Conducting Qualitative Interviews with AI

Jiugeng Sun, Jiaqing Xie

29/04/2024

ETH zürich

Overview

1. Background
2. Methodology
3. Experiment
4. Results
5. Remarks

Background

Qualitative research is a type of research that aims to gather and analyse **non-numerical** (descriptive) data in order to gain an understanding of individuals' social reality.

Qualitative vs Quantitative
Qualitative: interpretation-based, descriptive, and relating to language.
Quantitative: numbers-based, countable, or measurable.

Background

Qualitative interviews are a major source of knowledge in **social science**.

Pros of qualitative interviews:

explain the reasoning and motivations *in their own words.*
providing *uniquely rich insights* into people's decision-making processes.

image source: https://jayashreec.medium.com/interviews-qualitative-research-in-ux-design-29befec0157e

Background

The absence of qualitative interviews in economics:

limited scalability, high costs, and low generalizability from *small sample analyses* are typically associated with qualitative interviews.

Research Problem

How can we conduct *large-scale* qualitative interviews in the field of *economics*?

Qualitative Interviews with AI

Design an "AI interviewer" that helps us do qualitative interviews in a large-scale.

API integration with Open AI's GPT-4 model.

AI interviewer is a state-dependent probability distribution over the set of questions, with the *conversation history* as the primary state variable.
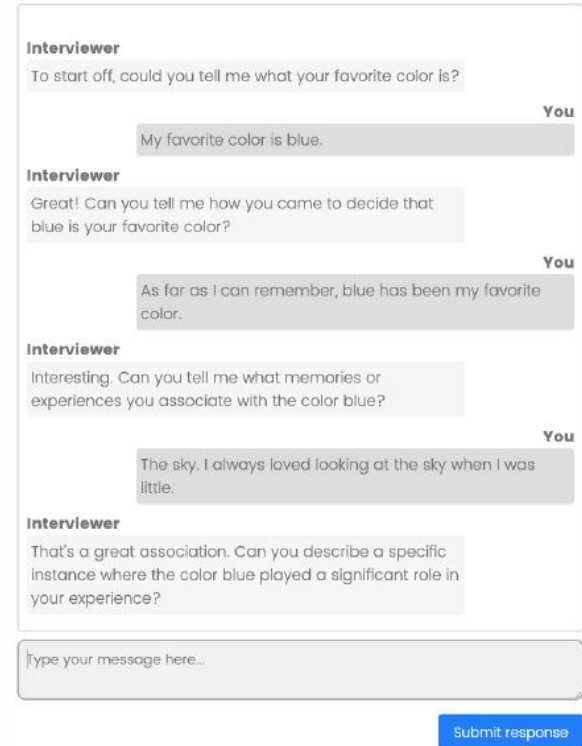

Figure 1: screenshot of the chat interface that respondents used as part of the qualitative interview with the AI.
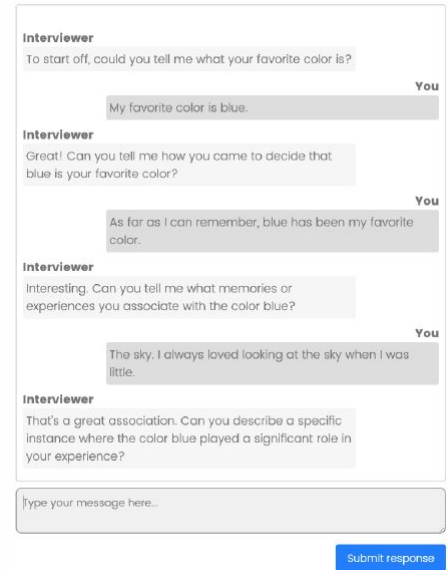
AI interviewer design choices

1. AI interviewer should adhere to methodological best practices for qualitative research.
2. AI interviewer must maintain consistent performance, both across interviewers as well as over extended conversations.
3. AI interviewer should be robust against attempts of malevolent interviewees to modify its behavior.
4. AI interviewer requires content moderation, ensuring that conversations remain within ethical boundaries.

AI interviewer design choices

FrontEnd:
- chat interface to mimic popular text messaging applications to reduce technological frictions.
- can be embedded into any survey design software that supports custom HTML and basic JavaScript functionalities such as HTTP requests.

AI interviewer design choices

BackEnd:

Figure 2: Flowchart of the AI agents responsible for the question generation process

AI interviewer design choices

BackEnd:
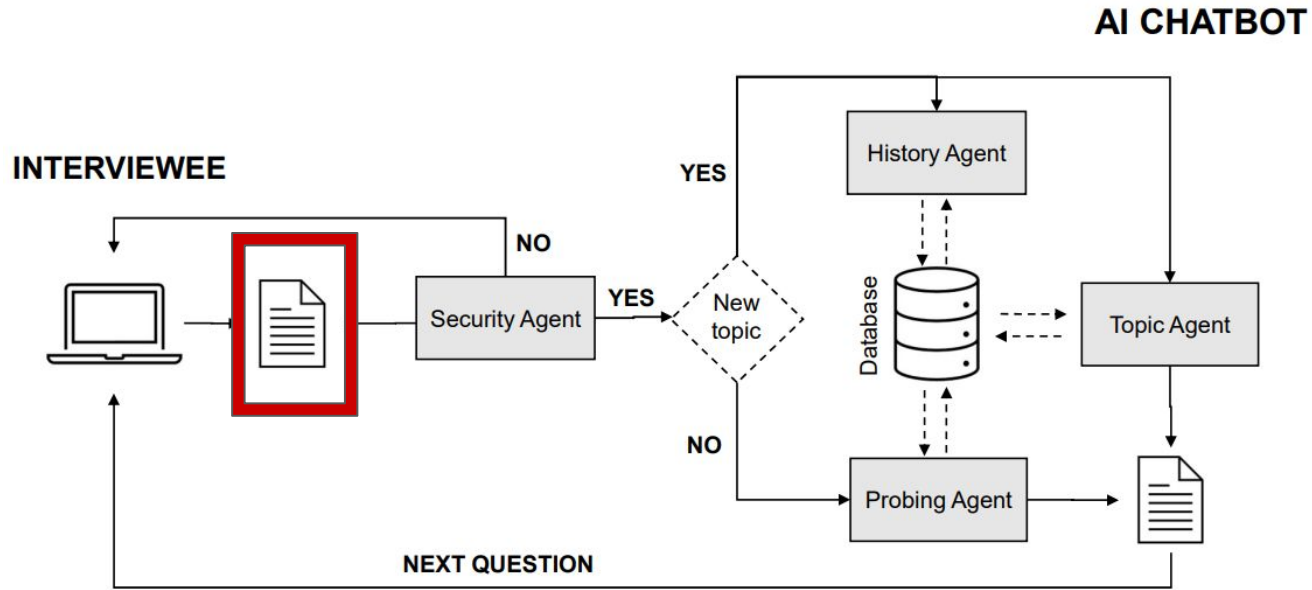
Figure 2: Flowchart of the AI agents responsible for the question generation process

Security Agent

An agent that determines whether the answer "fits into the context of an interview" by comparing it to the previous question.
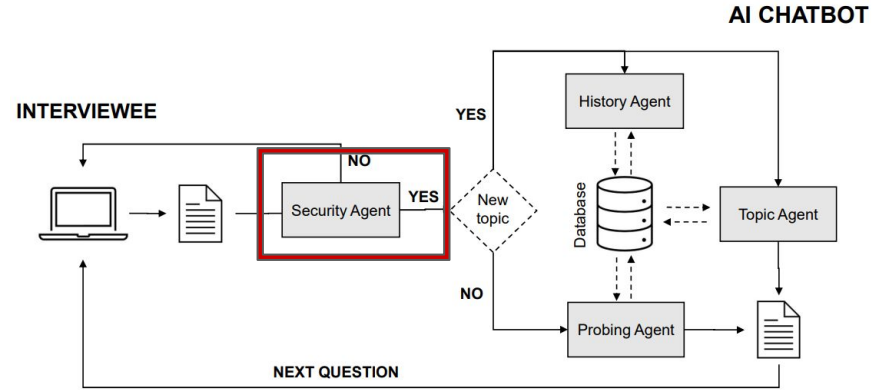
If Security Agent says NO, the interviewee receives a pre-determined message that with a gentle nudge to either rephrase the answer or decline to answer the question.
If Security Agent says YES, Chatbot proceeds to the next stage.

Second Stage

A binary decision tree on whether to continue with *additional probing questions* or to transition to a *new interview topic* from the interview plan or "topic guide".
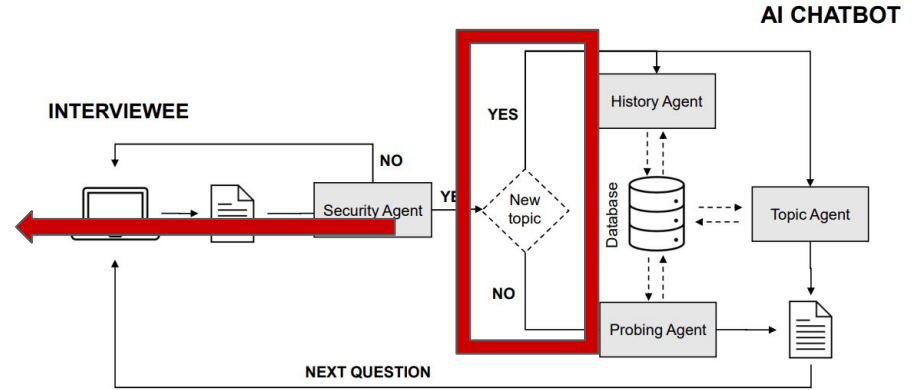
Interview Plan
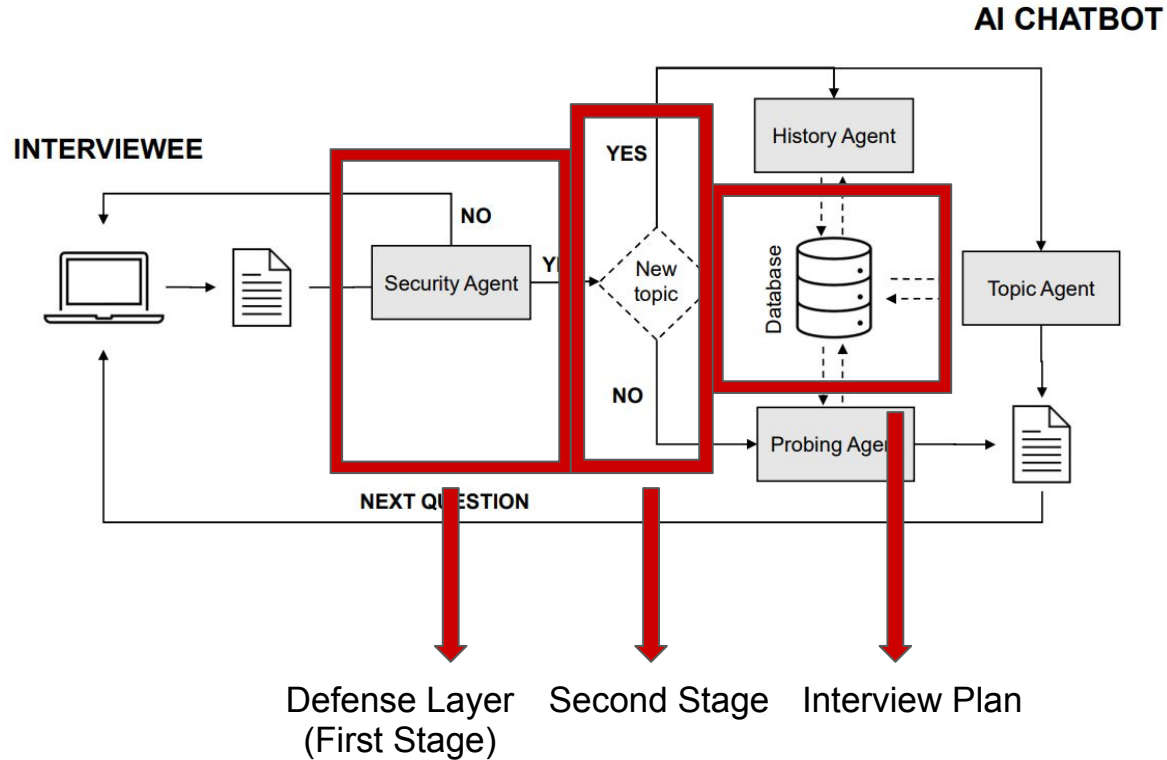
Probing Agent

Probing Agent is responsible for achieving breadth and depth of the interview.

Probing Agent receives:
1. a summary of the previous conversation history
2. the current topic of the Interview Plan
3. the conversation history within the current topic in the Interview Plan.
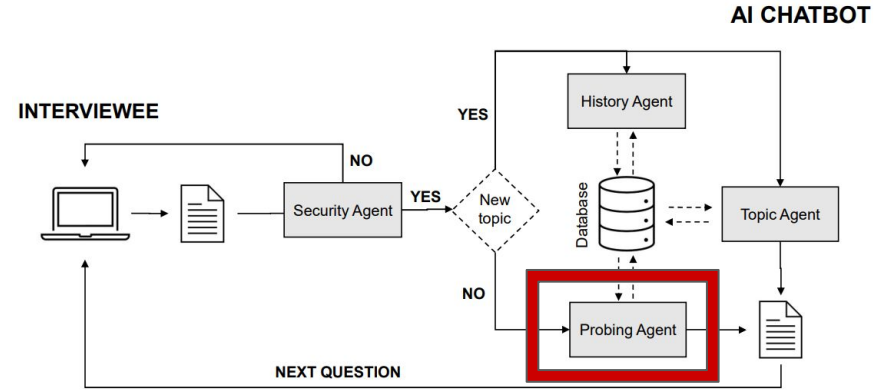
Both *general* guidelines as well as *specific* probing guidelines, and also *instructions* to clarify ambiguous answers and to pivot to new areas not covered in depth.

History Agent

History Agent is responsible for reviewing the conversation and creating an appropriate summary that can be passed on to the Probing and Topic agents.

History Agent receives:
1. the Interview Plan
2. the conversation summary from previous topics covered in the interview guide
3. the current topic of the Interview Plan
4. the conversation history that is not already covered by the previous conversation summaries
whenever the interview moves onto a new interview topic

History Agent updates the conversation summary.
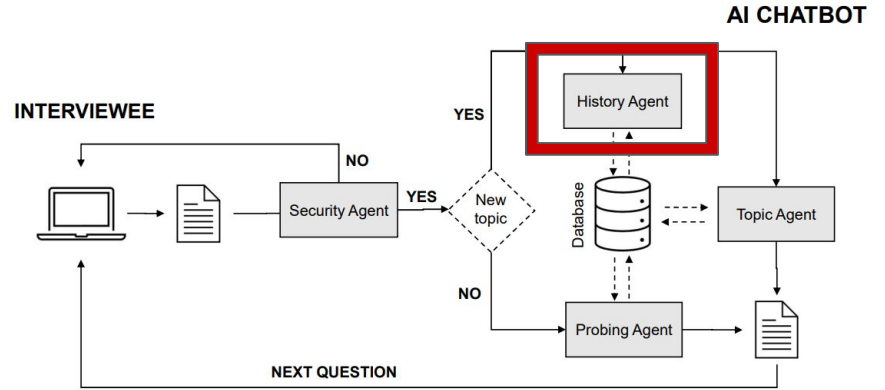
Topic Agent

Topic Agent is responsible for introducing the next topic in the Interview Plan.

Topic Agent receives:
1. the Interview Plan
2. the conversation summary from previous topics covered in the interview guide
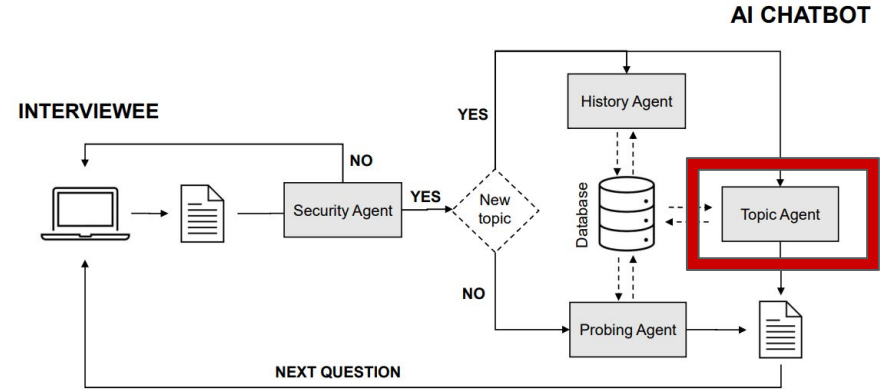
Respondents Selection

395 adult US respondents recruited from the research platform Prolific, a survey platform commonly used in economic research and associated with high data quality and attentive respondents.

Result 1: Interview experience and respondent effort

- 395 interviewees
- Interests did not decrease over time
- Number of written characters did not decrease over time

A. Interview duration
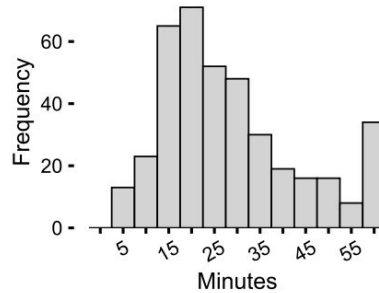
B. Avg. response times for individual questions
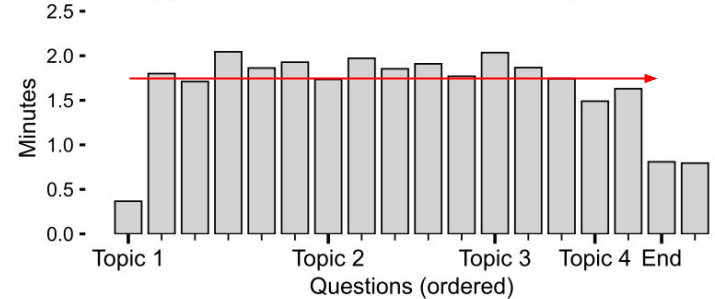
C. Total written characters
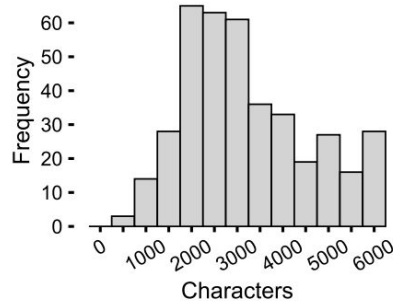
D. Avg. written characters for individual questions
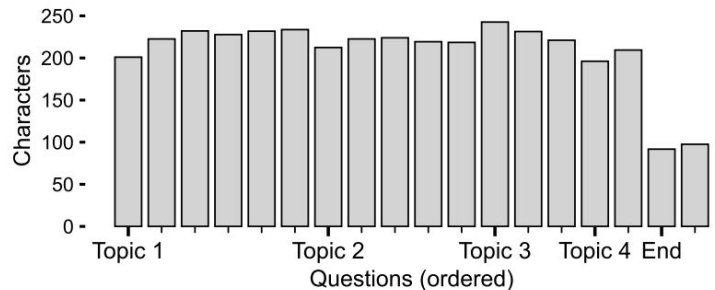
Result 1: Interview experience and respondent effort

Table 3: Analysis of response times and message length

	Message length		Response time (seconds)	
	(1)	(2)	(3)	(4)
Question number	-0.557		-0.826**	
	(0.382)		(0.398)	
Question number within topic		1.411***		0.225
		(0.487)		(0.562)
N	6,230	6,230	5,838	5,838
R^2	0.767	0.769	0.281	0.283
Dep. var. mean	221.696	221.696	109.432	109.432
Respondent fixed effect	Yes	Yes	Yes	Yes
Interview topic fixed effect		Yes		Yes

Panel Regression

Result 1: Interview experience and respondent effort

No ChatGPT evidence detected for AI interview

Figure 5: ChatGPT usage is not associated with lower effort during the interview

How often do you use ChatGPT in a typical week?
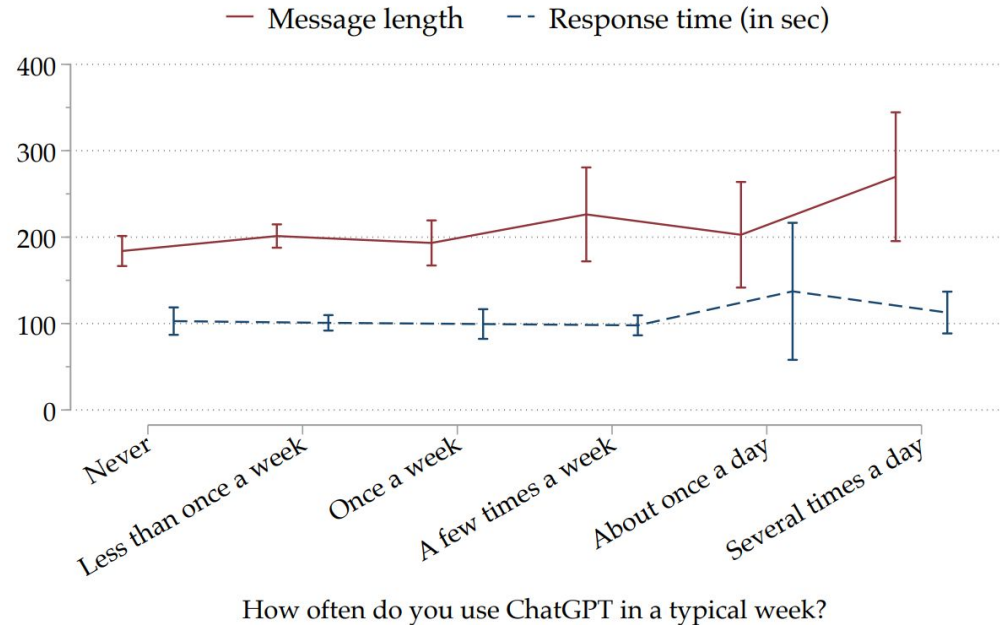
Result 2: Evaluation

Findings

- ❏ Enjoyed the interview (most)
- ❏ Natural with AI chatbot (most)
- ❏ Prefer texting with AI chatbot (half)
- ❏ Interview with AI chatbot again (most)

Figure 4: Respondents positively evaluate interviews with an AI chatbot

A. How would you rate your overall experience with the interview conducted by the AI chatbot?

Extremely negative — 6% | 9% | 23% | 36% | 23% — Extremely positive

B. How natural did the conversation with the AI chatbot feel?

Extremely unnatural — 5% | 13% | 6% | 33% | 25% | 15% — Extremely natural

C. If you were to participate in a future study involving a qualitative interview conducted through a chat interface, would you prefer texting with the same AI chatbot or an actual human interviewer?

Strongly prefer human — 7% | 7% | 7% | 26% | 15% | 19% | 19% — Strongly prefer AI

D. Would you be interested in participating in an interview with an AI chatbot again?

Yes — 95%
No — 5%

Result 3: Qualitative analysis

➢ **Surface and Depth**
  ○ Example: Stock Investing Non-participation
  ○ "Surface" explanation: low income (funds)
  ○ In-"Depth" explanation: fear making loss
  ○ Probe on some questions to get more
  ○ precise contextual results

Money available actually (most)

Table 1: Summary statistics

	Min	Mean	Median	Max	N
A. Demographics					
Age	19.00	39.32	36.00	78.00	395
Female	0.00	0.62	1.00	1.00	395
College education	0.00	0.54	1.00	1.00	395
Full-time employment	0.00	0.47	0.00	1.00	395
White	0.00	0.76	1.00	1.00	395
African American/Black	0.00	0.12	0.00	1.00	395
Hispanic	0.00	0.11	0.00	1.00	395
Region					
Northeast	0.00	0.18	0.00	1.00	395
Midwest	0.00	0.24	0.00	1.00	395
West	0.00	0.16	0.00	1.00	395
South	0.00	0.41	0.00	1.00	395
Household size	1.00	2.98	3.00	10.00	395
Number of children	.	.	.	.	0
B. Finances					
Household income ($)	35,000.00	73,037.97	65,000.00	212,500.00	395
Total financial assets ($)	0.00	82,857.47	17,500.00	550,000.00	395
Non-mortgage debt ($)	0.00	27,534.56	7,500.00	400,000.00	395
Housing					
Homeowner	0.00	0.50	1.00	1.00	395
Home value ($)	12,500.00	239,384.42	225,000.00	525,000.00	199
Any mortgage debt	0.00	0.27	0.00	1.00	395
Total mortgage debt ($)	12,500.00	123,247.66	87,500.00	475,000.00	107
Two months liquid savings	0.00	0.61	1.00	1.00	395

Result 3: Qualitative analysis

➢ Mental model of stock market risks
➢ In other words, one's stereotypes about market risks
➢ Example responses:
  ○ the market can "make or break"
  ○ devastating losses
  ○ gambling money / lottery ticket

Lack of understanding and knowledge prevents the participation in stock market

Result 3: Qualitative analysis

- ➤ Misconceptions about investing
- ➤ Example
  - ○ Monitor stock price regularly
  - ○ Conduct thorough research into companies to make informed trading decisions
  - ○ Predict which stocks will increase or decrease in value ahead of time

Result 4: Quantitative analysis

➢ Coding Assignment (by GPT-4)
➢ Coding Frequencies
  ○ Average 5.9 codes per respondent
  ○ Heterogeneous and multidimensional
  ○ Informational barriers dominated

Figure 6: Reasons for stock market non-participation

Category	Percentage
Financial Reasons	74%
Financial Constraints	72%
Financial Goals	18%
Risk Concerns	69%
Pessimistic Beliefs About Returns	20%
Low Risk Tolerance	40%
Rare Disaster Risk	16%
Perception of Market as Gambling	29%
Generic Risk Concerns	8%
Informational barriers	84%
Perceived Lack of Knowledge	83%
Complexity and Confusion	29%
Perceived Costs	29%
Learning and Initiation Costs	28%
Portfolio Management Costs	5%
Other	56%

Result 4: Quantitative analysis

➤ Co-occurrence of reasons for non-participation

➤ Most common co-occurrence is financial constraints and a perceived lack of knowledge

➤ The strong co-occurrence of many codes highlights the need for a nuanced analysis to understand the real barriers behind stock market non-participation

Figure 7: Co-occurrence of reasons for stock market non-participation

Result 4: Quantitative analysis

➢ Can small samples uncover the same patterns?
➢ Yes.
➢ Small sample variability in co-occurrences of codes across interviews

Figure 8: Co-occurrence of reasons for stock market non-participation: Small sample variability in random interview subsets of size 20

Result 4: Quantitative analysis

➢ A high degree of instability of the co-occurrence matrix in small samples

➢ Conducting qualitative interviews at scale has the opportunity to recover many insights that might be lost when conducting qualitative interviews with typical sample sizes of around 12 to 20 respondents

Figure 9: Co-occurrence of reasons for stock market non-participation: Instability of the co-occurrence matrix in small samples – Bootstrap exercise

Result 4: Quantitative analysis

➢ Can simple open-ended questions uncover the same patterns?
➢ Compare the first answer results with full interview
➢ Big difference: Perceived Lack of Knowledge, Lack of Trust in Market and Institutions
➢ Small difference: Generic Risk Concerns

Figure 10: Reasons for stock market non-participation: Full interview vs first open-ended question
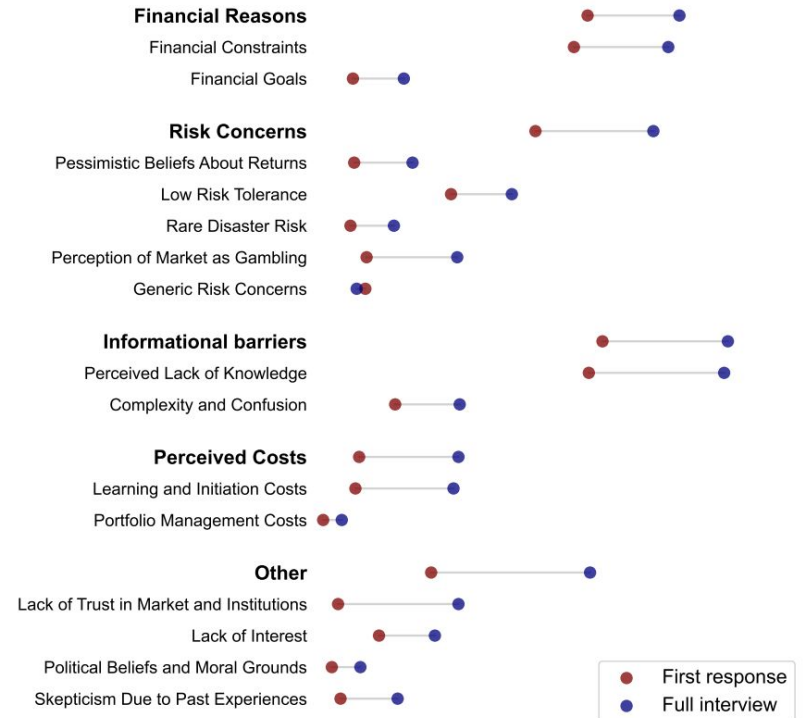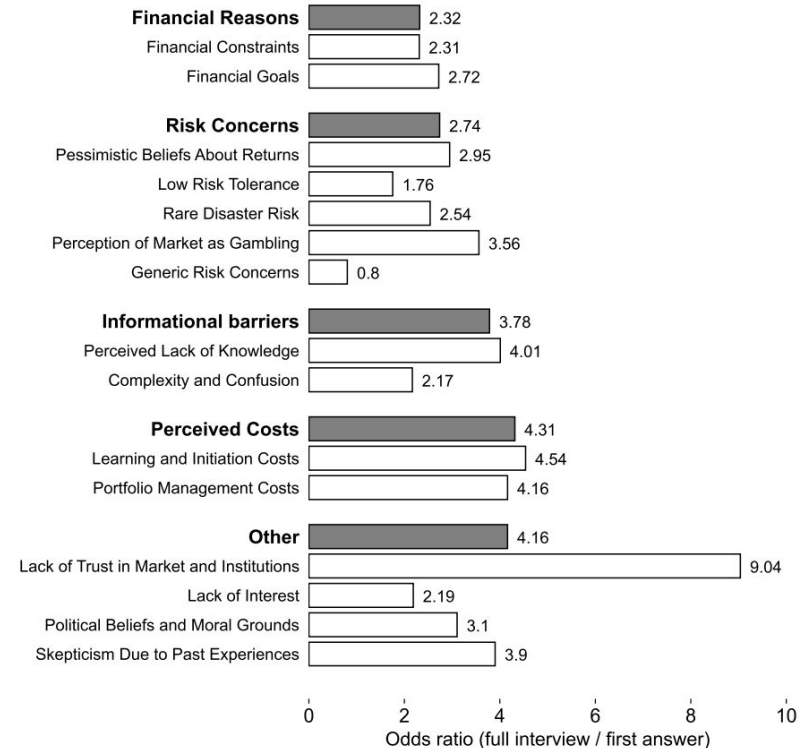
Result 4: Quantitative analysis

➢ Odd ratio
➢ Big ratio: Lack of Trust in Market and Institutions
➢ Small ratio: Generic Risk Concerns

Figure 11: Reasons for stock market non-participation: Odds ratio for a code appearing in the full interview vs the first response

|	Odds ratio
Financial Reasons	2.32
Financial Constraints	2.31
Financial Goals	2.72
Risk Concerns	2.74
Pessimistic Beliefs About Returns	2.95
Low Risk Tolerance	1.76
Rare Disaster Risk	2.54
Perception of Market as Gambling	3.56
Generic Risk Concerns	0.8
Informational barriers	3.78
Perceived Lack of Knowledge	4.01
Complexity and Confusion	2.17
Perceived Costs	4.31
Learning and Initiation Costs	4.54
Portfolio Management Costs	4.16
Other	4.16
Lack of Trust in Market and Institutions	9.04
Lack of Interest	2.19
Political Beliefs and Moral Grounds	3.1
Skepticism Due to Past Experiences	3.9

Odds ratio (full interview / first answer)

Conclusion: Single open-ended responses are unable to replicate the richness of full interviews

Figure 7: Co-occurrence of reasons for stock market non-participation

Figure 12: Co-occurrence of reasons for stock market non-participation: First answer only
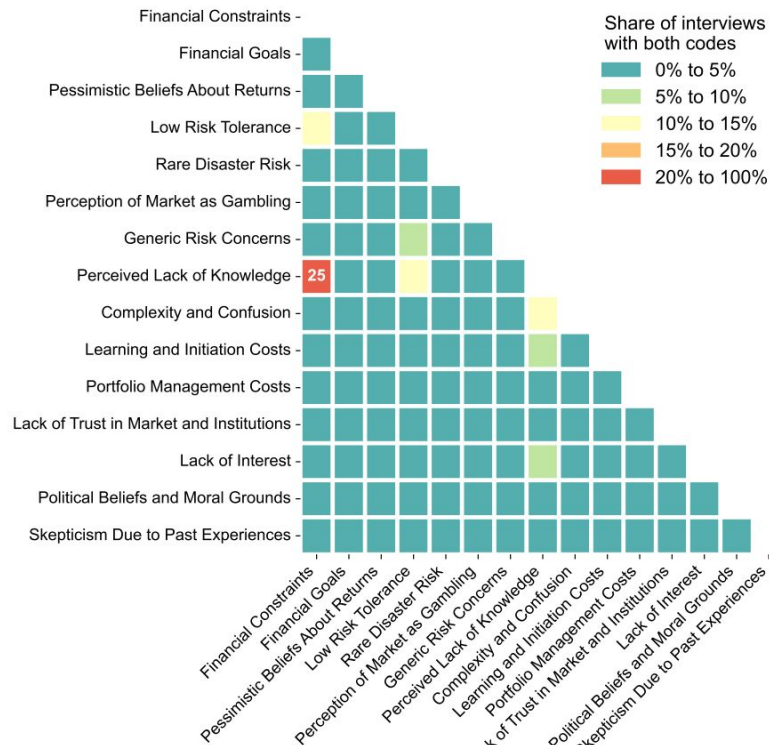
Remark 1

Strong domain bias against the sample.

Assumption irrelevant to the interview topic on the interview participants:
Able or willing to use mobile devices like computers/cell phones.
Feel comfortable or frequently using chat apps.

Advice:
Set up a comparative study on conducting the same interviews through traditional methods and analyse the results between two methods (with AI and without AI).

Remark 2

Agents generated from the prompting LLMs.

Probing agents lack domain knowledge potentially.

Advice:
Fine-tuning current probing agents with human interview transcript.