:

```
import openai
```

```
openai.api_key = "your-api-key"
```

```
def classify_response(response_text, coding_scheme):  
    prompt = f"""  
    You will classify text responses based on the coding scheme below.  
    Each response can fall into multiple categories or none.  
  
    {coding_scheme}
```

```

Response to classify: "{response_text}"
"""
response = openai.ChatCompletion.create(
    model="gpt-4o", # Use the latest recommended model
    messages=[{"role": "user", "content": prompt}],
    max_tokens=150,
    temperature=0
)
return response['choices'][0]['message']['content'].strip()

# Example usage
response_text = "I use this platform because all my friends are on it."
coding_scheme = """
- FOMO: Respondent mentions fear of missing out, feeling out of the loop,
  or justifies usage through others' usage.
Examples:
- "I feel compelled to keep 'in touch' with what I perceive as being
  the culturally relevant 'thing' at the moment. It breeds a sense
  of FOMO when you don't use it."
- "Everyone else uses it so I feel that I will be missing out if I don't."
"""
classification = classify_response(response_text, coding_scheme)
print(f"Classification: {classification}")

```

Finally, classification results can be stored in structured formats such as CSV files or databases for further analysis. Researchers should ensure accuracy by comparing the model's classifications against manual coding, iteratively refining the prompt and coding scheme. This iterative approach enhances the method's robustness and applicability across diverse research contexts.