Do AI-generated follow-up questions elicit more informative responses than static questions alone?

Charlotte Doidge, Abbie-Louise Smith, Jasmeet Kaler

School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, LE12 5RD, UK

Introduction

The qualitative online survey is becoming an increasingly common approach to qualitative data generation due to its flexibility in addressing research questions, its capacity to capture diverse perspectives across geographically dispersed populations, and its ability to offer participants a sense of anonymity (Braun et al., 2021). These surveys prioritise open-ended questions to understand respondents' experiences, perceptions, and narratives. As respondents may feel a greater sense of anonymity with qualitative surveys compared with interviews and focus groups, they present a particularly useful data generation approach for sensitive topics such as the topic researched in this project, animal welfare, specifically, public perceptions around dairy calves' quality of life in the UK. Several studies have shown that the public are either unaware of, or reject, several practices related to calf husbandry (Placzek et al., 2021). One of the most contentious practices is the act of cow-calf separation, in which a calf is removed from their mother shortly after birth.

Whilst qualitative surveys can provide rich qualitative data, in some cases the data can be thin compared to other qualitative data generation approaches due to the lack of ability to probe or provide follow-up questions to respondents. For example, in a qualitative survey aiming to understand cattle veterinarians' experiences of technology use on farms, some respondents with limited experience of technology use provided relatively thin answers (Doidge et al., 2024). AI-powered surveys have been suggested as a potential solution to the current limitations in qualitative surveys (Williams, 2025). These AI-powered surveys integrate natural language processing (NLP) within a standard survey platform. Questions are presented in a familiar survey format, but responses to open-ended questions are interpreted in real time to generate relevant follow-up prompts, potentially enabling deeper, more conversational exploration of respondents' perspectives. This means that AI-powered surveys are a combination of researcher generated static questions and AI-generated follow-up questions.

Whilst it is suggested that AI-powered surveys could offer deeper insights, there is currently limited exploration on how respondents answer to AI-generated questions and whether these AI-generated questions elicit additional insights than static researcher-generated questions alone. Therefore, the aim of this study was to explore whether the AI-follow up questions provide us with more information than the static questions alone.

Study design, sample, and ethics

The Glaut platform was used to host the survey. This survey platform has the capability to develop AI-generated follow up questions according to the specific prompt used to instruct the AI on how to interpret participants' first reply. The topic of the survey was public perceptions of dairy calf welfare. In total there were six open ended, researcher-generated, questions in the survey, which we will call static questions for the purpose of this paper. For each of these static questions, AI-generated follow up questions were enabled. We provided prompts for these AI-generated follow-up questions and instructed that only 1 follow-up question was asked per static question. Some closed static questions were also included in the survey to capture the respondent demographics. The participants could choose to respond either via voice or text.

The Prolific platform was used to recruit a representative sample of respondents from the UK, based on gender, age, and region of residence. There was a total of 296 respondents who completed the survey. The average time for completing the survey was 13:04 minutes.

The study was approved by the University of Nottingham School of Veterinary Medicine and Science Ethics Committee (no. 4479 161225).

The data were analysed using several techniques to enable us to understand whether the AI-follow up questions provide us with more information than the static questions alone. This included:

- **Verbosity:** Analysis to understand how many extra words the response to the AI-follow up question generates compared with the static question.
- **Lexical diversity:** Analysis to understand the ratio of different words (types) to total words (tokens) the response to the AI-follow up question generates compared with the static question. (Lexical diversity)
- **Keyness**: Analysis to identify words that are disproportionately frequent in the responses to the AI-generated questions compared to the responses to the static questions (Schweinberger, 2026).

AI moderation elicits detailed elaboration with low repetition

Verbosity and lexical diversity were analysed using linear mixed-effects models. For these analyses, responses to the initial static question were compared with the combined responses (i.e., the initial static response plus the subsequent AI-generated follow-up response), allowing us to assess the extent to which the addition of the AI-generated question increased overall response length and diversity. Word count (tokens) and corrected type–token ratio (CTTR) were specified as outcome variables in separate models. The CTTR compares the number of unique words (types) to the total word count (tokens), and accounts for the total number of words in the text. Condition (static vs. static + AI-generated) was included as a fixed effect. For the lexical diversity analysis, word count was additionally included as a covariate to account for its influence on token-type ration. Random intercepts were specified for respondent and question to account for clustering. Model-adjusted marginal means were estimated for each condition. All analyses were conducted in Stata 19 (Stata SE/19.5; StataCorp, College Station, TX, USA).

Verbosity

The addition of the AI-generated question increased the word count by an average of 30 words ($p<0.001$; 95% CI 28.178, 31.727), compared with the initial researcher-posed static question (Table 1). After accounting for clustering at respondent and question level, the static question has an average word count of 41 and the addition of the AI-moderated question has an average word count of 71 (Table 2). This means that the inclusion of the AI-generated question led to a 75% increase in verbosity in the response compared to the static question alone.

Table 1: Linear mixed-effects model predicting word count

Component	Variable	Estimate	Std. Error	p-value	95% CI
Fixed effects	AI-moderated (vs static only)	29.952	0.905	<0.001	28.178, 31.727
	Intercept	41.159	2.625	<0.001	36.014, 46.304
Random effects	Respondent	1834.739	162.362	-	1542.584, 2182.227
	Question	474.297	32.946	-	413.927, 543.472
	Residual	701.482	24.011	-	655.965, 750.156

Table 2: Model-adjusted marginal means of word count by condition

Condition	Predicted mean tokens	Std. Error	p-value	95% CI
Static response	41.16	2.63	<0.001	36.01, 46.30
Static + AI-moderated	71.11	2.62	<0.001	65.98, 76.25

Lexical diversity

The addition of the AI-generated question increased the CTTR by an average of 0.318 points ($p<0.001$; 95% CI: 0.291, 0.344), compared with the initial researcher-posed static question (Table 3). After holding response length constant and account for clustering at respondent and questions level – the initial researcher-posed question has an average CTTR of 3.43 and the addition of the AI-generated question has an average CTTR of 3.75 (Table 4). Therefore, the addition of the AI-moderated question increased the lexical diversity by 9% on average. This indicates that the **respondents were using significantly richer and more varied vocabulary with the addition of the response to the AI-generated questions.**

Table 3: Linear mixed-effects model predicting lexical diversity (corrected type-token ratio - CTTR)

Component	Predictor / Parameter	Estimate	Std. Error	p-value	95% CI
Fixed effects	AI-moderated (vs initial)	0.318	0.013	<0.001	0.291, 0.344
	Word count (tokens)	0.010	<0.001	<0.001	0.009, 0.010
	Intercept	2.880	0.026	<0.001	2.829, 2.932
Random effects	Respondent	0.140	0.014	—	0.115, 0.172
	Question	0.109	0.007	—	0.096, 0.123
	Residual	0.117	0.004	—	0.109, 0.126

Table 4: Model-adjusted marginal means of lexical diversity (CTTR) by condition

Condition	Predicted CTTR	Std. Error	p-value	95% CI
Initial response	3.43	0.025	<0.001	3.38, 3.48
Initial + AI-moderated	3.75	0.025	<0.001	3.70, 3.80

Taken together, the analyses of verbosity and lexical diversity shows that **the addition of the AI-generated question increases the word count significantly, and this is not due to respondents just repeating their response to the static question.** Instead, it indicates that the **AI-generated questions were able to elicit more detailed elaboration by introducing new diverse vocabulary.**

AI moderation elicits discussion of new topics (sometimes)

To further understand how the AI-generated questions introduce new, diverse vocabulary, we conducted a keyness analysis on each question. This was used to identify words that were disproportionately frequent in AI-generated follow-up question responses compared to static question responses. This was conducted in R (version 4.5.3.) using the quanteda package (Benoit et al., 2018). To calculate keyness, we used the log-likelihood ratio statistic (G^2), which is the keyness measure that is the most generally recommended in corpus linguistics (Schweinberger, 2026). Prior to analysis, standard text preprocessing was applied, including lowercasing, removal of punctuation and numbers, and removal of common English stopwords. Very low-frequency terms (n<3) were trimmed to reduce noise.

Words with the highest keyness values for the AI-generated question condition were examined and grouped into semantically coherent clusters representing candidate topics. To assess whether candidate topics reflected substantive topic expansion rather than increased verbosity, we calculated the proportion of respondents mentioning each topic in both conditions. Topics were considered "new" or "newly salient" only when they were expressed by substantially more respondents following the AI-generated follow-up and were qualitatively absent or rare in original responses.

We will now present the keyness results for each question. This includes reporting the top 20 words that were used significantly more in the responses to the AI-generated question compared with the static question and a description of any new topics that appeared to be expressed in the responses to the AI-generated question.

Question 2: What do you think a good life for a dairy calf looks like?

For the question "What do you think a good life for a dairy calf looks like?", the top 20 words that were used more in the responses to the AI-generated question compared with the static question are presented in Figure 1.

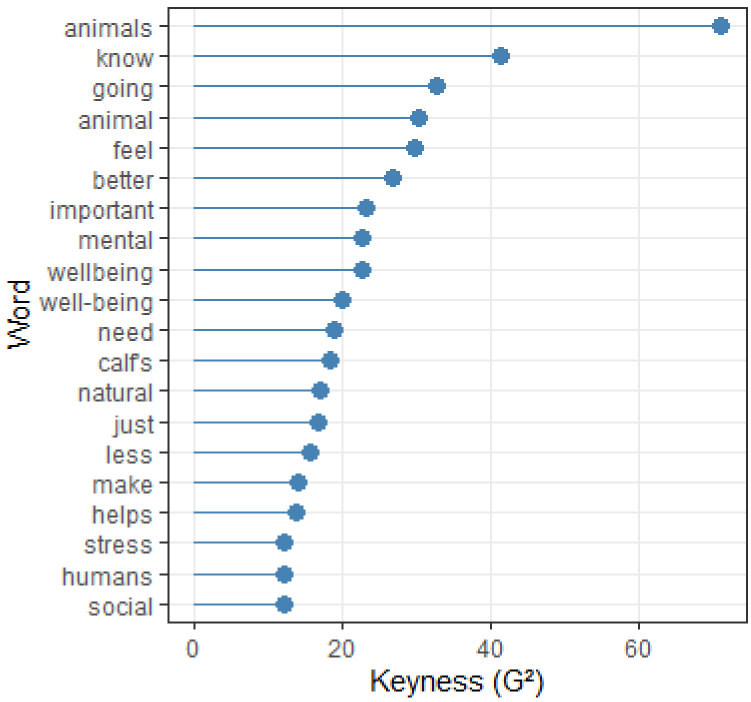

Figure 1: Top 20 keywords in the AI follow up responses for question 2, compared to the static responses and sorted by G^2 (log-likelihood)

Wellbeing-related considerations (e.g. wellbeing, mental, stress) were rarely present in static question responses but became prominent following the AI-generated follow-up. While a small number of respondents referenced terms related to wellbeing in the static condition (n = 10), this topic was expressed by substantially more respondents after the AI follow-up (n = 110), suggesting the **prompt elicited reflection in a domain that was otherwise under-articulated.**

We believe that the AI-generated question was able to elicit this new subject matter as the question "What do you think a good life for a dairy calf looks like?" was very broad and the topic of calf welfare is probably not salient to members of the public in the UK. The prompt we used to develop the AI-generated questions included asking to obtain examples of what dairy calves need in order for them to have a good life, how these examples contribute to a good life, and reasons for their answers. Therefore, respondents were likely to expand on their initial answer by giving reasons related to the wellbeing of the calf.

Question 3: Why is it important that dairy calves are provided with a good life (if at all)?

For the question "Why is it important that dairy calves are provided with a good life (if at all)?" there were thirteen words that were used more in the responses to the AI-generated question compared with the static question (Figure 2).

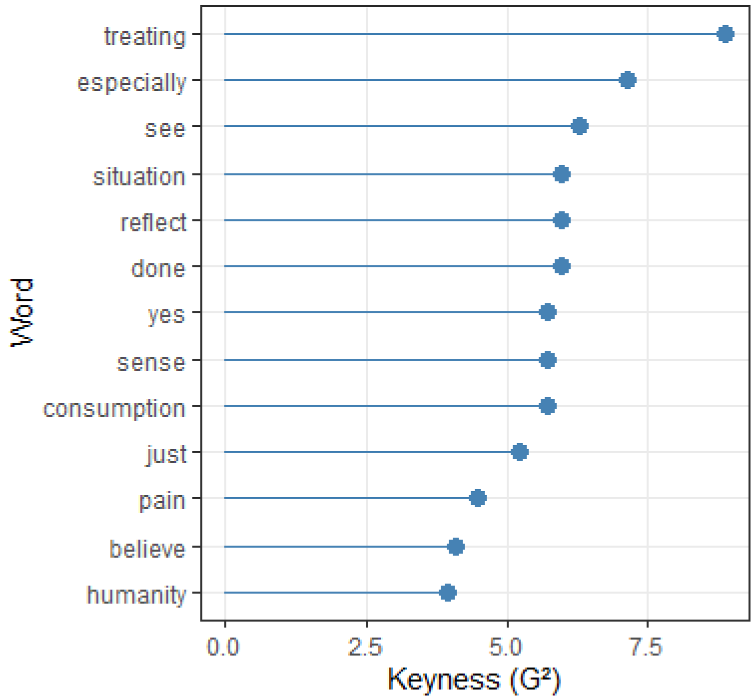

Figure 2: The thirteen keywords that feature significantly more in the AI follow up responses for question 3, compared to the static responses. Sorted by G2 (log-likelihood).

No new thematic considerations were presented in the AI-generated question follow up responses. Whilst there were some words that did appear more in the AI follow up responses (e.g. treating, especially, see), a common topic was not identified across these words.

For this question, the AI follow-up prompts were designed to explore the underlying reasons and motivations behind respondents' answers. We believe that this did not elicit much new information because the static question already asks "why" the respondents thought a specific way, which implies that the respondents would give reasons and motivations as their answer to the static question.

Question 4: Who do you believe is primarily responsible for ensuring that dairy calves are cared for so that they have a good life?

For the question "Who do you believe is primarily responsible for ensuring that dairy calves are cared for so that they have a good life?", the top 20 words that were used more in the responses to the AI-generated question compared with the static question are presented in

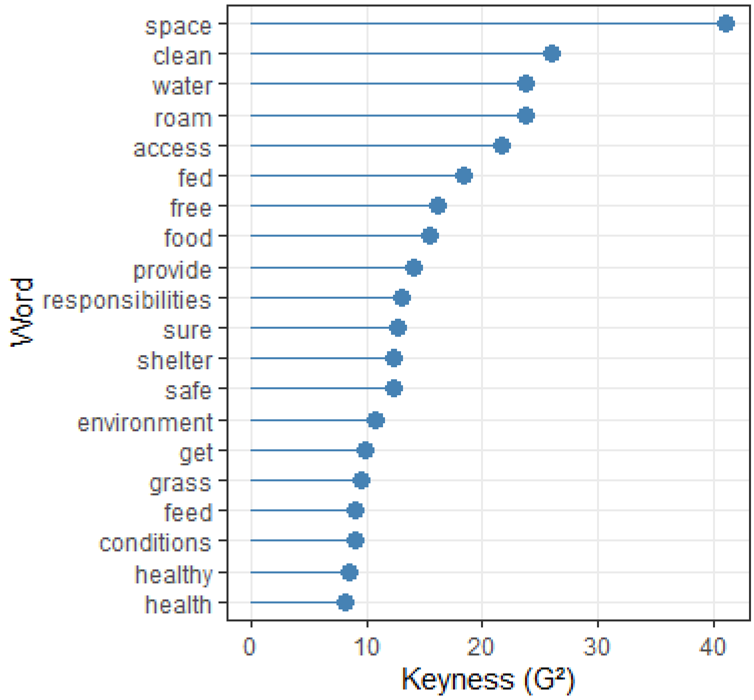

Figure 3: Top 20 keywords that feature significantly more in the AI follow up responses for question 4, compared to the static responses. Sorted by G2 (log-likelihood).

Nutrition related considerations (e.g. feed, water, fed, eat) were more prominent in the response to the AI-generated question compared to the static question responses. While a small number of respondents referenced nutrition related considerations in the static condition (n = 48), these themes were expressed by substantially more respondents after the AI follow-up (n = 150), suggesting **the prompt elicited reflection in a domain that was otherwise under expressed.**

Environment related considerations (e.g. space, shelter, environment) were more prominent in the response to the AI-generated follow up question compared to the static question responses. While a small number of respondents referenced nutrition related considerations in the static condition (n = 6), these themes were expressed by more respondents after the AI follow-up (n = 81).

We believe that this new information was elicited as the AI follow-up prompts were designed to explore the underlying reasons and motivations behind respondents' answers. **These follow-up questions may have elicited additional information on factors such as nutrition and the environment, as they sought different types of information compared to the initial static question, which primarily identified the stakeholders involved in calf care (e.g. farmers, veterinarians).**

Question 5: How do you think the quality of life of dairy calves in the United Kingdom compares with dairy calves in other countries?

For the question "How do you think the quality of life of dairy calves in the United Kingdom compares with dairy calves in other countries?", the top 20 words that were used more in the responses to the AI-generated question compared with the static question are presented in Figure 4.

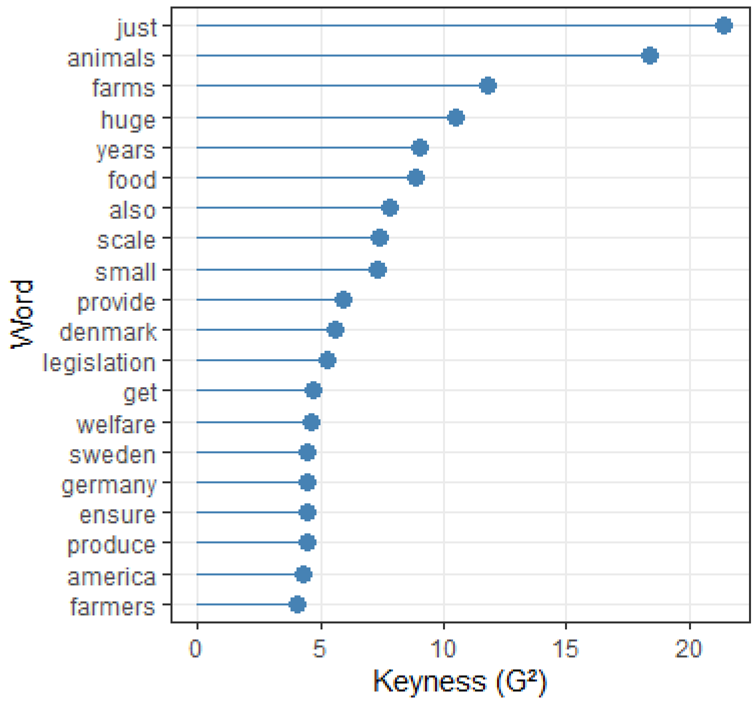

Figure 4: Top 20 keywords that feature significantly more in the AI follow up responses for question 5, compared to the static responses. Sorted by G2 (log-likelihood).

There were slightly more country related considerations (e.g. America, Sweden, Denmark, Germany) in the AI-generated follow up question compared to the static responses. Country related considerations were referenced by 11 respondents in the static condition and 26 respondents in the AI follow-up condition.

One of the instructions to the AI to generate the question was "Identify whether participants think other countries are better or worse than the UK at providing dairy calves with a good life" and this likely led to respondents providing more names of countries after the AI-generated follow up question. **The effectiveness of the AI-generated follow up question to elicit new information likely depends on the instructions provided to the software by the researcher to develop the AI-generated question.**

Question 8: Do you think this practice (cow-calf separation) has a positive or negative influence on the calf's ability to have a good life? Please provide reasons for your answer

For the question "Do you think this practice (cow-calf separation) has a positive or negative influence on the calf's ability to have a good life?", the top 20 words that were used more in the responses to the AI-generated question compared with the static question are presented in Figure 5.

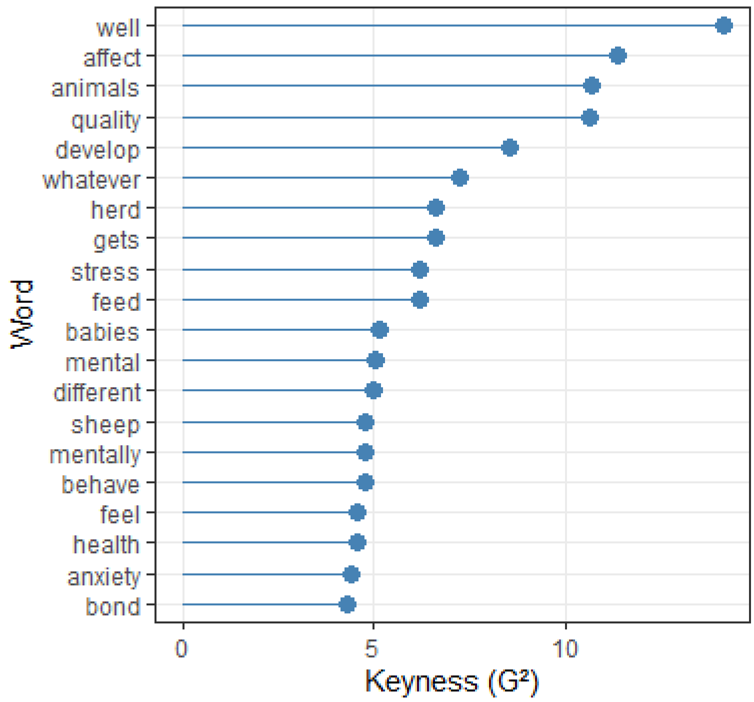

Figure 5: Top 20 keywords that feature significantly more in the AI follow up responses for question 8, compared to the static responses. Sorted by G2 (log-likelihood).

For this question, a similar wellbeing topic to question 2 was identified through keyness analysis (e.g. anxiety, mental, stress). However, when the qualitative data were inspected at the respondent level, the AI-generated follow-up did not meaningfully increase the proportion of respondents engaging with the topic (AI = 72, static = 71), suggesting **reinforcement rather than thematic expansion.**

Question 10: Do you think this practice (disbudding) has a positive or a negative influence on the calf's ability to have a good life? Please provide reasons for your answer

For the question "Do you think this practice (disbudding) has a positive or negative influence on the calf's ability to have a good life?", the top 20 words that were used more in the responses to the AI-generated question compared with the static question are presented in Figure 6.

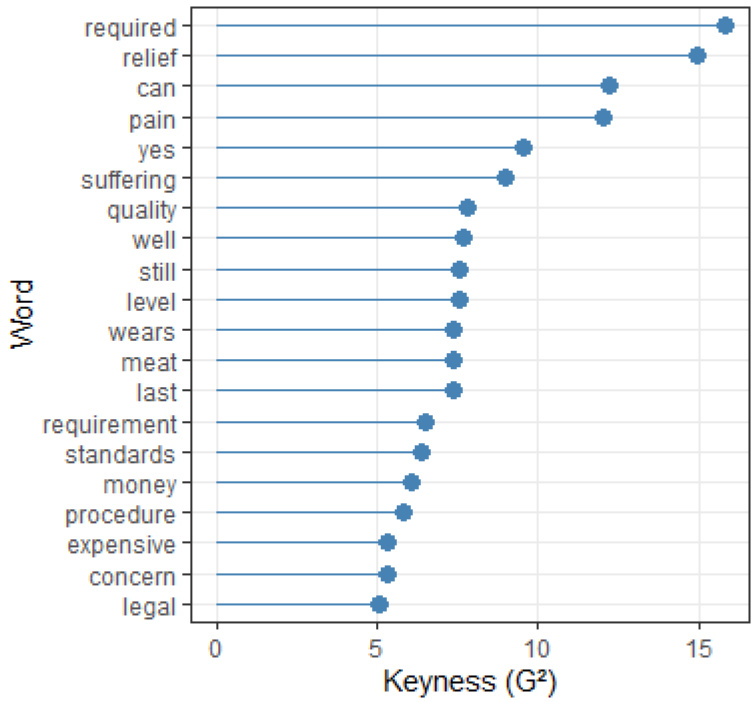

Figure 6: Top 20 keywords that feature significantly more in the AI follow up responses for question 10, compared to the static responses. Sorted by G2 (log-likelihood).

There were slightly more requirement related considerations (e.g. legal, standards, requirements) in the AI-generated follow up question compared to the static responses. Requirement related considerations were referenced by 14 respondents in the static condition and 24 respondents in the AI follow-up condition. However, this difference was not statistically significant.

We believe that the AI-generated questions **for question 8 and 10 elicited topics that reinforced the responses to the static questions, rather than eliciting new information,** was because we asked the respondents to provide their reasons to their answer in the initial static question. The prompt we used to develop the AI-generated questions also asked to obtain reasons for the respondents' answers and examples of how these practices contribute to a calf's good life. This may have led to tautological follow-up questions in some instances. This might especially be the case for these last two questions because they asked respondents about their views on much more specific topics (i.e. specific calf management practices) than the first few questions.

Take home message

Overall, the analysis of word count (verbosity) and lexical diversity showed that the addition of the AI-generated questions elicited longer responses that were richer and more diverse in vocabulary. This indicates that the respondents were willing to share their views on a potentially sensitive, and in some cases, unfamiliar, topic. The keyness analysis explored this diversity further by identifying keywords in the responses to the AI-generated questions that had a significantly higher frequency than in the responses to the static questions. We show that the AI-generated questions were able to elicit new topics of discussion for some, but not all, questions. This led us to the conclusion that the effectiveness of the AI-generated questions to provoke more informative responses depends on how the static researcher-generated question is worded and the instructions that are given to the AI. We can make the following recommendations:

AI-generated questions work well when:

- The initial question is broad or abstract
- The topic under investigation is not familiar to the respondents
- The AI is instructed to explore novel areas
- Respondents have not yet explained their reasoning for their response

AI-generated questions are less informative when:

- The initial question is already very specific
- The AI is instructed to ask questions that are tautological to the initial question
- Respondents have already fully articulated their reasoning

This means that:

- Question design is very important for the success of an AI moderation study
- Researchers should carefully consider how their instructions for the AI-generated questions complement their researcher-posed questions to ensure they make a meaningful contribution.

We suggest that future research into this method should focus on how to make the most effective and meaningful combination of static researcher-posed and AI-generated follow-up questions.

References

Benoit, K., K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo. 2018. quanteda: An R package for the quantitative analysis of textual data. Journal of Open Source Software 3(30):774-774.

Braun, V., V. Clarke, E. Boulton, L. Davey, and C. McEvoy. 2021. The online survey as a qualitative research tool. International Journal of Social Research Methodology 24(6):641-654.

Doidge, C., A. Burrell, G. van Schaik, and J. Kaler. 2024. A qualitative survey approach to investigating beef and dairy veterinarians' needs in relation to technologies on farms. animal:101124.

Placzek, M., I. Christoph-Schulz, and K. Barth. 2021. Public attitude towards cow-calf separation and other common practices of calf rearing in dairy farming—a review. Organic Agriculture 11(1):41-50.

Schweinberger, M. 2026. Keyness and Keyword Analysis in R. Vol. 2026.

Williams, R. T. 2025. The online survey as a qualitative research tool: Can AI be a probing tool? Frontiers in Research Metrics and Analytics 10:1519008.