Informational Quality and Completion Experience in Online Interviews: A Comparison Between Traditional Questionnaire and AI-Based Conversational Interface

ABSTRACT

This white paper analyzes the contribution of AI-based conversational interfaces to the informational quality of online surveys, comparing them with a traditional questionnaire administered via panel. Based on two independent surveys, weighted by gender, age, and geographic area, we examine the effects of AI on textual production, the cognitive depth of responses, and the overall respondent experience.

Quantitative analysis shows that the AI modality promotes greater verbosity, richer lexical variety, and a higher number of distinct concepts. AI mode produces responses that are on average 30% longer than those obtained with the traditional method (32.78 words versus 25.25).

AI mode recorded an average of 25.00 distinct lemmas per response, compared to 23.03 in the traditional mode. This increase corresponds to a rise of approximately 8.55%, indicating that participants activate a more varied vocabulary.

AI mode generated an average of 9.73 distinct concepts, higher than the 7.84 recorded in the traditional questionnaire. This difference corresponds to an increase of approximately 24.11%.

When voice is used in combination with the conversational interface, these effects are further amplified, generating responses that are longer, more coherent, and more articulated than both the traditional questionnaire and the text-only version of the same AI. Measures of semantic cohesion and argumentative depth confirm that AI does not merely produce more content, but actually stimulates more extended and logically organized reasoning.

Semantic cohesion, in particular, is significantly higher in the AI modality than in the traditional one: the average value increases from 0.213 to 0.338 (a delta of +0.125, corresponding to roughly +12.5 points on a 0–100 scale), a difference greater than both the 90% margin of error (±4.6) and the 95% margin (±5.5), indicating at least 95% significance. Therefore, **semantic cohesion is approximately 58.7% higher in AI mode compared to the traditional mode.**

The voice-only modality shows a further jump compared to text-only (0.558 vs. 0.272), with a difference of about +28.6 points against margins of error of ±8.6/±10.2, again significant at the 95% level. This means that **semantic cohesion in the Voice-only mode is more than double (approximately 105% higher) than in the Text-only mode**. The magnitude of this increase suggests that spoken interaction allows participants to develop their reasoning more naturally, producing responses structured like short speeches, which are significantly more coherent than typed ones.

The differences, however, do not concern only the informational dimension: the AI modality significantly improves the perceived experience, increasing the sense of engagement, being listened to, and ease of completion. In the comparison between AI and the traditional questionnaire, the proportions of final ratings ≥7, ≥8, and ≥9 are all statistically higher for AI (Welch's t test, $p \approx 0.000$ for all thresholds and for the overall mean), while the differences between text-only and voice-only, though suggesting an advantage for voice, do not reach conventional significance in most cases ($p > 0.05$), with a borderline tendency for the share of ratings ≥8 ($p \approx 0.053$).

Qualitative analysis further shows that, while the thematic structure of opinions remains stable, the conversational mode intensifies the narrative dimension and reduces signals of cognitive fatigue.

Overall, the results suggest that conversational AI is a promising solution to increase the value of open-ended questions in quantitative research, bringing the questionnaire closer to forms of expression typical of qualitative interviews, without compromising standardization and scalability.

INTRODUCTION

Online surveys are an indispensable tool for social and market research thanks to their ability to quickly collect large amounts of standardized data. However, the very operational standardization that enables their scalability often limits the informational quality of open-ended responses, which tend to be concise, fragmented, or weakly argued. The lack of dynamic interaction prevents probing, elaboration, or guided reasoning, as happens in qualitative interviews.

In recent years, the evolution of language models and AI-based conversational interfaces has opened new opportunities to bridge this gap. These systems are able to interpret the participant's response and generate tailored follow-up messages that stimulate contextual and immediate elaboration. The introduction of the voice channel, moreover, makes it possible to leverage forms of expression that are less controlled and more spontaneous, traditionally associated with greater narrative richness.

This study aims to systematically assess the impact of conversational AI on the informational quality of responses and on the participant experience, comparing it with a traditional online questionnaire mode. The goal is not only to verify whether AI produces "more words," but to understand whether it generates content that is more relevant, articulated, and useful for analysis, and whether it significantly changes the way people experience the participation process.

METHODOLOGICAL FRAMEWORK

2.1 Research Design

The research consists of two separate surveys. The traditional questionnaire was administered to the OpLine panel in October 2025, while the conversational AI modality was delivered to the PureSpectrum panel in November 2025. While the traditional questionnaire was only answerable via text replies, in the AI modality participants could respond via text, voice, or a combination of the two channels.

To ensure comparability between the two samples, cross-weights for gender × age × geographic area were applied, in order to align demographic structures and minimize any panel-related differences.

The research is based on a total sample of 1,003 participants, distributed across the different modes of administration as follows:

- **Traditional questionnaire:** 503 cases, all text- only replies
    - Valid open-ended responses: 474
    - Invalid (unusable) open-ended responses: 29 (excluded from analyses)

- **Conversational AI modality (total):** 500 cases, self selected in the following answer modalities:
    - AI text-only: 360 cases
    - AI voice-only: 114 cases
    - AI text + voice: 7 cases
- **Non-cooperative:** 19 cases (excluded from analyses)

The overall sample size is therefore well balanced between the traditional path and the AI path, allowing a solid and statistically robust comparison between modes.

Internal distribution within the AI modality

Within the AI group (N = 500):

- Text-only accounts for the majority share (72% of total AI)
- Voice-only involves 22.8% of participants
- Text + voice is residual (1.4%)

This distribution reflects a clear preference among users for the text channel, with a nonetheless significant share choosing voice as the primary response modality.

2.2 Questionnaire Structure

Both surveys included two main open-ended questions on the value of online reviews in purchase decisions; 3 sociodemographic questions, 2 closed questions about holiday habits and 3 closed questions to evaluate the compilation experience.

In the **traditional modality**, respondents provided an open-ended response and, through a piping mechanism, two follow-up questions were presented, directly anchored to the content of the initial response.

[Example of the piping logic:
Question 1: When evaluating hotels, restaurants, or tourist attractions, what specific characteristics of a review (for example, content, style, details, reviewer profile) make it truly useful in guiding your decision?
Answer 1: I'm interested in detailed reviews that also include the period of stay. Recent reviews, of course.
Question 2: You told us that what makes a review useful is: *"I'm interested in detailed reviews that also include the period of stay. Recent reviews, of course."* If you had to explain to a friend why these review characteristics are important, what would you say?]

In the **AI modality**, by contrast, the system analyzed the first response in real time and generated two personalized follow-up prompts, designed to promote articulation of thought and explicit expression of motivations

---

2.3 Metrics and Data Sources

The analyses are based on the following metrics:

- verbosity (total and average words);
- distinct lemmas and concepts;
- semantic cohesion and argumentative depth;

- distribution and frequency of themes;
- overall satisfaction and experiential items;
- differences between text and voice responses.

Description of the metrics:

Verbosity (word count)
Verbosity refers to the length of a response, measured by the total number of words used, calculated excluding Italian stopwords. It is a quantitative indicator used to assess how concise or verbose a participant is when expressing their thoughts.
[measurement methodology: ai software (Chat Gpt)]

Lemma
A lemma is the base or canonical form of a word (for example, go for goes, went, going). The number of distinct lemmas used in a response is an indicator of lexical diversity and reflects the breadth and variety of the vocabulary employed.
[measurement methodology: AI software (proprietary of Human Highway)]

Concept
A concept is a distinct semantic unit representing a meaningful idea expressed in a text. Concepts abstract away from specific word forms and capture the underlying meaning conveyed by one or more related terms or expressions. The number of distinct concepts identified in a response reflects the diversity of ideas addressed and the semantic richness of the content and variety of the vocabulary employed.
[measurement methodology: measurement made by researchers]

Semantic cohesion
Semantic cohesion is the degree to which the words and sentences in a text are connected to one another to form a coherent and integrated meaning. It measures how well the discourse “holds ideas together,” rather than being a list of disconnected concepts. value 0-1
[measurement methodology: measurement made by researchers]

Argumentative depth
Argumentative depth is a text’s ability to explain its claims, articulate cause-and-effect relationships, justify choices, and provide examples or conditions. It measures how much reasoning is contained within the response. scale from 1 to 5
[measurement methodology: measurement made by researchers]

Theme

Thematic analysis is a qualitative method used to identify, analyze, and interpret recurring themes or patterns of meaning within a set of textual data. It involves grouping related concepts and expressions into broader thematic categories
[measurement methodology: ai software (Chat Gpt)]

QUANTITATIVE RESULTS

3.1 Verbosity and Informational Variety

The first differences between modes emerge very clearly in the volume and variety of the content produced. The three main metrics are summarized in the following table.

Table 1 – Verbosity and Informational Richness

Metric	Traditional	AI (total)	AI text-only	AI voice-only
Words per response	25.25	32.78	24.94	58.65
Distinct lemmas	23.03	25.00	21.00	38.00
Distinct concepts	7.84	9.73	8.64	13.36

The results in Table 1 show that the AI modality not only increases the volume of responses, but also transforms their informational structure. The increase in words per response is not accompanied by simple textual stretching; rather, it is associated with greater semantic richness: both distinct lemmas and unique concepts rise consistently compared to the traditional questionnaire.

In particular, the **AI modality produces responses that are on average 30% longer than those in the traditional mode** (32.78 vs. 25.25 words), but what is striking is that this increase translates into more real content, not into repetition or generic formulas.

Distinct lemmas and concepts grow proportionally, indicating that respondents are activating a more varied vocabulary and a broader range of ideas. The increase in verbosity (+30%) is accompanied by an increase of 8.55% in lexical variety and of over 24.11% in the range of ideas expressed (distinct concepts).

When comparing the text-only conditions (traditional mode and AI text-only mode), the difference in the average number of lemmas is statistically significant at the 95% level ($p = 0.034$), with slightly higher mean values observed in the traditional mode compared to the AI text-only mode. However, it should be noted that the effect size is small (approximately two lemmas difference) and may have limited practical relevance despite being statistically significant.

Differences between the questionnaire completion modes were analyzed using Welch's t-test for independent samples, which was chosen to compare means in the presence of potentially unequal variances and sample sizes. The test was applied to quantitative indicators of textual production (number of words, number of lemmas, and number of concepts), using a two-tailed approach and a 95% significance level.

The comparison between the conversational AI mode and the traditional mode reveals statistically significant differences across all the indicators considered. Specifically, the conversational mode shows a significantly higher average number of words compared to the traditional mode (33 vs. 25; $p = 0.0000015$), greater lexical richness in terms of lemmas (25 vs. 23; $p = 0.040$), and higher conceptual density (10 vs. 8; $p < 0.000000000001$).

The voice modality further amplifies this effect. Responses given via voice leads to more than double the number of words per response compared to the traditional mode (on average 132.28% longer than those obtained with the traditional questionnaire.) and show even greater semantic enrichment: 38 distinct lemmas and over 13 unique concepts on average. This is a significant qualitative leap, suggesting that oral production reduces the "cognitive friction" typical of typing

(corrections, attention to form, interruptions in the flow of thought), enabling participants to articulate their reasoning in a more natural and fluid way.

In other words, voice not only frees up cognitive resources but also encourages more spontaneous and less fragmented expression. The result is responses that are denser, richer in nuance, and closer to the language of qualitative interviews.

Overall, the combination of AI and voice introduces a substantial change in how individuals describe their thoughts:
– more words,
– greater lexical variety,
– more concepts,

indicating a real increase in the informational value of open-ended responses rather than mere prolixity.

3.2 Semantic Cohesion and Argumentative Depth

Indicators of discursive quality paint an equally clear picture.

Table 2 – Semantic Cohesion and Argumentative Depth

Indicator	Traditional	AI (total)	AI text-only	AI voice-only
Semantic cohesion	0.213	0.338	0.272	0.558
Argumentative depth	1.86	2.40	2.20	3.10

The results in Table 2 show a change that is not only quantitative but structural in the quality of discourse produced by respondents. AI does not merely “make people write more”; it intervenes in the form of the expressed thought, making it more coherent and logically articulated.

Semantic cohesion is the first indicator to reveal this leap: the value rises from 0.213 in the traditional modality to 0.338 in the AI modality, an increase of about +12.5 percentage points (on a 0–100 scale). This is a substantial difference, well above both the 90% and 95% margins of error (*Two-proportion z-test - p value=0,0000044)*, confirming the statistical robustness of the effect. In other words, AI succeeds in eliciting responses in which concepts are linked to one another more smoothly and consistently, resulting in less fragmented content and more micro-narrative structures.

The result is even more pronounced in the voice modality, where semantic cohesion reaches 0.558: a value more than twice that of the traditional mode and significantly higher even than the text-based AI. This suggests that spoken interaction allows respondents to develop their reasoning more naturally, without the cognitive interruption imposed by typing. Voice accompanies the reasoning and makes it more continuous, leading to responses structured as short speeches rather than isolated statements.

Semantic cohesion is 59% higher in AI mode (overall), compared to static and 162% higher than static, when voice only answers are considered.

Argumentative depth confirms and enriches this picture. On this indicator, too, AI shows a significant improvement: values increase from 1.86 in the traditional mode to 2.40 in the AI modality, with a p-value of 0.0006 (Welch's t test), showing an **increase of 29%**. This figure indicates that the observed differences are not random but reflect a genuine change in how participants construct their answers.

Once again, the voice modality represents the most advanced point on the continuum: with a value of 3.10, it almost doubles argumentative depth compared with the traditional mode (**+67%**). Spoken responses more frequently include justifications, examples, conditions, comparisons, and evaluation criteria-elements characteristic of explicit, well-organized reasoning. This suggests that voice not only facilitates expression but activates cognitive modes closer to natural conversational thinking.

Overall, these results indicate that conversational AI transforms how individuals develop and articulate their thoughts during a survey. The higher semantic cohesion and greater argumentative depth are not mere stylistic improvements but crucial methodological components: more structured, coherent, and argued responses provide richer insights, easier to interpret and with higher analytical value.

3.3 Stability of Thematic Content

To assess whether the use of AI moderation introduces thematic bias, responses collected in the different conditions were subjected to the same thematic coding procedure. All open-ended answers were analyzed and assigned to a common set of thematic categories using ChatGPT as a coding tool. The resulting themes are reported in Table 3.

Table 3 – Main Themes of Useful Reviews

Theme (Question 1)	**Traditional**	**AI**
Content and details	1	1
Cleanliness and services	1	2
Price / Value	3	3
Reviewer credibility	4	4
Style and tone	5	5

Theme (Question 2)	**Traditional**	**AI**
Positive experience	1	1=
Conditions of the experience	2	1=

Theme (Question 2)	Traditional	AI
Habitual behavior	3=	4
Negative experience	3=	3
Altruistic motivation / usefulness to others	5	5

The analysis shows that no new or unique themes emerge in either group. The same set of themes is identified in responses collected with the traditional modality and with the AI-mediated modality. Moreover, the number of themes is identical across conditions, indicating that AI moderation does not lead to the emergence or suppression of specific topics.

The themes also appear with comparable relative importance across modes. While small variations in ranking are observed, the overall hierarchy of themes remains substantially stable, meaning that the same topics are consistently the most and least salient across conditions. These shifts in ranking reflect minor differences in relative prominence rather than structural changes in thematic content and do not alter the substantive interpretation of the results. Importantly, the thematic analysis captures which themes are present, not how extensively or vividly they are articulated. Qualitative differences observed in the AI-mediated condition-particularly in the voice modality, where responses more frequently include contextual details and concrete examples-do not translate into different thematic codes, but rather represent a richer expression of the same underlying themes. Such differences are therefore observable at the level of verbatim responses, not in the thematic structure itself.

Overall, the findings indicate that AI moderation does not affect the nature, number, or distribution of themes, supporting the conclusion that thematic content remains stable across traditional, text-based AI, and voice-based AI modes.

3.4 Overall Evaluation of the Experience

The overall evaluation of the compilation experience was assessed using a 10-point satisfaction scale [from 1 (not important at all) to 10 (very important)]

Differences in perceived experience are equally significant and confirm AI's ability to generate a more engaging interaction. Table 4 reports mean rating of the evaluation items individually analysed and specified in Table 5.

Table 4 – Final Interview Rating

Indicator	AI	Traditional	AI text-only	AI voice-only
Mean rating	8.84	8.09	8.81	8.99
Rating ≥ 7	95.2%	84.3%	96.1%	93.0%

Indicator	AI	Traditional	AI text-only	AI voice-only
Rating ≥ 8	85.6%	69.1%	83.9%	91.2%
Rating ≥ 9	66.2%	43.4%	65.0%	72.8%

The data in Table 4 show unambiguously that the AI modality offers a survey experience perceived as significantly superior to the traditional questionnaire. This advantage emerges for all indicators considered and is confirmed both descriptively and statistically.

The mean interview rating increases from 8.09 in the traditional mode to 8.84 in AI-a difference that goes beyond a marginal change and represents a substantial shift in overall evaluation. The fact that Welch's t test yields $p \approx 0.000$ indicates that this difference is highly significant and not attributable to chance.

Looking at the distribution of ratings, the AI modality not only shifts the mean upwards but also concentrates ratings more strongly in the higher bands. The proportion of respondents assigning a score of 7 or higher, 8 or higher, and 9 or higher is clearly higher than in the traditional administration, with differences ranging between 10 and 23 percentage points. This dynamic suggests an improvement not only in the general perception of the tool but also in deeper satisfaction with the experience offered.

A particularly noteworthy element concerns the comparison between AI text-only and AI voice-only. Although both channels perform well above the traditional mode, the voice modality shows a further increase in the share of high ratings (especially ≥8 and ≥9), indicating that spoken interaction can be perceived as more natural, fluid, and engaging. However, the p-values associated with these differences ($p = 0.1705$ for ≥7; $p = 0.0526$ for ≥8; $p = 0.1231$ for ≥9) suggest that the advantage of voice, though consistent and interesting, does not always reach conventional statistical significance.

Overall, these results highlight that AI not only improves the informational quality of responses but also elevates the participant experience, making the process more pleasant, intuitive, and rewarding. The systematic increase in high ratings and the significance of the observed differences indicate that the conversational mode changes how the questionnaire is perceived: from a technical form-filling task to a guided interaction, closer to a structured conversation than to traditional completion.

The differences between AI text-only and AI voice-only, by contrast, although showing a pattern consistently favorable to voice, are less clear-cut:

- $p = 0.1705$ for rating ≥7
- $p = 0.0526$ for rating ≥8
- $p = 0.1231$ for rating ≥9
- $p = 0.235$ for the mean rating.

3.5 Evaluation of Specific Aspects of the Experience

Table 5 – Top-Box Shares

Item	AI	Traditional	AI text-only	AI voice-only
I felt at ease	89.8%	80.4%	90.3%	92.9%
I like answering questionnaires of this kind	89.0%	70.2%	87.7%	87.7%
I had fun answering	89.8%	70.0%	89.7%	90.4%
I felt listened to / understood	86.6%	67.5%	86.1%	86.8%
It was easy to answer	92.6%	75.7%	92.5%	92.9%
I had the impression that my opinion really mattered	86.7%	77.9%	90.0%	86.8%
I did not enjoy taking this questionnaire (negative item)	16.9%	19.9%	16.9%	12.3%
I was not able to express myself well (negative item)	25.6%	32.6%	25.6%	21.1%

The evaluation of experiential items clearly confirms the advantage of the AI modality over the traditional questionnaire, with differences that are not only descriptive but also consistent with the qualitative patterns emerging from open-ended responses.

First, indicators related to comfort and emotional engagement show a substantial improvement: the share of people who felt “at ease” rises from 80.4% in the traditional mode to 89.8% in AI, reaching 92.9% among voice users. This suggests that the conversational interaction reduces the perceived distance between interviewer and respondent, fostering a more natural and less formal participation climate.

Similar results emerge for items measuring the pleasantness of the experience (“I like answering questionnaires of this kind” and “I had fun answering”), in which AI exceeds the traditional mode by nearly 20 percentage points. The voice modality adds a further element of spontaneity that contributes to making the experience more fluid and less “mechanical.”

Differences in relational indicators, such as the feeling of being listened to and understood, are also very relevant. Here, AI records values almost 20 points higher than the traditional mode (86.6% vs. 67.5%), confirming that conversational follow-ups are interpreted by users as signals of attention and interest in their opinions.

Cognitive accessibility, as measured by the item “It was easy to answer,” shows one of the sharpest differences: AI exceeds 92%, while the traditional mode stops at 75.7%. This suggests that the conversational interface, by guiding thought formulation and implicitly clarifying what is expected from a response, reduces the effort required of participants.

The negative items confirm the trend: the share of users stating that they did not enjoy the questionnaire or were not able to express themselves well is noticeably lower in the AI modality and decreases further in the voice modality. In particular, only 12.3% of voice users say they did not enjoy the experience, compared with 19.9% in the traditional mode.

Overall, the AI modality-and even more so the voice modality-not only improves the informational quality of responses but also transforms the very perception of the interview, making it more welcoming, engaging, and cognitively accessible. This experiential improvement is methodologically relevant: a more comfortable respondent is likely to be more motivated, more accurate, and less prone to cognitive fatigue, with positive effects on the quality of the data collected.

QUALITATIVE ANALYSIS

Qualitative analysis of the content, conducted through a systematic comparison of responses from the two modes, reveals a rich, multi-layered picture that goes well beyond the simple length differences observed in quantitative measures. The two modes generate surprisingly similar thematic sets, confirming that conversational technology does not alter the nature of the opinions expressed. What changes radically is the form of discourse: the AI modality, both in its text and voice variants, induces an evolution in the cognitive and linguistic structure of responses.

A first element concerns the depth and clarity of content. Responses obtained in the AI modality-and even more so those provided via voice-show a spontaneous tendency to make explicit the reasons behind a statement [AI mode - voice: "*Yes, I read a review on TripAdvisor from a guy-he seemed like a guy-who said the place was very nice but it was far away and not well signposted. Well, this is an example of a very positive review because it shows that you might like the service, the product, so to speak. At the same time, however, you should know that it has some significant flaws, which are good to know at the outset so you can be prepared or feel a little safe.*"]. Instead of merely indicating a desired attribute in a review, respondents explain how they interpret it, which risks they intend to avoid, which benefits they seek, and how that piece of information fits into their decision-making process. The text version of the AI path reflects this improvement, but voice amplifies it further, likely because spoken language favors the natural articulation of reasoning and enables more fluid discursive production. What remains an implicit link in text often becomes an explicit logical chain in voice.

The variety of perspectives displays a similar pattern. The AI prompting system encourages respondents to explore interpretative angles that remain latent in the traditional questionnaire. In the AI text modality, usage conditions, interpretation criteria, and situational distinctions emerge. However, it is in the voice modality that these perspectives become more layered: people tend to evoke personal experiences, recount specific episodes, and mention exceptions and edge cases, introducing contextual elements that enrich understanding of the phenomenon. Voice promotes a more natural narrative, with a weave of details and nuances that the written mode-although improved compared to traditional-rarely matches.

A particularly interesting aspect is the capacity of the two AI variants to generate actionable insights. While the traditional questionnaire often produces lists of attributes or concise assessments, the AI modality encourages the formulation of operational criteria. Once again, voice responses are richer: respondents explain how they weigh conflicting reviews, what convinces them of a reviewer's credibility, which signals they consider "red flags," and what type of information they deem sufficient to move to action. In the traditional path, insights are present but less articulated: the absence of follow-ups often prevents reaching a level of detail sufficient to transform responses into operational recommendations. [AI mode - text: "The review must be relevant to what we're talking about. For example, if I read a one-star review, but it was given because the manager is unpleasant, I'll ignore it. It makes no sense. A place can be high quality, and liking is subjective. *"* - *Traditional mode - "Reviewer profile, details, content"*]

Argumentative quality represents another point of divergence between modes. Voice responses show a strong tendency toward the construction of causal chains, justification of choices, and articulation of the conditions that influence a judgment. The AI text modality follows the same principle but in a more structured and less narrative form. In the traditional path, by contrast, arguments are more frequently reduced to isolated statements, lacking the logical links that characterize AI-mediated discourse. AI thus appears to favor the transformation of thought into a fully developed argumentative discourse, while voice, once again, enhances this dynamic and makes it more spontaneous. [AI mode - voice: "I left a very positive review at a hotel I stayed at in Tuscany because the service was excellent, which I wasn't expecting given the low-star rating. However, it was extremely clean and the management was very courteous. I left a negative review at a restaurant I booked through TheFork, which was very rude to us. The management and waiters were rude." - *Traditional mode - " a good experience pushes me to write a good review"*]

Finally, signals of cognitive fatigue reveal a sharp divide between modes. In the traditional questionnaire, vague responses, formulas such as "I don't know," truncated statements, or sequences of unconnected words occur more frequently. The AI modality drastically reduces these signals. Conversational interaction mitigates uncertainty and encourages respondents to reformulate even initially vague answers, turning content of low informational value into a clearer, more detailed expression of their thinking. Voice takes this mechanism to its highest level: vague responses, indecision, and signs of confusion that appear in the traditional questionnaire are almost completely absent. The act of speaking-free from the need to type and simultaneously monitor form-allows thought to be articulated naturally, producing richer, more coherent, and more authentic content.

DISCUSSION

Taken together, the results consistently show that conversational AI introduces a structural change in the completion experience and in the quality of the data produced. The benefits do not lie simply in increasing the length of responses but in activating cognitive processes that make thinking more explicit, more articulated, and more contextualized.

The voice modality amplifies these effects further, bringing the survey experience closer to that of an interview while remaining within the constraints of quantitative methods. This is particularly relevant for researchers who want rich data at contained costs: voice, integrated with an AI interface, offers a unique compromise between the spontaneity of spoken language and the standardization of the questionnaire.

The experiential dimension also appears to be profoundly transformed. The widespread perception of being listened to and understood is not just a subjective evaluation but has methodological implications, as it can reduce respondent fatigue and increase the motivation to provide accurate answers. AI does not simply "collect data"; it creates a conversational context that encourages participants to engage cognitively, improving the quality of the final data.

The statistical significance of the main indicators-semantic cohesion, argumentative depth, and overall satisfaction-strengthens the interpretation of a structural, rather than contingent, effect of conversational AI: the observed differences are not compatible with simple sampling fluctuations but point to a different way of thinking and narrating activated by interaction with the interface.

LIMITATIONS

The comparison between modes is based on two different panels, and although weighting reduces demographic distortions, selection effects cannot be completely ruled out. The voice modality was also chosen spontaneously by a limited portion of the AI sample, which calls for caution in interpreting some results. Future studies could incorporate randomized designs and larger samples to better isolate the specific effects of conversation and the voice channel.

The significance analyses conducted on semantic cohesion and overall satisfaction provide robust evidence on the main indicators but do not systematically cover all experiential items and all possible interactions between socio-demographic profiles and response modality. In particular, profile differences between those who choose text and those who choose voice, although sometimes significant on individual variables, are not sufficient to support strong causal inferences about channel preferences; more controlled experimental designs are needed to separate self-selection effects from those purely attributable to the administration mode.

CONCLUSIONS

The study shows that the introduction of an AI-based conversational interface can substantially improve the quality of open-ended responses in online surveys without altering the thematic structure of expressed content. The voice modality, in particular, produces richer, more natural, and more coherent discourse, bringing quantitative surveys closer to qualitative interviews.

The combination of cognitive depth, greater ease of expression, and improved overall participant experience suggests that AI could become a structural component of future research methodologies. This is not merely a technological evolution but a rethinking of how people interact with the survey tool and how their opinions can be collected in a more authentic, effective, and informative way.

Appendix 1. Questionnaire

[The questionnaire was administered in Italian]

Q. A1
In the past year, have you organized any vacations outside your city?
[Please consider only vacations where you spent at least 2 nights away from home and **not** staying at friends' or relatives' homes]

- I took at least one vacation outside Europe
- I took at least one vacation in Europe
- I took at least one vacation in Italy
- I did not take any vacations of this type in the past year

[Rotate – "I did not take any vacations" always last and exclusive – If "I did not take any vacations" → TERMINATE]

Q. A2
Let's now talk about how you organize your vacations (booking accommodation, restaurants, activities, etc.).
How important are reviews from previous customers when choosing a service?

[Please express your opinion on a scale from 1 (not important at all) to 10 (very important)]

Q. A3
Let's continue talking about vacation planning and reviews.
What are the first websites/apps that come to mind where you read/write reviews?

The website/app that comes to mind is: _____
And then this one: _____
And also this one: _____

☐ I can't think of any website/app

[It is mandatory to enter text in the first text box, or to select "I can't think of any website/app"]

Below are the open-ended questions. In this version of the questionnaire, they are illustrated with static follow-up prompts.
In the conversational AI branch, the same wording will be kept for questions A4 and A5, but follow-ups will be handled automatically by Glaut's AI.

Q. A4
When you evaluate hotels, restaurants, or tourist attractions, which specific characteristics of a review (for example: content, style, level of detail, reviewer profile) make it truly useful in guiding your decision?

Text box _____

[If “Static” branch]

Q. A4b
You told us that what makes a review useful is:
“$Piped text from A4$.”
If you had to explain to a friend why these characteristics of reviews are important, what would you tell them?

Text box _____

Q. A4c
You have described the most important characteristics. Which characteristics of a review do you consider less important instead?

Text box _____

Q. A5
After a trip, what determines whether or not you leave a review for hotels, restaurants, or attractions you visited? [3]

Text box _____

[If “Static” branch]

Q. A5b
You told us that what determines whether or not you leave a review is:
“$Piped text from A5$.”
Can you give us a recent, concrete example to better clarify what you mean? [4]

Text box _____

Q. A5c
Comparing this episode with your overall experience, how representative is it of your typical behavior, and what instead changes from time to time?

Text box _____

Q. A6
Thank you for your answers so far!
We would now like to ask about your opinion of the questionnaire you have just completed.
How would you rate your overall experience so far in completing this questionnaire?

[Please express your opinion on a scale from 1 (lowest score) to 10 (highest score)]

Q. A7
Where are you right now?

- At home

- At friends'/relatives' home

- At work / school or university
- Out and about (on the street, etc.)
- On public transportation
- In a public place
- Somewhere else, namely: _____
- I prefer not to answer

Q. A8

Please indicate your level of agreement with the following statements. [5] [6]

	Not at all agree	Slightly agree	Neither agree nor disagree	Quite agree	Completely agree
I felt comfortable sharing information in this questionnaire	O	O	O	O	O
I like answering questionnaires with this type of questions	O	O	O	O	O
I enjoyed answering the questions	O	O	O	O	O
I felt listened to / understood	O	O	O	O	O
It was easy to answer the questions	O	O	O	O	O
I felt that my opinion really mattered	O	O	O	O	O
I did not enjoy doing this questionnaire [7] [8]	O	O	O	O	O
I felt that I was not able to express my opinions well	O	O	O	O	O