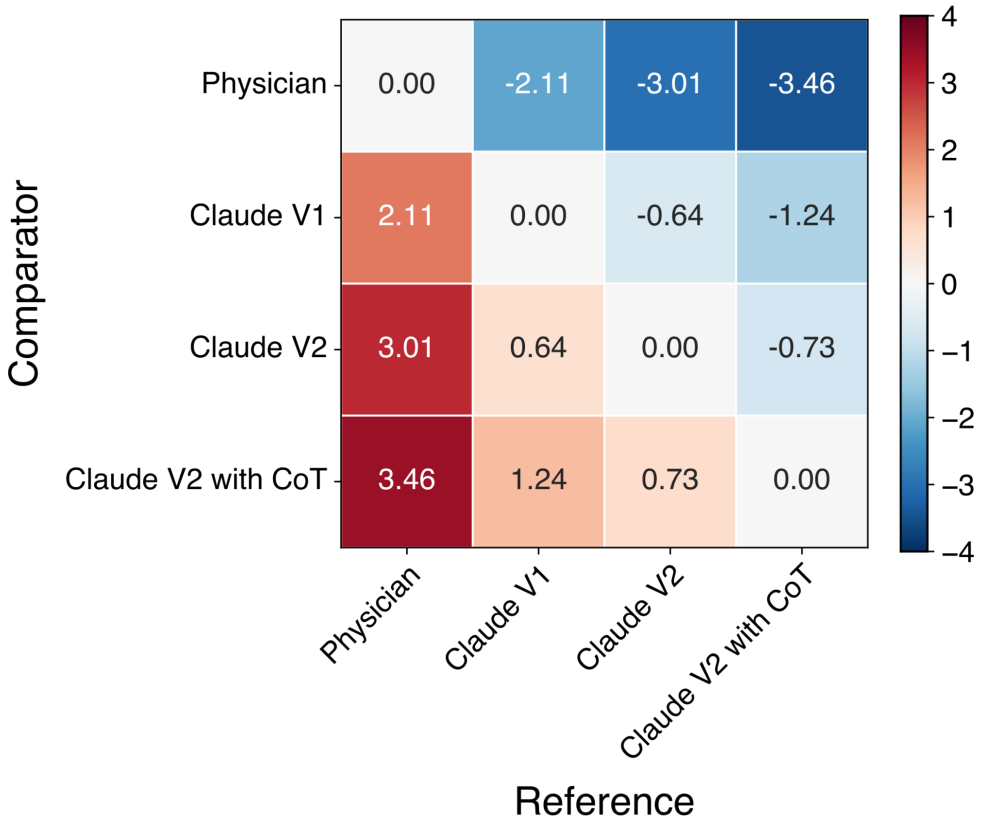
Supplementary Online Content

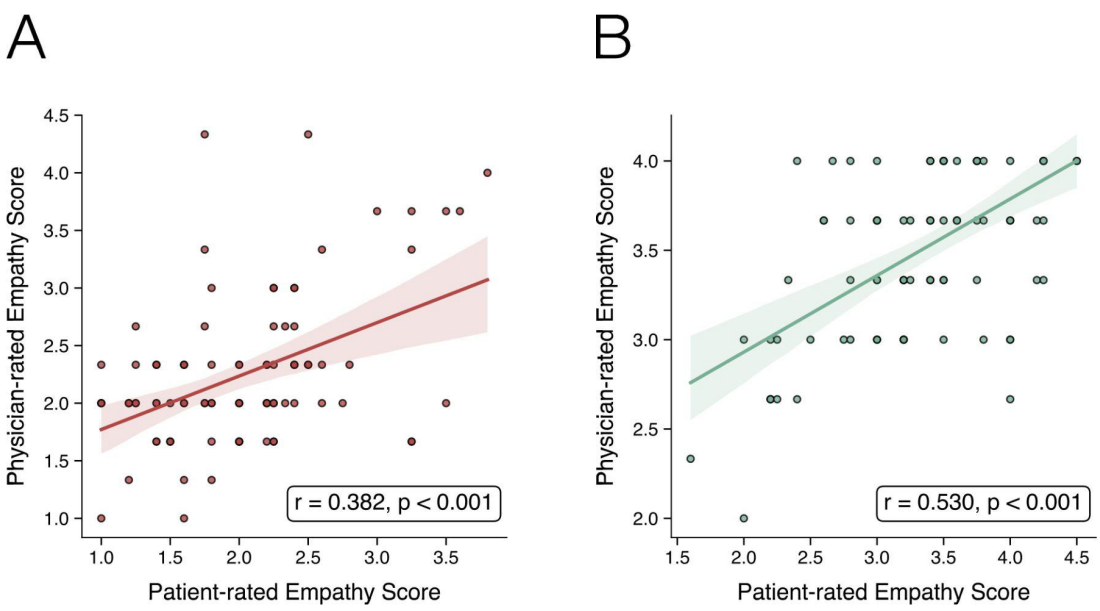
Chen, D., Chauhan, K., Parsa, R., Liu, A., Liu, F., Mak, E., Eng, L., Hannon, B.L., Croke, J., Hope, A., Fallah-Rad, N., Wong, P., Raman, S. Patient Perceptions of Empathy in Physician and Artificial Intelligence Chatbot Responses to Patient Questions about Cancer. *npj Digital Medicine.* Prepared on April 11, 2025.

This supplementary material has been provided by the authors to give readers additional information about their work.

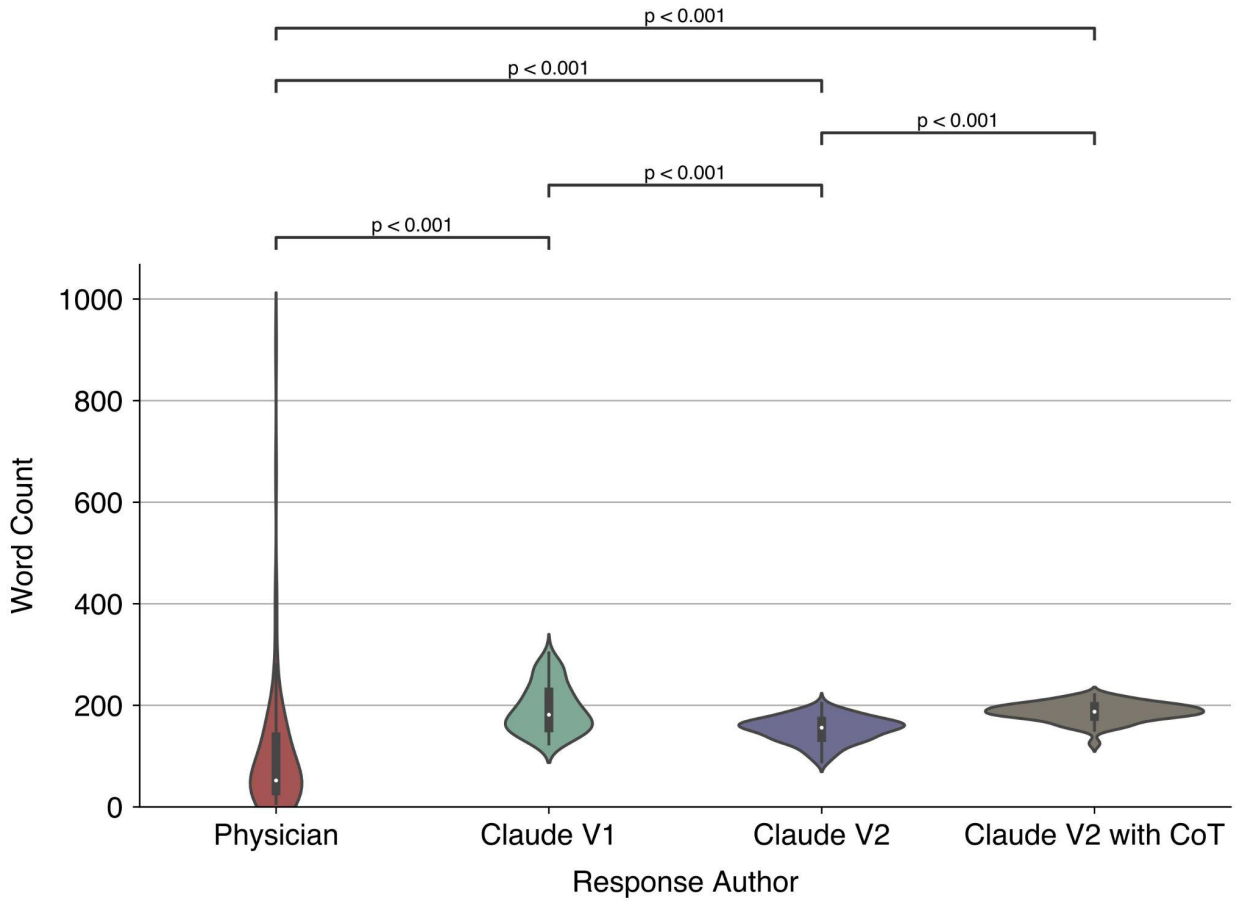
Supplementary Figure 1. Heatmap of Cohen's D effect size of pairwise comparisons of participant-rated empathy scores between chatbot and physician responses.



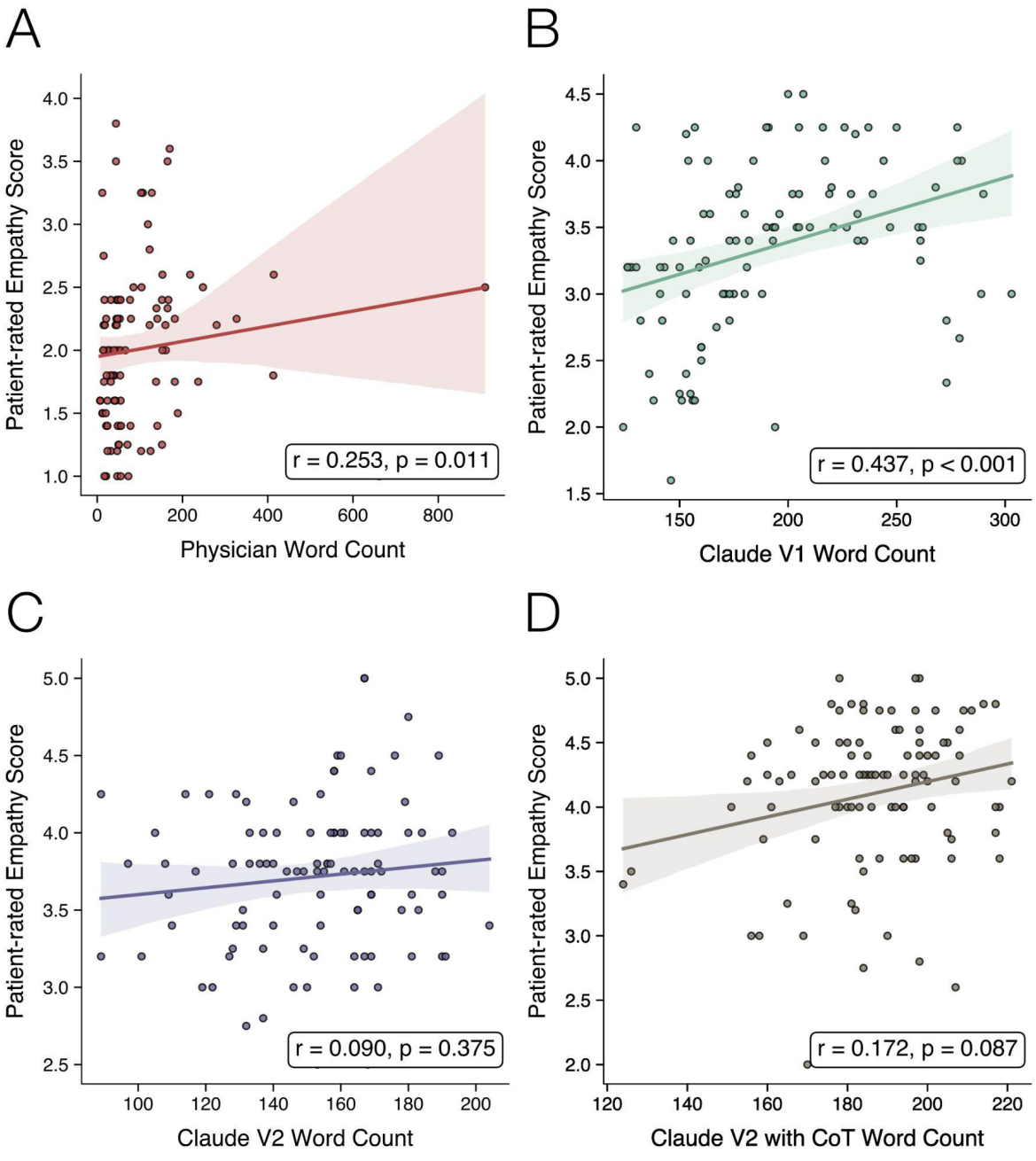
Supplementary Figure 2. Correlation between participant-rated and physician-rated empathy scores. A) Correlation between participant-rated and physician-rated empathy scores for physician responses. B) Correlation between participant-rated and physician-rated empathy scores for Claude V1 responses.



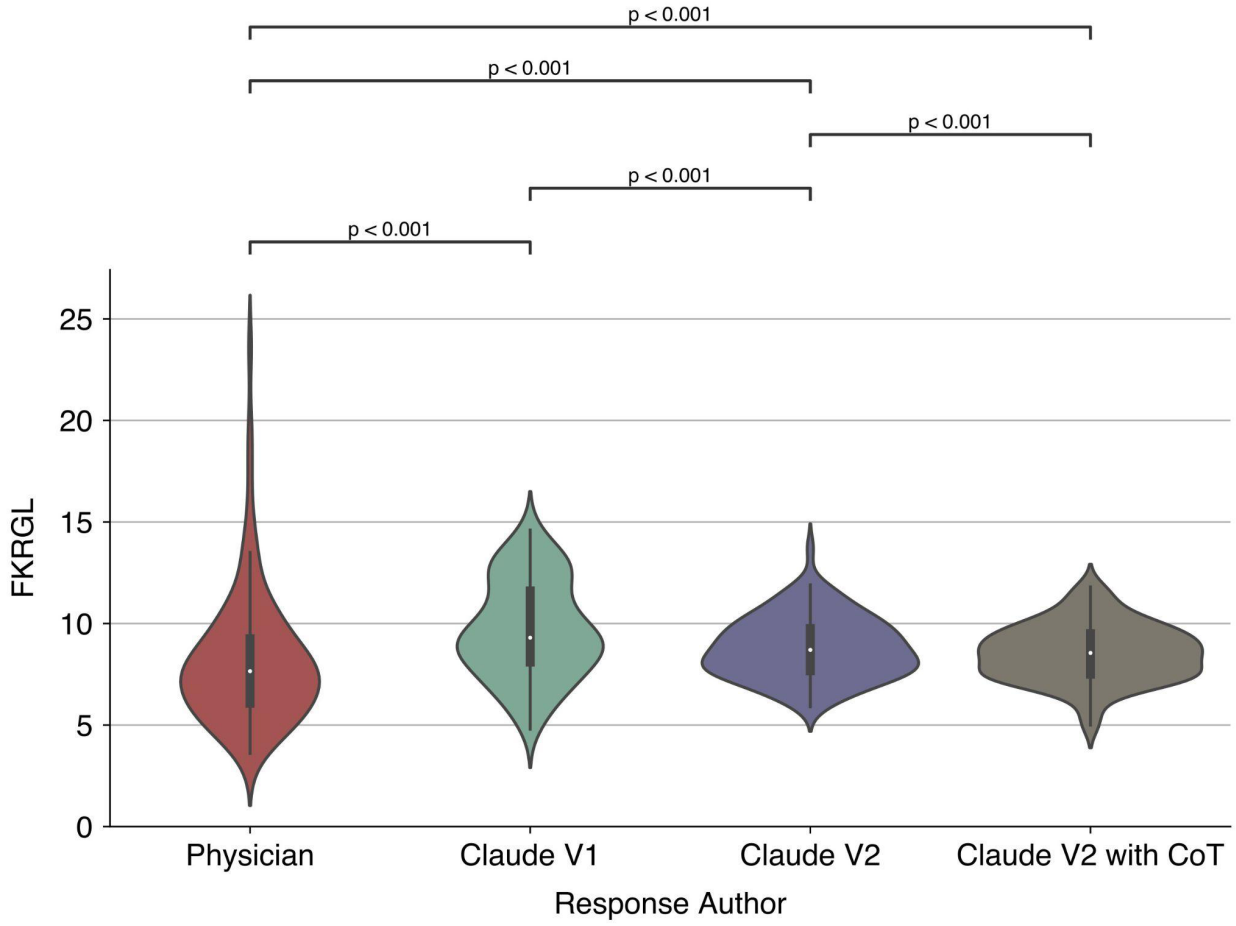
Supplementary Figure 3. Word count of physician and chatbot responses.



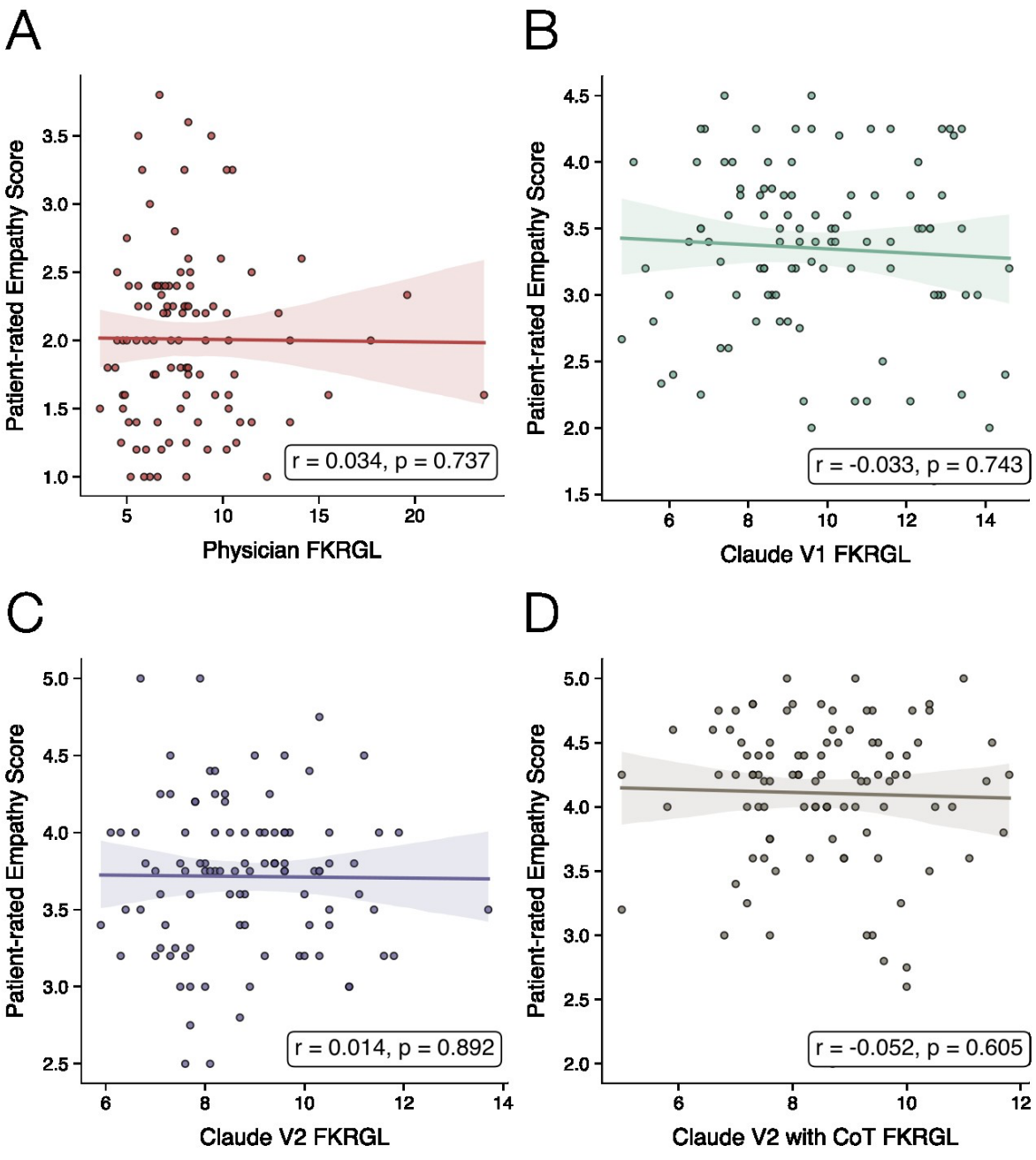
Supplementary Figure 4. Correlation between word count and participant-rated empathy scores. A) Correlation of word count and empathy scores of physician responses. B) Correlation of word count and empathy scores of Claude V1 responses. C) Correlation of word count and empathy scores of Claude V2 responses. D) Correlation of word count and empathy scores of Claude V2 with CoT responses.



Supplementary Figure 5. Flesch Kincaid Reading Grade Level (FKRGL) of physician and chatbot responses.



Supplementary Figure 6. Correlation between Flesch Kincaid Reading Grade Level (FKRGL) and participant-rated empathy scores. A) Correlation of FKGRL and empathy scores of physician responses. B) Correlation of FKGRL and empathy scores of Claude V1 responses. C) Correlation of FKGRL and empathy scores of Claude V2 responses. D) Correlation of FKGRL and empathy scores of Claude V2 with CoT responses.



Supplementary Table 1. Chain-of-thought (CoT) prompt-engineering procedure.

Prompt ID	Prompt
CoT-1	You are an oncology medical professional. Your goal is to generate high quality, empathetic, and readable responses to patients' questions about cancer. The intended audience is patients or their associates who are experiencing cancer.
CoT-2	Identify any word in the patient's question about cancer that represents the patient's emotion.
CoT-3	Emotional empathy is defined as the ability to share the affective experience of someone else. Cognitive empathy is defined as the ability to represent the mental states of someone else and their perspective in the resulting recommendation. Based on the patient's emotion, generate a response to the patient's question about cancer that is high in emotional empathy and cognitive empathy. Limit your response to 100 words.

Supplementary Table 2. Survey participant demographics.

Variable	Category	Count	Proportion
Age	18-39	1	2.22%
	40-64	11	24.44%
	65+	32	71.11%
	Undisclosed	1	2.22%
Income	<$100,000	10	22.22%
	>$100,000	29	64.44%
	Undisclosed	6	13.3%
Education	High School	6	13.33%
	Baccalaureate	17	37.78%
	Graduate	18	40.0%
	Undisclosed	4	8.89%
Race	White	40	88.89%
	Black	1	2.22%
	South Asian	1	2.22%
	Indigenous	1	2.22%
	Undisclosed	2	4.44%
Sex	Male	33	73.33%
	Female	10	22.22%

	Intersex	1	2.22%
	Undisclosed	1	2.22%
Gender	Male	34	75.56%
	Female	10	22.22%
	Undisclosed	1	2.22%

Supplementary Table 3. Descriptive statistics of participant empathy ratings, physician empathy ratings, word count, and Flesch Kincaid Reading Grade Level (FKRGL) of physician and chatbot responses.

Response Author	Participant-rated Empathy (Mean, 95% CI)	Physician-rated Empathy (Mean, 95% CI)†	Word Count (Mean, 95% CI)	FKRGL (Mean, 95% CI)
Physician	2.01 (1.88-2.13)	2.24 (2.11-2.37)	99.71 (74.34-125.08)	8.13 (7.50-8.76)
Claude V1	3.35 (3.23-3.48)	3.51 (3.42-3.60)	192.45 (183.35-201.55)	9.66 (9.19-10.13)
Claude V2	3.72 (3.62-3.81)	N/A	152.31 (147.46-157.16)	8.79 (8.49-9.08)
Claude V2 with CoT	4.11 (3.99-4.22)	N/A	186.72 (183.08-190.36)	8.55 (8.27-8.82)

† Oncology physician ratings of Claude V1 but not Claude V2 responses were analyzed in this study because only Claude V1 was released at the time of publication of the external physician rating dataset (PMID: 38753317).

Supplementary Table 4. Ordinary Least Squares (OLS) regression results evaluating the association between word count and participant empathy ratings of responses.

Response Author	Coefficient	Standard Error	T-statistic	P-value	CI Lower (95%)	CI Upper (95%)	R-squared
Physician	0.000602	0.000499	1.206935	0.230362	-0.000388	0.001592	0.014646
Claude V1	0.005092	0.001352	3.767084	0.000282	0.00241	0.007775	0.126489
Claude V2	0.002141	0.00201	1.06498	0.289502	-0.001848	0.00613	0.011441
Claude V2 with CoT	0.006929	0.003082	2.248569	0.026779	0.000814	0.013045	0.049061

Supplementary Note 1. Instructions to Patient Raters.

Study: Evaluating the Empathy of Physician and Artificial Intelligence Chatbot Responses from the Patient Perspective

You are being invited to participate in a research study.

The purpose of this study is to evaluate the empathy of the responses to patient questions by AI (Artificial Intelligence) chatbots as compared to physician responses from the perspective of patients who had or have a cancer diagnosis. The study will also compare the evaluation of chatbot empathy between physicians and patient-partners.

Taking part in the study is voluntary. If you agree to take part in the study, then you will follow the link below to complete a survey. Completing the survey will take up to one (1) hour. Please try to answer all the questions; however, you are free to omit any questions. The survey will ask you to rate responses to patient questions about their cancer-related healthcare on a scale. These responses can be from AI chatbots or physicians. You will be asked to rate responses based on their overall empathy. You will be asked to rate the overall empathy of the response on a Likert scale (1-5), where higher scores indicate greater empathy. Perception of overall empathy is based on emotional empathy (ability to share the affective experience of someone else) and cognitive empathy (ability to represent the mental states of someone else and their perspective in the resulting recommendation).

The data will be collected in an anonymous manner. The data will be stored in digital format at [redacted] for 1 year. Completing and submitting the survey will indicate your agreement to take part in the study.

There is no risk or benefit to you from taking part in the study. The study may help to improve responses by AI and physicians to patients in the future. Researchers have an interest in completing this study. Their interests should not influence your decision to participate in this study. Your data may be used for future publications. Your name and personal identifiers will not appear in any publication.

Please do not take this survey more than once.