

1 Exploring factors influencing user 2 perspective of ChatGPT as a 3 technology that assists in healthcare 4 decision making: A cross sectional 5 survey study.

6 Avishek Choudhury^{1a*}, Safa Elkefi^{2b}, and Achraf Tounsi^{3c}

7

8 ¹Industrial and Management Systems Engineering, Benjamin M. Statler College of Engineering and Mineral
9 Resources, West Virginia University, 1306 Evansdale Drive, Morgantown, West Virginia 26506, USA

10 ²Columbia University School of Nursing, Columbia University Irving Medical Center, New York, NY 10032,
11 USA

12 ³ADP, New Jersey, USA

13 ^a avishek.choudhury@mail.wvu.edu; [0000-0002-5342-0709](tel:0000-0002-5342-0709)

14 ^b se2649@cumc.columbia.edu;

15 ^c ash.tounsi@adp.com;

16

17

18 *** Corresponding:**

19 Avishek Choudhury (Ph.D.)

20 Assistant Professor

21 Industrial and Management Systems Engineering

22 Benjamin M. Statler College of Engineering and Mineral Resources

23 West Virginia University

24 1306 Evansdale Drive

25 Morgantown, West Virginia 26506

26 USA

27 avishek.choudhury@mail.wvu.edu;

28 <https://orcid.org/0000-0002-5342-0709>

29

30 Abstract

31 As ChatGPT emerges as a potential ally in healthcare decision-making, it is imperative to investigate
32 how users leverage and perceive it. The repurposing of technology is innovative but brings risks,
33 especially since AI's effectiveness depends on the data it's fed. In healthcare, where accuracy is critical,
34 ChatGPT might provide sound advice based on current medical knowledge, which could turn into
35 misinformation if its data sources later include erroneous information. Our study assesses user
36 perceptions of ChatGPT, particularly of those who used ChatGPT for healthcare-related queries. By
37 examining factors such as competence, reliability, transparency, trustworthiness, security, and
38 persuasiveness of ChatGPT, the research aimed to understand how users rely on ChatGPT for health-
39 related decision-making. A web-based survey was distributed to U.S. adults using ChatGPT at least once
40 a month. Data was collected from February to March 2023. Bayesian Linear Regression was used to
41 understand how much ChatGPT aids in informed decision-making. This analysis was conducted on
42 subsets of respondents, both those who used ChatGPT for healthcare decisions and those who did not.
43 Qualitative data from open-ended questions were analyzed using content analysis, with thematic coding
44 to extract public opinions on urban environmental policies. The coding process was validated through
45 inter-coder reliability assessments, achieving a Cohen's Kappa coefficient of 0.75. Six hundred and
46 seven individuals responded to the survey. Respondents were distributed across 306 US cities of which
47 20 participants were from rural cities. Of all the respondents, 44 used ChatGPT for health-related queries
48 and decision-making. While all users valued the content quality, privacy, and trustworthiness of
49 ChatGPT across different contexts, those using it for healthcare information place a greater emphasis on
50 safety, trust, and the depth of information. Conversely, users engaging with ChatGPT for non-healthcare
51 purposes prioritize usability, human-like interaction, and unbiased content. In conclusion our study

52 findings suggest a clear demarcation in user expectations and requirements from AI systems based on
53 the context of their use.

54 Introduction

55 In the discourse on the emergence and integration of artificial intelligence (AI) in daily life, the rise of
56 generative pre-trained transformers (GPT) like ChatGPT stands as a hallmark of innovation. As an AI
57 model developed by OpenAI, ChatGPT has garnered widespread attention and adoption for its ability to
58 generate human-like text, engaging in conversations and answering queries with a semblance of
59 understanding previously reserved for human intellect.

60 The application of ChatGPT extends well beyond its initial conception, echoing a common narrative in
61 the evolution of technology where tools are repurposed in manners unforeseen by their creators.
62 ChatGPT, while intended for conversational assistance, has been appropriated for diverse purposes, from
63 drafting legal and academic documents to creating artistic compositions [1-5]. The recontextualization of
64 technology, while innovative, surfaces inherent risks. As with any tool, the efficacy of AI is contingent
65 upon the parameters of its operation—parameters that are, in the case of AI like ChatGPT, defined by
66 data [6]. Such dependencies on data inputs introduces a temporal dimension to its reliability. The AI's
67 performance today may not be indicative of its performance in the future. The temporal variability is
68 crucial when considering ChatGPT's role in domains where accuracy is vital, such as healthcare. Just as
69 a treatment's efficacy may change over time with new medical discoveries, ChatGPT's responses are
70 subject to the ebb and flow of the data it consumes. A user may receive sound medical advice one
71 month, only to be misinformed the next, should the AI's data sources become tainted with erroneous
72 information [7].

73 User trust in technology is often built over time through consistent and reliable performance [8].

74 However, in the context of AI, this trust may become a liability if users become complacent, overlooking

75 the potential for AI responses to degrade as the data landscape shifts. User trust in such AI and the extent
76 to which they perceive it to be a helpful decision-making assistant depends on multiple factors such as
77 socio-ethical considerations, technical and design features, user characteristics, and expertise [9, 10].
78 When users are well-versed in the mechanics of ChatGPT and the principles guiding its responses, they
79 can navigate its capabilities with discernment, appropriately integrating it into their decision-making
80 processes. Conversely, misunderstanding of ChatGPT's functioning can result in hyped expectations and
81 distorted perception of the technology, leading to unwanted consequences when leveraged for critical
82 applications like healthcare. In other words, if the user is not qualified to validate ChatGPT's response,
83 the risk or probability of decision errors increases substantially. Furthermore, Trust in AI can also be
84 influenced by personal, organizational, and policy factors [7], as well as properties of the AI system,
85 including controllability, model complexity, embedded biases, and reliability [11]. Transparency is often
86 viewed as a prerequisite for trust in society, and the positive correlation between AI system transparency
87 and trust has been confirmed by previous empirical studies [12, 13].

88 ChatGPT's ability to deliver information persuasively also determines how and to what extent people
89 use it. While a convincing articulation can enhance user confidence in the AI's suggestions, it must be
90 carefully calibrated with the accuracy of the content provided. Persuasiveness without the foundation of
91 reliable and accurate information can lead to misplaced trust and potential misjudgments, especially in
92 high-stakes scenarios such as healthcare decision-making.

93 The convergence of reliable performance and security protocols also consolidates user trust. On
94 December 1, 2023, Google noted a critical flaw in ChatGPT highlighting the possibility of breaching its
95 training data [14]. GPT models, including ChatGPT, tend to memorize training data, which can lead to
96 privacy concerns [15]. This is particularly concerning if AI models are trained with personal
97 information, as it could lead to the exposure of sensitive data. While efforts are made to align AI models

98 like ChatGPT to prevent the release of large amounts of training data, there is still a risk of data breaches
99 through targeted attacks.

100 As ChatGPT emerges as a potential ally in healthcare decision-making, it is imperative to investigate
101 how users leverage and perceive it. Our study assesses user perceptions of ChatGPT, particularly of
102 those who used ChatGPT for healthcare-related queries. By examining factors such as competence,
103 reliability, transparency, trustworthiness, security, and persuasiveness of ChatGPT, the research aimed to
104 understand how users rely on ChatGPT for health-related decision-making. Additionally, this study
105 incorporates content analysis explore user concerns with ChatGPT. Our objective is particularly
106 important in the broader context of AI integration into healthcare systems. As AI technologies like
107 ChatGPT become more prevalent, understanding user perspectives on their effectiveness,
108 trustworthiness, and security becomes crucial. These insights are essential for developing AI tools that
109 are not only technically competent but also aligned with the expectations and concerns of end-users,
110 particularly in sensitive areas such as healthcare.

111 **Methods**

112 **Ethics statement**

113 The study, bearing the Institutional Review Board (IRB) protocol number 2302725983 and classified as
114 a flex protocol type, received approval from West Virginia University. No identifiers were collected
115 during the study. In compliance with ethical research practices, informed consent was obtained from all
116 participants before initiating the survey. Attached to the survey was a comprehensive cover letter
117 outlining the purpose of the study, the procedure involved, the approximate time to complete the survey,
118 and assurances of anonymity and confidentiality. It also emphasized that participation was completely
119 voluntary, and participants could withdraw at any time without any consequences. The cover letter also
120 included contact information of the researchers for any questions or concerns the participants might have

121 regarding the study. Participants were asked to read through this information carefully and were
122 instructed to proceed with the survey only if they understood and agreed to the terms described,
123 effectively providing their consent to participate in the study.

124 Data collection

125 We distributed a web-based semi structured survey to adults in the United States who actively use
126 ChatGPT at least once a month. We collected the data from February 1st, 2023, through March 30th,
127 2023. We conducted a soft launch of the survey and collected 40 responses. A soft launch is a small-
128 scale test of a survey before it is distributed to a larger audience. This soft launch aimed to identify any
129 potential issues with the survey, such as unclear or confusing questions, technical glitches, or other
130 problems that may affect the quality of the data collected. The survey was then distributed to a larger
131 audience across the US.

132 Instrument

133 The survey was designed on Qualtrics and was distributed by Centiment, an paid audience-paneling
134 service to reach a broader population [16]. The survey consisted of 17 Likert scale questions and two
135 open ended questions as reported in Table 1. To ensure response quality, we included a checking
136 question “We would like to ensure you are reading each question and responding thoughtfully. Please
137 select "Green" as your answer.” Questions assessing competence, reliability, transparency,
138 trustworthiness, and integrity of ChatGPT were informed by constructs in technology acceptance models
139 and trust research in information systems.

140 **Table 1.** Survey questions

Questions	Variable name
How frequently do you use ChatGPT	Use frequency
What is the main reason you use ChatGPT?	Use purpose
To what extent do you agree or disagree with the following:	
ChatGPT is competent in providing the information and guidance.	Competent
ChatGPT is reliable in providing consistent and dependable information.	Reliable
ChatGPT is transparent.	Transparent
ChatGPT is trustworthy.	Trustworthy
ChatGPT will not manipulate its responses.	Not manipulate

ChatGPT is secure and protects my privacy and confidential information	Secure
I am willing to use ChatGPT for healthcare-related queries.	ChatGPT for healthcare queries
Benefits of using ChatGPT outweigh any potential risks.	Benefits outweigh risks
ChatGPT helps me make informed and timely decisions.	Help make decisions
I am willing to take decisions based on the recommendations provided by ChatGPT.	Willing to take decisions
I am satisfied with ChatGPT.	Satisfaction
I am willing to use ChatGPT in the future.	Intent to use
ChatGPT can replace human-to-human interaction.	Replace human interaction
ChatGPT is persuasive.	Persuasive
What is the highest degree or level of education you have completed?	Education
*Describe in your own words, the improvements you would like to see in ChatGPT.	Not applicable
*Highlight your concerns with ChatGPT.	Not applicable
<hr/>	
<i>*Open ended questions</i>	

141

142 **Statistical analyses**

143 The analysis was divided into quantitative and qualitative. The quantitative analyses involved all the
144 Likert scale variables. We first calculated the data's descriptive statistics and reported the data's central
145 tendency, dispersion, and inference.

146 We then focused on the dataset corresponding to those who used ChatGPT for making healthcare-related
147 decisions. We calculated the multivariate and pairwise normality of the subset of the data. Given the data
148 distribution and normality violation, we conducted the Bayesian correlation test [17, 18]. In the
149 Bayesian Pearson correlation analyses, the strength and direction of the relationships between various
150 perceived attributes of ChatGPT are quantified using Pearson's correlation coefficient (r) quantifying the
151 evidence for a correlation against the null hypothesis of no correlation. The Bayesian Linear Regression
152 with Jeffreys-Zellner-Siow (JZS) priors for coefficients was conducted to ascertain the extent to which
153 respondents agree that ChatGPT helps them make informed and timely decisions, denoted as "Help
154 make decisions" [19]. The JZS prior is a default prior representing a compromise between
155 informativeness and non-informativeness, providing a reference analysis less sensitive to the prior
156 choice [19]. The analysis was predicated on a uniform prior distribution, reflecting the absence of a
157 priori preferences or knowledge about the importance of the predictors. This non-informative prior
158 ensures that the posterior distributions are primarily influenced by the data rather than by subjective

159 prior beliefs [19, 20]. Model comparison was achieved by contrasting each model with the best-
160 performing model as a reference. The Bayes Factor (BF_{10}) was used to quantify the evidence for each
161 model against the best model, and the coefficient of determination (R^2) was calculated to assess the
162 proportion of variance explained by the models. BF_{10} provides a measure of evidence of alternative
163 hypothesis (H_1) over null hypothesis (H_0), where H_0 suggests that there is no effect or no difference
164 between the variables being studied, and H_1 suggests that there is an effect or a difference [21]. Lastly,
165 the posterior distribution of the regression coefficients was summarized, showing the mean, standard
166 deviation (SD), and 95% credible intervals for each factor.

167 We repeated the same analysis on a subset of our data corresponding to respondents who did not use
168 ChatGPT for health-related decision-making.

169 The open-ended qualitative data was analyzed using content analysis [22]. The question was designed to
170 gather insights on public opinions regarding urban environmental policies. First the unit of analysis was
171 identified as thematic phrases within each response followed by the development of a coding scheme to
172 categorize the survey responses. The primary coding was conducted by one researcher, with periodic
173 validation checks by a second researcher to maintain coding consistency. All coders engaged in iterative
174 discussions to refine and finalize a set of consensus codes, ensuring that the identified themes accurately
175 captured the essence of the participants' experiences and perspectives. Inter-coder reliability was
176 assessed using Cohen's Kappa, yielding a coefficient of 0.75, indicative of substantial agreement [23].

177 Result

178 Data description and understanding

179 Six hundred and seven individuals responded to the survey. As illustrated in Fig 1, respondents were
180 distributed across 306 US cities. Only 20 participants were from rural cities. Of all the respondents, 44

181 used ChatGPT for health-related queries. Other uses of the technology were information gathering
 182 (n=219), entertainment (n=203), and problem-solving (n=135), and fun activities (n=6). Note:

183
 184 **Figure 1.** The figure shows the geographic distribution of study participants from 306 cities across the US. The blue circle
 185 size corresponds to the number of responses from each location. The data reveals a high concentration of participants in
 186 urban areas across the Eastern Seaboard, parts of the Midwest, and the West Coast, particularly in California, with sparser
 187 distribution in the central United States and rural regions.

188
 189 Table 2 shows the descriptive statistics of study variables. The statistics include the mean, standard error
 190 of the mean, a 95% confidence interval for the mean (with upper and lower bounds), standard deviation,
 191 Shapiro-Wilk test results, p-value of the Shapiro-Wilk test for normality, and the minimum and
 192 maximum values observed for each variable. Perceptions of ChatGPT’s attributes such as competence,
 193 reliability, transparency, trustworthiness, and security were rated positively, with all means exceeding
 194 3.0. The Shapiro-Wilk test uniformly indicated non-normality in the data distributions ($p < .001$),
 195 suggesting the presence of skewness or kurtosis across variables.

196
 197 **Table 2.** Descriptive statistics

	Mean	Std. error of mean	95% confidence interval mean		Std. Deviation	Shapiro- Wilk	Min	Max
			Upper	Lower				
Competent	3.199	0.034	3.265	3.134	0.825	0.794***	1	4
Reliable	3.161	0.032	3.225	3.098	0.800	0.812***	1	4
Transparent	3.119	0.035	3.187	3.050	0.859	0.817***	1	4
Trustworthy	3.168	0.034	3.235	3.101	0.837	0.807***	1	4
Not manipulate	3.099	0.036	3.169	3.029	0.881	0.823***	1	4
Secure	3.273	0.033	3.338	3.209	0.811	0.775***	1	4
ChatGPT for healthcare queries	3.097	0.035	3.165	3.029	0.855	0.826***	1	4
Benefits outweigh risks	3.204	0.033	3.268	3.140	0.801	0.798***	1	4
Helps make decisions	3.257	0.032	3.320	3.194	0.790	0.782***	1	4
Satisfaction	3.244	0.031	3.305	3.183	0.764	0.792***	1	4
Willing to take decisions	3.130	0.033	3.195	3.065	0.815	0.818***	1	4
Intent to use	3.379	0.031	3.439	3.319	0.757	0.751***	1	4
Replace human interaction	2.292	0.037	3.001	2.857	0.901	0.847***	1	4
Persuasive	2.979	0.033	3.044	2.913	0.819	0.846***	1	4
Education	2.855	0.036	2.925	2.785	0.884	0.873***	1	5

*** $p < 0.001$

198

199 **Quantitative findings using subset of the data who used ChatGPT for healthcare related**
 200 **queries.**

201 **Correlation analyses**

202 This sections reports finding based on 44 respondents who used ChatGPT for healthcare-related queries.

203 Table 4 presents a nuanced landscape of significant correlations among various perceived attributes of

204 ChatGPT as evaluated by users employing the AI for health-related inquiries. The significance of these

205 correlations is underscored by BF_{10} .

206 **Table 3.** The Bayesian Pearson Correlation of subset participants who use ChatGPT for health-related queries (n=44). The
 207 table details the correlation coefficients (Pearson's r) among various perceived attributes of ChatGPT.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
<i>Competent (A)</i>														
<i>Reliable (B)</i>	0.116													
<i>Transparent (C)</i>	0.469*	0.119												
<i>Trustworthy (D)</i>	0.436*	0.169	0.623***											
<i>Not manipulate (E)</i>	0.342	0.174	0.496**	0.322										
<i>Secure (F)</i>	0.443*	0.287	0.518**	0.425*	0.583***									
<i>ChatGPT for healthcare queries (G)</i>	0.289	0.016	0.437*	0.463*	0.481**	0.297								
<i>Benefits outweigh risks (H)</i>	0.415	0.132	0.287	0.274	0.384	0.390	0.532***							
<i>Helps make decisions (I)</i>	0.513*	0.084	0.467*	0.682***	0.446*	0.453*	0.538***	0.430*						
<i>Satisfaction (J)</i>	0.367	0.074	0.498*	0.467*	0.285	0.569***	0.354	0.355	0.359					
<i>Willing to take decisions (K)</i>	0.382	0.160	0.317	0.420	0.300	0.415	0.579***	0.337	0.453*	0.438*				
<i>Intent to use (L)</i>	0.425*	0.149	0.323	0.277	0.365	0.481**	0.334	0.236	0.429*	0.522***	0.632***			
<i>Replace human interaction (M)</i>	0.361	-0.046	0.438*	0.265	0.406	0.200	0.424*	0.263	0.369	0.098	0.272	0.280		
<i>Persuasive (N)</i>	0.393	0.104	0.541***	0.327	0.466*	0.348	0.454*	0.589***	0.249	0.461*	0.353	0.323	0.336	
<i>Education (O)</i>	0.030	0.044	-0.174	-0.044	-0.251	-0.210	0.043	0.087	-0.175	0.113	0.232	0.173	-0.180	0.086

*Bayes Factor (BF_{10})>10, ** BF_{10} >30, *** BF_{10} >100

208

209 A notable significant correlation ($r=0.623***$) exists between 'Transparent' and 'Trustworthy', suggesting

210 a robust relationship where the clarity of ChatGPT's processes and intentions is strongly associated with

211 users' trust. 'Trustworthy' also shares a substantial correlation ($r=0.682***$) with 'Helps make decisions',

212 indicating that trust in ChatGPT is crucial for users considering its advice in making health-related

213 decisions.

214 Additionally, 'Secure' exhibits very strong evidence of correlation ($r=0.583***$) with 'ChatGPT for

215 healthcare queries', pointing to security as a pivotal factor for users when consulting ChatGPT for health

216 concerns. The significant correlation between 'Secure' and 'Not manipulate' ($r=0.583***$) also highlights

217 the interdependence of security and the non-manipulative nature of responses in fostering a safe

218 environment for health-related interactions.

219 Furthermore, 'Benefits outweigh risks' has a very strong correlation ($r=0.532^{***}$) with 'ChatGPT for
220 healthcare queries', which could indicate that users who perceive higher benefits than risks are more
221 likely to engage with ChatGPT for health-related purposes. 'Intent to use' shows a strong positive
222 correlation ($r=0.632^{***}$) with 'Willing to take decisions', suggesting that users who intend to use
223 ChatGPT are also more likely to trust it for decision-making. The correlation of 'Intent to use' with
224 'Satisfaction' ($r=0.522^{***}$) and 'ChatGPT for healthcare queries' ($r=0.481^{**}$) emphasizes the connection
225 between future use intentions, current satisfaction levels, and the perceived utility of ChatGPT in health-
226 related matters. Lastly, 'Persuasive' (N) demonstrates very strong evidence of correlation ($r=0.589^{***}$)
227 with 'ChatGPT for healthcare queries', indicating that the ability of ChatGPT to influence user opinions
228 or actions is significantly related to its use for health queries.

229 As illustrated in Figure 2, in the subsequent in-depth analysis of selected pairwise correlation, we
230 investigated the relationship between 'Help make decisions' with 'Competence,' 'Transparent,' 'Benefits
231 outweigh risks,' and 'Persuasive'. We also explored pairwise correlation of 'Trustworthy' with
232 'Transparent' and 'Persuasive'. Prior and posterior distributions for Pearson's correlation coefficient
233 were generated to elucidate the relationship between variables. A robustness check was conducted to
234 ensure the stability of the Bayes Factor across varying priors. This analysis confirmed the strength of the
235 evidence for the alternative hypothesis (H1) (see S1).

236

237 **Figure 2.** Bayesian Correlation Sequential Analyses of ChatGPT's Attributes and Their Influence on Decision-Making
238 Assistance. **A:** Correlation between perceived competence of ChatGPT and its assistance in decision-making, indicating very
239 strong evidence for the positive association ($BF_{10} = 86.0414$); **B:** Association between perceived transparency of ChatGPT
240 and its aid in decision-making, demonstrating strong evidence for the correlation ($BF_{10} = 26.0618$); **C:** Relationship between
241 the perceived benefits outweighing risks when using ChatGPT for decision-making, showing strong evidence for the
242 correlation ($BF_{10} = 11.4741$); **D:** Correlation between perceived persuasiveness of ChatGPT and its impact on decision-
243 making, with anecdotal evidence for the association ($BF_{10} = 0.677$); **E:** Correlation between perceived trustworthiness and
244 persuasiveness of ChatGPT, suggesting anecdotal evidence for their combined influence on decision-making assistance (BF_{10}
245 = 0.544); **F:** Analysis of the relationship between transparency and trustworthiness in ChatGPT, with extreme evidence
246 supporting a very strong correlation ($BF_{10} = 3690$).

247

248 The sequential analysis further reinforced our Table 4 findings. As additional data points were
249 sequentially integrated into the analysis, the evidence for a positive correlation between the variable
250 remained robustly within the threshold for extreme, very strong, and anecdotal evidence. This pattern
251 persisted across the accumulation of data points, indicating that the observed correlation was not a
252 consequence of sample size but a persistent trend within the data.

253 For the attribute of competence, the analysis indicated very strong evidence ($BF_{10} = 86.0414$) for the
254 hypothesis that ChatGPT's competence positively influences its ability to help users make decisions.
255 This was further corroborated by the robustness checks, which consistently showed substantial support
256 for this relationship across a spectrum of priors. The transparency of ChatGPT also appeared to be a
257 significant factor, with a Bayes Factor suggesting strong evidence ($BF_{10} = 26.0618$) for its correlation
258 with aiding decision-making. This aligns with the notion that transparency in the functioning of AI
259 systems may bolster user trust and reliance on their decision-making capabilities.

260 In the case of the perceived benefits outweighing risks, there was strong evidence ($BF_{10} = 11.4741$)
261 supporting the relationship with decision-making assistance. Users who felt that the advantages of using
262 ChatGPT surpassed any potential risks were more likely to consider it helpful in making decisions.
263 However, when examining the attribute of persuasiveness, the evidence was merely anecdotal ($BF_{10} =$
264 0.677), indicating a weak relationship with decision-making aid. This suggests that while ChatGPT's
265 ability to persuade may be noticed by users, it does not significantly influence their reliance on the
266 system for decision-making support.

267 The trustworthiness attribute, when analyzed in conjunction with persuasiveness, showed a similarly
268 modest level of evidence ($BF_{10} = 0.544$), again suggesting that these factors alone do not strongly
269 predict ChatGPT's perceived utility in decision-making. The most compelling result was observed in the
270 correlation between transparency and trustworthiness, where an extreme Bayes Factor ($BF_{10} = 3690$)

271 was noted, indicating an exceptionally strong relationship between these attributes. This underscores the
272 integral role of transparent operations and trust in the perceived effectiveness of AI systems like
273 ChatGPT in supporting user decisions. This conclusion aligns with the overarching narrative of the
274 primary analysis and adds a layer of depth to the understanding of user perceptions of ChatGPT within
275 the specific context of health-related inquiries.

276 Bayesian Linear Regression Analysis

277 Table 5 presents a Bayesian linear regression comparing various models to ascertain the factors
278 influencing decision-making when ChatGPT is not used for healthcare-related queries. $P(M)$ represents
279 the prior probability of each model before data observation, assuming a non-informative or uniform
280 prior. $P(M|data)$ denotes the posterior probability, reflecting the model's probability after considering the
281 observed data. The Bayes Factor (BF_M) and Bayes Factor₁₀ (BF_{10}) provide evidence strength for each
282 model against a baseline model, with BF_M referencing the null model and BF_{10} comparing each
283 alternative hypothesis to the null hypothesis. The coefficient of determination (R^2) indicates the
284 proportion of variance in decision-making that is predictable from the independent variables in each
285 model. The models are ranked by the strength of evidence.

286 As shown in Table 5 the model comprising "Trustworthy + Benefits outweigh risks + Intent to use +
287 Education" emerged as the most robust, reflected by the highest Bayes Factor in relation to the null
288 model ($BF_{10} = 106.540$), suggesting substantial evidence in favor of this model compared to the
289 alternative models considered. The coefficient of determination (R^2) for this model was 0.617, indicating
290 that approximately 61.7% of the variance in the use of ChatGPT for making decisions was accounted for
291 by the predictors included in this model.

292 Table 4. Comparative Bayesian Analysis of Decision-Making Models (n=44).

Models → Help make decision	P(M)	P(M data)	BF_M	BF₁₀	R²
Trustworthy + Benefits outweigh risks + Intent to use + Education	6.104 x 10 ⁻⁵	0.006	106.540	1.000	0.617
Trustworthy + ChatGPT for healthcare queries + Intent to use + Education	6.104 x 10 ⁻⁵	0.004	63.255	0.595	0.606

Competent + Trustworthy + ChatGPT for healthcare queries	6.104×10^{-5}	0.004	59.841	0.563	0.573
Trustworthy + Benefits outweigh risks + Intent to use	6.104×10^{-5}	0.004	59.065	0.556	0.572
Competent + Trustworthy + ChatGPT for healthcare queries + Education	6.104×10^{-5}	0.003	54.670	0.515	0.603
Trustworthy + Benefits outweigh risks + Satisfaction + Intent to use + Education	6.104×10^{-5}	0.003	51.690	0.478	0.631
Competent + Trustworthy + ChatGPT for healthcare queries + Intent to use + Education	6.104×10^{-5}	0.003	49.168	0.463	0.630
Trustworthy + ChatGPT for healthcare queries + Benefits outweigh risks + Intent to use + Education	6.104×10^{-5}	0.003	49.076	0.462	0.630
Trustworthy + Benefits outweigh risks + Intent to use + Persuasive + Education	6.104×10^{-5}	0.003	45.529	0.429	0.628
Trustworthy + Intent to use + Education	6.104×10^{-5}	0.003	43.929	0.414	0.565

P(M)= prior probability of the model before observing the data (non-informative)
P(M|data) = posterior probability of the model given the observed data.
BF_M= Bayes Factor of the model relative to a baseline (null) model
F₁₀= Bayes Factor₁₀ refers to the comparison between the alternative hypothesis (H1) and the null hypothesis (H0)

293

294 The posterior summaries (Table 6) of coefficients revealed that 'Trustworthy' had a posterior mean
295 estimate of 0.454 with a 95% credible interval ranging from 0.156 to 0.684, indicating that as ChatGPT
296 is perceived as more trustworthy, participants are more likely to use it for healthcare-related decision-
297 making. P(incl) and P(excl) are the prior inclusion and exclusion probabilities, respectively, assuming an
298 initial 50/50 split without evidence. P(incl|data) and P(excl|data) represent the posterior probabilities of
299 inclusion and exclusion after data consideration. BF_{inclusion} stands for the Bayes Factor of inclusion,
300 providing the evidence ratio for each predictor's effect. The mean column represents the average
301 posterior estimate of the effect size, while the standard deviation (Std. Dev) quantifies the uncertainty
302 associated with the mean estimates. The 95% credible interval provides the range within which the true
303 effect size is likely to fall with 95% certainty.

304

Table 5. Posterior Analysis of Predictors Influencing Decision-Making Using Bayesian Linear Regression (n=44).

Coefficient	P(incl)	P(excl)	P (incl data)	P (excl data)	BF _{inclusion}	Mean	Std. Dev	95% Credible interval	
								Lower	Upper
Intercept	1.000	0.000	1.000	0.000	1.000	3.227	0.089	3.016	3.387
Competent	0.500	0.500	0.451	0.549	0.820	0.074	0.116	-0.042	0.366
Reliable	0.500	0.500	0.270	0.730	0.371	-0.018	0.079	-0.247	0.121
Transparent	0.500	0.500	0.285	0.715	0.399	-0.021	0.084	-0.259	0.101
Trustworthy	0.500	0.500	0.988	0.012	82.394	0.454	0.137	0.156	0.684
Not manipulate	0.500	0.500	0.343	0.657	0.521	0.038	0.090	-0.039	0.330
Secure	0.500	0.500	0.281	0.719	0.390	0.015	0.082	-0.116	0.299
ChatGPT for healthcare queries	0.500	0.500	0.458	0.542	0.844	0.089	0.137	-0.029	0.433
Benefits outweigh risks	0.500	0.500	0.516	0.484	1.065	0.097	0.129	-0.048	0.381
Satisfaction	0.500	0.500	0.290	0.710	0.409	-0.022	0.081	-0.254	0.113
Willing to take decisions	0.500	0.500	0.297	0.703	0.423	0.019	0.080	-0.080	0.247

Intent to use	0.500	0.500	0.554	0.446	1.245	0.119	0.144	-0.020	0.453
Replace human-to-human interaction	0.500	0.500	0.295	0.705	0.419	0.024	0.079	-0.094	0.261
Persuasive	0.500	0.500	0.306	0.694	0.440	-0.031	0.099	-0.343	0.088
Education	0.500	0.500	0.522	0.478	1.093	-0.079	0.103	-0.307	0.034

305

306 **Quantitative findings using subset of the data who did not use ChatGPT for healthcare**
 307 **related queries.**

308 **Bayesian correlation analysis**

309 This sections reports finding based on 563 respondents who did not use ChatGPT for healthcare-related
 310 queries. The correlation analysis (see S2) demonstrated a complex interplay of attributes with certain
 311 pairs, such as 'Competent' and 'Reliable' or 'Secure' and 'ChatGPT for healthcare queries', showing very
 312 strong evidence of a positive relationship. These robust statistical associations imply that perceptions of
 313 competence, reliability, transparency, trustworthiness, and security are significantly interrelated with the
 314 use of ChatGPT for healthcare queries, the perceived benefits outweighing risks, and the ability of
 315 ChatGPT to help make decisions. Additionally, satisfaction with ChatGPT, willingness to take decisions
 316 based on ChatGPT's advice, and intent to use ChatGPT in the future are also strongly interconnected,
 317 suggesting that positive user experience translates into continued use intentions.

318 **Bayesian Linear Regression Analysis**

319 Table 7 shows the model comparisons using BF_{10} identified the model with the highest evidence against
 320 the null model, consisting of the variables 'Trustworthy', 'Secure', 'Benefits outweigh risks', 'Satisfaction',
 321 'Willing to take decisions', 'Intent to use', 'Replace human-to-human interaction', and 'Persuasive'. This
 322 model achieved a BF_{10} of 828.416, suggesting very strong evidence in favor of this model over the null
 323 model, with an R^2 of 0.467, indicating it accounts for approximately 46.7% of the variance in the
 324 outcome.

325 Table 6. Comparative Bayesian Analysis of Decision-Making Models in Non-Healthcare Contexts (n=563).

Models → Help make decision	P(M)	P(M data)	BF_M	BF_{10}	R^2
Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Persuasive	6.104×10^{-5}	0.048	828.416	1.000	0.467
Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Replace human-to-human interaction	6.104×10^{-5}	0.045	778.833	0.943	0.471

+ Persuasive					
Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Replace human-to-human interaction	6.104 x 10 ⁻⁵	0.035	590.878	0.723	0.466
Competent + Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Replace human-to-human interaction	6.104 x 10 ⁻⁵	0.035	585.835	0.717	0.470
Competent + Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Replace human-to-human interaction	6.104 x 10 ⁻⁵	0.026	437.215	0.540	0.466
Competent + Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Replace human-to-human interaction + Persuasive	6.104 x 10 ⁻⁵	0.024	407.002	0.504	0.474
Competent + Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Persuasive	6.104 x 10 ⁻⁵	0.021	346.347	0.430	0.470
Trustworthy + Not manipulate + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Persuasive	6.104 x 10 ⁻⁵	0.018	291.935	0.364	0.469
Competent + Trustworthy + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Replace human-to-human interaction + Persuasive	6.104 x 10 ⁻⁵	0.015	244.653	0.306	0.469
Trustworthy + Not manipulate + Secure + Benefits outweigh risks + Satisfaction + Willing to take decisions + Intent to use + Replace human-to-human interaction	6.104 x 10 ⁻⁵	0.014	240.968	0.301	0.469
P(M)= prior probability of the model before observing the data (non-informative)					
P(M data) = posterior probability of the model given the observed data.					
BF _M = Bayes Factor of the model relative to a baseline (null) model					
BF ₁₀ = Bayes Factor ₁₀ refers to the comparison between the alternative hypothesis (H1) and the null hypothesis (H0)					

326

327 Table 8 summarizes the posterior distributions of regression coefficients for each variable contributing to
 328 decision-making efficacy. The posterior summaries of coefficients revealed that the credible intervals for
 329 'Secure,' 'Benefits outweigh risks,' and 'Willing to take decisions' have a significant positive impact on
 330 the dependent variable ('Help make decisions'). In other words, increases in these predictors are
 331 associated with an increase in the likelihood that ChatGPT helps users make informed and timely
 332 decisions.

333 Table 7. Posterior Analysis of Predictors Influencing Decision-Making Using Bayesian Linear Regression (n=563).

Coefficient	P(incl)	P(excl)	P		BF _{inclusion}	Mean	Std. Dev	95% Credible interval	
			(incl data)	(excl data)				Lower	Upper
Intercept	1.000	0.000	1.000	0.000	1.000	3.259	0.024	3.208	3.303
Competent	0.500	0.500	0.475	0.525	0.906	0.040	0.051	0.000	0.140
Reliable	0.500	0.500	0.124	0.876	0.142	0.002	0.016	0.000	0.059
Transparent	0.500	0.500	0.130	0.870	0.150	0.003	0.016	-0.014	0.051
Trustworthy	0.500	0.500	0.783	0.217	3.606	0.077	0.052	0.000	0.156
Not manipulate	0.500	0.500	0.293	0.707	0.414	0.015	0.030	0.000	0.090
Secure	0.500	0.500	0.992	0.008	128.914	0.154	0.044	0.063	0.229
ChatGPT for healthcare queries	0.500	0.500	0.180	0.820	0.220	0.007	0.022	-0.015	0.068
Benefits outweigh risks	0.500	0.500	0.995	0.005	209.320	0.163	0.044	0.080	0.247
Satisfaction	0.500	0.500	0.949	0.051	18.749	0.129	0.051	0.000	0.204
Willing to take decisions	0.500	0.500	0.986	0.014	69.047	0.141	0.043	0.062	0.233

Intent to use	0.500	0.500	0.739	0.261	2.838	0.076	0.057	-3.4×10^{-4}	0.165
Replace human-to-human interaction	0.500	0.500	0.666	0.334	1.995	0.050	0.044	0.000	0.122
Persuasive	0.500	0.500	0.598	0.402	1.487	0.055	0.055	-1.8×10^{-4}	0.154
Education	0.500	0.500	0.134	0.866	0.155	0.003	0.012	-0.007	0.045

334

335 Qualitative findings

336 Distinct patterns emerged when contrasting the concerns of users employing the ChatGPT for
337 healthcare-related inquiries against those using it for other purposes. Both cohorts expressed concerns
338 over privacy, a testament to the overarching necessity for robust data security protocols in AI
339 interactions. However, the nuances of their feedback reveal divergent emphases reflective of the context
340 of use.

341 Users engaging with ChatGPT for healthcare-related matters demonstrate a pronounced emphasis on
342 safety and trust. This is indicative of the critical nature of healthcare information and the consequential
343 outcomes dependent on its reliability. Concerns such as the AI becoming overly intelligent suggest
344 apprehensions about the delegation of health-related decision-making to artificial entities, raising
345 questions about the ethical bounds of AI in sensitive sectors. Moreover, while both sets of users
346 highlight the importance of content quality, the healthcare user group uniquely underlines the dual
347 dimensions of quantity and quality. This emphasis may underscore the necessity for comprehensive yet
348 reliable health information that AI platforms like ChatGPT are expected to deliver. In contrast, the
349 feedback from users not engaged in healthcare queries with ChatGPT tends to span a broader spectrum
350 of technical and usability enhancements. Calls for a more user-friendly interface, and improved load
351 times. Users also express the need for more human-like interactions, suggesting an aspiration for AI to
352 bridge the gap between technological functionality and human relationality. Notably, users from the non-
353 healthcare cohort voiced concerns about potential political or ideological biases in AI responses. This
354 highlights a distinct concern for objectivity and neutrality, which, while important in all contexts,
355 appears particularly salient for users seeking general information and assistance from ChatGPT.

356 The comparative analysis reveals that while the foundation of user trust in AI is built on privacy and
357 accuracy, the specific context of use—healthcare versus general inquiries—exerts a significant influence
358 on the nuances of user concerns. For healthcare-related AI applications, the implications are clear: there
359 is a critical need for heightened measures of accuracy, security, and ethical considerations, aligning with
360 the sensitive nature of healthcare information and decision-making processes.

361 Discussion

362 In the realm of healthcare decision-making, the prominence of trustworthiness, as indicated by our
363 study, is particularly instructive. Within our sample of 44 healthcare users of ChatGPT, trustworthiness
364 emerged as a critical factor influencing their reliance on the platform for healthcare-related queries.

365 Firstly, the critical nature of healthcare decisions inherently demands a measured degree of trust.
366 Decisions based on health-related information directly impact an individual's physical well-being,
367 making the accuracy and reliability of such information paramount. This is especially true in an era
368 where digital health advice is becoming increasingly prevalent, and users often must navigate a plethora
369 of information sources of varying credibility. In contrast, the non-healthcare user group, which is
370 notably larger, might prioritize other factors like security, user experience, or the ability to supplement or
371 replace human interaction. For these users, the stakes of the decision-making process, while still
372 important, do not typically bear the direct and immediate implications for personal health and well-
373 being. Consequently, their assessment of trustworthiness might be based more on factors like the
374 platform's ability to deliver satisfactory and efficient interactions.

375 Secondly, the complexity and validity of health information further elevates the need for trust.
376 Healthcare information is often laden with medical jargon and intricate details about conditions and
377 treatments, requiring users to place considerable trust in the source to ensure that the information is both
378 understandable and accurate. Moreover, the personal and confidential nature of health information calls

379 for a heightened level of trust in the platform's ability to handle such data ethically and securely. That
380 being said, it is often suggested to train GPT like ChatGPT on accurate, reliable, and perhaps patient
381 health data making the AI suitable for personalized recommendations and conversations [24]. Although
382 such discussions and recommendations are sound, they hold a critical latent risk. As recently brought to
383 notice, large language and GPT models tend to memorize training data, which can lead to privacy
384 concerns [14]. If ChatGPT or similar AI models are trained on sensitive healthcare data in the future to
385 provide personalized medical information, (with the hope of improving user experience and trust) this
386 could introduce a significant risk of large-scale data breaches [25, 26]. Such a situation could have far-
387 reaching consequences, as the breach of healthcare data not only compromises individual privacy but
388 also violates legal and ethical standards in medical confidentiality. It could lead to misuse of personal
389 health information, identity theft, and a loss of public trust in healthcare technology. Therefore, while the
390 advancement of AI in healthcare promises significant benefits, including personalized and efficient
391 patient care, it also necessitates robust security measures and ethical considerations to protect sensitive
392 patient data against potential breaches.

393 The recent discovery of ChatGPT's limitation [14] calls for urgent discussion about Technology
394 Readiness Levels (TRLs) [27]. TRLs are a systematic metric used to assess the development stage of a
395 technology, ranging from the conceptual stage (TRL 1) to full-scale deployment (TRL 9) [28]. This
396 metric is pivotal in determining whether a technology is sufficiently mature for public use, particularly
397 when it involves sensitive data. ChatGPT demonstrates immense potential in transforming healthcare
398 through personalized medicine and efficient data processing. However, the integration of AI in
399 healthcare necessitates a cautious approach, underscored by the rigorous assessment of TRLs. AI
400 systems dealing with healthcare data must not only demonstrate advanced functionality but also robust
401 security protocols to protect sensitive patient information. Assessing AI technologies against high TRL

402 standards ensures that they have undergone extensive testing and validation in real-world scenarios,
403 which includes evaluating their resilience against data breaches and extraction attacks.

404 Moreover, the importance of reaching higher TRLs before public deployment cannot be overstated in the
405 context of public trust and safety. Prematurely launched AI technologies that have not been thoroughly
406 vetted for security risks can lead to significant breaches of patient privacy. Such incidents not only erode
407 public confidence in AI healthcare applications but also pose severe ethical and legal challenges. The
408 healthcare sector, bound by stringent regulations and ethical standards, requires that any deployed
409 technology adheres to the highest levels of confidentiality and data protection.

410 In enhancing and extending the discussion on the limitations of our study, it's crucial to address both the
411 methodological constraints and the statistical considerations of our approach. The discrepancy in sample
412 sizes between the healthcare user group (n=44) and the non-healthcare user group (n=563) is a
413 limitation. We advocate for a cautious interpretation of the results, particularly in applying our findings
414 to a broader population of healthcare information seekers. Moreover, our sampling methodology,
415 primarily through social media channels, might introduce a selection bias. The demographic that
416 engages with healthcare-related AI through social media platforms may not accurately represent the
417 broader population seeking healthcare information. Reliance on self-reported data is another
418 methodological constraint. While self-reporting is a common and valuable tool in research, it inherently
419 carries the risk of inaccuracies. Participants' responses could be influenced by their perceptions,
420 experiences, or even social desirability bias, which might affect the authenticity of the data. To mitigate
421 some of these limitations, our study employed a Bayesian analytical approach. The Bayesian method is
422 advantageous in dealing with smaller sample sizes [29]. Despite these methodological efforts, further
423 research with a more diverse and larger sample, perhaps supplemented with objective usage metrics,

424 would be beneficial to validate and expand our findings and to develop a more comprehensive
425 understanding of how different user groups perceive and use AI in healthcare contexts.

426 In conclusion our study findings suggest a clear demarcation in user expectations and requirements from
427 AI systems based on the context of their use. In healthcare, where decisions have direct health
428 implications, trustworthiness emerges as a paramount concern, emphasizing the need for AI systems that
429 prioritize accuracy, clarity, and security. For non-healthcare contexts, while trustworthiness remains
430 important, it encompasses a broader range of factors including efficiency, user experience, and novelty.
431 This dichotomy provides valuable insights for AI developers and healthcare professionals. It suggests a
432 tailored approach to AI system design and communication, one that is acutely sensitive to the context of
433 use. Future research might focus on exploring these context-specific preferences in greater detail,
434 potentially leading to more personalized and effective AI applications across various domains.

435 Reference

- 436 1. Choi JH, Hickman KE, Monahan A, Schwarcz D. Chatgpt goes to law school. Available at SSRN. 2023.
- 437 2. Parikh PM, Shah DM, Parikh KP. Judge Juan Manuel Padilla Garcia, ChatGPT, and a controversial
438 medicolegal milestone. *Indian Journal of Medical Sciences*. 2023;75(1):3-8.
- 439 3. Lo CK. What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*.
440 2023;13(4):410.
- 441 4. Montenegro-Rueda M, Fernández-Cerero J, Fernández-Batanero JM, López-Meneses E. Impact of the
442 implementation of ChatGPT in education: A systematic review. *Computers*. 2023;12(8):153.
- 443 5. Guo C, Lu Y, Dou Y, Wang F-Y. Can ChatGPT boost artistic creation: The need of imaginative intelligence
444 for parallel art. *IEEE/CAA Journal of Automatica Sinica*. 2023;10(4):835-8.
- 445 6. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The Practical Implementation of Artificial Intelligence
446 Technologies in Medicine. *Nature Medicine*. 2019. doi: 10.1038/s41591-018-0307-0.
- 447 7. Choudhury A. Toward an Ecologically Valid Conceptual Framework for the Use of Artificial Intelligence in
448 Clinical Settings: Need for Systems Thinking, Accountability, Decision-making, Trust, and Patient Safety
449 Considerations in Safeguarding the Technology and Clinicians. *JMIR Hum Factors*. 2022;9(2):e35421. doi:
450 10.2196/35421.
- 451 8. Linkov I, Galaitsi S, Trump BD, Keisler JM, Kott A. Cybertrust: From Explainable to Actionable and
452 Interpretable Artificial Intelligence. *Computer*. 2020. doi: 10.1109/mc.2020.2993623.
- 453 9. Bach TA, Khan AN, Hallock H, Beltrão G, Sousa S. A Systematic Literature Review of User Trust in AI-
454 Enabled Systems: An HCI Perspective. *International Journal of Human-Computer Interaction*. 2022. doi:
455 10.1080/10447318.2022.2138826.
- 456 10. Choudhury A. Factors influencing clinicians' willingness to use an AI-based clinical decision support
457 system. *Frontiers in Digital Health*. 2022;4. doi: 10.3389/fgdth.2022.920662.

- 458 11. Sebastian G, George A, Jackson G. Persuading Patients Using Rhetoric to Improve Artificial Intelligence
459 Adoption: Experimental Study. *Journal of Medical Internet Research*. 2023. doi: 10.2196/41430.
- 460 12. Haresamudram K, Larsson S, Heintz F. Three Levels of AI Transparency. *Computer*. 2023. doi:
461 10.1109/mc.2022.3213181.
- 462 13. Yu L, Li Y, Fan F. Employees' Appraisals and Trust of Artificial Intelligences' Transparency and Opacity.
463 *Behavioral Sciences*. 2023. doi: 10.3390/bs13040344.
- 464 14. Nasr M, Carlini N, Hayase J, Jagielski M, Cooper AF, Ippolito D, et al. Scalable Extraction of Training Data
465 from (Production) Language Models. *arXiv preprint arXiv:231117035*. 2023.
- 466 15. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, et al., editors. Extracting training data
467 from large language models. 30th USENIX Security Symposium (USENIX Security 21); 2021.
- 468 16. Tambling RR, D'Aniello C, Russell BS. Health insurance literacy among caregivers of substance using
469 young adults. *Alcoholism Treatment Quarterly*. 2021;39(4):505-16.
- 470 17. Marsman M, Wagenmakers E-J. Bayesian benefits with JASP. *European Journal of Developmental*
471 *Psychology*. 2017;14(5):545-55.
- 472 18. Nuzzo RL. An introduction to Bayesian data analysis for correlations. Elsevier; 2017. p. 1278-82.
- 473 19. Bayarri M, García-Donato G. Generalization of Jeffreys divergence-based priors for Bayesian hypothesis
474 testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2008;70(5):981-1003.
- 475 20. Sarma A, Kay M, editors. Prior setting in practice: Strategies and rationales used in choosing prior
476 distributions for Bayesian analysis. *Proceedings of the 2020 chi conference on human factors in computing*
477 *systems*; 2020.
- 478 21. Kass RE, Raftery AE. Bayes factors. *Journal of the american statistical association*. 1995;90(430):773-95.
- 479 22. Stemler SE. Content analysis. *Emerging trends in the social and behavioral sciences: An Interdisciplinary,*
480 *Searchable, and Linkable Resource*. 2015:1-14.
- 481 23. Banerjee M, Capozzoli M, McSweeney L, Sinha D. Beyond kappa: A review of interrater agreement
482 measures. *Canadian journal of statistics*. 1999;27(1):3-23.
- 483 24. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for
484 electronic health records. *NPJ Digital Medicine*. 2022;5(1):194.
- 485 25. Huang H, Zheng O, Wang D, Yin J, Wang Z, Ding S, et al. ChatGPT for shaping the future of dentistry: the
486 potential of multi-modal large language model. *International Journal of Oral Science*. 2023;15(1):29.
- 487 26. Shoeybi M, Patwary M, Puri R, LeGresley P, Casper J, Catanzaro B. Megatron-Lm: Training multi-billion
488 parameter language models using model parallelism. *arXiv preprint arXiv:190908053*. 2019.
- 489 27. Mankins JC. Technology readiness levels. *White Paper*, April. 1995;6(1995):1995.
- 490 28. Martínez-Plumed F, Gómez E, Hernández-Orallo J. Futures of artificial intelligence through technology
491 readiness levels. *Telematics and Informatics*. 2021;58:101525.
- 492 29. McNeish D. On using Bayesian methods to address small sample problems. *Structural Equation*
493 *Modeling: A Multidisciplinary Journal*. 2016;23(5):750-73.

494

495

496

497

498 **Supporting information**

499 **S1. Supplementary file.** Extra analysis and results

500 **S2. Supplementary file.** Correlation analysis

501