1 Fact Check: Assessing the Response of ChatGPT to Alzheimer's Disease Statements with 2 **Varying Degrees of Misinformation** 3 Sean S. Huang M.D.*a, Qingyuan Song M.E.*c, Kimberly J. Beiting M.D.a, Maria C. Duggan 4 5 M.D. a, Kristin Hines M.D. a, Harvey Murff M.D. a, Vania Leung M.D. e, James Powers M.D. a,d, T.S. Harvey Ph.D.f, Bradley Malin Ph.D.b,c,g, Zhijun Yin Ph.D.b,c 6 7 8 ^a Department of Medicine, Division of Geriatrics, Vanderbilt University Medical Center, 9 Nashville, TN, USA 10 ^b Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, 11 TN, USA ^c Department of Computer Science, Vanderbilt University, Nashville, TN, USA 12 ^d Geriatric Research Education and Clinical Center (GRECC), Department of Veteran 13 14 Affairs, Tennessee Valley Healthcare System, Nashville, TN, USA 15 ^e Department of Academic Internal Medicine and Geriatrics, University of Illinois at 16 Chicago, Chicago, IL, USA 17 ^f Department of Anthropology, Vanderbilt University, Nashville, TN, USA 18 g Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA 19 * These authors contributed equally to the work 20 21 Funding Disclosure: Research reported in this paper was supported by the National Institutes of Health under award number U54HG012510. 22 23 24 Dr. Beiting is supported by the Health Resources and Services Administration (HRSA) of the U.S. 25 Department of Health and Human Services (HHS) under grant number K01HP49070. This information or 26 content and conclusions are those of the author and should not be construed as the official position or 27 policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government. 28

29 **Corresponding Author:** 30 Sean S. Huang, MD 31 Assistant Professor of Medicine 32 Division of Geriatric Medicine, Vanderbilt University Medical Center 33 2147 Belcourt Ave, Suite 100 34 Nashville, TN 37212 35 Phone: 615-875-2468 36 sean.huang@vumc.org 37 Twitter: @sshuangmd 38 39 **Word Counts:** 40 Abstract – 297 41 **Main text – 1845** Tables -1, Figures -242 43 44 **Impact Statement** 45 There are many statements regarding Alzheimer's disease (AD) diagnosis, management, and 46 treatment circulating on the Internet, each exhibiting varying degrees of accuracy, inaccuracy, and 47 misinformation. Large language models are a popular topic currently, and many patients and 48 caregivers may turn to LLMs such as ChatGPT to learn more about the disease. This study aims 49 to assess ChatGPT's ability to identify and address AD myths with reliable information. We certify 50 that this work is novel. 51 **Key Points** 52 - Geriatricians acknowledged the potential value of ChatGPT in mitigating misinformation in 53 Alzheimer's Disease 54 - There remain nuanced cases where ChatGPT explanations are not as refined or appropriate. 55 - Why does this matter? Large language models such as ChatGPT are very popular nowadays 56 and patients and caregivers often may use them to learn about their disease. The paper seeks to 57 determine whether ChatGPT does an appropriate job in moderating understanding of 58 Alzheimer's Disease myths.

Keywords: Alzheimer's Disease, Misinformation, Large Language Models, ChatGPT

Abstract

Background

There are many myths regarding Alzheimer's disease (AD) that have been circulated on the Internet, each exhibiting varying degrees of accuracy, inaccuracy, and misinformation. Large language models such as ChatGPT, may be a useful tool to help assess these myths for veracity and inaccuracy. However, they can induce misinformation as well. The objective of this study is to assess ChatGPT's ability to identify and address AD myths with reliable information.

Methods

We conducted a cross-sectional study of clinicians' evaluation of ChatGPT (GPT 4.0)'s responses to 20 selected AD myths. We prompted ChatGPT to express its opinion on each myth and then requested it to rephrase its explanation using a simplified language that could be more readily understood by individuals with a middle school education. We implemented a survey using Redcap to determine the degree to which clinicians agreed with the accuracy of each ChatGPT's explanation and the degree to which the simplified rewriting was readable and retained the message of the original. We also collected their explanation on any disagreement with ChatGPT's responses. We used five Likert-type scale with a score ranging from -2 to 2 to quantify clinicians' agreement in each aspect of the evaluation.

Results

The clinicians (n=11) were generally satisfied with ChatGPT's explanations, with a mean (SD) score of $1.0(\pm0.3)$ across the 20 myths. While ChatGPT correctly identified that all the 20 myths were inaccurate, some clinicians disagreed with its explanations on 7 of the myths.

Overall, 9 of the 11 professionals either agreed or strongly agreed that ChatGPT has the potential to provide meaningful explanations of certain myths.

Conclusions

The majority of surveyed healthcare professionals acknowledged the potential value of
ChatGPT in mitigating AD misinformation. However, the need for more refined and detailed
explanations of the disease's mechanisms and treatments was highlighted.

89 Introduction

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

Over 80% of American adults now use the Internet to obtain health information. This democratization has led to an increase in access to information as well as an escalation in the spread of health-related misinformation, posing severe challenges to public health and patient decision-making.² Broad or lay public awareness of a health issue or a disease, including Alzheimer's disease (AD), often coexists with certain misconceptions.³ In this respect, there are widely accessed AD-related statements that have been posted online, each demonstrating varying degrees of accuracy, inaccuracy, and misinformation. While aware of the disease, many patients and caregivers, whether through experience or training, lack certain knowledge regarding its diagnosis and management.⁴ Through public access to misinformation circulated and gathered from non-clinical sources, not only is AD often unfairly stigmatized but medical professionals are under consulted. This trend has led to some clinicians who treat AD to call for action to help distinguish fact from fiction.⁵ The emergence of large language models (LLMs) has created a new opportunity for people to search for relevant medical knowledge. LLMs comprise a complex neural network with billions of parameters estimated on large quantities of data. This platform and training process enables LLMs to capture and generate complex linguistic patterns and dependencies, thus supporting an ability to convey and express human-like text based on input prompts. ⁶ The commonly known LLMs, such as ChatGPT (or GPT-4⁷) developed by OpenAI have shown remarkable capabilities in language understanding and contextually relevant text generation.⁸⁻¹² However, LLMs are not infallible and do not exist outside of human culture, as such they can mirror the assumptions and prejudices inherent in their training data, leading to biases and

misinformation in their responses, raising ethical and safety concerns.¹³ Thus, to leverage LLMs

to combat AD misinformation, it is necessary to examine their performance through verification with domain or subject matter experts. Notably, ChatGPT has been shown to effectively recognize myths in cancer, ¹⁴ but no study has shown its benefits in AD. Therefore, the objective of this study is to assess ChatGPT's ability to accurately address AD myths and provide reliable information.

119 Methods

We conducted a cross-sectional study of clinicians' evaluation of ChatGPT's opinion of 20 AD myths (see Table 1). In this investigation, "myths" are understood as sets of ideas that become accepted and/or otherwise understood by communities according to dominant ideals that may themselves be based on varying degrees accuracies, inaccuracies, and misinformation. We explain how AD myths were selected and validated in the Supplemental Information (SI). The study took place between April 2023 and May 2023 and was deemed to not be human subjects research by the Vanderbilt University Medical Center Internal Review Board.

We prompted ChatGPT to generate an opinion for each myth and then rewrite the technical explanation using non-specialist terms such that recipients with a middle school education-level could understand it. We then developed and administered a survey to inquire about the degree to which clinicians agreed with the accuracy of each original explanation and the extent to which the rewritten (simplified) explanation was readable, and retained the information in the original.

Figure 1 provides a sample of the survey questions for Myth #3. The survey begins with a "To ChatGPT:" component to inquire about ChatGPT's opinion on the myth. Next, the survey

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

asks the clinician if they are satisfied with ChatGPT's opinion. The clinician could express their disagreements with the ChatGPT's opinion through a binary 'yes/no' question, complemented by an optional text field to discuss the disagreements. The second "To ChatGPT:" component contains ChatGPT's rewritten explanation and two questions for the clinicians about their assessment of its information retention and readability. In the survey, these questions use a fivepoint Likert-type scale that ranged from either "Very Unsatisfied" or "Strongly Disagree" to "Very Satisfied" or "Strongly Agree". Each option was associated with an according integer score ranging from -2 to 2. As such, a positive score indicates that clinicians favor ChatGPT's explanation or rewriting. In addition, we asked clinicians before and after the survey, "Do you agree that ChatGPT can be a useful tool to provide explanations and clarifications of misinformation about Alzheimer's disease?". This question also uses a five-point Likert-type-type scale ranging from "Strongly Disagree" to "Strongly Agree," with scores from -2 to 2. Each clinician was asked to review the responses for all 20 statements, which were presented in a randomized order to mitigate potential ordering effects. The complete survey for the assessment is provided in the Supplemental Information. The clinicians were recruited from a convenience sample to complete the ChatGPT assessment survey in REDCap. 16 The set of clinicians included attending geriatricians, geriatric trainees, nurse practitioners, and physician assistants who primarily treat and care for geriatric patients from three medical centers (Vanderbilt University Medical Center, University of Illinois Chicago, Rush Medical Center). Survey respondents received an Amazon gift card with a

monetary amount proportional to the number of responses completed.

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

Results We sent an email to 34 practitioners/clinicians who practice primarily in geriatrics, and 15 agreed to complete the survey. Finally, we received survey responses from 11 of them: six attending physicians, three geriatric trainee physicians, one nurse practitioner, and one physician assistant. Only one of these clinicians had used ChatGPT occasionally, nine clinicians had heard of it but never used it, and one other clinician had never heard of it. Figure 2 illustrates the participants' evaluation of ChatGPT's opinions. Generally, among the 20 myths, the clinicians were satisfied with these explanations, with (mean [standard deviation] score of 1.0 [±0.3]. The majority of clinicians selected "Agree" or "Strongly Agree" for each statement (green in Figure 2). Some statements (e.g., statement #7, AD is not fatal) had more "Disagree" or "Strongly Disagree" responses (red in Figure 2). With respect to the overall results of information retention in the rewritten explanations, 6% selected "Strongly Agree", 76% selected "Agree", and 9% selected "Disagree". No clinicians selected "Strongly Disagree". For readability, 7% selected "Strongly Agree", 74% selected "Agree", and 3% selected "Disagree". Again, none of the clinicians selected "Strongly Disagree". In both situations the mean Likert-type-type score was found to be 0.8. More detailed results are provided in the Supplemental Information. When comparing the clinicians' assessment of ChatGPT's potential to clarify and explain misinformation about AD before and after reviewing the ChatGPT's responses, three (27%) shifted their responses from "Disagree" or "Neutral" to "Agree" or "Strongly Agree." In total, 82% of clinicians agreed or strongly agreed, 9% were neutral, and 9% disagreed or strongly disagreed.

180 Discussion

This investigation suggests that ChatGPT exhibits a satisfactory ability to identify and explain misinformation about AD and to generate audience-specific readable explanations. Still several explanations were not entirely accurate. To gain insight into the scenarios where ChatGPT excelled, as well as those that posed a challenge, four statements can be used as examples: two with high total Likert-type scores, one with a high neutral Likert-type score, and another with a low Likert-type score.

#5: Alzheimer's disease is a shameful diagnosis.

Average Likert-type score: 1.6

AD often carries a social stigma that produces deleterious effects on the well-being of patients afflicted with the disease and their families. These attitudes often lead to delays in care-seeking, diagnosis, and accessing resources for treatment. The language learning model prefaces its statement by noting that it does not have a personal opinion. Because ChatGPT is not a 'person' this statement is technically true, however, it can be interpreted as amounting to a claim of objectivity. That said, ChatGPT expresses the opinion that AD is not a shameful diagnosis, and those affected should be treated with compassion, understanding, and support. While no diagnosis should every be 'shameful', it is important to note that technically what makes a diagnosis "shameful" is the social context in which the disease is experienced and not the diagnosis itself. Its anticipated response echoes recent suggestions that ChatGPT and similar models can mirror empathy equivalent to a human's. While it certainly cannot replace a clinician trained in the management of AD or caregiver communities of practice, working large

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

language models may be one day be used to establish more personalized and potentially less stigmatized care. #3: All types of dementia are Alzheimer's disease. Average Likert-type score: 1.4 ChatGPT makes a very emphatic declaration, appropriately noting that this statement is incorrect. It notes that AD is the most common type but then demonstrates that not all types of dementia are AD by identifying other types of dementia, including vascular dementia, Lewy body disease (LBD), and frontotemporal dementia.²⁰ The distinctions are important as they may affect treatment decisions. For example, cholinesterase inhibitors are commonly effective in LBD but can often worsen frontotemporal dementia. Additionally, ChatGPT only mentions that symptoms of LBD can resemble Parkinson's disease but fails to distinguish that dementia with Lewy Bodies and Parkinson's disease dementia are two different diseases with different underlying etiologies, pathologies, and diagnostic criteria. 21,22 #1: Alzheimer's disease symptoms are a normal part of aging. Average Likert-type score: 0.9 One of the most common misconceptions of AD is that its symptoms are a common part of aging. ^{23–25} ChatGPT starts by stating that its symptoms are not considered part of normal aging. However, as one clinician noted, the phrasing seems to suggest that the cognitive decline of normal aging is part of the spectrum of AD, which should be clarified. The normative decline in brain reserve with aging is not discussed, nor are the dementia warning signs that help with differentiation. ChatGPT does not discuss distinctions between normal aging, mild cognitive

impairment, and AD. Understanding what specific cognitive abilities are being tested requires more extensive evaluation.²⁶ Such nuance may not be communicated well by ChatGPT.

#7: Alzheimer's disease is not fatal.

Average Likert-type score: 0.2

This statement produced the lowest Likert-type score tested. Alzheimer's is generally considered a terminal neurodegenerative disease that breaks down neurologic processes with a predominant loss of cognitive function. There is no cure, and in 2023, it was considered the seventh leading cause of death, with deaths from the disease more than doubling between 2000 and 2019.²⁷ Additionally, many deaths linked to the disease are underreported.²⁸ ChatGPT notes that AD is not a direct cause of death but only contributes to complications, explicitly citing the inability to swallow, among other complications. This statement is misleading, as it may cause people with dementia and their caregivers to think feel that death can be avoided if they prevent complications.²⁹ As one provider noted, "This may be one of the biggest and most harmful fallacies in care for dementia." Still, ChatGPT correctly notes that the most common causes of death from AD are failure to thrive and infection.³⁰

Conclusion

Like many diseases, AD, given its prevalence, is prone to misinformation communicated, a problem that is exacerbated and accelerated on online platforms. Though large language models, like ChatGPT, are not directly responsible for creating or propagating misinformation, their utilization may inadvertently amplify such misinformation due to potential biases or misinterpretations embedded within their training data. While ChatGPT can provide useful

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

information for most readers, there could be more refined and detailed explanations of the disease's mechanisms and treatments. Acknowledgements Conflict of Interest: There are no conflicts of interest to disclose. Author Contributions: SH and ZY conceived the paper and designed the analysis. SH and QS collected the data. QS performed the analysis. KB, MD, KH, HM, VL, JP contributed data and helped interpret the results. TH and BM provided crucial edits. SH, QS, ZY wrote the paper. Sponsor's Role: Zhijun Yin helped with the design, methods, analysis, and preparation of paper. References Swire-Thompson B, Lazer D., Public health and online misinformation: challenges and 1. recommendations. Annu Rev Public Health. 2020;41(1):433-451. doi: 10.1146/annurevpublhealth-040119-094127. 2. Chou WYS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA. 2018;320(23):2417-2418. doi:10.1001/jama.2018.16865 3. Carpenter DM, Geryk LL, Chen AT, Nagler RH, Dieckmann NF, Han PKJ. Conflicting health information: a critical research need. *Health Expectations*. 2016;19(6):1173-1182. doi: 10.1111/hex.12438 4. Anderson LA, Day KL, Beard RL, Reed PS, Wu B. The public's perceptions about cognitive health and Alzheimer's disease among the US population: a national review. Gerontologist. 2009;49(S1):S3–S11. doi: 10.1093/geront/gnp088

- 271 5. Rosin, E. R., Blasco, D., Pilozzi, A. R., Yang, L. H., & Huang, X. (2020). A Narrative
- 272 Review of Alzheimer's Disease Stigma. Journal of Alzheimer's disease: JAD, 78(2),
- 273 515–528. Doi: 10.3233/JAD-200932
- Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Guyon I, Luxburg
- U Von, Bengio S, et al., eds. Advances in Neural Information Processing Systems. Vol
- 276 30. Curran Associates, Inc.; 2017.
- 7. OpenAI. GPT-4 Technical Report. Published online March 15, 2023.
- 8. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. In: Larochelle
- H, Ranzato M, Hadsell R, Balcan MF, Lin H, eds. Advances in Neural Information
- 280 Processing Systems. Vol 33. Curran Associates, Inc.; 2020:1877-1901.
- 281 9. Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a
- unified text-to-text transformer. The Journal of Machine Learning Research.
- 283 2020;21(1):5485-5551. doi: 10.48550/arXiv.1910.10683
- 284 10. Devlin J, Chang MW, Lee K, Toutanova K. DERT: Pre-training of deep bidirectional
- transformers for language understanding. arXiv preprint arXiv:181004805. Published
- 286 online 2018. doi: 10.48550/arXiv.1810.04805
- 287 11. Dong L, Yang N, Wang W, et al. Unified language model pre-training for natural
- language understanding and generation. Adv Neural Inf Process Syst. 2019;32. doi:
- 289 10.48550/arXiv.1905.03197
- 290 12. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le Q V. Xlnet: Generalized
- autoregressive pretraining for language understanding. Adv Neural Inf Process Syst.
- 292 2019;32. doi: 48550/arXiv.1906.08237

313

10.1136/jnnp.51.6.745.

293 13. Stiff H, Johansson F. Detecting computer-generated disinformation. *Int J Data Sci Anal*. 294 2022;13(4):363-383. doi: 10.1007/s41060-021-00299-5 295 14. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to 296 evaluate cancer myths and misconceptions: artificial intelligence and cancer information. 297 JNCI Cancer Spectr. 2023;7(2):pkad015. doi: 10.1093/jncics/pkad015 298 15. Crowston K. Amazon mechanical turk: A research tool for organizations and information 299 systems scholars. In: Shaping the Future of ICT Research. Methods and Approaches: 300 IFIP WG 8.2, Working Conference, Tampa, FL, USA, December 13-14, 2012. 301 Proceedings.; 2012:210-221. 302 16. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international 303 community of software platform partners. J Biomed Inform. 2019;95:103208. 304 17. Lopez RP, Rose KM, Kenney L, Sanborn V, Davis JD. Managing shame: A grounded 305 theory of how stigma manifests in families living with dementia. J Am Psychiatr Nurses 306 Assoc. 2020;26(2):181-188. doi: 10.1177/1078390319832965. 307 18. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence 308 chatbot responses to patient questions posted to a public social media forum. AMA Intern 309 Med. 2023 Jun 1;183(6):589-596. doi: 10.1001/jamainternmed.2023.1838. 310 19. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Vol 5. 311 20. Gibb WR, Lees A. The relevance of the Lewy body to the pathogenesis of idiopathic 312 Parkinson's disease. J Neurol Neurosurg Psychiatry. Jun;51(6):745-52. doi:

314	21.	Gomperts SN. Lewy body dementias: dementia with Lewy bodies and Parkinson disease
315		dementia. Continuum (Minneap Minn). 2016;22(2 Dementia):435-463.
316		doi:10.1212/CON.000000000000000000000000000000000000
317	22.	Sengoku R. Aging and Alzheimer's disease pathology. Neuropathology: official journal
318		of the Japanese Society of Neuropathology, 40(1), 22–29. doi: 10.1111/neup.12626
319	23.	Blay SL, Peluso É de TP. The public's ability to recognize Alzheimer disease and their
320		beliefs about its causes. Alzheimer Dis Assoc Disord. 2008;22(1):79-85.
321	24.	Lee SE, Lee HY, Diwan S. What do Korean American immigrants know about
322		Alzheimer's disease (AD)? The impact of acculturation and exposure to the disease on
323		AD knowledge. Int J Geriatr Psychiatry. 2010;25(1):66-73. doi:10.1002/gps.2299
324	25.	Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment,
325		MoCA: a brief screening tool for mild cognitive impairment. [published correction
326		appears in J Am Geriatr Soc. 2019 Sep;67(9):1991]. J Am Geriatr Soc. 2005;53(4):695-
327		699. doi:10.1111/j.1532-5415.2005.53221.x
328	26.	2023 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2023;19(4):1598-
329		1695. doi:10.1002/alz.13016
330	27.	Stokes AC, Weiss J, Lundberg DJ, et al. Estimates of the association of dementia with
331		US mortality levels using linked survey and mortality records. JAMA Neurol.
332		2020;77(12):1543-1550. doi:10.1001/jamaneurol.2020.2831
333	28.	Meier DE, Ahronheim JC, Morris J, Baskin-Lyons S, Morrison RS. High short-term
334		mortality in hospitalized patients with advanced dementia: lack of benefit of tube feeding.
335		Arch Intern Med. 2001;161(4):594-599. doi:10.1001/archinte.161.4.594

336 29. Mitchell SL, Teno JM, Kiely DK, et al. The clinical course of advanced dementia. N Engl

J Med 2009; 361:1529-1538 doi: 10.1056/NEJMoa0902234

Table 1. The 20 Alzheimer's disease myths used in this study.

Index	Myth
1	Alzheimer's disease symptoms are a normal part of aging
2	Symptoms of memory loss must be Alzheimer's disease
3	All types of dementia are Alzheimer's disease
4	All Alzheimer's disease patients are disoriented, look confused, and act outside
4	social norms
5	Alzheimer's disease is a shameful diagnosis
6	An at-home genetic test can tell me if I have (or will have) Alzheimer's disease
7	Alzheimer's disease is not fatal
8	Alzheimer's disease patients are close to death
9	Alzheimer's disease patients cannot get depressed
10	I will develop Alzheimer's disease if my parent has it
11	Only older people get Alzheimer's disease
12	Supplements found online can prevent or cure Alzheimer's disease
13	Specific diets can reduce the risk of Alzheimer's disease
14	Prescription medications can reverse the progression to Alzheimer's disease
15	Non-steroidal anti-inflammatory drugs (NSAIDs) can prevent Alzheimer's
13	disease
16	Ginkgo biloba can prevent Alzheimer's disease
17	Hormone therapy prevents and treats Alzheimer's disease
18	Vaccines increase the risk of Alzheimer's disease
19	Aspartame increases the risk of Alzheimer's disease
20	Statins increase the risk of Alzheimer's disease

337

338

339

