Conversations at Scale

Robust Al-led Interviews with a Simple Open-Source Platform

Friedrich Geiecke (LSE) Xavier Jaravel (LSE)

July 16, 2025

Motivation

► Canonical approach to link theory to data in economics: revealed preference

- Complementary approach: eliciting expectations, knowledge, beliefs, attitudes, etc. through surveys (e.g., Stantcheva 2023)
 - Two types of surveys: close-ended (more specific, easier to analyze) or open-ended (less priming)

Motivation

- Alternative to close-ended/open-ended surveys: qualitative interviews
 - Sometimes used in economics to generate hypotheses (e.g., Bewley (1999)) or better understand mechanisms (Bergman et al. (2024), Duraj et al. (2024))
 - ► Large language models (**LLM**) provide an opportunity to conduct qualitative interviews at a large scale, with thousands of respondents at low cost

Motivation

- Alternative to close-ended/open-ended surveys: qualitative interviews
 - Sometimes used in economics to generate hypotheses (e.g., Bewley (1999)) or better understand mechanisms (Bergman et al. (2024), Duraj et al. (2024))
 - ► Large language models (**LLM**) provide an opportunity to conduct qualitative interviews at a large scale, with thousands of respondents at low cost
- By asking non-leading questions, Al-led interviews retain the advantage of open-ended surveys (no "priming")
- Could offer three potential additional benefits through follow-up questions:
 - 1. Elicit more precise/complete views
 - 2. Helps the respondent refine and clarify their own thoughts (lower cognitive load)
 - 3. Respondents may enjoy the conversational style and be more engaged



Preview of Open-Source Platform

- ► Simple architecture
 - ▶ Both text and audio interviews, with a range of LLMs, including open-weight models
 - ▶ Incorporate best practices from sociology following Small and Calarco (2022)
 - Can be adapted to new interview topics almost instantaneously (single agent)
 - Open-source code on GitHub
- ► Full pipeline for interviews and textual analysis of transcripts

Key challenge: no ground truth

Key challenge: no ground truth

To assess the quality of interviews, combine three simple approaches:

Key challenge: no ground truth

To assess the quality of interviews, combine three simple approaches:

1. Do experts like Al-led interviews? How do they compare to human-led interviews?

Key challenge: no ground truth

To assess the quality of interviews, combine three simple approaches:

- 1. Do experts like Al-led interviews? How do they compare to human-led interviews?
- 2. Do respondents like the interview? Do they share more than with open text fields?

Key challenge: no ground truth

To assess the quality of interviews, combine three simple approaches:

- 1. Do experts like Al-led interviews? How do they compare to human-led interviews?
- 2. Do respondents like the interview? Do they share more than with open text fields?
- 3. Can we use the transcripts to generate hypotheses that are relevant out of sample?

Four Applications

Investigate quality of Al-led interviews in four complementary applications:

- 1. Eliciting **political preferences**: France's snap legislative elections
- 2. Eliciting deeply personal views: meaning in life
- Eliciting key factors in economic decision-making: educational and occupational choices
- 4. Eliciting **mental models**: views of the potential impacts of recent policies of the Trump administration

Literature

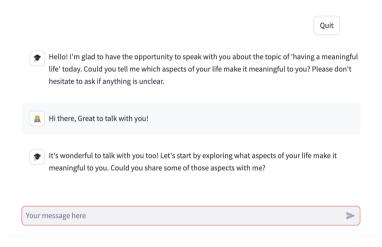
- Qualitative interviews in economics: Bewley (1999), Bergman et al. (2024), Duraj et al. (2024)
- Interviews with LLMs: Chopra and Haaland (2024), Cuevas et al. (2025), Wuttke et al. (2025)
- ► Surveys: Stantcheva (2023), Haaland et al. (2024)
- Voting: Gerber et al. (2008), Buttice and Stone (2012), Allcott and Gentzkow (2017), Hansen and Treul (2021), Enke (2020)
- Mental models: Shiller (2017), Eliaz and Spiegler (2020), Andre et al. (2022), Andre et al. (2023)
- Meaning in life: Crumbaugh and Maholick (1964), Antonovsky (1993), Steger et al. (2006), Oishi and Diener (2014), King and Hicks (2021)
- Educational and occupational choice: Roy (1951), Goldin (2006) Rothstein and Rouse (2011), Hoxby and Turner (2015) Bursztyn et al. (2017), Bell et al. (2019)



Outline

- 1. Methodology
 - ► Al-led interviews
 - Analysis of transcripts
- 2. Empirical applications
 - ► French snap legislative elections
 - Perceptions of meaning in life
 - Educational and occupation choice
 - Perceptions of policies of the Trump administration

Interface



Prompt

- ► Instructions are passed to the LLM (GPT-4o) via a "system prompt" in two main parts:
 - ► Interview Outline
 - General Instructions

Prompt: Interview Outline

Ask up to 30 questions to find out the underlying factors that contribute to the respondent's voting decision. If the respondent does not intend to vote, understand why. If the respondent does intend to vote, understand who their favorite candidate is and why they want to vote for that candidate. Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about the topic of the upcoming U.S. presidential election. Do you intend to vote in this election? Please don't hesitate to ask if anything is unclear.'.

Prompt: General Instructions (1/3), based on Small and Calarco (2022)

Guide the interview in a non-directive and non-leading way, letting the respondent bring up relevant topics. Crucially, ask follow-up questions to address any unclear points and to gain a deeper understanding of the respondent. Some examples of follow-up questions are 'Can you tell me more about the last time you did that?'. 'What has that been like for you?'. 'Why is this important to you?', or 'Can you offer an example?', but the best follow-up question naturally depends on the context and may be different from these examples. Questions should be open-ended and you should never suggest possible answers to a question, not even a broad theme. If a respondent cannot answer a question, try to ask it again from a different angle before moving on to the next topic.

Prompt: General Instructions (2/3), based on Small and Calarco (2022)

Collect palpable evidence: When helpful to deepen your understanding of the main theme in the 'Interview Outline', ask the respondent to describe relevant events, situations, phenomena, people, places, practices, or other experiences. Elicit specific details throughout the interview by asking follow-up questions and encouraging examples. Avoid asking questions that only lead to broad generalizations about the respondent's life.

Prompt: General Instructions (3/3), based on Small and Calarco (2022)

Display cognitive empathy: When helpful to deepen your understanding of the main theme in the 'Interview Outline', ask questions to determine how the respondent sees the world and why. Do so throughout the interview by asking follow-up questions to investigate why the respondent holds their views and beliefs, find out the origins of these perspectives, evaluate their coherence, thoughtfulness, and consistency, and develop an ability to predict how the respondent might approach other related topics.

- Run interviews with US participants on a range of topics, swapping only the outline in the prompt
 - ► Topics: voting, meaning in life, occupational choice, housing, trust in institutions, cost of living, view on climate change
- Each transcript is analyzed by sociology PhD students (Harvard/LSE)
 - ► Comparison to **hypothetical human expert** using an online text chat interface

How good was the AI Interviewer compared to what a human expert (academic working with qualitative interviews) could have achieved with the same respondent using an online text chat interface? [1 = worst human expert, 3=average human expert, 5=best human expert]

Average grade:	2.95 (0.138)	
Grade distribution:		
1	1 (2.50%)	
2	13 (32.50%)	
3	13 (32.50%)	
4	13 (32.50%)	
5	` 0	
N	40	

- ▶ Run face-to-face and online interviews with PhD-level sociologists at the LSE Behavioral Lab, as well as Al-led interviews
- Ask for grades from a separate team of sociologists (Cambridge/Harvard/Johns Hopkins/LSE/Oxford)
 - Comparison to hypothetical human expert, assigning two grades
 - a grade relative to what an expert in the field could have achieved in a 30-minute online text chat interview;
 - a grade relative to what an expert could have achieved in a 30-minute face-to-face interview

How good was the Interviewer compared to what a human expert could have achieved with the same respondent in a thirty-minute interview using an online text chat interface? 1 to 5 [1 = worst human expert, 3=average human expert, 5=best human expert]

	Face-to-face interview,	Online interview,	Online interview,	Online interview,
	human interviewer	human interviewer, text	Al interviewer, text	AI interviewer, voice
	(1)	(2)	(3)	(4)
Average grade:	3.6	2.8	3.2	3.9
	(0.32)	(0.31)	(0.33)	(0.26)
N	17	16	16	17

How good was the Interviewer compared to what a human expert could have achieved with the same respondent in a thirty-minute in-person interview?

1 to 5 [1 = worst human expert, 3=average human expert, 5=best human expert]

	Face-to-face interview,	Online interview,	Online interview,	Online interview,
	human interviewer	human interviewer, text	Al interviewer, text	AI interviewer, voice
	(1)	(2)	(3)	(4)
Average grade:	3.6	2.3	2.8	3.4
	(0.27)	(0.26)	(0.30)	(0.31)
Ν	17	16	16	17

Analysis of transcripts

▶ Pipeline for automated analysis of transcripts with LLMs in two steps

Analysis of transcripts

- Pipeline for automated analysis of transcripts with LLMs in two steps
- ► Step 1: Overview and hypothesis generation
 - ▶ Paste transcript summaries into the context of an LLM, ask questions to inquire about broad themes, surprising findings, etc.
 - Validate accuracy with sociology PhD students

Analysis of transcripts

- ▶ Pipeline for automated analysis of transcripts with LLMs in two steps
- Step 1: Overview and hypothesis generation
 - Paste transcript summaries into the context of an LLM, ask questions to inquire about broad themes, surprising findings, etc.
 - Validate accuracy with sociology PhD students
- ► <u>Step 2</u>: Hypothesis testing, transcript by transcript
 - ▶ Iterate over each transcript with an LLM prompted to detect if a certain concept is contained in the particular transcript
 - Validate accuracy with research assistants

Outline

- 1. Methodology
 - Interviews
 - Analysis of transcripts
- 2. Empirical applications
 - ► French snap legislative election
 - Perceptions of meaning
 - Educational and occupation choice
 - Perceptions of policies of the Trump administration

Voting decisions

- Which factors drive voting decisions across the political spectrum?
- ▶ Sample of 442 respondents before France's snap legislative elections in June 2024

Key part of the interview outline (translated to English)

With around 20 questions, please explore the motivations behind the choice of the party to vote for; in particular, assess the importance of the new public policies proposed by the party (both their general philosophy and specific measures) or other factors (e.g., trust in the party's leaders). Evaluate whether the participant's main motivation is adherence to the ideas of the party they decide to vote for, or rather the rejection of the ideas of other parties.

Quality metrics

	Fraction of Respondents
In the future, would you rather take the interview with	
An Al	49%
A human	15%
I do not mind	36%
How well does the interview capture your views? 1 ("poorly") to 4 ("very well")	3.31

Top 5 reasons for electoral choice: Left

Left	
Trait	% of transcripts
Block other parties	73
Reduce inequalities	49
Increase the minimum wage	36
Green transition	32
Tax the rich	28

Top 5 reasons for electoral choice: Far Right

Far Right	
Trait	% of transcripts
Reduce immigration	77
Reduce crime	47
Block other parties	38
Policies favoring the French over foreigners	30
Increase purchasing power	19

Top 5 reasons for electoral choice: Center

Center		
Trait	% of transcripts	
Block other parties	71	
Continuity of ongoing policies	24	
Economic stability	20	
Pro-European policies	17	
Pension reform	14	

Voting Decisions: Takeaways

- ► Al-led interviews on electoral choice during France's snap legislative elections highlight that:
 - 1. Participants are very comfortable sharing their views on sensitive political questions with an Al
 - 2. The AI performs well in a language other than English

Outline

- 1. Methodology
 - Interviews
 - Analysis of transcripts
- 2. Empirical applications
 - ► French snap legislative election
 - ► Perceptions of meaning
 - ► Educational and occupation choice
 - Perceptions of policies of the Trump administration

Perceptions of meaning

- ▶ Which factors give people a sense of "having a meaningful life"?
- ► Sample of 466 US respondents
- ▶ Between-subjects design: randomize participants into Al-led interview vs. open-ended questions

Key part of the interview outline

Ask up to around 30 questions to explore different dimensions of life and find out the underlying factors that contribute to the respondent's sense of meaning in life. Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about the topic of 'having a meaningful life' today. Could you tell me which aspects of your life make it meaningful to you? Please don't hesitate to ask if anything is unclear.'.

Al vs human interviewer vs open-ended questions

	Fraction of Respondents
In the future, would you rather take the interview with	
An Al	43%
A human	19%
I do not mind	38%
Would you have preferred to answer open-ended questions instead?	
Yes	12%
No	76%
I do not mind	12%

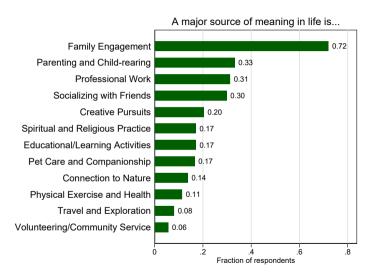
Respondent evaluations

	Al-Led Interview (1)	Open Text Fields (2)
How well does it summarize		
what gives you a sense of meaning?	3.58	3.45
1 ("poorly") to 4 ("very well")	(s.e. 0.045)	(s.e. 0.039)
Are you able to clearly identify		
sources of meaning in your life?		
My thoughts are still evolving	34%	42%
I can clearly pinpoint sources of meaning in my life	52%	41%
I am somewhere in between	14%	17%
Number of words	460 (+142%)	190

Expert Assessment: Interviews vs. Open-Text Fields

Which response provides a deeper understanding of the aspects of life the respondent finds meaningful?	Fraction
Al-led interview transcript Open-text field response Indifferent	75% 2.5% 22.5%

Activities associated with meaning in life



Meaning in Life: Takeaways

- ► Al-led interviews on meaning in life highlight that:
 - 1. Al-led interviews can perform very well for highly complex topic
 - 2. Respondents shared richer information than with standard open text fields

Outline

- 1. Methodology
 - Interviews
 - Analysis of transcripts
- 2. Empirical applications
 - ► French snap legislative elections
 - Perceptions of meaning in life
 - ► Educational and occupation choice
 - Perceptions of policies of the Trump administration

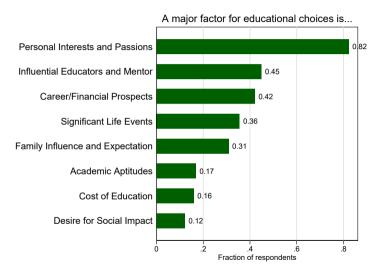
Educational and occupational choices

- How do people decide which degrees to study and which jobs to work in?
- ► Sample of 107 US respondents
- ▶ At the end of the interview, the LLM provides a summary to the respondent

Quality metrics

	Fraction of Respondents
In the future, would you rather take the interview with	
An Al	35.90%
A human	15.38%
I do not mind	48.72%
How well does the summary capture your reasons? 1 ("poorly") to 4 ("very well")	3.78

Educational choice



Educational and occupational choices: Takeaways

- ► Al-led interviews on educational and occupational choices highlight that:
 - Compared to political/personal topics, participants are less likely to strictly prefer Al over human interviewer
 - 2. Respondents give excellent grades to the interview summary

Outline

- 1. Methodology
 - Interviews
 - Analysis of transcripts
- 2. Empirical applications
 - ► French snap legislative elections
 - Perceptions of meaning
 - ► Educational and occupation choice
 - ▶ Perceptions of policies of the Trump administration

Narratives about Policies

- ▶ How are new policies perceived and understood across the political spectrum?
- ► Sample of 800 U.S. respondents in April 2025
- ► Elicit people's views (positive/negative/neutral) and mental models (policy ⇒ likely impact) about recent decisions of the Trump administration

Narratives about Policies

- How are new policies perceived and understood across the political spectrum?
- ► Sample of 800 U.S. respondents in April 2025
- ► Elicit people's views (positive/negative/neutral) and mental models (policy ⇒ likely impact) about recent decisions of the Trump administration
- ▶ Use a follow-up close-ended survey to assess the relevance and "completeness" of top 35 mental models elicited in qualitative interviews
 - ► Today, report results of the follow-up close-ended survey (300 respondents)

Negative Narrative about Tariffs

Do you agree or disagree that this chain of thought is a major reason to believe the Trump administration's decisions will lead to negative outcomes?

Tariffs on Imported Goods \Rightarrow Increased Prices for Essentials \Rightarrow Financial Strain and Reduced Quality of Life

Tariff drive up the cost of goods like groceries, electronics, and clothing, leading to financial hardship and lifestyle adjustments.

Negative Narrative about Tariffs

Do you agree or disagree that this chain of thought is a major reason to believe the Trump administration's decisions will lead to negative outcomes?

Tariffs on Imported Goods \Rightarrow Increased Prices for Essentials \Rightarrow Financial Strain and Reduced Quality of Life

Tariff drive up the cost of goods like groceries, electronics, and clothing, leading to financial hardship and lifestyle adjustments.

Fraction who agree/disagree, %		
Fully agree	87.9	
Partially agree	8.7	
Partially disagree	3	
Fully disagree	4	

Sample restricted to respondents who believe the decisions of the Trump administration were mostly negative



Positive Narrative about Tariffs

Do you agree or disagree that this chain of thought is a major reason to believe the Trump administration's decisions will lead to positive outcomes?

Tariffs on Foreign Goods \Rightarrow Boosted Domestic Manufacturing \Rightarrow Job Creation and Economic Growth

Tariffs can encourage domestic production, creating jobs and fostering economic growth, which can benefit individuals and the national economy.

Positive Narrative about Tariffs

Do you agree or disagree that this chain of thought is a major reason to believe the Trump administration's decisions will lead to positive outcomes?

Tariffs on Foreign Goods \Rightarrow Boosted Domestic Manufacturing \Rightarrow Job Creation and Economic Growth

Tariffs can encourage domestic production, creating jobs and fostering economic growth, which can benefit individuals and the national economy.

Fraction who agree/disagree, %		
Fully agree	47	
Partially agree	36.5	
Partially disagree	11.8	
Fully disagree	2.3	
Unsure	2.3	

Sample restricted to respondents who believe the decisions of the Trump administration were mostly positive



Completeness of Narratives?

How well did the survey cover your reasons to think the decisions of the Trump administration so far have been mostly positive/negative?

Completeness of Narratives?

How well did the survey cover your reasons to think the decisions of the Trump administration so far have been mostly positive/negative?

	Mostly Negative	Mostly Positive
The survey covers all major reasons.	82	76.5
The survey partially covers the major reasons.	18	23.5
The survey does not cover the major reasons.	0	0

Takeaways

- ► Al-led interviews on perceptions of the decisions of the Trump administration highlight that:
 - 1. Al-led interviews are a useful tool to map people's mental models / narratives about policies and their impacts
 - 2. Mental models about the same policies are very heterogeneous across individuals

Conclusion

Conclusion

- ▶ Flexible open-source platform to conduct qualitative interviews with Al at scale
 - incorporates best practices from sociology
 - adaptable to new topics within minutes
- Evaluations show this approach to interviews is reliable and effective

Try It Out!

- ► Full code available on Github, in Python
 - Can set up the full platform in about an hour

- ► Testing platform
 - lacktriangle Test your own interview topics online in < 10 minutes, without writing a line of code
 - Google Colab Notebook link: https://colab.research.google.com/drive/1sYl2BMiZACrOMlyASuT-bghCwS5FxHSZ

Thank you!

References I

- **Allcott, Hunt and Matthew Gentzkow**, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 2017, 31 (2), 211–236.
- Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective models of the macroeconomy: Evidence from experts and representative samples," *The Review of Economic Studies*, 2022, 89 (6), 2958–2991.
- __ , Ingar Haaland, Christopher Roth, and Johannes Wohlfart, "Narratives about the Macroeconomy," 2023.
- **Antonovsky, Aaron**, "The structure and properties of the Sense of Coherence scale," *Social Science & Medicine*, 1993, *36* (6), 725–733.
- **Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen**, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *The Quarterly Journal of Economics*, 2019, 134 (2), 647–713.
- Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer, "Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice," *American Economic Review*, 2024, 114 (5), 1281–1337.



References II

- Bewley, Truman, "Why Wages Don't Fall during a Recession," Harvard University Press, 1999.
- Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, 107 (11), 3288–3319.
- Buttice, Matthew K. and Walter J. Stone, "Candidates Matter: Policy and Quality Differences in Congressional Elections," *The Journal of Politics*, 2012, 74 (3), 870–887.
- Chopra, Felix and Ingar Haaland, "Conducting qualitative interviews with AI," Working Paper, 2024.
- **Crumbaugh, James C. and Leonard T. Maholick**, "An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis," *Journal of Clinical Psychology*, 1964, 20 (2), 200–207.
- Cuevas, Alejandro, Jennifer V Scurrell, Eva M Brown, Jason Entenmann, and Madeleine IG Daepp, "Collecting Qualitative Data at Scale with Large Language Models: A Case Study," Proceedings of the ACM on Human-Computer Interaction, 2025, 9 (2), 1–27.

References III

- Duraj, Kamila, Daniela Grunow, Michael Chaliasos, Christine Laudenbach, and Stephan Siegel, "Rethinking the stock market participation puzzle: A qualitative approach," *IMFS Working Paper Series*, 2024.
- **Eliaz, Kfir and Ran Spiegler**, "A model of competing narratives," *American Economic Review*, 2020, 110 (12), 3786–3816.
- Enke, Benjamin, "Moral Values and Voting," *Journal of Political Economy*, 2020, 128 (10), 3679–3729.
- **Gerber, Alan S., Donald P. Green, and Christopher W. Larimer**, "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment," *The American Political Science Review*, 2008, *102* (1), 33–48.
- **Goldin, Claudia**, "The Quiet Revolution That Transformed Women's Employment, Education, and Family," *American Economic Review*, 2006, *96* (2), 1–21.
- Haaland, Ingar K, Christopher Roth, Stefanie Stantcheva, and Johannes Wohlfart, "Measuring what is top of mind," *National Bureau of Economic Research*, 2024.



References IV

- Hansen, Eric R. and Sarah A. Treul, "Inexperienced or anti-establishment? Voter preferences for outsider congressional candidates," *Research & Politics*, 2021, 8 (3), 1–7.
- **Hoxby, Caroline M. and Sarah Turner**, "What High-Achieving Low-Income Students Know about College," *American Economic Review*, 2015, *105* (5), 514–517.
- King, Laura A. and Joshua A. Hicks, "The Science of Meaning in Life," *Annual Review of Psychology*, 2021, 72 (Volume 72, 2021), 561–584.
- Oishi, Shigehiro and Ed Diener, "Residents of poor nations have a greater sense of meaning in life than residents of wealthy nations," *Psychological Science*, 2014, 25 (2), 422–430.
- **Rothstein, Jesse and Cecilia Elena Rouse**, "Constrained after college: Student loans and early-career occupational choices," *Journal of Public Economics*, 2011, *95* (1), 149–163.
- **Roy, A. D.**, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, 3 (2), 135–146.
- Shiller, Robert J., "Narrative Economics," American Economic Review, April 2017, 107 (4), 967â1004.



References V

- Small, Mario Luis and Jessica McCrory Calarco, Qualitative literacy: A guide to evaluating ethnographic and interview research, Univ of California Press, 2022.
- **Stantcheva, Stefanie**, "How to run surveys: A guide to creating your own identifying variation and revealing the invisible," *Annual Review of Economics*, 2023, *15* (1), 205–234.
- Steger, Michael F., Patricia Frazier, Shigehiro Oishi, and Matthew Kaler, "The meaning in life questionnaire: Assessing the presence of and search for meaning in life," *Journal of Counseling Psychology*, 2006, 53 (1), 80–93.
- Wuttke, Alexander, Matthias Aßenmacher, Christopher Klamm, Max M Lang, Quirin Würschinger, and Frauke Kreuter, "Al conversational interviewing: Transforming surveys with LLMs as adaptive interviewers," arXiv preprint arXiv:2410.01824, 2025.