Title: Voice vs. Text in AI-Moderated Interviews: A Comparative Study of Data Quality, Disclosure, and Participant Experience
Affiliation: Glaut Inc.
Date: September 2025

Abstract

How does response modality affect the quality of insights collected through AIMIs? Results from 252 AI-moderated interviews across three groups: voice, text, and hybrid answer modality show that voice responses are longer (**236% more overall words)**, more diverse (**138% more "unique" words)**, descriptive (**213% more "content" words)** and thematically richer (**28% more thematic codes assigned)** than text responses. Hybrid responses (free choice between Voice and Text) fall in between these extremes. Participants rated all modalities similarly well in terms of: ease of use, empathy of the AI moderator perceived, and willingness to disclose (open up) though most (55%) would rather use text modality, for comfort and privacy. These findings highlight trade-offs and offer companies guidance in balancing depth with respondent preference.

Keywords: voice-based interactions, chatbot experience, AI-moderated interview, answer modality, disclosure, engagement.

1. Introduction

The rise of conversational AI has expanded the use of chatbots and AI-moderated interviews (AIMIs) in market research and user experience testing. A central design decision in these systems is whether respondents interact via voice or text, or whether both options should be offered. Previous research indicates that when people speak instead of type, they naturally give more detail (brands, purposes, specifics), which leads to better results (Melumad, 2023; Melumad & Meyer, 2020). Voice is often associated with more natural, human-like communication, fostering openness and spontaneity. Text, by contrast, may encourage precision, reflection, and greater control over responses (Joinson, 2001; Schouten et al., 2020). Thus, the answer modality of an AIMI may activate different cognitive processes, influencing results and their managerial implications. To date, limited empirical work has systematically compared **data quality** and **participant experience** across response modalities in **structured AI-moderated interviews**.

The present study aims at answering the following research questions:

- **RQ1:** How does response modality (voice, text, hybrid) affect data quality in AI-moderated interviews?
- **RQ2**: How does response modality influence participant experience and willingness to disclose personal information?

Previous research and observation of real market research projects conducted on Glaut's platform, guided our derivation of the following hypotheses:

- **H1:** Voice responses will produce higher verbosity, lexical diversity, descriptiveness and thematic richness compared to text.
- **H2**: Participants' experience will not be substantially influenced by response modality as the overall experience is mainly driven by the personalisation of the interview rather than by voice modality.

2 . Methodology & Design

We employed a between-subjects one-way factorial design with three experimental conditions:

- **Group A (Voice):** respondents answered exclusively via spoken input.
- **Group B (Text):** respondents answered exclusively via typed input.
- **Group C (Hybrid):** respondents could freely choose between voice and text throughout the interview.

Respondents were randomly assigned to one of the three groups and completed the task via desktop, tablet, or smartphone.

2.1 Sample & Fieldwork

A total of 252 UK participants, between 18 and 65 (sourced through panel Prolific), completed AI-moderated interviews investigating the topic of "living alone in the UK". We chose this topic because it can elicit rich, qualitative answers, trigger diverse emotional responses and provide business relevant insights. The original questionnaire can be found in appendix (6. 2) and was composed of 8 "key" open questions (the focus of our data quality comparison) and additional closed questions about the topic, to reflect most-like scenarios in typical market research studies. In addition, 4 feedback questions were asked at the end of the main set to collect self-reported evaluation of the experience.

2.2 Interview Procedure and Platform Consistency

All interviews were conducted using the Glaut AIMI platform, which was specifically designed to ensure methodological consistency across participants and modalities.

Moderator Standardization

- The AI moderator was programmed to use a *neutral, professional tone* and delivered questions with identical wording across all modalities.
- Voice and text prompts were presented simultaneously, with no adaptive changes, to minimize bias.

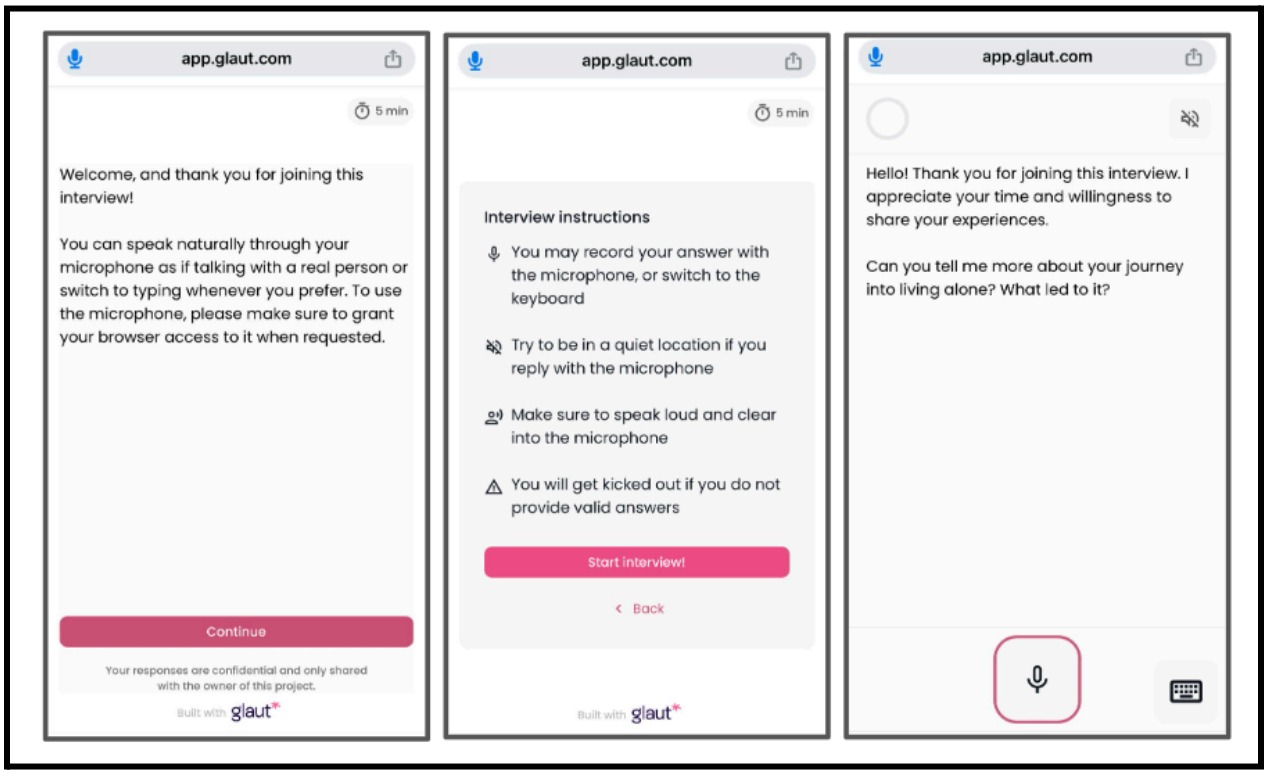
Instructions to Participants

- Before the interview began, participants received a standardized introduction explaining the purpose of the session (research about living alone), confidentiality of their responses, and how to use the platform.
- Respondents were explicitly told that there were *no right or wrong answers* and that they should answer as naturally as possible.
- Depending on condition, participants were informed of the response modality available (voice only, text only, or hybrid). Hybrid participants were told they could switch at any time.

Technical Consistency

- Participants could use a desktop, laptop, tablet, or smartphone. The interface adapted responsively to device type to ensure comparable usability.
- The platform automatically handled voice-to-text transcription for spoken answers, using the same transcription engine for all respondents. This ensured that thematic and lexical analyses were applied on a standardized textual dataset.
- Respondents were asked to complete the interview in a *quiet and private environment*. However, environmental factors (e.g., background noise, interruptions) were not experimentally controlled and thus represent a naturalistic element of the study.

Figure1: Welcome Page, instructions and first question flow of the AI-moderated interview (Mobile view) (Hybrid group, with free choice between voice and text answer)



3. Results

3.2 Data quality comparison between voice, text and hybrid

To evaluate whether voice responses provide greater value than text or hybrid responses, we compared answers to the eight focal open-ended questions (see Table 1) across multiple linguistic and thematic dimensions: verbosity, lexical diversity and descriptiveness.

Table 1: List of focal open ended questions analysed for groups comparison

Reasons for solo living	*Can you tell me more about your journey into living alone? What led to it?*
Benefits of solo living	*What do you enjoy most about living alone?*

Downsides of solo living	*What's the hardest or most frustrating thing about living alone?*
Effects on finances	*How does living alone affect your financial decisions?*
Food behaviours	*How does living alone shape how you shop, cook, or plan your meals?*
Leisure activities	*How do you prefer spending your free time and have fun?*
Positive actions by brands	*Which brands can you think of that feel well-suited to people who live alone? You can name brands in any product (or service) category*
Hopes and wishes	*What's one thing you wish brands, services, or your local area understood better about people who live alone?*

To determine the appropriate statistical test to use to compare these dimensions between groups, we first assessed the normality of the data using the Shapiro-Wilk test and checked for homogeneity of variance with a Brown–Forsythe test. To account for unequal variances and unequal group sizes, we used **Welch's ANOVA**, which adjusts the degrees of freedom accordingly. Where significant omnibus effects were observed, we conducted **Games–Howell post-hoc tests**, as this procedure is robust to violations of homogeneity of variances and does not require equal group sizes.

3.2.1 Verbosity. Verbosity was measured as the number of words used to answer all eight questions. A Welch's ANOVA indicated a significant effect of group on overall word production, $F(3, 9.69) = 35.10$, $p < .001$. **Group Voice** (M = 159.26, SD = 96.31) produced significantly more words than **Group Text** (M = 46.20, SD = 25.49; $p < .001$) and **Group Hybrid** (M = 93.15, SD = 88.22; $p < .001$). **Group hybrid** also produced more words than **Group Text** ($p < .001$). As clearly seen in chart 1**,** voice respondents produced significantly more words than those typing (**236% more words**), with hybrid respondents positioned between the two extremes.

3.2.2 Lexical Diversity (Unique Words). When speaking and typing, people might use repeated words (repetitions). Therefore, we assessed lexical diversity by counting unique words per respondent, after removing repetitions. Higher values indicate more varied vocabulary. For unique word count, Welch's ANOVA was again significant, $F(3, 9.83) = 38.94$, $p < .001$. **Group Voice** (M = 84.64, SD = 36.95) produced significantly more unique words than **Group Text** (M = 34.86, SD = 16.17; $p < .001$) and **Group Hybrid** (M = 56.07, SD = 37.61; $p < .001$). **Group Hybrid** also produced more unique words than **Group text** ($p < .001$). As shown in chart 2, Voice modality elicited a more diverse vocabulary than text (**138% more unique words**), with hybrid responses again falling in the middle.

3.2.3 Descriptiveness (content words). To capture the descriptiveness of responses, we counted content words (nouns, verbs, adjectives, adverbs). Welch's ANOVA revealed significant group differences in the number of content words, $F(3, 9.71) = 34.25$, $p < .001$. **Group Voice** (M = 70.67, SD = 41.62) produced significantly more content words than **Group Text** (M = 22.06, SD = 11.84; $p < .001$) and **Group Hybrid** (M = 42.60, SD = 38.80; $p < .001$). **Group Hybrid** also produced more content words than **Group Text** ($p < .001$). As shown in chart 3, Voice respondents produced a richer set of meaningful, descriptive terms (**213% more content words**), than text respondents, with hybrid respondents positioned between the two extremes.

Chart 1: Verbosity across groups

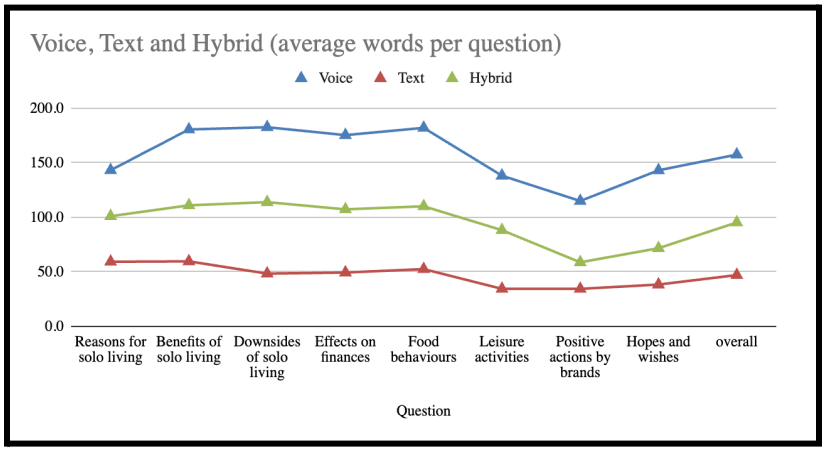


Chart 2: Lexical diversity across groups

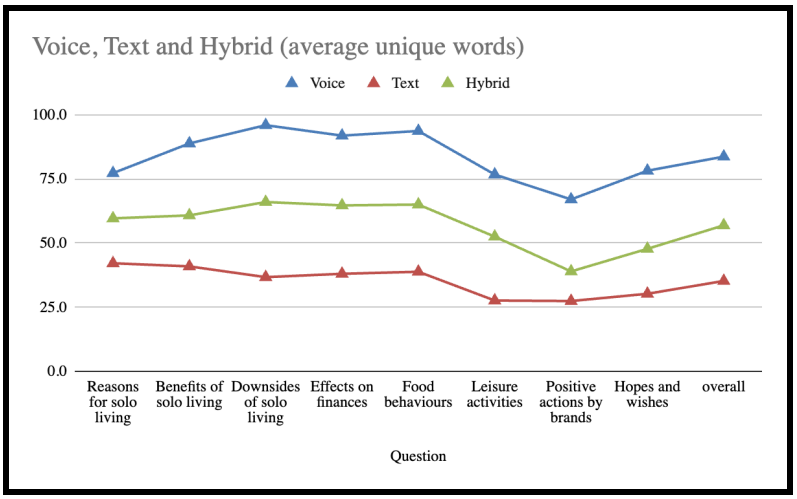
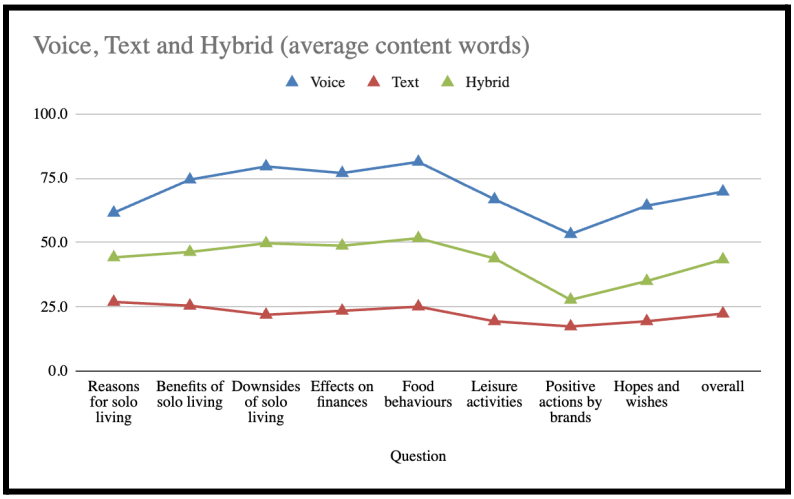


Chart 3: Content words across groups



3.2.4 Thematic variety: Using the Glaut thematic encoder[1] (with consistent prompts across the 3 groups), we analyzed the number of distinct **themes/codes** identified in each answer. A higher number of themes suggests a richer, more detailed, and more insightful answer. For the average number of codes, Welch's ANOVA indicated a significant effect, $F(3, 10.39) = 10.69$, $p = .002$. **Group Voice** (M = 3.44, SD = 1.04) had a significantly higher code count than both **Group Text** (M = 2.65, SD = 0.58; $p < .001$) and **Group Hybrid** (M = 2.86, SD = 0.81; $p < .001$). No significant differences were observed between **Groups Text and Hybrid** ($p = .18$)

When aggregated across the eight focal questions, **Voice responses produced an average of 27 themes**, compared to **21 themes for Text** and **23 themes for Hybrid**. Therefore, voice responses consistently enabled richer thematic coverage, yielding deeper and more varied insights.

[1] Details on the Glaut thematic encoder are included in Appendix 6.1

3.2.5. Concreteness and Imagery/Abstractness. To further evaluate the richness of open-text responses, we computed **concreteness** and **imagery** scores for each answer, two widely used psycholinguistic measures.

- **Concreteness** captures the degree to which a word refers to a tangible, perceptible concept (e.g., *table, car*), as opposed to an abstract or intangible idea (e.g., *freedom, strategy*). More concrete language is generally easier to process and conveys specific, actionable insights, whereas abstract language may signal general attitudes or high-level reflections (Paetzold & Specia, 2016; Pennebaker et al., 2007; Kacewicz et al., 2014).
- **Imagery** reflects the extent to which a word evokes mental images or sensory associations. Words with high imagery (e.g., *beach, chocolate*) are more vivid and experiential, often enhancing engagement and memorability, whereas low-imagery words (e.g., *system, policy*) tend to be more abstract and less evocative (Paetzold & Specia, 2016; Rubin, 1995; Chafe, 1982).

These measures provide a lens for assessing **how concrete and vivid respondents' answers are across answers modality**. For example, responses with higher concreteness and imagery may indicate a stronger connection to lived experience, richer descriptive detail, and greater emotional resonance. Conversely, lower scores may suggest more abstract reasoning, strategic framing, or distance from the topic. By comparing concreteness and imagery scores across modalities (voice, text, hybrid), we gain an additional dimension of insight into **how respondents express themselves** and **the type of information that each modality elicits**.

On these 2 metrics Shapiro–Wilk tests indicated that residuals deviated significantly from normality for both Abstractness ($W = 0.963$, $p < .001$) and Concreteness ($W = 0.963$, $p < .001$).

Concreteness: Welch's ANOVA revealed significant group differences in Concreteness, $F(2, 159.6) = 40.07$, $p < .001$. **Group Text** (M = 333.46, SD = 10.52) showed the highest mean score, followed by **Group Hybrid** (M = 328.38, SD = 8.88), and **Group Voice** (M = 322.53, SD = 5.64). Post-hoc Games–Howell tests confirmed that all pairwise differences were significant, with large effect sizes (text vs voice: $d \approx 1.3$; hybrid vs voice: $d \approx 0.8$; text vs hybrid: $d \approx 0.5$).

Imagery: Welch's ANOVA indicated significant group differences in imagery, $F(2, 159.6) = 40.07$, $p < .001$. Descriptive statistics showed the highest mean score in

Group Voice (M = 488.74, SD = 2.82), followed by **Group Hybrid** (M = 485.81, SD = 4.44), and the lowest in **Group Text** (M = 483.27, SD = 5.26). Post-hoc comparisons (Games–Howell) revealed that all pairwise differences were significant, with large standardized effect sizes (voice vs text: $d \approx 1.3$; voice vs hybrid: $d \approx 0.8$; text vs hybrid: $d \approx 0.5$).

In line with methodological suggestions from prior work on modality effects in qualitative data collection, our results indicate that the choice of elicitation mode systematically shapes the linguistic properties of responses. Text-based input produced more concrete lexical content, whereas voice-based input encouraged more image-rich descriptions. Hybrid responses consistently fell between these two extremes

Overall, in line with our expectations (H1), we observed that, across all metrics (verbosity, lexical diversity, content richness, and thematic variety), **voice responses outperformed text**, with hybrid responses consistently in between. This pattern aligns with previous research showing that voice encourages more spontaneous, elaborative communication (Melumad, 2023). Furthermore, the analyses on Concreteness and imagery revealed systematic differences between **voice, text, and hybrid responses** showing a **modality trade-off**:

- **Voice elicitation** promotes **richer, more vivid mental imagery**, possibly because oral expression allows respondents to narrate experiences more spontaneously and vividly.

- **Text elicitation** encourages **greater concreteness**, likely reflecting more deliberate lexical choice and editing when typing.

- **Hybrid responses** blend these tendencies, sitting between the two extremes on both measures.

3.3 Participants experience

Participants also evaluated their interview experience on perceived ease, empathy demonstrated by the AI moderator, and willingness to disclose (to open up and share with the AI moderator), each measured on a 5-point Likert scale. As often is the case with ordinal data (like 1-5 scales), the data does not follow a normal distribution (Shapiro–Wilk test performed on all 3 metrics for all groups, resulting in $p<0.05$). To

account for unequal variances and unequal group sizes, we used Welch's **ANOVA**. Chart 4 depicts an overview of the self-reported feedback.

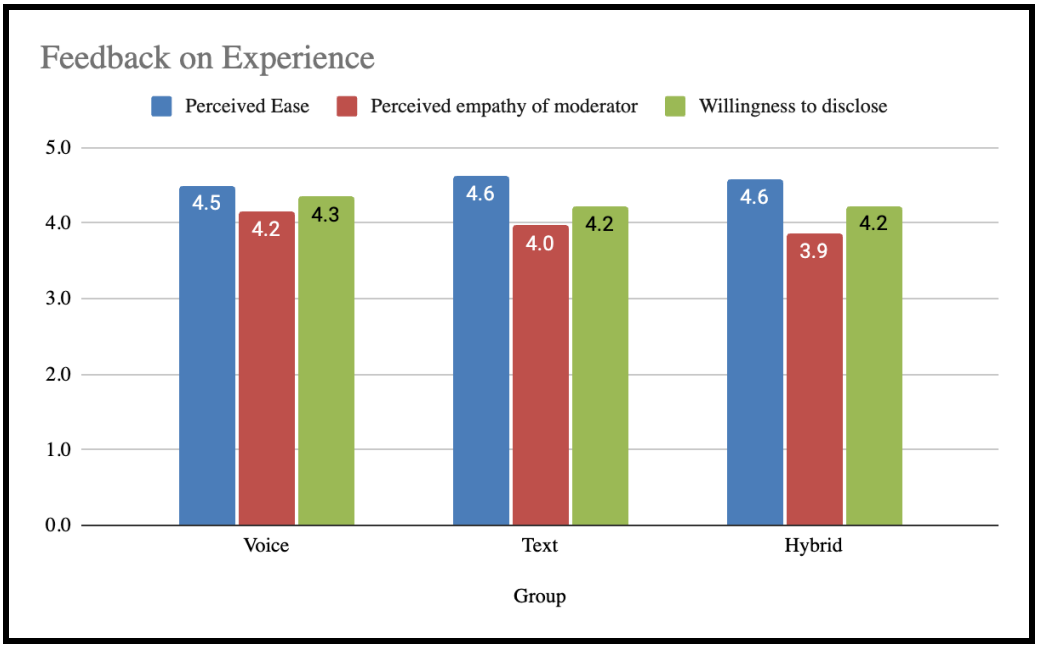
3.3.1 Perceived Ease: Respondents rated the ease of expressing their thoughts in the interview using a 1-5 scale (where 1 = very difficult, 5 = Very easy). Average scores show a **high ease of expression across all modalities**: Voice (M = 4.5), Text (M = 4.6), Hybrid (M = 4.6). A Welch's ANOVA showed no significant differences in perceived ease of use among the three groups, $F(2, 161.24) = 0.59$, $p = .556$)

3.3.2 Empathy: Respondents were asked to rate the extent to which they perceived that the AI interviewer listened and understood them during the interview (where 1= Not at all, 5 =Completely). Average scores demonstrate a **high perceived empathy of the AI moderator** across modalities **Voice (M = 4.2)**, **Text (M = 4.0)**, **Hybrid (M = 3.9)**. A Welch's ANOVA showed no significant group differences $F(2, 164.7) = 1.76$, $p = .175$.

3.3.3 Willingness to Disclose: Respondents were asked to rate the extent to which they felt open to share their personal thoughts with the AI moderator (where 1 = Not open at all, 5 = Very open). Average scores show a high openness to sharing personal thoughts across modalities: **Voice (M = 4.3)**, **Text (M = 4.2)**, **Hybrid (M = 4.2)**. A Welch's ANOVA indicated no significant differences in sharing/openness scores between groups, $F(2, 163.54) = 0.46$, $p = .632$.
We dove deeper with an open question to understand the reasons behind the score given on this last metric.

- **High disclosure (ratings 4–5, N = 209):** Many described the interview as *effortless* and *human-like* (52.6%), particularly among Voice respondents (56%) compared to Text respondents (43%). Voice also enhanced perceptions of *anonymity* and *judgment-free* expression (61% in Voice vs. 47% in Text). Hybrid respondents reported the *smoothest, least effortful* experience (59%).
- **Low disclosure (ratings 1–3, N = 42):** Barriers mentioned included *discomfort with AI* (60%), *lack of real-time feedback* (38%), and *low trust in data security* (33%). Concerns over data security were especially pronounced among Voice respondents (50%). Text respondents cited *rigidity and lack of personalization* more often (20%).

Chart 4: Self-reported feedback on interview experience



3.3.4 Modality perceptions:

When asked which modality they would prefer to answer another (future) AIMI, **text was slightly favored overall (55% vs. 45% for voice)**. However, preferences were strongly influenced by experimental condition, suggesting a priming effect:

- **Voice group:** 77% preferred voice.
- **Text group:** 76% preferred text.
- **Hybrid group:** 62% preferred text, 38% preferred voice.

Notably, most **Hybrid** participants (80%) did not switch modalities once they began. We only observed 2 respondents who started with voice and swap to text around the middle of the interview because of a technical issue with the microphone. Another 3 respondents chose one answer modality, tried the other for only 1 question, and switched back to their first choice. These behaviours reveal that **respondents choose the answer modality they prefer right at the beginning and DO NOT change it over time**, unless there is some technical issue.

Qualitative deep dives uncovering drivers of each modality reveal that:

- Respondents choose **Voice** for its quickness and ease (38%), typing is seen as uncomfortable by some (5%) (“*you don't have to be sitting looking at your keyboard, you can sort of move around a bit more.*”). Respondents seek a more

authentic and conversational experience (29%), where speaking feels more natural and emotionally connected than typing (*“I think it's easier and it feels more natural and more authentic”*). Preference for **Voice** is also dependent on situational factors (11%) (“*it depends on what time of the day and the noise levels in my block of flats.* “)

- On the contrary, they choose **Text** for its ability to enhance clarity and allow *thoughtful,* revised responses (26%) (“*I'd probably be more considered, think about things a bit more. And also, you know, have second thoughts. If you start to write something, you can delete it, whereas if, you know, you've said something, you can't take it back*.”) and because it’s a familiar modality and therefore easy and comfortable to use (19%). The desire for *privacy* and *anonymity* also drives preference for **Text** modality (12%). Text seems to be *reducing anxiety and pressure* (10%), providing a more comfortable communication experience.

4. Discussion & Business Relevance

This study contributes to the growing literature on AI-moderated interviews (AIMIs) by providing the first systematic, quantitative comparison of response modalities. Results indicate that **voice** responses substantially outperform **text** in terms of verbosity, lexical diversity, descriptiveness, and thematic richness. These findings align with prior work on conversational media showing that **speech activates more spontaneous and elaborative cognitive processes compared to writing** (Melumad & Meyer, 2020; Melumad, 2023).

As expected, participant experience did not significantly differ across modalities. All three groups rated perceived ease, empathy of the AI moderator, and willingness to disclose highly, suggesting that **AIMIs can provide a consistently positive user experience regardless of answer modality**. This supports findings in the human–computer interaction (HCI) literature that perceived empathy and comfort are often shaped more by conversational design and personalization than by input modality alone (Zhou et al., 2019; Cranshaw & Caine, 2022). However, preference patterns revealed a tendency toward preferring text, consistent with Glaut’s internal metrics and prior reports that users perceive **text as less socially risky and more controllable** (Joinson, 2001; Schouten et al., 2020). This points to a fundamental trade-off: while

voice maximizes data quality, text maximizes adoption and comfort. Hybrid interviews, which allow modality switching, provide a pragmatic middle ground delivering data quality above text-only while supporting user choice and situational flexibility.

From a managerial perspective, these findings suggest that researchers and brands should strategically align AIMI modality with study objectives. If the priority is richer, more detailed insights (such as in exploratory or generative phases) voice should be emphasized. If the priority is respondent comfort, scale, or sensitive topics, text may be preferable (confirmatory phases). Offering the hybrid choice may provide the best compromise option by accommodating diverse respondent needs.

4.1 Limitations and suggestions for future research

A few limitations temper the generalizability of these findings. First, the study was conducted with a UK sample on a single topic (solo living), which may not capture modality effects across different cultural, demographic, or thematic contexts. Second, the experimental design relied on a between-subjects allocation, meaning individual-level preferences and trade-offs could not be directly observed. Finally, technical and environmental factors (e.g., microphone quality, noisy settings, multitasking) were not controlled for, yet may strongly shape modality performance in real-world deployments.
Building on these findings, future investigations could delve into:

1. **Longitudinal adoption:** Examine how modality preferences and disclosure patterns evolve with repeated AIMI exposure. Does comfort with voice grow over time, or do initial preferences persist?
2. **Topic sensitivity:** Explore whether modality effects differ across sensitive vs. neutral domains (e.g., health, mental wellbeing, workplace grievances), where disclosure dynamics are likely to vary.
3. **Cross-cultural replication:** Investigate modality differences across linguistic and cultural contexts, as norms around voice vs. text communication differ globally.
4. **Situational constraints:** Assess how environmental conditions (e.g., private vs. public space, mobile vs. desktop) moderate modality choice and perceived ease.

4.2 Conclusion

This study demonstrates that modality is not a neutral design choice in AI-moderated interviews but a determinant of the depth, diversity, and descriptive richness of collected data. While voice yields more elaborated and thematically rich responses, text seems to

be preferred by users. Hybrid options offer a balanced path and can solve a trade off between data richness and participants needs. These findings underscore that modality design should be aligned with study objectives, respondent expectations, and contextual constraints.

Further Collaboration

We see this study as one step in advancing methodological understanding of AI-moderated interviews. If you are considering adopting or testing AIMIs in your own work, our team is available for dialogue and knowledge sharing. You can reach us at info@glaut.com or learn more at www.glaut.com for access to supporting materials and comparative studies.

5. References

Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and Written Language: Exploring Orality and Literacy* (pp. 35–53). Ablex.

Cranshaw, J., & Caine, K. (2022). *Understanding trust in conversational agents*. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 1–24.

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology, 31*(2), 177–192.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143.

Melumad, S. (2023). More than words: How speech disfluencies can signal cognitive effort and enhance consumer responses. *Journal of Consumer Research, 49*(6), 1051–1072.

Melumad, S., & Meyer, R. (2020). Full disclosure: How smartphones enhance consumer self-disclosure. *Journal of Marketing, 84*(3), 28–4

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count (LIWC)*. Austin, TX.

Paetzold, G. H., & Specia, L. (2016). *Inferring psycholinguistic properties of words*. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 435–440). Association for Computational Linguistics.

Rubin, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes*. Oxford University Press.

Schouten, A. P., Valkenburg, P. M., & Peter, J. (2020). Digital self-disclosure and well-being: The role of communication competence and inclusivity. *Computers in Human Behavior, 103*, 208–219.

Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2019). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics, 46*(1), 53–93.

6. Appendix (questionnaire)

6.1. Methodological Note

Open-ended responses were coded using the Glaut Thematic Encoder, an AI-based system for qualitative analysis of interview transcripts. The encoder operates through a structured, multi-stage pipeline. First, interview excerpts are processed in small batches (five interviews at a time), formatted in XML, and submitted to a large language model (LLM) that performs initial coding. At this stage, the model identifies relevant respondent quotes and assigns codes based on the specific analytic instructions.

Two coding modes are available. In thematic analyses, the system uses an adaptive codebook that allows the generation of new codes, following strict formatting rules (2–4 word noun phrases in the target language). In fixed-codebook analyses, coding is limited to a predefined set of categories. After the initial coding, a refinement stage is carried out by a second AI agent, which reviews the codebook, identifies conceptually overlapping codes, and merges them when appropriate. Each merger requires explicit reasoning and clear inclusion/exclusion criteria to ensure transparency.

To balance quality and efficiency, the system processes data in progressively larger batches (beginning with 25 interviews and scaling up to 200). All coding decisions are represented as structured “meaning units,” which link specific respondent quotes to category IDs. These are stored in XML format, ensuring full traceability and compatibility with downstream analyses. The encoder also supports multilingual datasets and includes XML sanitization protocols to handle parsing errors and edge cases, ensuring robust performance across diverse interview content.

6.2 Original questionnaire

Screening & Quotas

S1. In which of the following categories do you identify with?

- Male (50%)
- Female (50%)
- Prefer not to answer (OUT)

S2. What is your age?

- younger than 18 (0%)
- 18–24 (25%)
- 25–34 (25%)
- 35–49 (25%)
- 50–65 (25%)
- older than 65 (0%)

S3. How many people currently live in your household, including yourself?

- 1 (I live alone)
- 2 (OUT)
- 3 (OUT)
- more than 3 (OUT)

Questionnaire:

Q1. Can you tell me more about your journey into living alone? What led to it? —> *Follow up to understand if the decision was planned or resulted from something that happened due to life events*

Q2.How do you feel about living alone overall?

- 1= I really dislike it
- 2 =I don't enjoy it much
- 3= It's okay / I feel neutral about it
- 4 = I like it overall
- 5 = I love it and wouldn't want to live any other way

Q3: What do you enjoy most about living alone? —> *Follow up to understand the benefits of living alone*

Q4. What's the hardest or most frustrating thing about living alone? —> *Follow up to understand the cons of living alone*

Q5. How would you describe your current financial situation?

- I live comfortably and can afford extras
- I meet my needs but don't have much left over
- I often have to make careful trade-offs
- I struggle to cover basic expenses

Q6. How does living alone affect your financial decisions? —> *follow up to understand what they spend on, save for, or skip*

Q7. How does living alone shape how you shop, cook, or plan your meals? —> *Follow up to understand if they waste more or less food, cook differently or eat out more.*

Q8. How do you prefer spending your free time and have fun? —> *Follow up to understand specific activities and places they visit*

Q9. Which brands can you think of that feel well-suited to people who live alone? You can name brands in any product (or service) category —> *Follow up to understand what these brands do right*

Q10. What's one thing you wish brands, services, or your local area understood better about people who live alone? —> Follow up to understand what they wished these parties would do for their living situation

Q11. Thank you for your answers. I have another few questions for you, before closing. What is your current relationship status?

- Single, not currently dating
- In a relationship but living separately
- Divorced or separated
- Widowed
- Other

Q12. Where do you currently live?

- A large city or city centre
- A suburb or town near a city
- A small town or rural area

Q13. What is your current employment situation?

- Employed full-time
- Employed part-time
- Self-employed or freelance
- Retired
- Unemployed or between jobs
- Student

Q14. What is your approximate total annual personal income (before tax)?

- Under £15,000
- £15,000 – £29,999
- £30,000 – £44,999
- £45,000 – £64,999
- £65,000 or more
- Prefer not to say

And at very last, I would like to understand what you think about the format of this interview (AI moderated). (Q15)

Q16. Perceived Ease: On a scale from 1 to 5, how easy was it to express your thoughts using this format? (1 = very difficult, 5 = Very easy)

Q17: Empathy: On a scale from 1 to 5, how much did it feel like the interviewer was listening and understanding you? (1 = Not at all, 5 = Completely)

Q18: Willingness to Disclose: On a scale from 1 to 5, How open did you feel sharing personal thoughts in this format? (1 = Not open at all, 5 = Very open)

Q19: if Q18 = 1,2,3: What made it difficult for you to fully open up during this conversation?

Q20: if Q18= 4,5: What made it easy for you to naturally open up during this conversation?

F4: Group A only: In this interview you were allowed to answer the questions only by using your mic (by voice). How did it feel to answer by speaking out loud instead of typing?

F5: If you could choose whether to answer the next interview using your mic (by voice) or using your keyboard (by text), what would you choose? —> reasons

F6: Group B only: In this interview you were allowed to answer the questions only by using your keyboard (texting). How would you feel about the possibility to speak instead of typing?

F5: If you could choose whether to answer the next interview using your keyboard (by text) or using your mic (by voice), what would you choose? —> reasons

F7. Group C only: In the interview, you had the possibility to change the modality to answer the questions, either speaking through your mic (VOICE) or via typing on your keyboard (TEXT). Have you changed answer modality over the interview?

1. If yes: think about the way you expressed your answers in the interview (voice and text) can you tell me how you decided to swap modality? Was there any particular reason? (follow up: goal: understand why they swap modality)

2. if no: which option have you chosen for answering? How would you describe the experience you had through this modality?

F5: If you could choose whether to answer the next interview using your keyboard (by text) or using your mic (by voice), what would you choose? —> reasons