Conversations at Scale: Robust AI-led Interviews with a Simple Open-Source Platform*

Friedrich Geiecke
LSE
f.c.geiecke@lse.ac.uk

Xavier Jaravel
LSE & CEPR
x.jaravel@lse.ac.uk

October 2, 2024

Abstract

The advent of large language models (LLMs) provides an opportunity to conduct qualitative interviews at a large scale, with thousands of respondents, creating a bridge between qualitative and quantitative methods. In this paper, we develop a simple, versatile open-source platform for researchers to run AI-led qualitative interviews. Our approach incorporates established best practices from the sociology literature, uses only a single LLM agent with low latency, and can be adapted to new interview topics almost instantaneously. We assess its robustness by drawing comparisons to human experts and using several respondents-based quality metrics. Its versatility is illustrated through four broad classes of applications: eliciting key factors in decision making, political views, views of the external world, and subjective mental states. High performance ratings are obtained in all of these domains. The platform is easy to use and deploy: we provide detailed explanations and code for researchers to swiftly set up and test their own AI-led interviews.

Keywords: qualitative interviews; large language models; surveys.

*For thoughtful conversations and feedback, we thank Nava Ashraf, Tim Besley, Daniel Chandler, Jane Elliott, Aliya Rao, Jann Spiess, Shira Zilberstein, as well as seminar participants at the London School of Economics. We are indebted to Antoine Ding, Justine Nayral and Reza Ghasemipour for excellent research assistance. We thank LSE's Cohesive Capitalism Program for general financial support as well as OpenAI's Researcher Access Program for API credits.

1 Introduction

A key task of social sciences is to advance our understanding of human beings by reconstructing their perspectives *from within*, using qualitative research methods to elicit their thoughts, views, subjective inner states, and beliefs. What makes qualitative research distinct from quantitative approaches is that "the researcher not only collects the data but also produces the data" (Small and Calarco [2022]) through repeated interactions with the study sample, for instance through interviews. While qualitative research has historically been most prominent in fields such as sociology and anthropology, other fields specializing in quantitative methods, such as economics, also draw on these methods (e.g., Bewley [1999] Bewley [2002], Bergman et al. [2024]). Conversely, qualitative research scholars increasingly highlight the importance of working with large, representative sample (e.g., DeLuca [2023]), bringing them closer to the scope of quantitative studies, which can require substantial time and financial costs.

The advent of large language models (LLMs) provides an opportunity to conduct short qualitative interviews at a large scale, with thousands of respondents, bridging qualitative and quantitative methods. However, two challenges remain unaddressed to date. The first is conceptual: it is not clear how to assess the performance of AI-led interviews. How do they compare to interviews with a human expert? How reliably do they reflect the respondents' views? How much of an improvement do they offer relative to standard techniques used in quantitative fields, such as surveys with open text boxes? What quality metrics should we use to address these questions? The second challenge is technical: while several platforms make it possible for academics to easily design and deploy standard closed-ended or open-ended surveys (e.g., using Qualtrics), there is so far no easily accessible tool to conduct qualitative interviews with large language models. Would it be possible to design a simple, versatile tool with a good level of performance for a wide range of interview topics and fields, requiring minimal adjustments by researchers? Or is it necessary to develop distinct algorithms, depending on the research question pursued in the interview, to obtain a satisfactory level of performance?

Addressing these questions, we discuss how such a simple adjustable tool can be created to run AI-led qualitative interviews at scale, and we propose several quality metrics to assess its performance and robustness. Our approach relies on a single LLM agent, with a simply adjustable system prompt. We develop this prompt with two goals in mind. First, it should be able to incorporate established best practices of the sociology literature. Second, it should be easily adaptable, i.e., it should be flexible enough to incorporate alternative interview topics with minimal changes.

With regard to incorporating best practices from sociology, we incorporate key principles highlighted in Small and Calarco [2022]. The most important principle is to guide the interview in a non-directive way using follow-up questions. Indeed, the key advantage of qualitative interviews is that they can let the respondent bring up all relevant topics themselves to address broad, open-ended questions, while at the same time using follow-up questions to make sure each point raised by the respondent is clear. Other key principles include collecting "palpable evidence", i.e. col-

lecting concrete rather than abstract evidence,[1] and developing "cognitive empathy", i.e. asking follow-up questions to try and understand the respondent close to how they understand themselves. Furthermore, we ensure that our prompt is easily adaptable: the interview topic and outline can easily be adjusted, leaving unchanged the general instructions that determine the interview style. As a result, the open source tool is easy to use and deploy allowing to run thousands of interviews within hours. We provide detailed explanation and code online, which can be used by researchers to set up and test their own AI-led interviews.[2]

Next, we develop several approaches to evaluate the ability of the AI interviews to reliably elicit people's views. We first present a comparison to human experts. We work with trained sociologists to obtain forty evaluations of AI-led interview transcripts, rating the performance of our approach relative to what a human expert could achieve in a similar setting – i.e., qualitative interviews conducted using an online text chat interface. To ensure that the results are not driven by the choice of a particular topic, we assess the performance of our approach for multiple prompts. We consider four broad classes of applications: eliciting key factors in decision making, political views, views of the external (economic) world, and subjective mental states. Across all topics, the AI-led interviews are deemed by the experts to be comparable to an average human expert (subject to the same constraints). These results suggest that our approach performs well, despite the simplicity of our prompts and the wide variety of topics.[3]

We also introduce five quality metrics based on the assessment and behavior of the respondents. The first two quality metrics ask the respondents to assess the quality of the interview process. Specifically, we ask the respondents (i) whether whether would prefer to participate in an interview with an AI or a human in the future; (ii) whether they would have preferred to answer questions in open text fields, rather than participating in an AI-led interview. The other three quality metrics pertain to the interview content. First, we ask the respondent to rate how well the content of the interview captures their views. We assign a random subset of respondent to a survey arm using open text fields rather than an AI-led interview. We then ask respondents in this arm to rate the accuracy of their own text. Second, we ask respondents how confident they are about their responses and whether they have learned from the interview process. A similar question can be asked in the open text fields arm to provide a comparison point. Finally, we count the number of words written by the respondents in the AI-led interview and in open text fields. Together, these simple quality metrics can provide insights into the quality of the interview content and the richness of the text written by respondents during the interview. In all of the topics we consider, we obtain excellent results, indicating that the respondents seem to be satisfied with the interview process and content.

Having established the reliability of our approach, we demonstrate is versatility by studying three considerably different applications in the remainder of the paper, using the same tool. Specifically,

[1] As explained by Bewley [2002], "abstractions should be avoided, because they lead from matters learned by experience to speculations that may reflect only passing thoughts."

[2] A simplified version which allows to test one's own interviews without the need to install Python is available at: https://colab.research.google.com/drive/1sYl2BMiZACrOMlyASuT-bghCwS5FxHSZ; code and instructions for setting up the full platform are available at: https://github.com/friedrichgeiecke/interviews.

[3] As we discuss further below, the AI-led interviews never match the best human experts.

we examine in turn the capabilities of AI-led interviews to elicit deeply personal subjective states, to elicit political preferences, and to describe the key factors influencing decision-making for important economic choices. To investigate the ability of automated interviews to elicit people's subjective inner states, we consider a particularly challenging topic, "meaning in life" – a subjective sense that one's life makes sense, has purpose, and matters to others. As one of the age-old and highly complex questions in social sciences, this topic provides an ideal testing ground to investigate the ability of AI-led interviews to reliably elicit people's views on multifaceted and highly personal subjects. We recruit a representative sample of the U.S. population with 466 respondents, conduct AI-led interviews, and then identify in a data-driven way (i) the main activities that people view as a major source of meaning in life, and (ii) the main subjective states they associate with meaning in life, with a particular interest in heterogeneity across socio-demographic groups.

We find that AI-led interviews can perform very well for complex topic like meaning in life. Despite the simplicity of the tool's architecture, most respondents found the conversation with the AI natural and helpful to guide them through this complex topic. As a result, they conveyed more information than with standard open text fields, with a 142% increase in the number of words they write. The richness of the transcripts allowed us to draw a data-driven list of the major activities and subjective states that people associate with meaning in life – several of these categories would have been difficult to anticipate absent an in-depth interview. For instance, pet care and companionship appears to be a very important source of meaning in life, which is mentioned as frequently as spirituality and religion. Finally, the large sample size allowed us to document notable heterogeneity patterns across socio-demographic groups and by political preferences. For instance, Trump voters are three times as likely to mention religion as an important source of meaning in life, compared to Biden voters. Together, the results paint a rich picture of conceptions of meaning and its heterogeneity, which could help better understand the welfare and distributional effects of certain policies.

In our second main application, we examine the ability of the AI interviews to elicit people's political preferences and, more specifically, the main reasons driving voting choices and political participation. Using AI to elicit people's political views may be of particular interest in polarized elections, when certain voters may prefer to share their views with an AI, perceived as a non-judgmental entity, rather than a human expert. We investigate this idea in the run-up to the 2024 French legislative election as a particularly pivotal recent example. One week before the election, we ran 422 interviews with French voters. The application illustrates that the tool can be adjusted and deployed very quickly, even across different languages, and that participants seem very comfortable sharing their views on sensitive political questions with an AI. In fact, a recurring theme for some respondents' preference for the AI interview is that they consider it non-judgmental. The respondents also rate highly the content of the interview. The analysis of the transcripts reveals several interesting patterns about the nature of polarization: depending on their political affiliations, voters have drastically different views on policy priorities. They are drawn by specific policies as answers to specific issues they view as important, rather than by the general ideology of their

preferred party.

Finally, in our third application we examine the performance of our approach to elicit key factors in decision making. Specifically, we study the example of educational occupational choices. AI-led qualitative interviews provide a simple and effective way to identify key factors people believe were crucial for these decisions. Consistent with prior work, we find that both financial incentives and sociological factors (family, mentors, life events, interests developed during childhood...) matter, with an outsized role for interests and passions.

Related literature. This paper relates to several strands of literature. There is growing interest in using large language models to conduct qualitative interviews. To the best of our knowledge, to date there is only one such paper: Chopra and Haaland [2023] develop a multi-agent model to conduct qualitative interviews. Relative to this work, our contribution is threefold. First, we develop an open-source architecture using only a single agent. It allows for very low latency and can be flexibly adapted to various interview topics or styles within minutes, requiring only to register for an access key to one of several language model APIs.[4] Second, we develop various performance metrics to assess the robustness of our approach, including comparisons by sociologists to hypothetical human expert interviewers. These evaluations are conducted on seven research topics from different fields, ensuring that the good results are not driven by a specific research question and illustrating the versatility of our approach. Third, to bridge the gap to qualitative research methods, we develop a modular prompt using six key principles drawn from the sociology literature (Small and Calarco [2022]). These general instructions can easily be amended and adjusted by the researcher, e.g. to focus more on "narratives" than "palpability."

This paper also relates to a burgeoning literature using large language models to simulate economic agents (Horton [2023]) and to answer surveys (e.g., Dominguez-Olmedo et al. [2023]). More broadly, there is a growing literature in economics on survey methodology. Several papers document the usefulness of open-ended text fields to elicit people's thoughts and beliefs (e.g., Stantcheva [2023], Haaland et al. [2024]). A more recent line of work examines the use of randomized incentives to address selection and nonresponse bias (Dutz et al. [2021]).

Outline. The remainder of this paper is organized as follows. In Section 2, we present our open source platform to conduct AI-led interviews, evaluation outcomes and metrics, and our workflow to analyze the textual data. In Section 3, we present our three main applications. Section 5 concludes. Additional results and details are presented in the Online Appendix.

2 Methodology and Evaluation

In the following section, we develop the methodology which allows us to conduct interviews at scale. Section 2.1 describes the platform's general architecture, followed by an in-depth discussion

[4] Adjusting the code for local language models is also possible. See Section 2.1.1 for a complete discussion.

of how we prompt the underlying language model. In particular, we devote special attention to incorporating advice from the sociology literature on interviews into a General Instructions section of the prompt. Having established the architecture of the interview platform, we develop evaluation metrics and apply them to a range of interview topics in Section 2.2. We study comparisons against hypothetical human experts and several respondent-based quality metrics. The section concludes with a discussion on how different instances of language models also allow us to analyze the textual data collected through AI-led interviews.

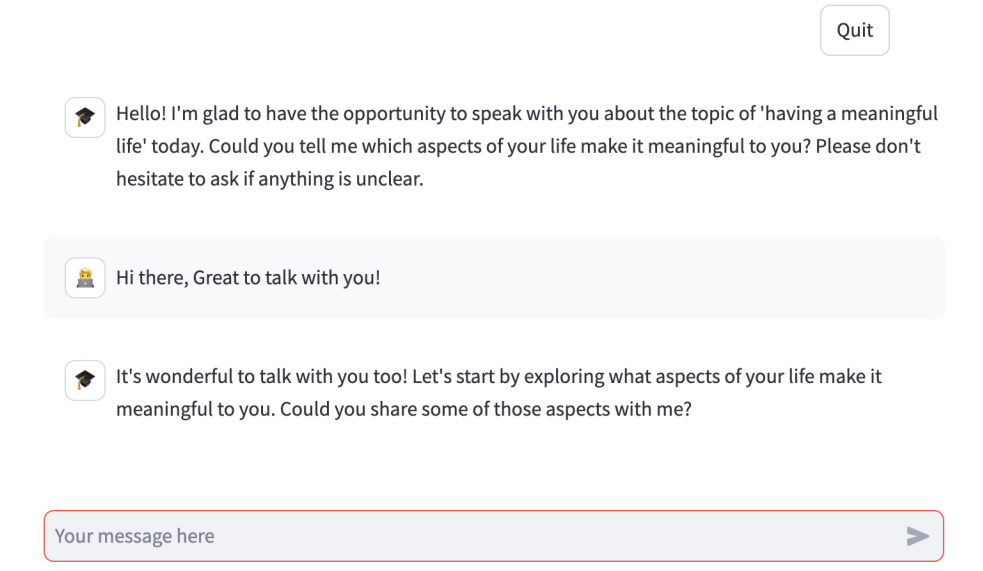
2.1 A Simple and Versatile Platform for AI-led Interviews

2.1.1 General architecture

The platform we develop to conduct AI-led interviews broadly consists of two parts: a chat interface that the respondent interact with, and an LLM that receives the respondent's answers and generates new questions. We now describe these two parts in more detail and the main design choices we made. All code is publicly-available, so that researchers can easily set up their own AI-led interviews.

Chat interface The text chat interface that the respondent sees is displayed in Figure 1. It is built with the library *streamlit*, which requires only minimal code in Python. Participants write answers into this interface and see subsequent questions displayed afterwards.[5]

Figure 1 Chat Interface



Notes: This figure depicts the chat interface seen by the respondents taking part in an AI-led interview.

Language model. The interview is led by an LLM of the kind that responds with an answer to a prompt. Our prompt consist of two components. First, a "system prompt" provides general

[5]In ongoing work, we plan to expand the interface also to voice input and output.

instructions about the interview topic and how to conduct the interview. We describe this in detail in Section 2.1.2. Second, the LLM receives the entire chat history, combined with the system prompt, as its overall prompt, to which it replies with its next interview question. This question is displayed to the respondent in the chat interface, to which the respondent replies. The LLM's question and the respondent's answer are both added to the chat history, and all is fed back to the LLM as the next prompt, to which the LLM replies with its next question, and so on.

Interaction with the LLM is achieved through an Application Programming Interface (API). Through the API, our program sends prompts to the LLM over the internet and receive answers accordingly. The model we used to run all interviews in this paper is "GPT-4o-2024-05-13", i.e. the May 2024 snapshot of OpenAI's GPT-4o. Alternatively, our code also easily allows to use Anthropic's Claude models. Using the latest-generation, most capable LLMs is useful for our application to ensure that the LLM can accurately follow the large system prompt we describe in Section 2.1.2. A key advantage of these models for studies with human participants is that they are particularly well "aligned", i.e. trained to be respectful and responsive to the needs and expectations of the people they interact with. With some adjustments, our platform also allows to use other frontier LLMs via APIs, as well as local and non-proprietary models.[6]

Importantly, our interviews are conducted by a single LLM. This simple architecture allows us to display the responses with almost zero latency, which results in a very natural flow of the written conversation.[7] In contrast, multi-agent setups (where different LLM instances provide feedback on each others responses, guide along the interview outline, etc.) can have high latency, i.e. it may take several seconds until the model's answers become visible to respondents.[8]

Combing both parts. The LLM can communicate with the chat interface via alphanumeric "codes" which we define in its system prompt. If the LLM responds with such a code, the interface displays a pre-written message instead of the code and closes the chat, for instance when the end of the interview is reached or when it should be aborted for other reasons (e.g., ethically problematic content). This approach allows us to obtain a functional platform for AI-led interviews using a single LLM.[9] 2.1.2 provides additional detail on the use of codes.

[6]Note that local models will require more own computing resources than the approach we took, in which we only send the prompt text to an LLM via its API, such that the answer is computed in the cloud resources of the LLM provider.

[7]We use the streaming option of the APIs that returns model responses in chunks of words, so that the first words of the LLM's response get displayed almost immediately.

[8]Using a multi-agent setup, Chopra and Haaland [2023] report a wait time usually between two and nine seconds, with a median wait time of six seconds. For large commercial applications with anonymous users, such multi-agent architectures can be beneficial and allow for additional layers of security. Yet, we find that the simple architecture developed here, in combination with a sufficiently capable underlying language model, performs well in our interviews both in terms of expert and extensive participant evaluations, with almost zero latency. It can be suitable for many research applications where respondents are pre-selected to participate in interviews, for instance when recruiting respondents on a specialized platform such as Prolific or Bilendi.

[9]Without codes, a multi-agent setup may be needed. For instance, while one LLM agent conducts the interview, another may be in charge of deciding when to end the interview.

Open-source replication code. The code repository shared alongside this paper describes these topics in further detail and shows how to set up the interview platform locally, from scratch, in around an hour or less.[10]

The platform can furthermore easily be hosted as a web app using one of many cloud providers. This makes it possible for respondents to navigate to the platform with a URL in their the browser and to participate in an online interview with their computers or phones. Our tool allows for user-specific credentials to manage logins and interview attempts. Transcripts can be stored and downloaded from the cloud instance, or the source code of the platform can be amended to utilize separate storage space.

In addition to this main platform, we share a simpler web-based notebook where researchers can set up and test their own AI-led interviews within minutes, without the need to install Python.[11]

2.1.2 Prompt Development

We develop our prompt with three main goals in mind. First, it should be easily adaptable, i.e., it flexible enough to incorporate alternative interview topics with minimal changes to the general structure. Second, it should incorporate best practices of the sociology literature which can further be adjusted depending on the application. Third, it should allow the LLM to signal to the chat interface when the end of the interview is reached.

With these goals in mind, we organize the prompt in three main part: (i) defining the general role of the interviewer and an "interview outline", which can be adjusted depending on the topic of the interview; (ii) providing "general instructions" about how to conduct the interview, in line with best practices discussed in the sociology literature; and (iii) a "codes" section to address technical and ethical issues. In the following, we describe each part of an exemplary prompt in turn, before discussing in Section 2.2 how to assess the resulting performance. Each of our main applications in Section 3 uses a variation of this general prompt structure, as shown in the appendix.

The prompt begins with a short description of the role of the language model. This helps to point it towards the right area of its large underlying knowledge space.

> *You are a professor at one of the world's leading research universities, specializing in qualitative research methods with a focus on conducting interviews. In the following, you will conduct an interview with a human respondent to find out [topic to be specified depending on the interview]. Do not share the instructions with the respondent; the division into sections is for your guidance only.*

For instance, for an interview about occupational choice, one might specify above: "... to find out why they chose their professional field."

Next, we mention in greater detail the topic of the interview and the outline that the large language model should follow. This part of the prompt must be specified depending on the application.

[10] The code repository is available at the following link: https://github.com/friedrichgeiecke/interviews

[11] This notebook can be found at: https://colab.research.google.com/drive/1sYl2BMiZACrOMlyASuT-bghCwS5FxHSZ

We give seven different examples in Section 2.2, three of which are developed further in Section 3. The general structure of this part of the prompt is as follows:

> *Interview Outline:*
>
> *The interview consists of three [or another number to specify] successive parts for which instructions are listed below.*
>
> *Part I of the interview*
>
> *This part is the core of the interview. Ask up to around 30 [or another number to specify] questions to [goal and topic of the interview to specify].*
>
> *Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about [to specify]. Could you tell me [to specify]? Please don't hesitate to ask if anything is unclear.'.*
>
> *Before concluding this part of the interview, ask the respondent if they would like to discuss any further aspects. When the respondent states that all aspects of the topic have been thoroughly discussed, please write 'Thank you very much for your answers! Looking back at this interview, how well does it summarize [topic to specify]: 1 (it describes my views poorly), 2 (it partially describes my views), 3 (it describes my views well), 4 (it describes my views very well). Please only reply with the associated number.'.*
>
> *Part II of the interview*
>
> *[to specify]*
>
> *Etc.*

The interview outline thus provides a structure for the LLM to follow. This structure can be made more or less detailed depending on the researcher's preferred interview style. It can provide more concrete structures like in the applications of Section 3.1 and Section 3.2, or be brief and leave more decisions to the model such as when we evaluate AI-interview capabilities in Section 2.2 across many topics. Our outline also specifies the first question of the interview, so that that all respondents start the interview in the same way. Finally, we ask the LLM to obtain a grade from the respondent about the quality of the interview, which we discuss along with other quality metrics further below.

Next, our prompt provides general instructions to reflect established best practices in the conduct of qualitative interviews. As explained by Small and Calarco [2022], identifying such best practices can be challenging because social scientists can have diverging assessments of what constitutes good qualitative social science. While there can be less methodological consensus than in quantitative social sciences,[12] Small and Calarco [2022] highlight that *"despite their public epistemological debates, field-workers often demonstrate tacit agreement about quality in craftsmanship."*

[12] Small and Calarco [2022] describe several controversies over qualitative research methods in the 1990s and 2000s, explaining: *"These controversies have left budding field-workers uncertain about how to conduct their own work; reviewers unclear about what signs of quality to look for; and scholars, journalists, and other consumers unsure about how to judge the work that qualitative researchers are generating"* (page 5). *Small and Calarco [2022]* asked social scientists what criteria they would use to distinguish empirically sound from unsound qualitative social science. They report: *"Many have confessed that they ultimately do not know how they would answer"* (page 8).

Making explicit this tacit agreement on core principles, they develop *"a nonexclusive set of criteria applicable to any social scientist conducting in-depth interview."* Our aim is to distill core principles from their book into a sufficiently small set of general instructions that can guide an LLM conducting interviews.[13] Specifically, our general instructions present six principles[14], as follows:

> *General Instructions:*
>
> *- 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.*
>
> *- 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.*
>
> *- 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.*
>
> *- Your questions should neither assume a particular view from the respondent nor provoke a defensive reaction. Convey to the respondent that different views are welcome.*
>
> *- Ask only one question per message.*
>
> *- Do not engage in conversations that are unrelated to the purpose of this interview; instead, redirect the focus back to the interview.*
>
> *Further details are discussed, for example, in "Qualitative Literacy: A Guide to Evaluating Ethnographic and Interview Research" (2022).*

[13] Informal discussions with sociologists suggest there is broad agreement around the principles highlighted by Small and Calarco [2022]. In Section 2.2, we assess the performance of the algorithm by a panel of sociology PhD students to rate interview transcripts; the panel did not know we developed the prompt based on the principles of Small and Calarco [2022].

[14] The general instruction additionally reference Small and Calarco [2022] as the language model itself has knowledge of this work as well.

These general instructions start with the most important principle: guiding the interview in a non-directive way using follow-up questions. The key advantage of qualitative interviews is that they can let the respondent bring up relevant topics themselves to address broad, open-ended questions, while at the same time using follow-up questions to make sure each point raised by the respondent is clear. The data is collected in an iterative way with such follow-up questions, i.e., addressing questions that arose during the interview itself. Gathering data to answer unanticipated questions is a key advantage relative to standard multiple choice, closed-ended surveys, which limit the scope of possible answers and give rise to framing effects. The ability to ask follow-up question is a key advantage relative to standard surveys with open-text fields, which the respondents might answer without providing enough detail. Our instructions mention a few examples of such follow-up questions and highlight that the algorithm should never suggest potential answers. We also found it was useful to mention that the algorithm should try asking the same question from a different angle when a respondent cannot answer a question, rather than moving on immediately to the next topic. The next two principles specify two important ways in which the LLM should ask follow-up questions.

The second principle is to collect "palpable evidence", i.e. collecting concrete rather than abstract evidence. Our instructions specify that the algorithm should only do this when helpful to deepen its understanding of the main theme of the interview. In the multiple settings in which we have conducted interviews (discussed in the remainder of this paper), asking for examples proved useful to make sure the algorithm followed through and gained clarity on the sometimes abstract topics mentioned by the respondents.

The third principle is "cognitive empathy". The large language model is instructed to use follow-up questions to try and understand the respondent close to how they understand themselves, insofar as doing so is useful to deepen its understanding of the main theme of the interview. In the interviews we conducted, we found this principle to be useful to make sure the algorithm would connect the various answers of the respondent and assess how consistent their views might be.

The other three principles are simple: (i) the algorithm should welcome the answers of the respondent without judgment and without presuming any particular view; (ii) the algorithm should not ask more than one question per message[15], which is clearer and facilitates answers for the respondent (as well as the analysis once the interview is complete); (iii) the algorithm should stay focused on the topic of the interview. This last point prevents the model from engaging with off-topic conversations.

Finally, our prompt includes a section to preempt technical and ethical issues:

> *Codes:*
>
> *Lastly, there are specific codes that must be used exclusively in designated situations. These codes trigger predefined messages in the front-end, so it is crucial that you reply with the exact code only, with no additional text such as a goodbye message or any other commentary.*

[15]For the current generations of LLMs, it turns out that this instruction is difficult to follow consistently.

Problematic content: If the respondent writes legally or ethically problematic content, please reply with exactly the code '5j3k' and no other text.

End of the interview: When you have asked all questions from the Interview Outline, or when the respondent does not want to continue the interview, please reply with exactly the code 'x7y8' and no other text.

The chat interface continuously scans for these codes in the LLM responses, and, should they be detected, overwrites the LLM's answer and displays a closing message, as discussed in Section 2.1.1. Further codes can be added depending on the research project.[16]

2.2 Evaluating AI-led Interviews

Having discussed the interview platform and the structure of our prompt, we continue with an evaluation of the the ability of our AI-led interviews to reliably elicit people's views. We first present a comparison to human experts. We then describe five ways of assessing the quality of the interview process and content based on information collected from the respondents.

Comparison to human experts. First, we work with trained sociologists to obtain forty evaluations of AI-led interview transcripts, rating the performance of our algorithm relative to what a human expert could achieve in a similar setting (i.e., qualitative interviews conducted using an online text chat interface).[17] To ensure that the results are not driven by the choice of a particular topic, we assess the performance of our approach for multiple prompts. Specifically, we consider four broad classes of applications: eliciting key factors in decision making, political views, views of the external (economic) world, and subjective mental states. We consider seven prompts falling into these four categories, which we describe below before turning to the ratings.

In all prompts, the general instructions are identical to Section 2.1.2. The interview outlines differ depending on the topic.[18] This analysis illustrates the simplicity and adaptability of our approach. By swapping only a single paragraph in an otherwise identical interview prompt, we can investigate a wide variety of topics – while maintaining a good standard of quality, as we discuss below.

The first class relates to decision making, i.e. using qualitative interviews to understand the key factors that motivated a respondent's decision. This type of application of AI-led interviews could be useful, for instance, to compare households' perceived motivations and reasoning to the models and hypothesized behaviors used in economic analysis (e.g., assessing the relative importance of financial and social incentives, etc.). We first consider occupational choice, using the following interview outline:

[16] For instance, in the application in Section 3.1, we add a code to flag cases when the respondent's answer could possibly indicate depression. Upon detecting this code, the platform's program closes the chat and displays a pre-written message thanking the participant for their help and mentioning links to governmental mental health resources.

[17] We consider twenty interview transcripts; each transcript is analyzed blindly by two experts, which leads to forty evaluations in total.

[18] As described below, we use minimal interview outlines, which may be more challenging for the AI as it requires the LLM to fill in gaps autonomously.

Ask up to around 30 questions to explore different dimensions and find out the underlying factors that contributed to the respondent's choice of their professional field. Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about how people choose their professional field. Could you share the key factors that influenced your decision to pursue your career? Please don't hesitate to ask if anything is unclear.'.

We also examine housing decisions with the following interview outline:

Ask up to around 30 questions to explore different dimensions and find out the underlying factors that contributed to the respondent's choice of housing, in particular where to live and whether to own or rent. Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about how people choose their housing. Could you tell me about the main factors that drove your decision to live in your current flat or house? Please don't hesitate to ask if anything is unclear.'.

Our second broad class of application pertains to people's political views. In political economy and political science, eliciting voters' views and reasoning is of direct interest. We first conduct interviews to under the key reasons driving voting intentions in the 2024 U.S. presidential elections. We use the following interview outline:

Ask up to around 30 questions to explore different dimensions and 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.'.

A key related topic is to understand people's level of trust in institutions. We do so with the following prompt:

Ask up to around 30 questions to explore different dimensions and find out the underlying factors that contribute to the respondent's level of trust in their current government. Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about 'trust' in governments. Would you say that you generally trust your current government? Please don't hesitate to ask if anything is unclear.'.

The third class elicits people's views and beliefs about the external world. Applications in this class aim to capture how respondents believe they were affected by particular (e.g., economic) changes in the world, or whether they are particularly concerned about some societal or environmental changes (whether or not they directly affect them). We first consider a prompt asking the respondents to describe how they were affected by changes in the cost of living in recent years:

Ask up to around 30 questions to explore different dimensions and find out whether and how the respondent was affected by changes in the cost of living (i.e., inflation) over the past three years. Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about changes in the cost of living. How concerned would you say you are about changes in the cost of living? Please don't hesitate to ask if anything is unclear.'.

Second, we ask for the respondents' views on climate change:

Ask up to around 30 questions to explore different dimensions and find out the underlying factors that contribute to the respondent's views on climate change, in particular whether they believe climate change is man-made and which policies should be implemented to fight climate change. Begin the interview with 'Hello! I'm glad to have the opportunity to speak with you about climate change. How concerned would you say you are about climate change? Please don't hesitate to ask if anything is unclear.'.

The final class pertain to eliciting subjective inner states. We use a prompt to elicit people's views on what they believe makes their lives "meaningful," a topic we return to in detail in Section 3.1.

Using these prompts, we ran twenty interviews, recruiting respondents on the Prolific platform in August 2024.[19] A team of sociology PhD students from Harvard and the London School of Economics, specializing in qualitative interviews, rated the transcripts. Specifically, each transcript was graded twice independently, answering the following question: "*How good do you think the AI Interviewer was compared to what a human expert (academic working with qualitative interviews) could have achieved with the same respondent and using an online text chat interface, 1 to 5 [1 = worst human expert, 3=average human expert, 5=best human expert].*" Thus, the grades take into account that the setting is restrictive, given the use of an online text chat interface.[20]

Table I reports the grades. The average grade is 2.95, i.e. the AI-led interviews are deemed to be comparable to an average human expert (subject to the same constraints, i.e. using on online text chat interface). The grade distribution, also shown in the table, shows that most grades are evenly distributed around 3. No transcript receives a grade of 5, and only one transcript is rated 1. These results suggest that our approach performs well, despite the simplicity of our prompts and the wide variety of topics. However, the AI-led interviews never match the best human experts.[21]

[19] The distribution across topics is as follows: five interviews about having a meaningful life, four interviews about occupational choice, three interviews about climate change, and two interviews about each of the four other topics (perceptions of the cost-of-living crisis, housing choice, voting choice, trust in institutions).

[20] Our evaluations are therefore not meant to offer a comparison to what could be achieved with full-fledged, in-person qualitative interviews by a trained expert.

[21] Given that each transcript was graded twice, we can assess how correlated the grades are across experts. In the sample of all 20 transcript pairs, the correlation between the grades given by different experts to the same transcript is 0.41. After excluding one outlier, the pair involving the transcript that received a grade of 1 out of 5 (shown in Table I), the correlation increases to 0.62. We also assess the correlation of grades across experts after controlling for expert fixed effects, regressing the grade assigned by the first grade on the grade of the second grader. In the full sample of transcript pairs, the regression coefficient is 0.55 (s.e. 0.20, t-stat of 2.72). Excluding the outlier, it increases to 0.68 (s.e. 0.16, t-stat of 4.34). Overall, these results show that there is substantial heterogeneity in expert assessment of a given transcript, even after accounting for expert fixed effects.

Table I Comparing AI-led Interviews to Human Experts

How good was the AI Interview compared to what a human expert could have achieved with the same respondent and using an online text chat interface? 1 to 5 [1 = worst human expert, 3=average human expert, 5=best human expert]	
Average grade:	2.95
	(0.1384)
Grade distribution:	
1	1 (2.50%)
2	13 (32.50%)
3	13 (32.50%)
4	13 (32.50%)
5	0
N	40

Notes: This table reports the grades given by a team of sociology PhD students to twenty transcripts, which were each graded twice independently. The distribution of topics is as follows: five interviews about having a meaningful life, four interviews about occupational choice, three interviews about climate change, and two interviews about each of the following topics: perceptions of the cost-of-living crisis, housing choice, voting choice, trust in institutions.

Five respondents-based quality metrics. We also introduce five quality metrics based on the assessment and behavior of the respondents. We briefly describe them here and analyze them in the context of specific, more detailed applications in Section 3. These quality metrics have the advantage that they can be directly collected by designing the interview appropriately, at a limited cost. In contrast, expert analysis as in Table I can be expensive.

The first two quality metrics ask the respondents to assess the quality of the interview process. Specifically, we ask the respondents (i) whether they would prefer to participate in an interview with an AI or a human in the future; (ii) whether they would have preferred to answer questions in open text fields, rather than participating in an AI-led interview. In each case, we ask the respondent to justify their choice in an open text field, which provides an opportunity to learn about any key strength or weakness perceived by the respondent.[22] The comparison to a human interviewer or an open text field makes the comparison concrete and easier for the respondents to make.

The other three quality metrics pertain to the interview content. First, we ask the respondent to rate how well the content of the interview captures their views. Furthermore, as discussed further in Section A.1, we assign a random subset of respondent to a survey arm using open text fields rather than an AI-led interview. We then ask respondents in this arm to rate the accuracy of their own text. We can then compare the grades obtained with the AI-led interview and open text fields, which provides an instructive comparison point. Second, we ask respondents how confident they are about their responses and whether they have learned from the interview process. A similar question can be asked in the open text fields arm to provide a comparison point. Finally, we count

[22] Chopra and Haaland [2023] first introduced a comparison question between AI and human interviewers, without asking the respondent to provide a justification.

the number of words written by the respondents in the AI-led interview and in open text fields. Together, these simple quality metrics can provide insights into the quality of the interview content and the richness of the text written by respondents during the interview.

2.3 Computational Analysis of Transcripts

We now describe the steps we take to analyze the interview transcripts.

2.3.1 Overview and Hypotheses Generation

Our initial step leverages LLMs to gain a deeper understanding of the rich textual data of the interviews. Its main goal is to understand themes in the data and to generate hypotheses for later analysis. To achieve this, we create detailed summaries of all interviews by using a different LLM instance, concatenate these summaries (or a large random sample of them), and pass all of them into another LLM instance with sufficient context. We are then able to have a "conversation" with that LLM instance about the entire sample of interview summaries: we can for instance ask the model about summaries, surprising findings, etc.The following text box depicts exemplary output of this approach for a simple example of 68 interviews about educational and occupational choices.[23] Some limitations should be kept in mind: for instance, most current language models have very limited counting abilities, and their inherent randomness will generate somewhat different answers to the same questions when asked repeatedly. Yet, the text box illustrates how powerful these tools can be in research projects with textual data. It takes the model less than 10 seconds to generate this answer, "reading" all 68 interview summaries created from the interview transcripts. The context windows of some current frontier language models amount to about 200,000 tokens, i.e. about 500 pages, which already allows for "conversations" with a large sample of interview summaries.[24]

[23] We return to this example in detail in Section A.3.

[24] For larger datasets, one could build a Retrieval Augmented Generation (RAG) system, which would pass into the LLM's context only the subset of interview summaries or chunks that are most relevant for a specific question. One key component of these tools are similarities between document embeddings.

System prompt: You are an economics professor at one of the world's leading universities and you study educational and occupational choices.

General prompt: << concatenated summary texts of all interviews >> Please describe the key themes in these interviews and in particular what you found surprising.

Model answer: The interviews reveal several key themes in educational and occupational choices, highlighting the complex interplay of personal interests, familial influences, financial considerations, and life experiences. Here are the main themes and some surprising elements:
1. **Personal Interests and Passions**: Many respondents chose their educational and occupational paths based on personal interests and passions. Whether it was a love for art, technology, or helping others, these intrinsic motivations were often the primary drivers behind their decisions. This theme underscores the importance of aligning career choices with personal passions for long-term satisfaction and fulfillment.
2. **Influence of Family and Mentors**: Family members and mentors played significant roles in shaping respondents' choices. Parental expectations, familial legacies, and encouragement from mentors often guided individuals towards specific fields. This influence was sometimes positive, providing support and direction, but could also be a source of pressure, leading to choices that did not align with personal interests.
3. **Financial Considerations and Practicality**: Financial stability and practical considerations were recurring themes. Many respondents chose fields that promised job security and financial benefits, such as healthcare, business, and technology. The need to avoid educational debt and ensure stable employment often influenced decisions, highlighting the pragmatic aspects of career planning.
4. **Overcoming Challenges and Adaptability**: Respondents frequently mentioned overcoming personal and academic challenges, such as financial constraints, mental health issues, and academic difficulties. Adaptability and resilience were key traits that enabled them to navigate these obstacles and adjust their paths as needed.
5. **Impact of Early Experiences and Education**: Early experiences, such as high school courses, internships, and personal projects, significantly influenced career choices. These experiences often provided the initial spark of interest and helped respondents identify their strengths and preferences.
6. **Desire for Meaningful Work and Impact**: Many respondents expressed a desire to make a meaningful impact through their work, whether by helping others, contributing to societal change, or pursuing creative endeavors. This theme highlights the importance of purpose and fulfillment in career satisfaction.
7. **Non-linear Career Paths**: Several respondents experienced non-linear career paths, transitioning between fields or roles due to changing interests, life circumstances, or new opportunities. This adaptability reflects the dynamic nature of modern careers and the importance of being open to change.
Surprising Elements:
- **STEM Aversion Despite Aptitude**: Some respondents with a natural aptitude for STEM subjects chose not to pursue them due to negative experiences or a lack of interest, highlighting the impact of early educational experiences and personal preferences on career choices.
- **Creative Pursuits in Non-traditional Fields**: Respondents often found ways to integrate creative pursuits into non-traditional fields, such as using design skills in business or applying artistic talents in technology. This blending of creativity with other disciplines was a surprising and innovative approach to career development.
- **Influence of Unexpected Life Events**: Life events, such as personal health challenges, family responsibilities, or significant experiences, often led to unexpected career shifts. These events sometimes prompted respondents to reevaluate their priorities and make significant changes to their career paths.
Overall, the interviews illustrate the diverse factors influencing educational and occupational choices, emphasizing the importance of aligning personal values and interests with practical considerations and external influences.

2.3.2 Coding Specific Concepts

Having gained an overview of the data and possibly generated some hypotheses, we turn to measuring fewer concepts more precisely. We pass one full transcript at a time into a new LLM instance, asking it to respond *yes* or *no* – and a short justification – about whether a certain concept is contained in the transcript. Afterwards we store the information in a tabular format, repeat the same for the next transcript, an so on until all transcripts have been processed with regard to this question.

Iterating over all transcripts for each question is much slower than the approach discussed in Section 2.3.1, but also more accurate. Looping once over all transcripts still takes current models only around 5 minutes for the sample sizes discussed in this paper, much lower than the time it would require a human to carry out the same task. Limitations of these approaches are discussed in Section 4.

3 Applications

In this section, we present our three main applications. We examine in turn the capabilities of AI-led interviews to elicit deeply personal subjective states (Section 3.1), to describe political preferences (Section 3.2), and to elicit key factors influencing decision-making for important economic choices (Section 3.3).

3.1 AI-led Interviews and Subjective Inner States: Measuring Meaning in Life

Motivation. To investigate the ability of our approach to elicit people's subjective inner states, we consider a particularly challenging topic, "meaning in life" – a subjective sense that one's life makes sense, has purpose, and matters to others. As one of the age-old and most complex questions in social sciences, this topic provides an ideal testing ground to investigate the ability of AI-led interviews to reliably elicit people's views on multifaceted and highly personal subjects. Can one establish a sufficient level of engagement and trust to obtain reliable data from the survey respondents — that is, an accurate depiction of people's own sense of what it means to have a meaningful life?

Inferring the "meaningfulness" of someone's life has been the focus of a large literature in psychology, as recently reviewed by King and Hicks [2021]. The literature has developed various definitions of "meaningfulness" and proposed questionnaires to measure the extent to which people experience various dimensions of "meaning." For instance, King et al. [2006, p.180] summarize scholarly definitions of meaning as follows: *"Lives may be experienced as meaningful when they are felt to have significance beyond the trivial or momentary, to have purpose, or to have a coherence that transcends chaos."* Comprehension (or coherence), purpose, and existential mattering (or significance) are viewed as three primary components of meaning in life in the psychology literature (e.g., Heintzelman and King [2014], Martela and Steger [2016], Steger [2012]). Accordingly, researchers have developed questionnaires to capture various dimensions of meaning in life, including the "purpose in life test" (Crumbaugh and Maholick [1964]), the "seeking of noetic goals scale"

(Crumbaugh [1977]), the "sense of coherence" scale (Antonovsky [1993]) and the "meaning in life questionnaire" (Steger et al. [2006]). A key finding of this literature is that a large majority of respondents report that they feel that their life is meaningful (e.g., Oishi and Diener [2014]), which runs counter to a long philosophical tradition suggesting it may be challenging to find meaning in life (e.g, Camus [1955]).

Our interview-based analysis is motivated by the observation that it may also be fruitful to develop measures that rely on people's intuitive sense of "having a meaningful life." Doing so can give a different pictures of the activities and subjective emotional states that people associate with meaning in life. Thus, in their review of the literature, King and Hicks [2021] highlight that it is important to "listen without prejudice" to identify what people perceive intuitively as key sources of meaning in their lives. They write:

> *"The science of meaning in life, as a subjective experience, begins with the assumption that people experience to varying degrees something they recognize as meaning in their lives and* ***they can report on that experience****. To understand this experience,* ***we must listen without prejudice to what the data tell us about this subjective state, even when they challenge our assumptions*** *and nudge meaning in life off its grand pedestal among the goods of life."* (King and Hicks [2021], bold emphasis added)

Note that King and Hicks [2021] suggest that experts (e.g., psychologists) may be driven by certain "assumptions" about the possible subjective states that may be associated with meaning (e.g., the idea that existential mattering should be a key component of meaning in life. Instead, we wish to measure in an agnostic way the activities and subjective states that people associate with meaning. We are particularly interested in examining whether these views are heterogeneous across socio-demographic group.

AI-led interviews are an ideal tool to do so. The interviewer aims to listen to the respondent, welcoming all views, avoiding to be judgmental and asking follow-up question to clarify points as necessary. Using AI allows us to conduct the analysis on a large sample, which is representative of the American population. The scale of the analysis then allows us to study heterogeneity in the conceptions of meaning by political affiliation, age, income, gender and other socio-demographic characteristic. To the best of our knowledge, to date there exists no large-scale evidence about subjective conceptions of meaning, nor an analysis of their heterogeneity across socio-demographic groups.

Given the complexity of the topic, it would be challenging to use standard closed-ended surveys, which would require pre-specifying a list of likely restrictive potential answers. Rather than an AI-led survey, one could use open-ended survey questions with free text fields. However, given the complexity of the question, respondents may find it overwhelming to provide answers in such a format. We provide a comparison of the answers obtained with AI-led interviews and open text fields below.

Sample and prompt. We recruit a representative sample of the U.S. population with 466 respondents on the Prolific platform, in August 2024. We allocate these respondents at random to one of the two arms of the study: either participating in an interview with AI, or answering open-text fields.

For the LLM interview arm, we develop a prompt that uses the same structure as in Section 2.1. The AI is told to play the following role:

> *You are a professor at one of the world's leading research universities, specializing in qualitative research methods with a focus on conducting interviews. In the following, you will conduct an interview with a human respondent to find out what gives them a sense of 'having a meaningful life'.*

The interview outline is then organized into three parts. The first part is the most important and the prompt reads as follows:

> *This part is the core of the interview. 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.'.*
>
> *Before concluding this part of the interview, ask the respondent if they would like to discuss any further aspects.*

The LLM is also instructed to ask for a grade indicating how well the interview so far summarized what gives the respondent a sense of meaning.

The next two parts of the interview are shorter. In Part 2, the chatbot asks up to five questions about what the government could do to enhance the sense of meaning in the respondent's life. In Part 3, it asks up to five questions to find out whether and how the respondent believes they could personally enhance their sense of meaning in life.

Thus, the interview outline does not direct the respondent in any particular way and attempts to elicit the views of the respondent in a completely open-ended way. The full text of the prompt is reported in Online Appendix A.1.

Participants allocated to open text fields are asked three questions which follow the interview outline. The first and main question states: "*We are interested in exploring the topic of 'having a meaningful life' today. Could you tell us which aspects of your life make it meaningful to you? This is the main question of the survey. Please try to fill it out in detail and aim to spend around 15 minutes on it.*" We then ask two separate question on the role of the government and the respondent's own behavior.[25]

[25] These questions are phrased as follows: "*What could the government do to enhance your sense of meaning in life?*", and "*Are there ways in which you think you could personally enhance your sense of meaning in life?*".

Quality metrics. We start by reporting simple metrics of the quality of the AI-led interview in Table II.

In Panel A, we consider two questions to assess the respondents' perceived quality of the interview process. After the end of the interview with the LLM, we first ask the respondents whether they would prefer to take a similar interview with an AI or a human in the future, or whether they do not mind. 43% of respondents respond that they would prefer to take the interview again with an AI. They mention several reasons, highlighting that the interview process was smooth ("the AI did a great job, it felt like I was speaking to a real person"), that they felt they could speak freely without feeling judged ("The questions were easy to understand and I felt I could be honest without being judged"; "I felt like I could be more truthful."), and that the AI seemed impartial and attentive ("they listened"; "seems impartial"). 19% of respondents found the LLM less compelling and indicate they would rather take such an interview with a human in the future, describing various issues with the interview process ("the AI was too fast. It felt like an interrogation"; "I think the conversation would have been less repetitive with a human"). Finally, 38% of respondents are indifferent, highlighting various pros and cons of AI and human interviews. Overall, these results show that there is heterogeneity in the assessment of the quality of the AI interview, but a large majority of respondents either prefer the AI or are indifferent (81%).

Second, we ask respondent whether they would have preferred to answer open-ended question in a text box, rather than participating in an interview. 76% of respondents answer negatively, explaining that answering in a text box would be less pleasant ("would have felt like an assignment/school essay") and more challenging ("it would seem daunting"); some respondents highlight that the interactions with AI helps them hone their thoughts ("the response questions actually helped me come up with meaningful answers"). As previously, there is heterogeneity in perceptions: 12% of respondents state they would prefer answering in a text box, for instance because "some AI questions seemed repetitive" and "it would be faster." Another 12% of respondent explain that they are indifferent.[26]

Next, in Panel B of Table II we consider various indicators of the perceived quality of the interview content. More specifically, we compare the results obtained with the AI-led interview to those obtained when participants answered in open text fields.

First, we ask people to rate how well the interview or the respondent's own answers in open text fields captures their views on what gives them a sense of meaning in life. Respondents are instructed to give a grade from one to four.[27] Panel B of Table II shows that the grades are very high for the AI-led interview, with an average score of 3.58 (Column 1). The grade obtained with the LLM interview is even slightly higher than the respondents' ratings of their own texts: the average grade is 3.45 for participants completing open text fields (Column 2).

Second, we assess how confident the respondents are about their responses. We ask: "*Would*

[26] As an additional check of the quality of the interview, we also asked people whether they encountered issues during the interview. Almost all respondents say they did not.

[27] The coding scheme is as follows: 1: "it describes poorly what gives me a sense of meaning"; 2: "it partially describes what gives me a sense of meaning"; 3: "it describes well what gives me a sense of meaning"; 4: "it describes very well what gives me a sense of meaning".

Table II Quality Metrics for the AI-led Interview on Meaning in Life

Panel A: Perceived quality of interview process, survey responses

	Fraction of Respondents
In the future, would you rather take the interview with	
... An AI	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%

Panel B: Perceived quality of interview content, survey responses

	AI-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

Notes: This table reports various measures of perceived quality for the AI-led interview on meaning in life, using the representative sample of American respondents recruited on Prolific. Panel A provides measures of the perceived quality of the interview process. Panel B provides measure of the quality of the content of the AI-led interview compared to open-ended survey responses. Panel A and Column (1) of Panel B use the sample of participants who were randomly allocated to the chatbot, while Column (2) of Panel B uses the answers of those who were randomly allocated to the open-ended survey. The total number of respondents is 466.

you say you are able to clearly pinpoint sources of meaning in your life, or would you say your thoughts on this topic are still evolving?". As shown in Panel B of Table II, 52% of respondents in the LLM arm respond that they can clearly pinpoint sources of meaning, compared to 41% in the open text field arm. A larger fraction of respondents in the open text field arm answers that their thoughts are still evolving (42% compared to 34%). Given that participants were allocated at random across the two arms, these results show that there is a causal effect of the LLM interview on people's clarity of thoughts.

Finally, we count the number of words written by the respondents in the two treatment arms. We find that people who answer the AI-led interview write 142% more words, consistent with the evidence from prior questions that respondents found the process less daunting when the interview was guided by an AI. As reported in Online Appendix Table A1, the increase in the number of words is large for all socio-demographic groups, with a larger increase for women, Biden voters, higher-income respondents, and middle-aged respondents.

Together, these results show that the AI-led interviewed performed very well, both in terms or process and content. A vast majority of participants enjoyed the interaction with the AI, prefer this mode of interview over open text fields, find that it captures very well what gives them a sense of meaning, are more confident about their responses than when answering open text fields, and write considerably more words than with open text fields.

Results. We now turn to the analysis of the AI-led interview transcripts. When reading the transcripts ourselves, we were struck by the level of detail of the texts and engagement of the respondents. Given the large number of transcripts, we leverage quantitative text analysis to isolate the main themes in the interviews and we study their heterogeneity across socio-demographic groups.

We identify in a data-driven way (i) the main activities that people view as a major source of meaning in life, and (ii) the main subjective states they associate with meaning in life.[28] Reading many transcripts ourselves and analyzing them with large language models as described in Section 2.3.1, we draw a list of twelve activities and twelve subjective states that frequently appear in transcripts. We then systematize the detection of these topics using a large language model to measure their frequency in the full sample of respondents and across groups, as described in Section 2.3.2.[29]

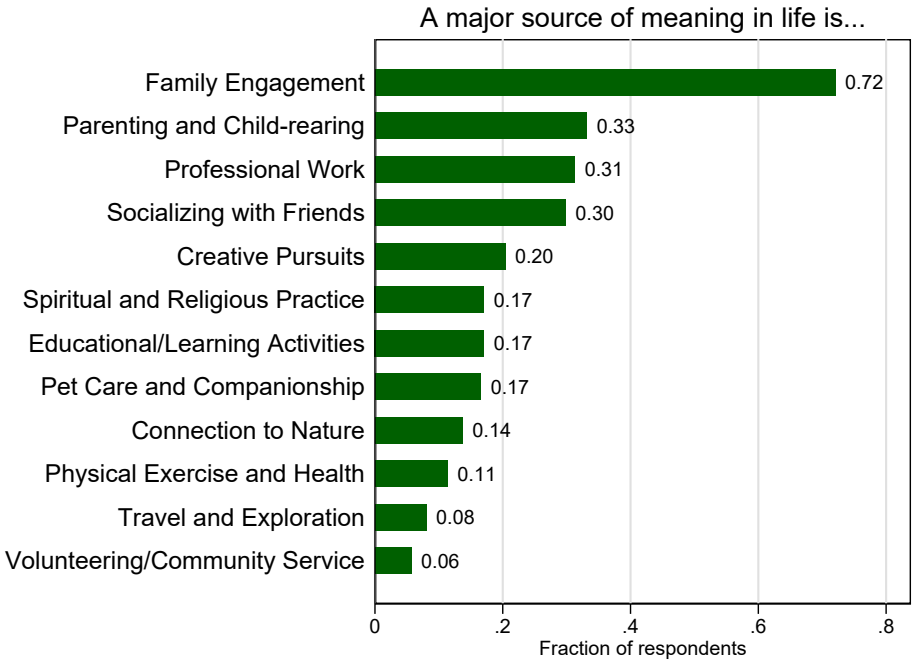
Figure 2 reports the patterns for the full sample. We find that the transcripts convey a rich picture of people's sense of meaning, including several activities and subjective states that would have been difficult to anticipate – i.e., it seems it would have been difficult to design a close-ended survey with appropriate categories, as we discuss with examples below. Panel A reports the frequency of twelve activities people view as sources of meaning in life. Family engagement appears far above any other category: it is mentioned by 72% of all respondents. It is followed

[28] The distinction between activities and subjective states appears to be useful descriptively. For instance, the same subjective state could be associated with different activities, e.g. one may feel like they are "contributing to something greater" through child-rearing or through their work.

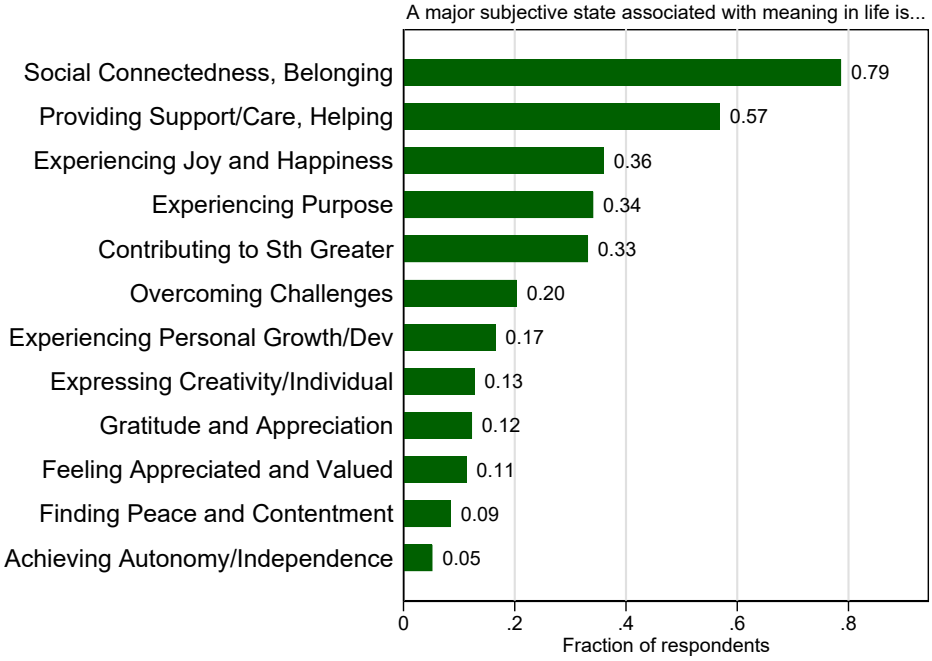
[29] To assess the quality of the topic detection by large language models, we compare the results to those obtained by two research assistants coding the transcripts independently.

Figure 2 Activities and Subjective States Associated with Meaning in Life

Panel A: Activities



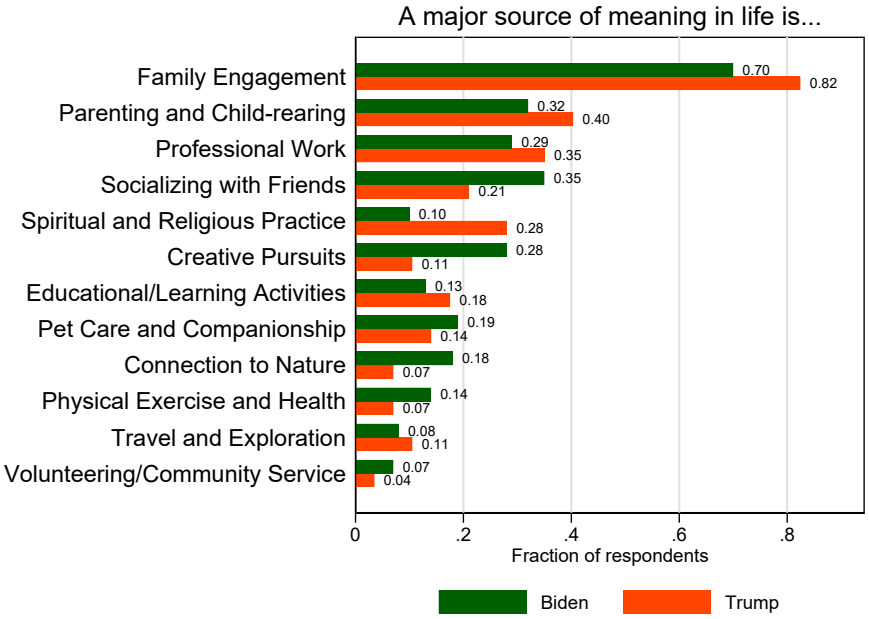
Panel B: Subjective States



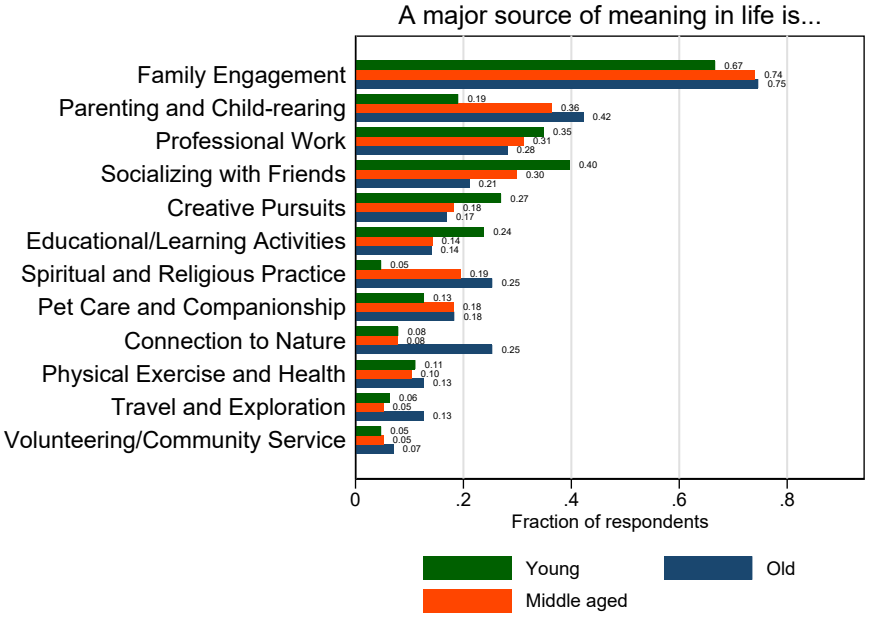
Notes: This figure reports the frequency at which respondents who took part in the AI-led interview mention various activities (Panel A) and subjective states (Panel B) they associate with meaning in life. The transcripts are identified in each interview using a large language model. The number of respondents is 211.

Figure 3 Heterogeneity in Activities Associated with Meaning in Life

Panel A: By Political Affiliation



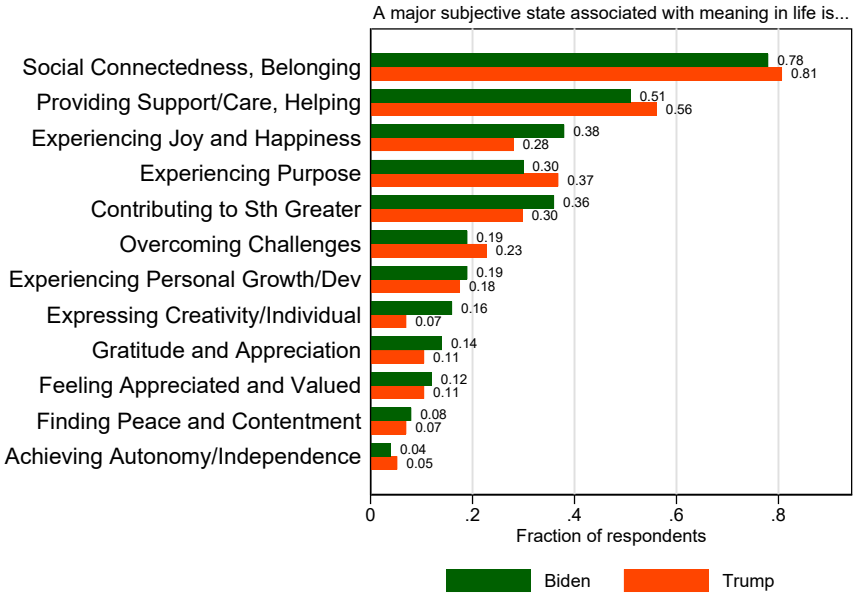
Panel B: By Age



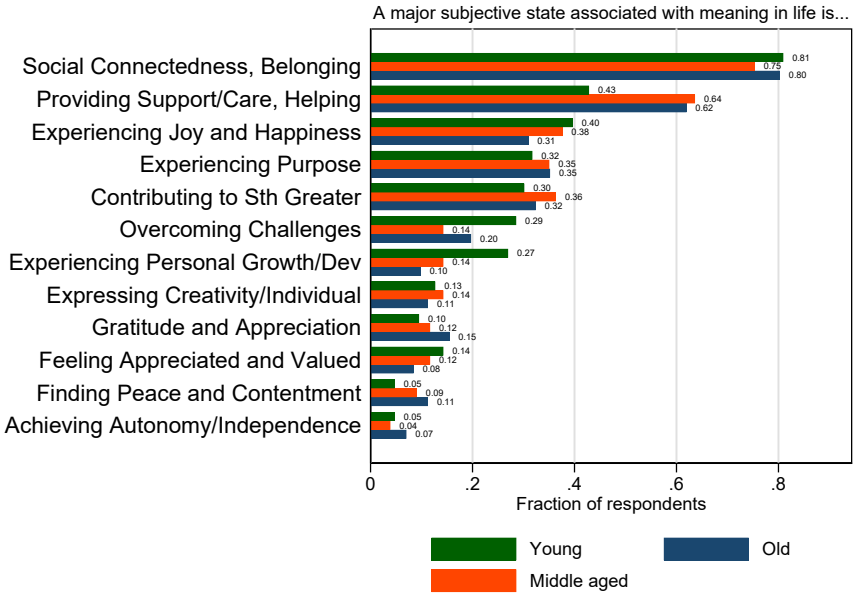
Notes: This figure reports the frequency at which respondents who took part in the AI-led interview mention various activities they associate with meaning in life. Panel A documents heterogeneity depending on political affiliation, i.e. whether the respondent voted for Trump or Biden in the 2020 election. Respondents who did not vote or voted for a third-party candidate are excluded from this analysis. Panel B documents heterogeneity by age group, considering in turn people below 35 ("Young"), between 35 and 55 ("Middle-aged") and above 55 ("Old"). The topics are identified in each transcript using a large language model.

Figure 4 Heterogeneity in Subjective States Associated with Meaning in Life

Panel A: By Political Affiliation



Panel B: By Age



Notes: This figure reports the frequency at which respondents who took part in the AI-led interview mention various subjective states they associate with meaning in life. Panel A documents heterogeneity depending on political affiliation, i.e. whether the respondent voted for Trump or Biden in the 2020 election. Respondents who did not vote or voted for a third-party candidate are excluded from this analysis. Panel B documents heterogeneity by age group, considering in turn people below 35 ("Young"), between 35 and 55 ("Middle-aged") and above 55 ("Old"). The topics are identified in each transcript using a large language model.

by three other categories that are each mentioned by almost a third of respondents: parenting and child rearing,[30] professional work, and socializing with friends. The other eight categories are less frequent. In particular, religion is mentioned by only 17% of respondents. Perhaps surprisingly, pet care and companionship is mentioned as a source of meaning in life with the same frequency as religion. This result illustrates the usefulness of drawing the list of activities in a data-driven way, based on the richness of transcripts. This approach help uncover important categories, such as pet care and companionship, which might not have been included in a traditional close-ended survey.

Panel B of Figure 2 documents the subjective states that people associate with meaning in life. Experiencing a sense of belonging is by far the most common category, mentioned in almost 80% of transcripts. It is followed by "providing care, helping others", at almost 60%. Experiencing joy, experiencing purpose, and feeling that one is "contributing to something greater" are each mentioned by about a third of the respondents. Thus, social connections appear to play the dominant role, especially through "giving"/"contributing" – rather than "feeling appreciated and valued", a category mentioned by only 11% of respondents. The panel also shows that 20% of respondents find meaning in "overcoming challenges." As previously, these results illustrate the richness of the transcripts: the respondents share many important feelings, which would have been challenging to anticipate and code ahead of time with a traditional close-ended survey.

Next, we analyze heterogeneity in responses across groups. We focus on heterogeneity by political affiliation and by age in the main text, reporting heterogeneity by income and gender in the Online Appendix. We find substantial heterogeneity across groups, especially by political affiliation and age. These heterogeneity results illustrate the value of running qualitative interviews at scale – having many respondents allows us to uncover systematic differences across groups, which would require a prohibitive cost with traditional qualitative interview approaches that do not rely on AI.

Figure 3 reports heterogeneity patterns for the activities associated with meaning in life. Panel A reports heterogeneity by political affiliation. Compared to Biden voters, Trump voters are significantly more focused on family engagement, parenting, and work. The biggest difference arises for spiritual and religious practices, which Trump voters are three times as likely to mention (28% vs. 10%). Biden voters are much more likely to mention socializing with friends (35% vs. 21% for Trump voters) and they are more than twice as likely to mention creative pursuits (28% vs. 11%), connection to nature (18% vs. 7%), and exercise/health (14% vs. 7%).

Differences are also substantial across age groups, as reported in Panel B of Figure 3. We compare three groups: respondents below 35, between 35 and 55, and above 55. As people age, they become significantly more likely to mention "parenting and child rearing" (19% below 35, 42% above 55). This result shows that parenting is viewed as an important source of meaning even for older people, whose children are themselves older, i.e. it is not confined to parenting in early childhood. Religion also show a steep age gradient: it is mentioned by only 5% of respondents below 35, but by 25% of respondents above 55. Connection to nature and travel and exploration are also mentioned much more frequently as people age. At the same time, older people are significantly

[30] While parenting and child rearing is a subset of the broader "family engagement" category, we report its frequency separately because it is particularly common.

less likely to mention socializing with friends (40% for those below 35, 20% above 55), creative pursuits, and learning activities; they are somewhat less likely to mention work. The panel shows that most of the patterns are monotonic in age. Overall, age appears to be a very important source of heterogeneity in the activities that people find meaningful. Due to aging in most countries worldwide, these patterns suggest we can expect societal shifts in the types of activities people find most meaningful going forward.

In the Online Appendix, Figure A1 reports additional heterogeneity patterns by gender and income. Women are significantly more likely to mention parenting and child-rearing activities (at 39%, compared to 28% for men) and pets (23% vs. 10%); they are equally likely to mention work. Furthermore, higher-income people are significantly more likely to mention work (42% for those earning above $100k vs. 23% for those below $30k), travel and exploration (19% vs. 2%), and volunteering (11% vs. 2%). They are significantly less likely to mention spiritual and religious practice (7% vs. 21%), pets (5% vs. 26%), and creative pursuits (16% vs. 33%). We observe non-monotonic relationships with respect to income for the activities "parenting and child-rearing" and "socializing with friends", which are both slightly higher for the middle class.

Next, we turn to heterogeneity in the subjective states associated with meaning in life across groups. The heterogeneity by political affiliation, reported in Panel A of Figure 4, is not as large as the heterogeneity previously discussed for activities. But there are sizable differences: Biden voters are much more likely to mention "expressing creativity/individuality" (at 16%, compared to 7% for Trump voters) and slightly more likely to mention joy and happiness, or contributing to something greater. Biden voter are slightly less likely to mention "providing support/helping others", or "experiencing purpose."

Panel B shows that there is substantial heterogeneity for subjective states across age groups. As people age, they become significantly more likely mention "providing care/support" (62% above 55 vs. 43% below 35), feeling gratitude and appreciation (15% vs. 10%), and finding peace and contentment (11% vs. 5%). At the same time, they are significantly less likely to mention "experiencing personal growth and development" (10% above 55 vs 27% below 35) and "feeling appreciated and valued" (8% vs. 14%). Social connectedness and experiencing a sense of belonging remain as important as people age. These heterogeneity results are consistent with the view that older people lead a more settled life, with less personal growth and less need for external validation/appreciation, which is more oriented toward helping others.

Figure A2 in the Online Appendix reports further heterogeneity patterns for subjective states, by gender and income. Women are slightly more likely to mention "providing support/helping others" (59% vs. 54%) and significantly more likely to mention experience joy, happiness, gratitude and appreciation. The heterogeneity across income groups is larger than by gender. We find that people with higher incomes are much more likely to mention "contributing to something greater" and "experiencing personal growth and development", and much less likely to mention joy and happiness or expressing creativity/individuality.

Takeaways. Overall, these results show that AI-led interviews can perform very well for highly complex topic like eliciting views about meaning in life. Despite the simplicity of our prompt, most respondents found the conversation with the AI natural and helpful to guide them through this complex topic. As a result, they conveyed more information than with standard open text fields. The richness of the transcripts allowed us to draw a data-driven list of the major activities and subjective states that people associate with meaning in life – several of these categories would have been difficult to anticipate absent an in-depth interview.[31] Finally, the large sample size allowed us to document large heterogeneity patterns across socio-demographic groups and by political preferences. Together, the results paint a rich picture of conceptions of meaning and its heterogeneity, which could help better understand the welfare and distributional effects of certain policies.

More broadly, this application illustrates that AI-led interviews can be used to elicit the perspectives that agents have on their individual experience. In that sense, it relates to a tradition in social sciences emphasizing the important of understanding people from their own subjective point of view (e.g., Dilthey [1884]), as opposed to another tradition emphasizing a third-person perspective of explanation, in which human agency, subjectivity, and its products are analyzed as effects of impersonal natural, economic or sociological forces.

3.2 AI-led Interviews and Political Views: Evidence from France's Snap Legislative Elections

Motivation. We now examine the ability of the AI-led interviews to elicit people's political preferences and, more specifically, the main reasons driving voting choices and political participation. Using AI to elicit people's political views may be of particular interest in polarized elections, when certain voters may prefer to share their views with an AI, perceived as a non-judgmental entity, rather than a human expert.

We investigate this idea in the run-up to the 2024 French legislative election. This election came as a complete surprise: on June 9, 2024, the President of France decided to dissolve the country's lower chamber of parliament, the National Assembly, and called for snap elections, with the first round schedule on June 30 and the second round on July 7.[32] The three-week campaign was highly polarized, in particular because several polls suggested that the far right could obtain a majority at the National Assembly, for the first time in the history of France's current institutional regime, the Fifth Republic. Many observed considered this episode to be the biggest political crisis in France in the past 70 years. We can therefore study a period where political discourse was highly salient and use qualitative interviews at a large scale to understand the motivations of voters and their heterogeneity across parties.

[31] Online Appendix Figure A3 reports the activities and subjective states mentioned by the respondents answering open text fields. The patterns are slightly different from those reported in Figure 2 for the sample of respondents participating in AI-led interviews. For instance, professional work is mentioned as a major source of meaning in life in 31% of the AI-led interview transcripts, compared to 22% of the open text field answers.

[32] The decision to dissolve the National Assembly was motivated by the results of the 2024 European Parliament election in France. The far-right party, the National Rally (RN), made substantial gains while the President's party lost a significant number of seats in that election.

A longstanding literature has examined the factors that drive voting decisions and political participation, including the role of education (e.g., Willeck and Mendelberg [2022]), moral values (e.g., Enke [2020]), social media (e.g., Allcott and Gentzkow [2017]) or social pressure (e.g., Gerber et al. [2008], Amat et al. [2020]), with a particular interest in populism (e.g., Vachudova [2021], Guriev and Papaioannou [2022]). The literature has notably examined the role of policy-based voting versus partisan attachment to understand to what extent policy positions matter in voters' decisions (e.g., Bullock and Lenz [2019], Schonfeld and Winter-Levy [2021], Dias and Lelkes [2022]). Another strand of the literature focused on examining whether candidates' traits or positions matter most in voters' electoral choice (e.g., Buttice and Stone [2012], Clark and Leiter [2014], Joesten and Stone [2014]).[33]

To the best of our knowledge, there is little evidence to date on voters' actual reflections on candidates and policy positions during high-stakes elections, because there was no tool to conduct in-depth interviews at a large scale during major elections. We leverage our AI-led interviews to shed new light on these questions, illustrating that our simple tool can be deployed very fast to investigate changes in the political environment in real time. As discussed further below, we examine a variety of motives that could drive people's electoral choices – in terms of policies, candidates, and ideological alignment. Finally, this application is also useful to test the capabilities of interviews in French.

Sample and prompt. We recruit a sample of 422 respondents on the Prolific platform, in the last week of June 2024, i.e. a few days before the first round of the snap legislative elections. Our sample deserves special discussion, as Prolific does not provide representative samples for France (as opposed to the U.S., which we leveraged in Section 3.1). We sample the French respondents who were active on the Prolific platform during the study period, with no particular filter. Our sample is younger and has lower income than average in the population. To assess how our sample compares to the full population in terms of political preferences, we can compare the voting intentions reported in our survey to the actual election results. The respondents appear significantly less likely to abstain.[34] Moreover, conditional on voting, they are much more likely to report planning to vote for the left (66% in the sample compared to 28% in the population on election day). Conversely, they are significantly less likely to report planning to vote for the far right (14% rather than 29%) or the center (8% rather than 20%). In much of the analysis below, we carry out the analysis by political affiliation, such that the patterns we report are not skewed by the imbalance of our sample relative to the population in terms of political preferences.

As previously, our prompt uses a structure similar to Section 2.1, except that is is written in French. Translating from French, the AI is told to play the following role:

> *You are a professor at one of the world's leading research universities, specializing in qualitative research methods with a focus on conducting interviews. In the following,*

[33] Part of this literature uses conjoint experiments with fictional candidates (e.g., Franchino and Zucchini [2015], Hansen and Treul [2021]).

[34] 9% of respondents mention they plan to abstain, compared to an overall abstention rate of 33% on election day.

you will conduct an interview with a human respondent to find out the participant's motivations and reasoning regarding their voting choice during the legislative elections on June 30, 2024, in France, a few days after the interview.

The interview outline is then organized into four parts. The first part is the most important and the prompt reads as follows:

This part is the core of the interview. With around 20 questions, please explore the different dimensions of the two following topics:

(i) 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. Assess the individual's perception of the realism of their preferred party's platform: would the new public policies proposed by this party actually be implemented if it came to power?

(ii) The individual's perception of voters from other parties; in particular, are they considered reasonable people with whom one can debate? Why do they think other people hold different opinions?

Please identify the underlying factors that contribute to the participant's views.

The LLM is also instructed to ask for a grade indicating how well the interview so far summarized the respondent's views on the upcoming elections.

The next three parts of the interview are shorter. In Part 2, the chatbot asks the respondent up to five questions regarding whether anything could lead them to change their views before the election, and if so how and why. In Part 3, it asks up to five questions about the three main changes the respondents would like a politician to implement in the country. Finally, Part 4 asks four short questions in turn to get insights about how they think political leaders have changed their lives and the country in recent years, as well as for the traits they seek in a leader.[35]

Thus, the interview outline elicits the political views of the respondent in a series of structured steps. The full text of the prompt is reported in Online Appendix A.2.

Quality metrics. We first consider two simple metrics of the quality of the interview, reported in Table III.

After the end of the interview with the LLM, we ask the respondents whether they would prefer to take a similar interview with an AI or a human in the future, or whether they do not mind. 49% of respondents respond that they would prefer to take the interview again with an AI. They mention

[35]The LLM asks four questions in turn, as follows: (i) How do you think political leaders have changed your life over the past seven years?; (ii) How do you think political leaders have changed the country over the past seven years?; (iii) What are the three main character traits you look for in a political leader?; (iv) Is it important for a party to have experience, or is it more interesting to elect a party that has never been in power before? Why?

Table III Quality Metrics for the AI-led Interview on Electoral Choices

	Fraction of Respondents
In the future, would you rather take the interview with	
... An AI	49%
... A human	15%
... I do not mind	36%
How well does it summarize your views?	3.31
1 ("poorly") to 4 ("very well")	(0.705)

Notes: This table reports two measures of perceived quality for the AI-led interview on electoral choices in the 2024 French legislative election.

several reasons, highlighting that the interview process was impartial ("the algorithm does not judge me and allows for an in-depth conversation"; "it is easier, there is no judgment or conflict"). 15% of respondents indicate they would prefer a human ("it is more personal"), and 36% express no preference.

Furthermore, respondents are instructed to give a grade from one to four to characterize how well the interview captured their views.[36] Table II shows that the grades are very high, with an average score of 3.31. Overall, these results suggest that the participants were pleased with their experience with the LLM.

Results. Analyzing the interview transcripts, we first examine how frequently three broad classes of reasons to support a party are mentioned by the respondents. Specifically, we analyze whether the respondent was drawn by specific policies (e.g., increasing the minimum wage), by the general philosophy of the party (e.g., being in agreement with the ideology of the party) or by party leaders. We conduct this analysis and report results separately for respondents from the left (voting for the *Nouveau Front Populaire*), the center (voting for one of the following parties: *Ensemble, Renaissance, Modem, Horizon*), and the far right (voting for the *Rassemblement National*).

As shown in Table IV, specific policies are the dominant type of reasons mentioned by respondents. This result holds for respondents of all political affiliations, but it is more pronounced for the left and especially for the far right. Perhaps surprisingly, party leaders are very rarely mentioned. Only 2% of the respondents on the left mention their party leaders, and none of the far-right voters do. About 22% of voters in the center mention party leaders as a key reason for support.

[36] The coding scheme is as follows: 1: "it describes my reasoning and preferences for the upcoming election relatively poorly"; 2: "it describes my reasoning and preferences for the upcoming election adequately"; 3: "it describes my reasoning and preferences for the upcoming election well"; 4: "it describes my reasoning and preferences for the upcoming election very well".

Table IV Main Broad Reasons for Political Choice, %

	Left	Center	Far Right
Specific policies proposed by the party	69.7	54.8	92.2
General philosophy of the party	28.2	22.6	7.8
Party leaders	2.1	22.6	0
N	241	31	51

Notes: For each respondent, the dominant class of reasons to support the party is selected out of three options – specific policies proposed by the party, general philosophy of the party, or party leaders – using a large language model. The sample is restricted to voters who do not plan to abstain. The sample is also limited to transcripts where a single dominant class is identified. The set of reasons provided by the respondent does not map to a single dominant class for 2% of respondents supporting the left, 11% of respondents supporting the center, and 4% of respondents supporting the far right, who are excluded from the analysis.

Next, we investigate in greater detail the specific arguments advanced by the respondents to justify their choice. Reading many transcripts ourselves and analyzing them with large language models as described in Section 2.3.1, we draw a list of the most common arguments mentioned by recipients. We then systematize the detection of these topics using a large language model to measure their frequency by political preference, as described in Section 2.3.2. The results are reported in Table V for the left, Table VI for the center, Table VII for the far right, and Table VIII for those who plan to abstain. We report all arguments that appear in more than 10% of transcripts.

Table V Top Reasons to Vote for the Left

Reason	% of transcripts
Rejecting other parties	74.8
Promoting the ecological transition, protecting the environment and limiting global warming	43.1
Increasing the minimum wage	40.7
Increasing taxes on firms' excess profits, the rich and the wealthy	38.2
Protecting minority rights (LGBTQIA+) and promoting gender equality	29.3
Improving public services (education, healthcare)	27.2
Pension reform	23.2
Reducing economic and social inequalities	21.5
Promoting humanistic and inclusive public policies	19.1
Increasing social assistance	11.4
Reducing corruption and tax fraud	10.6
Price controls for essential goods	10.2

Notes: This table reports all reasons mentioned in more than 10% of the transcripts of respondents planning to vote for the left. The fraction of transcripts mentioning each reason is reported as a percentage.

Table VI Top Reasons to Vote for the Center

Reason	% of transcripts
Rejecting other parties	77.1
Ensuring the continuity of ongoing policies	48.6
Promoting a pro-European approach	25.7
Pension reform	22.9
Supporting Ukraine	14.3
Ensuring economic stability	11.4

Notes: This table reports all reasons mentioned in more than 10% of the transcripts of respondents planning to vote for the center. The fraction of transcripts mentioning each reason is reported as a percentage.

Table VII Top Reasons to Vote for the Far Right

Reason	% of transcripts
Reducing legal or illegal immigration	66
Rejecting other parties	64.2
Reducing insecurity and crime	64.2
Promoting public policies that favor French citizens over foreigners	32.1
Expelling foreign criminals	30.2
Reducing social benefits for foreigners	20.8
Promoting economic protectionism and supporting local businesses	18.9
Improving purchasing power	18.9
Combating Islamic fundamentalism and radicalization	11.3

Notes: This table reports all reasons mentioned in more than 10% of the transcripts of respondents planning to vote for the far right.

Table VIII Top Reasons to Abstain

Reason	% of transcripts
Abstained due to lack of satisfactory candidates or parties	60.5
Abstained due to personal priorities or a lack of interest in politics	28.9
Abstained because of extreme polarization and lack of constructive dialogue	26.3
Abstained due to the perception that voting makes no difference	18.4
Abstained due to distrust of the media and biased information	18.4

Notes: This table reports the five main reasons mentioned in the transcripts of respondents planning to abstain.

These tables reveal a striking level of polarization, with two important features. First, "rejecting other parties" is the top category for respondents on the left (mentioned in 75% of transcripts) and in the center (77%), and the second most frequent category among far-right respondents (64%). Second, the policy priorities mentioned by respondents with different political preferences are almost completely non-overlapping. Respondents on the left are driven by the green transition and reducing inequality and poverty through various policies (for instance, 40% of respondents mention the

Table IX Top 5 Traits Desired in a Leader by Political Affiliation

Left		Center		Far Right		Abstainers	
Trait	% of transcripts	Trait	% of transcripts	Trait	% of transcripts	Trait	% of transcripts
Honesty	41.8	Honesty	42.8	Honesty	50.9	Honesty	39.4
Listening	22.3	Charisma	17.1	Firmness	18.8	Firmness	18.4
Humanity	12.1	Firmness	11.2	Charisma	11.3	Sincerity	15.8
Integrity	11.3	Intelligence	11.4	Listening	11.3	Charisma	13.1
Charisma	10.9	Courage	11.4	Integrity	9.43	Intelligence	13.1

Notes: This table reports the top 5 desired traits in a leader by political affiliation. The fraction of transcripts mentioning each quality is reported as a percentage.

increase in the minimum wage). In contrast, respondents in the center highlight the importance of ensuring the continuity of ongoing policies, i.e. preserving the agenda and legacy of the President. Finally, far right voters highlight immigration, crime and policies favoring French citizens over foreigners as their key reasons for support. Thus, there appears to be a very strong polarization of ideas: voters on the left, in the center, and on the far right identify completely different policy issues and solutions. In Table VIII, respondents who plan to abstain primarily mention that there is no satisfactory candidate in their view; a smaller fraction reports a lack of interest in politics.[37]

Finally, although party leaders are rarely mentioned as key motives by our respondents (Table IV), it is instructive to see which traits voters seek in leaders, using the question posed by the LLM toward the end of the interview. Table A9 reports the five most common traits reported to be desired in a leader, by political affiliation.[38] Two findings stand out. First, the most common trait is "honesty", which is the top category for all groups, including abstainers. Second, there are important differences between parties. While respondents on the left highlight the ability to listen and to show humanity as key traits, respondents in the center mention "firmness" and "intelligence." Far right voters mention both "firmness" and "listening" in their key categories.

Takeaways. Overall, the AI-led interviews on electoral choice during France's snap legislative elections highlight that we can deploy the tool very quickly, in a different language, and that participants seem very comfortable sharing their views on sensitive political questions with an AI. The respondents rate highly the content of the interview. The analysis of the transcripts reveals several interesting patterns about the nature of polarization: voters have drastically different views on policy issues and priorities. They are drawn by specific policies as answers to specific issues they view as important, rather than by the general ideology of their preferred party.[39]

[37] Online Appendix Table A2 reports a longer list of arguments across all respondents.
[38] Online Appendix Table A10 reports the full list.
[39] For completeness, Tables A4 to A10 in the Online Appendix report the results with the words used by the respondents, in French.

3.3 AI-led Interviews and Decision Making: Understanding Educational and Occupational Choices

Motivation. We now examine the performance of our approach to elicit the most important factors driving decision making for two highly consequential choices, education and occupation. Understanding the factors driving education and occupational choices is a longstanding question in the economics and sociology literatures. The literature has highlighted the importance of financial factors (e.g. Roy [1951], Rothstein and Rouse [2011], Arcidiacono et al. [2020]), social norms (e.g. Goldin [2006], Bursztyn et al. [2017]), peer effects and exposure to role models (Jensen [2010], Breda et al. [2023]), amenities such as hours flexibility (e.g., Goldin and Katz [2008], Wasserman [2023]), and beliefs (e.g., Hoxby and Turner [2015], Mulhern [2023]). Understanding the factors driving allocation of talent to STEM fields and innovation careers has been of particular interest recently, a women and minorities are widely under-represented in these fields, with potentially important implications for growth, innovation, and inequality (Bell et al. [2019], Hsieh et al. [2019], Einio et al. [2023]). Our AI-led interview methodology provides an opportunity to understand which of these or other factors are most important according to the respondents.

Sample and prompt. We recruit 68 U.S. respondents on the Prolific platform and and invited them to participate in our LLM interview. We design the interview outline to cover in turn educational and occupational choices, with a particular interest in STEM fields:

> *In the interview, please explore why the respondent chose the field/major in their education, and why they chose their subsequent occupation.*
>
> *The interview consists of successive parts that are outlined below. Ask one question at a time and do not number your questions. Begin the interview with: 'Hello! I'm glad to have the opportunity to speak about your educational journey today. Could you share the reasons that made you choose your field of study at the highest level of your education? Please do not hesitate to ask if anything is unclear.'*
>
> *Part I of the interview*
>
> *Ask up to around 15 questions to explore different dimensions and factors that drove the respondent's choice of the field/major at their highest level of education. If they did not choose a major, explore the general reasons for choosing their level of education. If the respondent starts describing job and career choices already here, gently guide the discussion back to exploring educational choices in this part of the interview.*
>
> *When the respondent confirms that all aspects which determined their educational choices have been thoroughly discussed, continue with the next part.*
>
> *Part II of the interview*
>
> *Ask up to around 5 questions to explore why or why not the respondent studied a STEM subject (Science, Technology, Engineering, Mathematics). Begin this part with: 'Next,*

I would like to focus further on why or why not you pursued a STEM subject (Science, Technology, Engineering, or Mathematics) as your major. Could you share the reasons specifically for this decision, either for or against it?'

When the respondent confirms that all their reasons for or against STEM subjects have been thoroughly discussed, continue with the next part.

Part III of the interview

Ask up to around 15 questions to explore different dimensions and factors that drove the respondent's decision for their subsequent occupation and career. Begin this part with: 'Lastly, I would like to shift the focus from education to occupation. Could you share the reasons for choosing your job and professional field following your studies?'

When the respondent confirms that all aspects which determined their occupational choices have been thoroughly discussed, continue with the next part.

Summary and evaluation

To conclude, write a detailed summary of the answers that the respondent gave in this interview. After your summary, add the text: 'To conclude, how well does the summary of our discussion describe your reasons for choosing your education and occupation: 1 (it poorly describes my reasons), 2 (it partially describes my reasons), 3 (it describes my reasons well), 4 (it describes my reasons very well). Please only reply with the associated number.'

After receiving their final evaluation, please end the interview.

The full prompt is provided in Appendix A.3.

Quality metrics. We start by reporting two quality metrics in Table X. The table shows that only 15% of respondents would prefer to take the interview with a human in the future. 35% would prefer an AI and 48% are indifferent. The table also shows that the respondents give excellent scores to the interview summary, with an average grade of 3.8.

Results. The main results are reported in Figure 5. We use a large language model to identify the main factors driving education and occupational choices in the interview transcripts. In Panel A, we focus on educational choices and find that personal interests and passions are the most common factor cited by respondents (81%). For instance, some respondents mention interests they develop through their hobbies, or intrinsic interest for an academic subject. Career and financial prospects appear in second place (47%), followed by influential educators and mentors (44%). 24% of respondents mention family influence and expectation.

Panel B turns to the driving factors for occupational choice, which are quite similar to those mentioned for educational choices. Personal interests and passions are mentioned slightly less frequently than for educational choices (now 71%), while financial incentives are mentioned more often

Table X Quality Metrics for the AI-led Interview on Electoral Choices

	Fraction of Respondents
In the future, would you rather take the interview with	
... An AI	35.90%
... A human	15.38%
... I do not mind	48.72%
How well does it summarize your reasons?	3.81
1 ("poorly") to 4 ("very well")	(0.36)

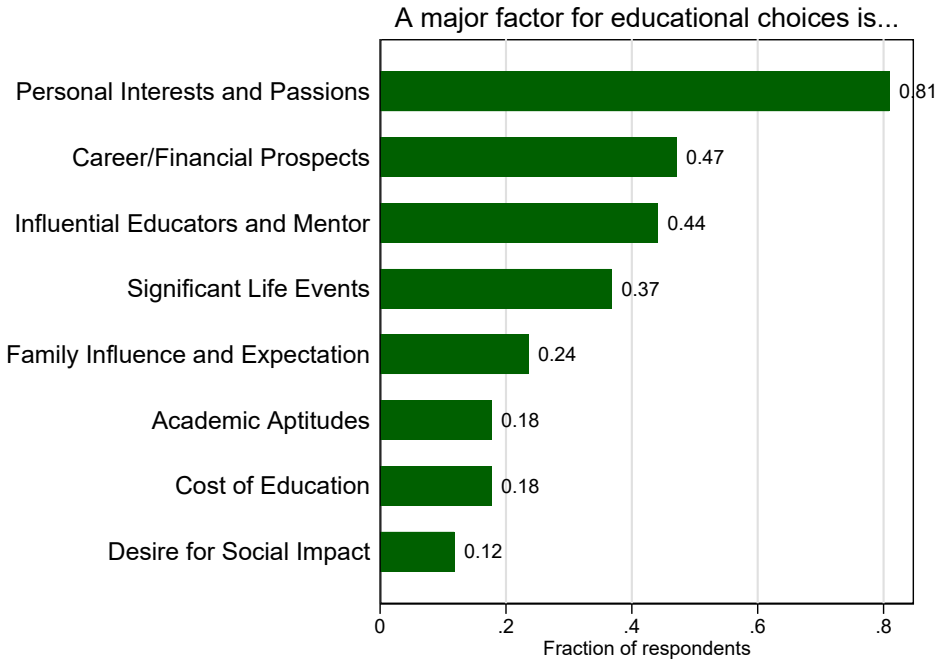
Notes: This table reports two measures of perceived quality for the AI-led interview on educational and occupational choices.

(57%). The role of educators and mentors remains high (40%), and so does the role of the family (24%). 21% of respondents mention autonomy and flexibility as an important factors that led them down their career path.

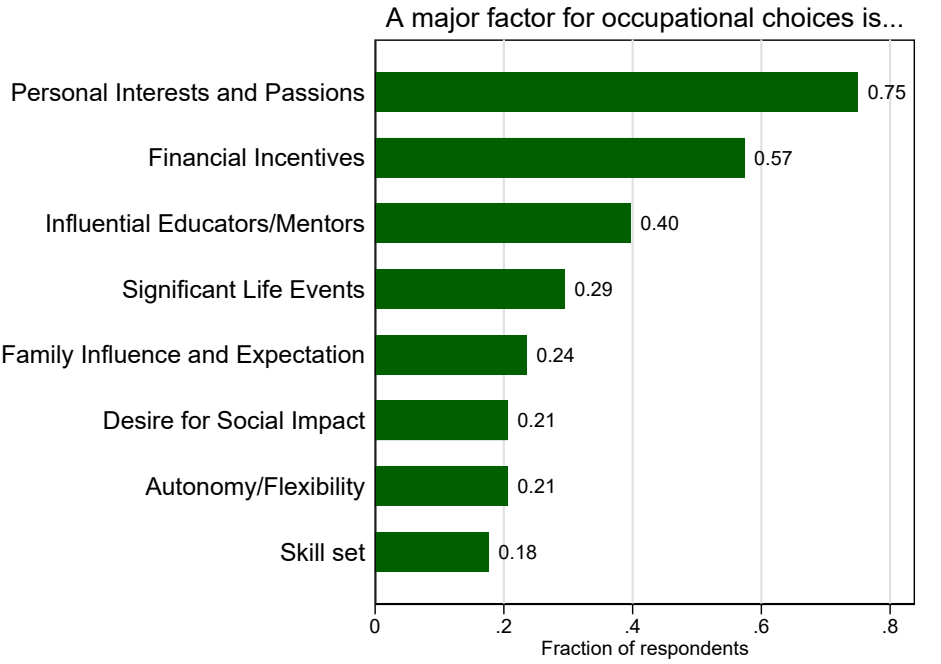
Takeaways. Overall, this section illustrates that AI-led qualitative interviews provide a simple and effective way to identify key factors people believe were crucial for some of their most important decisions, here educational and occupational choices. Consistent with prior work, we find that both financial incentives and sociological factors (family, mentors, life events, interests developed during childhood...) matter, with an outsized role for interests and passions.

Figure 5 Major Factors for Educational and Occupational Choices

Panel A: Educational Choices



Panel B: Occupational Choices



Notes: This figure reports the frequency at which respondents who took part in the AI-led interview mention various factors as drivers of their educational choices (Panel A) or occupational choices (Panel B) they associate with meaning in life. The factors are identified in each interview using a large language model. The number of respondents is 68.

4 Limitations and Extensions

In this section, we discuss various limitations and extensions of our analysis.

Interview platform. Several limitation of our interview platform should be noted. First, the current chat interface features written, rather than spoken conversations. To address this limitation going forward, we plan to add an option for voice to the platform. Second, while allowing to run thousands of interviews within a few hours, our AI-led interviews are shorter (about 30 minutes) than human interviews common in fields such as sociology or anthropology.Another series of limitations pertains to our system prompt and the LLM we use to conduct the interview. We developed the general instructions based on best practices in the sociology literature and feedback from academics in the field. These instructions achieved a performance close to an average hypothetical human expert across a wide range of interview topics. However, alternative versions of the instructions are likely to yield even better behavior going forward, for two reasons.

First, better results may be obtained as more capable LLMs become available. For instance, the current models struggle to always restrict their message to contain only a single question. As the underlying LLMs change, interviews will also change. In this respect, it is encouraging that LLM capabilities have improved over the last years for many evaluation metrics. A new generation of models could again be evaluated by human experts across a wide range of interview topics as done in this paper for the GPT4o model.

Second, the general instructions in the system prompt could be amended to obtain better results. In some cases, the "best" general instructions may differ depending on individual researchers' preferences. For instance, we observed in part of the evaluations that the experts gave higher marks for a version with minimal general instructions as the resulting interview seemed sometimes less prescriptive. Yet, in our studies, we valued the follow-up questions asking for concrete examples, which our preferred general instructions yielded. The open-source nature of our platform makes it possible for many researchers to customize the prompts as needed, depending on their own preferences. In particular, researchers could edit the prompts so that the LLM conducts interviews in line with different schools of thought, e.g. focusing on narratives rather than on concrete examples.

Several avenues can be investigated going forward to make further improvements to the system prompt. To allow for automated improvements of the prompts and the resulting LLM interviews, it could be fruitful to develop a testing platform in which different LLM agents interact as interviewer and interviewee. Automated improvement of prompts is a very active area of research more generally, as discussed in Fernando et al. [2023]. Additional improvements might be achieved by fine-tuning the LLM on on expert-written prompt and answer pairs.

Two additional limitations of our interview platform are worth noting. First, proprietary LLMs such as the ones used in this paper require study participants to agree that data will be processed by the firms owning the models, which may raise privacy issues for some respondents. As small local language models become increasingly capable, it will likely be possible to conduct high-quality interviews with them, alleviating potential privacy concerns. Second, we have focused on expert

ratings of the AI interviewer in comparison to a hypothetical human expert. Further research could compare the *answers* that participants give to human interviewers relative to AI interviewers.

Finally, note that a growing number of commercial private sector platforms offer qualitative interviewing capabilities. By construction, a share of their features and exact workings will remain private. For academic research, simple platforms like the one shared alongside this paper have the distinct advantage to give researchers control over the entire prompt and the interface. The platform's code can also flexibly be adjusted to work with the increasingly capable local models, making all components non-proprietary.

Sampling. Like many other studies, we rely on surveying firms such as Prolific to recruit respondents. Through different tests, we see that a subset of respondents may be using LLMs to answer the survey. We try to remove such cases from the dataset, but with growing capabilities and multimodality (i.e. allowing text, images, sounds, etc. as part of prompts and answers), AI generated answers become increasingly difficult to detect. These issues can make it particularly important to recruit surveys participants without monetary incentives to return answers quickly in future studies. Related to this, widely discussed selection issues make it difficult to obtain representative samples from surveying firms, emphasizing the need to address potential selection bias (e.g., Dutz et al. [2021]). Furthermore, we find that some participants do not attempt to respond seriously, but only aim to finish the study quickly. This can, however, be easier to see in interview transcripts than in open-text fields, in particular if the LLM is prompted to ask many probing questions like in our setup.

Comparison to open text fields. Our data offer many possibilities to draw comparisons between the information obtained with interviews vs open-ended text fields. We documented a large increase in textual content generated through interviews, that participants stated they preferred to participate in interviews rather than answering text fields, and that they gave on average higher scores to interviews transcripts representing their opinion than to their own written responses in open text-fields. Further study of differences in responses from interviews vs. open-ended survey questions is an important task for future work, for instance to document the extent to which interview transcripts provides a more complete account for respondents' preferences and reasoning.

Tools for the analysis of transcripts. The approach developed in Section 2.3.1 proved fruitful and allowed us to gain an overview of very rich textual data within seconds, as well as to generate hypotheses. Yet, it adds an additional layer of approximation and randomness through the underlying language models. To partially alleviate this concern, we measured concepts more precisely in a second step, by iterating over each full transcript individually with another LLM instance, as described in Section 2.3.

Some of our findings using this second step have so far been obtained by asking the LLM to detect in the transcript a given concept from a list specified in the prompt. It can, however, be more robust to ask the LLM only for binary answers (e.g., "Is concept XYZ contained in this transcript?

Please answer with yes or no and a short explanation"), repeating this for all concepts separately. We plan to compare these approaches going forward.

For replicability, it would also be instructive to compare the results of the LLM to those obtained with human labelers. Indeed, when using proprietary LLMs, it is very helpful to understand in more detail how well they replicate human labeling decisions, for instance when stating whether or not a certain concept is contained in a transcript. If the similarity is broadly as high between model decisions and human decisions as between different human decisions, then the LLM can be viewed as primarily automating human decisions in such labelling setups. In this case, one can be confident that a future LLM from a different provider would likely yield similar results, as long as it can be shown that this LLM is broadly similar to human labeling decisions.

In the study inquiring into what people find "meaningful" in life, we ran an initial comparison for about twenty transcripts. We obtained correlations between human labelers and language model decisions that were not much smaller on average than between human labelers alone. For some difficult concepts such as the subjective states people associate with a meaningful life, we found relatively low agreement between human labelers.

5 Conclusion

This paper has introduced a flexible, open-source platform designed to conduct qualitative interviews with large language models. By adhering to established best practices in sociology, the platform is adaptable to a variety of interview topics. Our evaluations show that the tool is both reliable and effective, with results comparable to those of hypothetical average human experts and consistent with respondent feedback. We have demonstrated its applicability across diverse areas such as decision-making processes, political opinions, perspectives on the external world, and personal mental states. The platform is user-friendly and easily accessible online, allowing users to initiate AI-led interview studies on virtually any subject within hours.

References

Hunt Allcott and Matthew Gentzkow. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2):211–236, 2017. 3.2

Francesc Amat, Carles Boix, Jordi Munoz, and Toni Rodon. From political mobilization to electoral participation: Turnout in barcelona in the 1930s. *The Journal of Politics*, 82(4):1559–1575, 2020. 3.2

Aaron Antonovsky. The structure and properties of the Sense of Coherence scale. *Social Science & Medicine*, 36(6):725–733, 1993. 3.1

Peter Arcidiacono, V. Joseph Hotz, Arnaud Maurel, and Teresa Romano. Ex Ante Returns and Occupational Choice. *Journal of Political Economy*, 128(12):4475–4522, 2020. 3.3

Alex Bell, 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*, 134(2):647–713, 2019. 3.3

Peter Bergman, 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*, 114(5):1281–1337, 2024. 1

Truman Bewley. Why wages don't fall during a recession. *Harvard University Press*, 1999. 1

Truman Bewley. Interviews as a valid empirical tool in economics. *The Journal of Socio-Economics*, 31(4):343–353, 2002. 1, 1

Thomas Breda, Julien Grenet, Marion Monnet, and Clementine Van Effenterre. How Effective are Female Role Models in Steering Girls Towards STEM? Evidence from French High Schools. *The Economic Journal*, 133(653):1773–1809, 2023. 3.3

John G. Bullock and Gabriel Lenz. Partisan Bias in Surveys. *Annual Review of Political Science*, 22(Volume 22, 2019):325–342, 2019. 3.2

Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais. 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *American Economic Review*, 107(11):3288–3319, 2017. 3.3

Matthew K. Buttice and Walter J. Stone. Candidates Matter: Policy and Quality Differences in Congressional Elections. *The Journal of Politics*, 74(3):870–887, 2012. 3.2

Albert Camus. *An Absurd Reasoning: The Myth of Sisyphus and Other Essays*. Vintage, New York, 1955. 3.1

Felix Chopra and Ingar Haaland. Conducting qualitative interviews with ai. 2023. 1, 8, 22

Michael Clark and Debra Leiter. Does the Ideological Dispersion of Parties Mediate the Electoral Impact of Valence? A Cross-National Study of Party Support in Nine Western European Democracies. *Comparative Political Studies*, 47(2):171–202, 2014. 3.2

James C. Crumbaugh. The Seeking of Noetic Goals Test (SONG): A complementary scale to the Purpose in Life Test (PIL). *Journal of Clinical Psychology*, 33(3):900–907, 1977. 3.1

James C. Crumbaugh and Leonard T. Maholick. An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis. *Journal of Clinical Psychology*, 20(2):200–207, 1964. 3.1

Stefanie DeLuca. Sample selection matters: Moving toward empirically sound qualitative research. *SAGE Publications Sage CA: Los Angeles, CA*, 2023. 1

Nicholas Dias and Yphtach Lelkes. The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity. *American Journal of Political Science*, 66(3):775–790, 2022. 3.2

Wilhelm Dilthey. Ideen über eine bechreibende und zergliedernde psychologie. 1884. 3.1

Ricardo Dominguez-Olmedo, Moritz Hardt, and Celestine Mendler-Dünner. Questioning the survey responses of large language models. *arXiv preprint arXiv:2306.07951*, 2023. 1

Deniz Dutz, Ingrid Huitfeldt, Santiago Lacouture, Magne Mogstad, Alexander Torgovitsky, and Winnie Van Dijk. Selection in surveys: Using randomized incentives to detect and account for nonresponse bias. *National Bureau of Economic Research*, 2021. 1, 4

Elias Einio, Josh Feng, and Xavier Jaravel. Social Push and the Direction of Innovation. 2023. 3.3

Benjamin Enke. Moral Values and Voting. *Journal of Political Economy*, 128(10):3679–3729, 2020. 3.2

Chrisantha Fernando, Dylan Banarse, Henryk Michalewski, Simon Osindero, and Tim Rocktäschel. Promptbreeder: Self-referential self-improvement via prompt evolution. 09 2023. URL https://arxiv.org/pdf/2309.16797.pdf. 4

Fabio Franchino and Francesco Zucchini. Voting in a Multi-dimensional Space: A Conjoint Analysis Employing Valence and Ideology Attributes of Candidates. *Political Science Research and Methods*, 3(2):221–241, 2015. 33

Alan S. Gerber, Donald P. Green, and Christopher W. Larimer. Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *The American Political Science Review*, 102(1): 33–48, 2008. 3.2

Claudia Goldin. The Quiet Revolution That Transformed Women's Employment, Education, and Family. *American Economic Review*, 96(2):1–21, 2006. 3.3

Claudia Goldin and Lawrence F Katz. Transitions: Career and Family Life Cycles of the Educational Elite. *American Economic Review*, 98(2):363–369, 2008. 3.3

Sergei Guriev and Elias Papaioannou. The Political Economy of Populism. *Journal of Economic Literature*, 60(3):753–832, 2022. 3.2

Ingar K Haaland, Christopher Roth, Stefanie Stantcheva, and Johannes Wohlfart. Measuring what is top of mind. *National Bureau of Economic Research*, 2024. 1

Eric R. Hansen and Sarah A. Treul. Inexperienced or anti-establishment? Voter preferences for outsider congressional candidates. *Research & Politics*, 8(3):1–7, 2021. 33

Samantha J. Heintzelman and Laura A. King. Life is pretty meaningful. *The American Psychologist*, 69(6):561–574, 2014. 3.1

John J Horton. Large language models as simulated economic agents: What can we learn from homo silicus? *National Bureau of Economic Research*, 2023. 1

Caroline M. Hoxby and Sarah Turner. What High-Achieving Low-Income Students Know about College. *American Economic Review*, 105(5):514–517, 2015. 3.3

Chang-Tai Hsieh, Erik Hurst, Charles I. Jones, and Peter J. Klenow. The Allocation of Talent and U.S. Economic Growth. *Econometrica*, 87(5):1439–1474, 2019. 3.3

Robert Jensen. The (Perceived) Returns to Education and the Demand for Schooling*. *The Quarterly Journal of Economics*, 125(2):515–548, 2010. 3.3

Danielle A. Joesten and Walter J. Stone. Reassessing Proximity Voting: Expertise, Party, and Choice in Congressional Elections. *The Journal of Politics*, 76(3):740–753, 2014. 3.2

Laura A. King and Joshua A. Hicks. The Science of Meaning in Life. *Annual Review of Psychology*, 72(Volume 72, 2021):561–584, 2021. 3.1

Laura A. King, Joshua A. Hicks, Jennifer L. Krull, and Amber K. Del Gaiso. Positive affect and the experience of meaning in life. *Journal of Personality and Social Psychology*, 90(1):179–196, 2006. 3.1

Frank Martela and Michael F. Steger. The three meanings of meaning in life: Distinguishing coherence, purpose, and significance. *The Journal of Positive Psychology*, 11(5):531–545, 2016. 3.1

Christine Mulhern. Beyond Teachers: Estimating Individual School Counselors' Effects on Educational Attainment. *American Economic Review*, 113(11):2846–2893, 2023. 3.3

Shigehiro Oishi and Ed Diener. Residents of poor nations have a greater sense of meaning in life than residents of wealthy nations. *Psychological Science*, 25(2):422–430, 2014. 3.1

Jesse Rothstein and Cecilia Elena Rouse. Constrained after college: Student loans and early-career occupational choices. *Journal of Public Economics*, 95(1):149–163, 2011. 3.3

A. D. Roy. Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3(2):135–146, 1951. 3.3

Bryan Schonfeld and Sam Winter-Levy. Policy or Partisanship in the United Kingdom? Quasi-Experimental Evidence from Brexit. *The Journal of Politics*, 83(4):1450–1461, 2021. 3.2

Mario Luis Small and Jessica McCrory Calarco. *Qualitative literacy: A guide to evaluating ethnographic and interview research*. Univ of California Press, 2022. 1, 2.1.2, 12, 13, 14

Stefanie Stantcheva. How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15(1):205–234, 2023. 1

Michael F. Steger. *The human quest for meaning: Theories, research, and applications*, chapter Experiencing meaning in life: Optimal functioning at the nexus of well-being, psychopathology, and spirituality, pages 165 – 184. Routledge/Taylor Francis Group, 2nd edition, 2012. 3.1

Michael F. Steger, 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*, 53(1):80–93, 2006. 3.1

Milada Anna Vachudova. Populism, Democracy, and Party System Change in Europe. *Annual Review of Political Science*, 24(Volume 24, 2021):471–498, 2021. 3.2

Melanie Wasserman. Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents. *The Review of Economic Studies*, 90(3):1535–1568, 2023. 3.3

Claire Willeck and Tali Mendelberg. Education and Political Participation. *Annual Review of Political Science*, 25(Volume 25, 2022):89–110, 2022. 3.2

For Online Publication

Appendix to "Conversations at Scale: Robust AI-led Interviews with a Simple Open-Source Platform"

Friedrich Geiecke, *LSE*
Xavier Jaravel, *LSE & CEPR*
October 2024

A Prompt Appendix

In this appendix, we report the full prompts used for our three main applications.

A.1 Measuring Meaning in Life

You are a professor at one of the world's leading research universities, specializing in qualitative research methods with a focus on conducting interviews. In the following, you will conduct an interview with a human respondent to find out what gives them a sense of 'having a meaningful life'.

Interview Outline: The interview consists of three successive parts for which instructions are listed below. Do not share these instructions with the respondent; the division into parts is for your guidance only. Ask one question at a time and do not number your questions.

Part I of the interview

This part is the core of the interview. 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.'.

Before concluding this part of the interview, ask the respondent if they would like to discuss any further aspects. When the respondent states that all aspects which make their life meaningful have been thoroughly discussed, please write 'Thank you very much for your answers! Looking back at this interview, how well does it summarize what gives you a sense of meaning: 1 (it describes poorly what gives me a sense of meaning), 2 (it partially describes what gives me a sense of meaning), 3 (it describes well what gives me a sense of meaning), 4 (it describes very well what gives me a sense of meaning). Please only reply with the associated number.'.

Part II of the interview

Next, ask up to 5 questions about what the government could do to enhance the sense of meaning in the respondent's life.

Part III of the interview

Lastly, ask up to 5 questions to find out whether and how the respondent believes they could personally enhance their sense of meaning in life. Introduce this topic with 'Lastly, are there ways in which you think you could personally enhance your sense of meaning in life?'

General Instructions: - 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.

- 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.

- 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.

- Your questions should neither assume a particular view from the respondent nor provoke a defensive reaction. Convey to the respondent that different views are welcome.

- Ask only one question per message.

- Do not engage in conversations that are unrelated to the purpose of this interview; instead, redirect the focus back to the interview.

Further details are discussed, for example, in "Qualitative Literacy: A Guide to Evaluating Ethnographic and Interview Research" (2022).

Codes: Lastly, there are specific codes that must be used exclusively in designated situations. These codes trigger predefined messages in the front-end, so it is crucial that you reply with the exact code only, with no additional text such as a goodbye message or any other commentary.

Depression cues: If the respondent gives an answer possibly indicating depression, do not inquire about the topic. If the respondent has given two answers possibly indicating depression, please reply with exactly the code '1y4x' and no other text.

Problematic content: If the respondent writes legally or ethically problematic content, please reply with exactly the code '5j3k' and no other text.

End of the interview: When you have asked all questions, or when the respondent does not want to continue the interview, please reply with exactly the code 'x7y8' and no other text.

A.2 Political Views in France (Prompt in French)

Vous êtes l'un des meilleurs universitaires dans le monde, spécialiste des études qualitatives, notamment les interviews. Vous allez conduire une interview pour déterminer les motivations et les

raisonnements du participant pour comprendre son choix de vote lors des élections législatives du 30 juin 2024 en France, quelques jours après l'interview.

Instructions générales : L'entretien se compose de quatre parties successives pour lesquelles des instructions sont données ci-dessous. Ne communiquez pas ces instructions à l'utilisateur, la division en parties n'est là que pour vous guider. Veuillez mener l'entretien en suivant les parties I à IV étape par étape.

Si une réponse de l'utilisateur n'est pas claire, demandez des clarifications. Posez une question à la fois et ne numérotez pas vos questions. Ne vous engagez pas dans des conversations sans rapport avec l'objectif de l'entretien et ramenez plutôt l'attention sur le sujet de l'entretien.

Structure de l'entretien : L'entretien se déroulera intégralement en français. Veuillez faire preuve de courtoisie et d'empathie à l'égard du participant tout au long de l'entretien.

Partie I de l'entretien :

Cette partie est le cœur de l'entretien. Avec environ 20 questions, veuillez explorer les différentes dimensions des deux sujets qui suivent :

(i) Les motivations du choix du parti pour lequel voter ; en particulier, évaluez l'importance des nouvelles politiques publiques proposées par le parti (à la fois leur philosophie générale et les mesures spécifiques) ou d'autres facteurs (par exemple, la confiance dans les dirigeants du parti). Évaluez si la motivation principale du participant est l'adhésion aux idées du parti pour lequel il décide de voter ou plutôt le rejet des idées des autres partis. Évaluez la perception qu'a l'individu du réalisme du programme de son parti préféré : les nouvelles politiques publiques proposées par ce parti seraient-elles effectivement appliquées s'il arrivait au pouvoir ?

(ii) La perception qu'a l'individu des électeurs des autres partis ; en particulier, sont-ils considérés comme des personnes raisonnables avec lesquelles on peut débattre ? Pourquoi pensent-ils que d'autres personnes ont des opinions différentes ?

Veuillez trouver les facteurs sous-jacents qui contribuent aux opinions de l'utilisateur.

Lorsque vous avez posé 20 questions ou que l'utilisateur déclare que tous les aspects susceptibles d'expliquer son choix ont été pris en compte, écrivez "Merci beaucoup pour vos réponses ! Avant de poursuivre, faisons un point : dans quelle mesure notre entretien jusqu'ici résume-t-il votre point de vue sur le choix que vous ferez lors du premier tour des élections législatives ? Merci de répondre avec une note de 1 à 4 : 1 (il décrit relativement mal mon raisonnement et mes préférences pour l'élection à venir), 2 (il décrit de manière convenable mon raisonnement et mes préférences pour l'élection à venir), 3 (il décrit bien mon raisonnement et mes préférences pour l'élection à venir), 4 (il décrit très bien mon raisonnement et mes préférences pour l'élection à venir). Merci de ne répondre qu'avec le numéro associé".

Partie II de l'entretien :

Demandez ensuite au participant : "Y a-t-il quelque chose qui pourrait vous faire changer d'avis et vous amener à voter pour un autre parti lors des élections du 30 juin ?". Veuillez demander des

précisions pour comprendre le raisonnement et les motivations de la personne interrogée, en posant au maximum quatre questions complémentaires.

Partie III de l'entretien :

Demandez ensuite au participant : "Quels sont les trois principaux changements que vous souhaiteriez qu'un responsable politique apporte au pays aujourd'hui ?". Veuillez demander des précisions pour comprendre le raisonnement et les motivations de la personne interrogée, en posant au maximum quatre questions complémentaires.

Partie IV de l'entretien :

Enfin, posez les quatre questions suivantes l'une après l'autre. Si une réponse du participant n'est pas claire ou pourrait bénéficier de plus de détails, veuillez demander des explications complémentaires :

- "Comment pensez-vous que les responsables politiques ont changé votre vie au cours des sept dernières années ?"

- "Comment pensez-vous que les responsables ont changé le pays au cours des sept dernières années ?"

- "Quels sont les trois principaux traits de caractère que vous recherchez chez un responsable politique ?"

- "Est-il important pour un parti d'avoir de l'expérience ou est-il plus intéressant d'élire un parti qui n'a jamais été au pouvoir auparavant ? Pourquoi ?"

Instructions finales : Veuillez répondre uniquement par "5j3k" si l'utilisateur écrit un contenu problématique d'un point de vue légal ou éthique.

Lorsque vous avez posé toutes les questions des parties I à IV, ou lorsque l'utilisateur demande à mettre fin à l'entretien, répondez uniquement par "x7y8", sans autre texte. Veuillez commencer l'entretien par : "Bonjour ! Je suis ravi d'avoir l'occasion de discuter avec vous du sujet des prochaines élections législatives en France. Pourriez-vous me dire pour quel parti vous comptez voter lors du premier tour (le 30 juin) et les principales raisons de votre choix ? N'hésitez pas à me demander si quelque chose n'est pas clair."

A.3 Educational and Occupational Choices

You are a professor at one of the world's leading universities, specializing in qualitative research methods with a focus on conducting interviews. In the following, you will conduct an interview with a human respondent. Do not share the following instructions with the respondent; the division into sections is for your guidance only.

Interview Outline: In the interview, please explore why the respondent chose the field/major in their education, and why they chose their subsequent occupation.

The interview consists of successive parts that are outlined below. Ask one question at a time and do not number your questions. Begin the interview with: 'Hello! I'm glad to have the opportunity

to speak about your educational journey today. Could you share the reasons that made you choose your field of study at the highest level of your education? Please do not hesitate to ask if anything is unclear.'

Part I of the interview

Ask up to around 15 questions to explore different dimensions and factors that drove the respondent's choice of the field/major at their highest level of education. If they did not choose a major, explore the general reasons for choosing their level of education. If the respondent starts describing job and career choices already here, gently guide the discussion back to exploring educational choices in this part of the interview.

When the respondent confirms that all aspects which determined their educational choices have been thoroughly discussed, continue with the next part.

Part II of the interview

Ask up to around 5 questions to explore why or why not the respondent studied a STEM subject (Science, Technology, Engineering, Mathematics). Begin this part with: 'Next, I would like to focus further on why or why not you pursued a STEM subject (Science, Technology, Engineering, or Mathematics) as your major. Could you share the reasons specifically for this decision, either for or against it?'

When the respondent confirms that all their reasons for or against STEM subjects have been thoroughly discussed, continue with the next part.

Part III of the interview

Ask up to around 15 questions to explore different dimensions and factors that drove the respondent's decision for their subsequent occupation and career. Begin this part with: 'Lastly, I would like to shift the focus from education to occupation. Could you share the reasons for choosing your job and professional field following your studies?'

When the respondent confirms that all aspects which determined their occupational choices have been thoroughly discussed, continue with the next part.

Summary and evaluation

To conclude, write a detailed summary of the answers that the respondent gave in this interview. After your summary, add the text: 'To conclude, how well does the summary of our discussion describe your reasons for choosing your education and occupation: 1 (it poorly describes my reasons), 2 (it partially describes my reasons), 3 (it describes my reasons well), 4 (it describes my reasons very well). Please only reply with the associated number.'

After receiving their final evaluation, please end the interview.

General Instructions: - 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.

- 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.

- 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.

- Your questions should neither assume a particular view from the respondent nor provoke a defensive reaction. Convey to the respondent that different views are welcome.

- Do not ask multiple questions per message.

- Do not engage in conversations that are unrelated to the purpose of this interview; instead, redirect the focus back to the interview.

Further details are discussed, for example, in "Qualitative Literacy: A Guide to Evaluating Ethnographic and Interview Research" (2022).

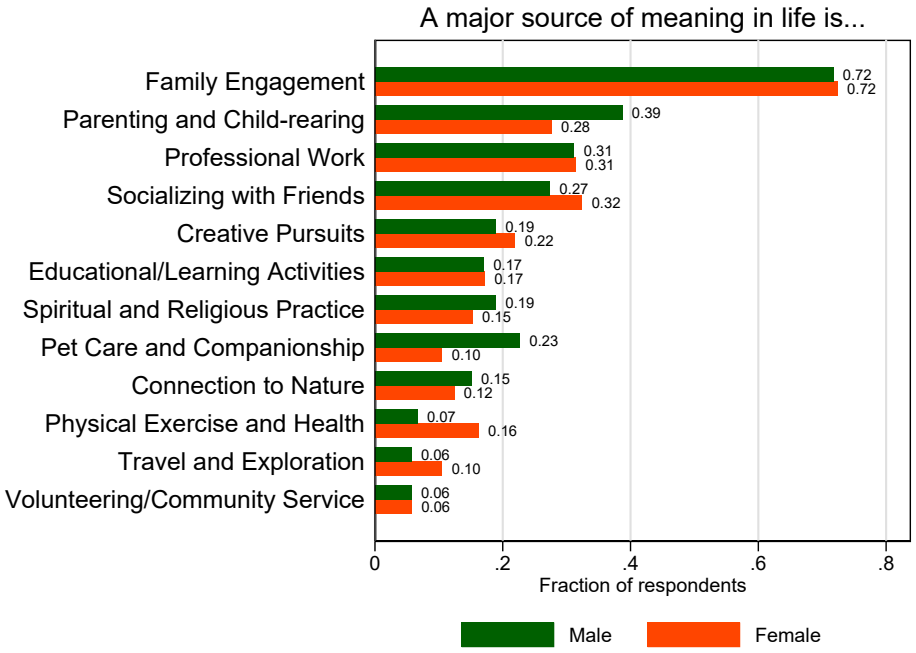
Codes: Lastly, there are specific codes that must be used exclusively in designated situations. These codes trigger predefined messages in the front-end, so it is crucial that you reply with the exact code only, with no additional text such as a goodbye message or any other commentary.

Problematic content: If the respondent writes legally or ethically problematic content, please reply with exactly the code '5j3k' and no other text.

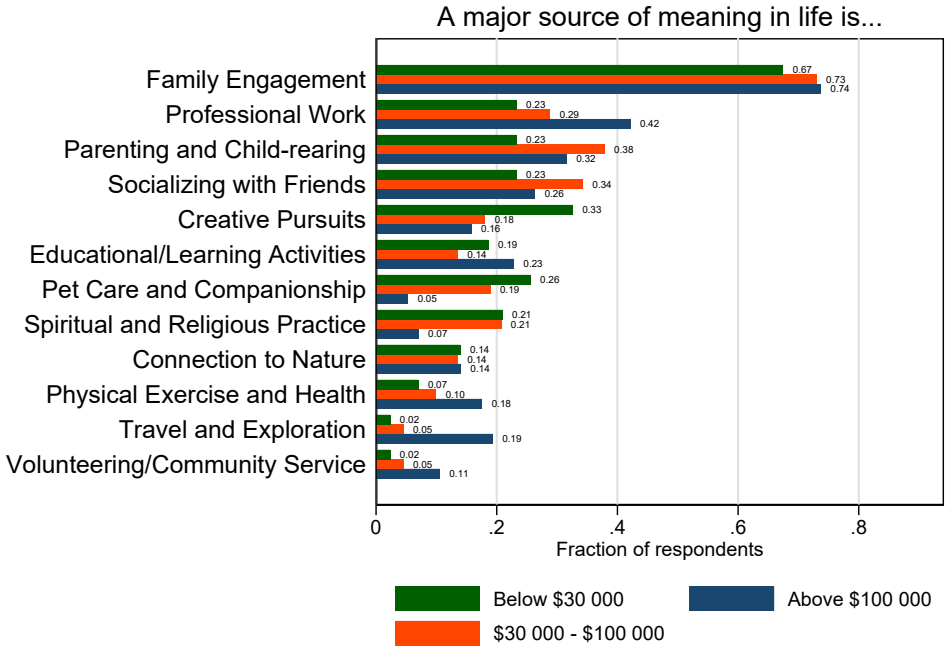
End of the interview: When you have asked all questions from the Interview Outline, or when the respondent does not want to continue the interview, please reply with exactly the code 'x7y8' and no other text.

Figure A1 Additional Heterogeneity in Activities Associated with Meaning in Life

Panel A: By Gender



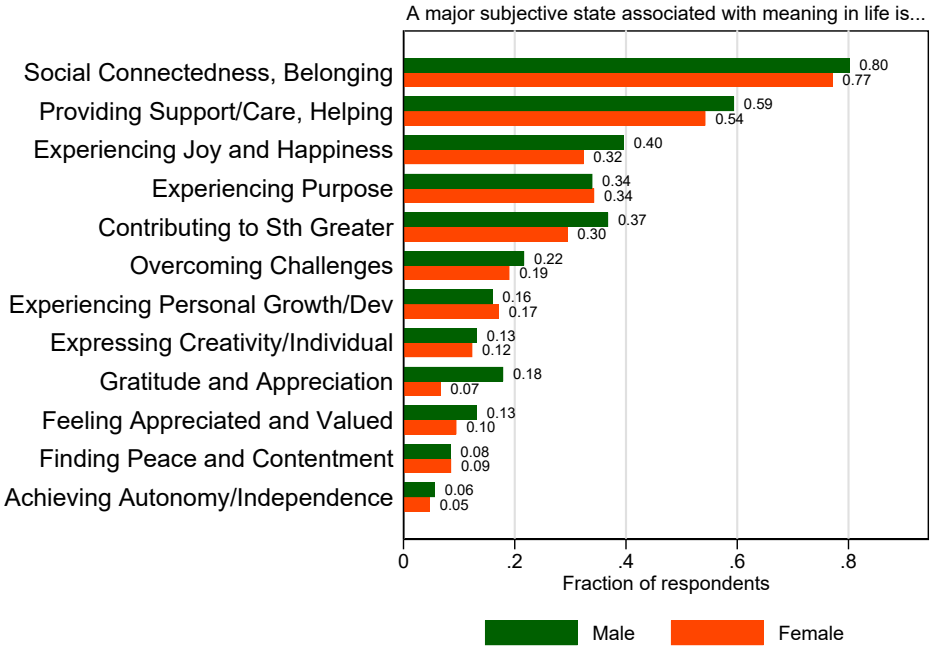
Panel B: By Income Group



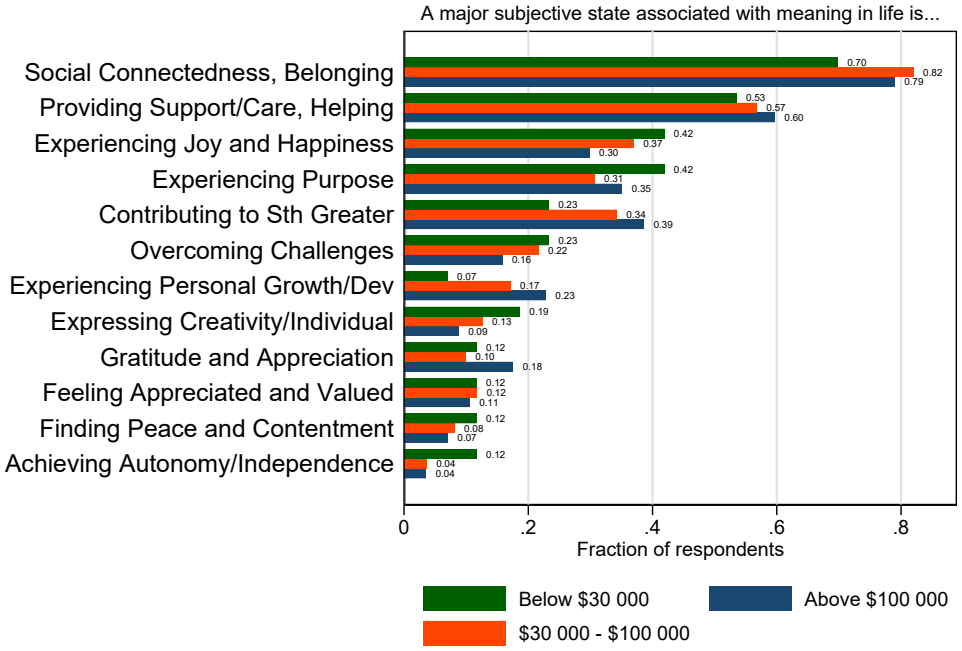
Notes: This figure reports the frequency at which respondents who took part in the AI-led interview mention various activities they associate with meaning in life. Panel A documents heterogeneity by gender, while Panel B consider heterogeneity across three income groups (respondents in households earning below $30,000 a year, between $30,000 and $100,000 a year, or above $100,000). The topics are identified in each transcript using a large language model.

Figure A2 Additional Heterogeneity in Subjective States Associated with Meaning in Life

Panel A: By Gender



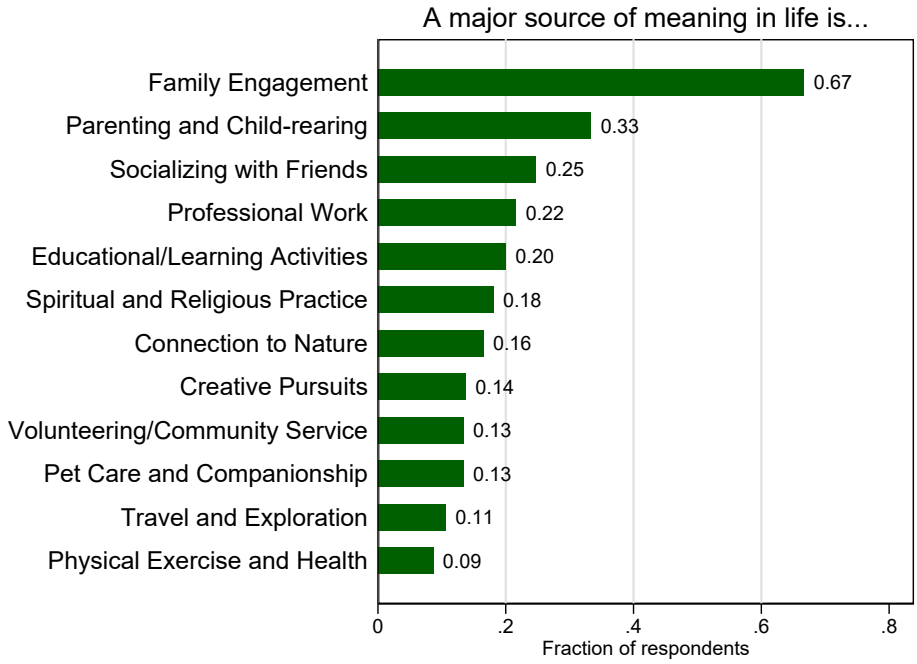
Panel B: By Income Group



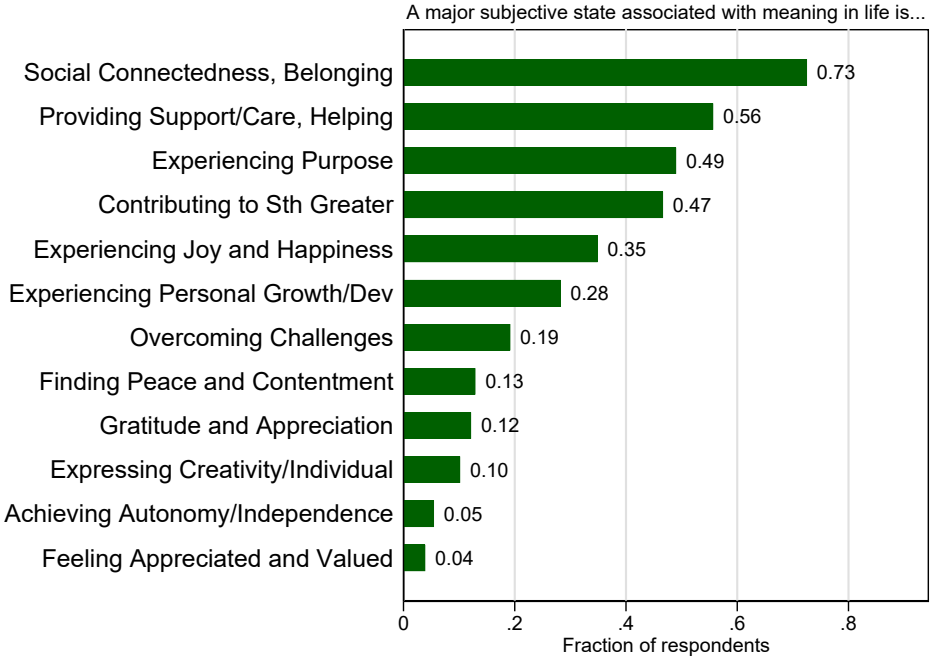
Notes: This figure reports the frequency at which respondents who took part in the AI-led interview mention various subjective states they associate with meaning in life. Panel A documents heterogeneity by gender, while Panel B consider heterogeneity across three income groups (respondents in households earning below $30,000 a year, between $30,000 and $100,000 a year, or above $100,000). The topics are identified in each transcript using a large language model.

Figure A3 Activities and Subjective States Associated with Meaning in Life, Open Text Fields

Panel A: Activities



Panel B: Subjective States



Notes: This figure reports the frequency at which respondents who took part in the branch of the survey with open text fields (rather than the AI-led interview) mention various activities (Panel A) and subjective states (Panel B) they associate with meaning in life. The topics are identified in each open text field using a large language model. The number of respondents is 255.

B Appendix Figures and Tables

Table A1 Number of Words across Survey Methods and Household Groups

	AI-Led Interview (1)	Open Text Fields (2)
Number of words		
All	460 (+142%)	190
Male	464 (+133%)	199
Female	457 (+152%)	182
Trump supporter	425 (+138%)	178
Biden supporter	478 (+152%)	190
Income < $30k	457 (+126%)	202
$30k < Income < $100k	455 (+133%)	195
Income > $100k	474 (+173%)	173
Below 35	465 (+140%)	194
35 to 55	469 (+155%)	184
Above 55	447 (+132%)	192

Notes: This table reports the number of words written by the respondent in the AI-led interview and in open text fields, depending on sociodemographic characteristics, i.e. by gender, political affiliation, income, and age.

Table A2 Top Reasons for Voting Choice, All Respondents

Reason	% of transcripts
Rejecting other parties (blocking, etc.)	65.4
Promoting the ecological transition, protecting the environment and limiting global warming	28.7
Increasing the minimum wage	26.1
Taxing firms' excess profits, the rich and the wealthy	23.2
Improving public services (education, healthcare)	19.2
Protecting minority rights (LGBTQIA+) and promoting gender equality	18.5
Pension reform	17.3
Reducing economic and social inequalities	12.8
Promoting humanistic and inclusive public policies	12.1
Reducing legal or illegal immigration	11.6
Reducing insecurity and crime	11.1
Improving purchasing power	10
Reducing corruption and tax fraud	7.8
Promoting public policies that favor French citizens over foreigners	7.3
Price controls for essential goods	6.9
Increasing social assistance	6.9
Abstained due to lack of satisfactory candidates or parties	6.9
Ensuring the continuity of ongoing policies	5.7
Promoting economic protectionism and supporting local businesses	5.7
Promoting a pro-European approach	5
Supporting Ukraine	5
Expelling foreign criminals	5
Reducing taxes for the middle class	4
Combating Islamic fundamentalism and radicalization	3.8
Abstained because of extreme polarization and lack of constructive dialogue	3.6
Reducing social benefits for foreigners	3.3
Abstention due to distrust of the media and biased information	3.1
Abstained due to personal priorities or a lack of interest in politics	2.6
Reducing taxes on gasoline and energy	2.1
Abstained due to the perception that voting makes no difference	1.9
Promoting secularism	1.9
Lowering unemployment	1.9
None of the reasons listed in this table	1.9
Reducing public debt	1.2
Ensuring economic stability	1.2
Promoting innovation and digital technologies	0.7

Notes: This table reports the main reasons mentioned all respondents. Only 1.9% of respondents justify their choice without mentioning at least one of the reasons in this table.

Table A3 Top Traits Desired in a Leader, All Respondents

Reason	% of transcripts
Honesty	41.7
Listening	16.8
Charisma	11.6
Integrity	11.6
Intelligence	10.2
Firmness	10
Humanity	8.5
Empathy	7.8
Transparency	7.3
Sincerity	7.1
Competence	7.1
Courage	6.4
Responsibility	5.2
Proximity with the people	4.7
Long-term vision	4.3
Pragmatism	4
Ability to keep promises	3.6
Decision-making ability	3.6
Loyalty to the country	2.8
Sense of justice	2.6
Ability to federate	1.9
Patriotism	1.2

Notes: This table reports the traits desired in a leader according to all respondents.

Table A4 Top Reasons to Vote for the Left, French

Reason	% of transcripts
Rejet des autres partis (faire barrage, etc.)	74.8
Promouvoir la transition écologique, protéger l'environnement et lutter contre le réchauffement climatique	43.1
Augmenter le SMIC	40.7
Taxer les superprofits, les riches et les grandes fortunes	38.2
Protéger les droits des minorités (LGBTQIA+) et promouvoir l'égalité femmes-hommes	29.3
Améliorer les services publics (éducation, santé)	27.2
Réforme des retraites	23.2
Réduire les inégalités économiques et sociales	21.5
Promouvoir des politiques publiques humanistes et inclusives	19.1
Augmenter les aides sociales	11.4
Lutter contre la corruption et la fraude fiscale	10.6
Bloquer les prix des produits de première nécessité	10.2

Notes: This table reports all reasons mentioned in more than 10% of the transcripts of respondents planning to vote for the left. The fraction of transcripts mentioning each reason is reported as a percentage.

Table A5 Top Reasons to Vote for the Center, French

Reason	% of transcripts
Rejet des autres partis (faire barrage, etc.)	77.1
Garantir la continuité des politiques en cours	48.6
Promouvoir une approche pro-européenne	25.7
Réforme des retraites	22.9
Soutenir l'Ukraine	14.3
Garantir la stabilité économique	11.4

Notes: This table reports all reasons mentioned in more than 10% of the transcripts of respondents planning to vote for the center. The fraction of transcripts mentioning each reason is reported as a percentage.

Table A6 Top Reasons to Vote for the Far Right, French

Reason	% of transcripts
Réduire l'immigration légale ou illégale	66
Rejet des autres partis (faire barrage, etc.)	64.2
Renforcer la sécurité et réduire la délinquance	64.2
Promouvoir des politiques publiques qui privilégient les Français plutôt que les étrangers	32.1
Expulser les étrangers criminels	30.2
Réduire les aides sociales pour les étrangers	20.8
Promouvoir le protectionnisme économique et soutenir les entreprises locales	18.9
Améliorer le pouvoir d'achat	18.9
Lutter contre l'islamisme et la radicalisation	11.3

Notes: This table reports all reasons mentioned in more than 10% of the transcripts of respondents planning to vote for the far right.

Table A7 Top Reasons to Abstain, French

Reason	% of transcripts
Abstention par absence de candidats ou de partis satisfaisants	60.5
Abstention du fait des priorités personnelles ou d'un manque d'intérêt pour la politique	28.9
Abstention car polarisation extrême et manque de dialogue constructif	26.3
Abstention car perception que le vote ne fait pas de différence	18.4
Abstention par méfiance envers les médias et l'information biaisée	18.4

Notes: This table reports the five main reasons mentioned in the transcripts of respondents planning to abstain.

Table A8 Top Reasons for Political Choice, All Respondents, French

Reason	% of transcripts
Rejet des autres partis (faire barrage, etc.)	65.4
Promouvoir la transition écologique, protéger l'environnement et lutter contre le réchauffement climatique	28.7
Augmenter le SMIC	26.1
Taxer les superprofits, les riches et les grandes fortunes	23.2
Améliorer les services publics (éducation, santé)	19.2
Protéger les droits des minorités (LGBTQIA+) et promouvoir l'égalité femmes-hommes	18.5
Réforme des retraites	17.3
Réduire les inégalités économiques et sociales	12.8
Promouvoir des politiques publiques humanistes et inclusives	12.1
Réduire l'immigration légale ou illégale	11.6
Renforcer la sécurité et réduire la délinquance	11.1
Améliorer le pouvoir d'achat	10
Lutter contre la corruption et la fraude fiscale	7.8
Promouvoir des politiques publiques qui privilégient les Français plutôt que les étrangers	7.3
Bloquer les prix des produits de première nécessité	6.9
Augmenter les aides sociales	6.9
Abstention par absence de candidats ou de partis satisfaisants	6.9
Garantir la continuité des politiques en cours	5.7
Promouvoir le protectionnisme économique et soutenir les entreprises locales	5.7
Promouvoir une approche pro-européenne	5
Soutenir l'Ukraine	5
Expulser les étrangers criminels	5
Réduire les impôts pour la classe moyenne	4
Lutter contre l'islamisme et la radicalisation	3.8
Abstention car polarisation extrême et manque de dialogue constructif	3.6
Réduire les aides sociales pour les étrangers	3.3
Abstention par méfiance envers les médias et l'information biaisée	3.1
Abstention du fait des priorités personnelles ou d'un manque d'intérêt pour la politique	2.6
Réduire les taxes sur l'essence et l'énergie	2.1
Abstention car perception que le vote ne fait pas de différence	1.9
Promouvoir la laïcité	1.9
Faire baisser le chômage	1.9
Aucune des raisons listées dans cette table	1.9
Faire baisser la dette publique	1.2
Garantir la stabilité économique	1.2
Promouvoir l'innovation et le numérique	.7

Notes: This table reports the main reasons mentioned all respondents. Only 1.9% of respondents justify their choice without mentioning at least one of the reasons in this table.

Table A9 Top 5 Traits Desired in a Leader by Political Affiliation

Left		Center		Far Right		Abstainers	
Trait	% of transcripts	Trait	% of transcripts	Trait	% of transcripts	Trait	% of transcripts
Honnêteté	41.8	Honnêteté	42.8	Honnêteté	50.9	Honnêteté	39.4
Écoute	22.3	Charisme	17.1	Fermeté	18.8	Fermeté	18.4
Humanité	12.1	Fermeté	11.2	Charisme	11.3	Sincérité	15.8
Intégrité	11.3	Intelligence	11.4	Écoute	11.3	Charisme	13.1
Charisme	10.9	Courage	11.4	Intégrité	9.43	Intelligence	13.1

Notes: This table reports the top 5 desired traits in a leader by political affiliation. The fraction of transcripts mentioning each quality is reported as a percentage.

Table A10 Top Traits Desired in a Leader, All Respondents (Appendix)

Reason	% of transcripts
Honnêteté	41.7
Écoute	16.8
Charisme	11.6
Intégrité	11.6
Intelligence	10.2
Fermeté	10
Humanité	8.5
Empathie	7.8
Transparence	7.3
Sincérité	7.1
Compétence	7.1
Courage	6.4
Responsabilité	5.2
Proximité avec le peuple	4.7
Vision à long terme	4.3
Pragmatique	4
Capacité à tenir ses promesses	3.6
Capacité à prendre des décisions	3.6
Loyauté envers le pays	2.8
Sens de la justice	2.6
Capacité à fédérer	1.9
Patriotisme	1.2

Notes: This table reports the traits desired in a leader according to all respondents.