Research Report

ChatGPT as a Diagnostic Aid in Alzheimer's Disease: An Exploratory Study

Mohamad El Haj[a,b,*], Claire Boutoleau-Bretonnière[c], Karim Gallouj[d], Nathalie Wagemann[c], Pascal Antoine[e], Dimitrios Kapogiannis[f] and Guillaume Chapelet[b,g]

[a]*Institut Universitaire de France, Paris, France*
[b]*Clinical Gerontology Department, CHU Nantes, Nantes, France*
[c]*CHU Nantes, Inserm CIC04, Nantes, France*
[d]*Unité de Gériatrie, Centre Hospitalier de Tourcoing, Tourcoing, France*
[e]*Universite de Lille, CNRS, CHU Lille, UMR 9193 SCALab – Sciences Cognitives et Sciences Affectives, Lille, France*
[f]*Laboratory of Clinical Investigation, National Institute on Aging, Baltimore, MD, USA*
[g]*Université de Nantes, Inserm, TENS, The Enteric Nervous System in Gut and Brain Diseases, IMAD, Nantes, France*

Abstract.
Background: The potential of ChatGPT in medical diagnosis has been explored in various medical conditions.
Objective: We assessed whether ChatGPT can contribute to the diagnosis of Alzheimer's disease (AD).
Methods: We provided ChatGPT with four generated cases (mild, moderate, or advanced stage AD dementia, or mild cognitive impairment), including descriptions of their complaints, physical examinations, as well as biomarker, neuroimaging, and neuropsychological data.
Results: ChatGPT accurately diagnosed the test cases similarly to two blinded specialists.
Conclusions: While the use of generated cases can be a limitation to our study, our findings demonstrate that ChatGPT can be a useful tool for symptom assessment and the diagnosis of AD. However, while the use of ChatGPT in AD diagnosis is promising, it should be seen as an adjunct to clinical judgment rather than a replacement.

Keywords: Alzheimer's disease, artificial intelligence, ChatGPT, diagnosis, mild cognitive impairment

INTRODUCTION

The diagnosis of Alzheimer's disease (AD) involves the comprehensive assessment of cognitive symptoms and performance metrics, functional abilities, and neurological signs by specialized physicians. Cognitive evaluations typically entail neuropsychological tests covering general cognitive function, memory, language, and executive function. Functional and neurological assessments are typically conducted through clinical examinations and require receiving information from third-parties, typically family members from patients. The diagnosis of the underlying neurobiological entity of AD in living individuals requires neuroimaging and biomarkers, which may provide evidence on the accumulation of amyloid-beta and tau and neurodegeneration [1]. In the diagnostic process, the presence of medical conditions such as cardiovascular disease, obesity, and diabetes is also considered due to their association

---

*Correspondence to: Mohamad El Haj, Faculté de Psychologie, LPPL – Laboratoire de Psychologie des Pays de la Loire, Université de Nantes, Chemin de la Censive du Tertre, BP 81227, 44312 Nantes Cedex 3, France. E-mail: mohamad.elhaj@univ-nantes.fr.

with increased risk of AD and multifactorial dementia [2]. The diagnosis of AD also includes categorizing patients into different stages, most commonly as belonging to the mild, moderate, or advanced stages, with many patients seeking consultation during the mild stage. Furthermore, the diagnostic process also considers preclinical and early clinical or prodromal stages. Patients may present with mild cognitive impairment (MCI), which involves objectively verifiable cognitive changes while, crucially, are still being able to perform daily tasks independently [3]. MCI can be a precursor to AD, in which case it may also be called prodromal AD, although not all individuals with MCI ultimately develop AD or other dementia [3]. To summarize, the diagnosis of AD involves a battery of assessments, including history intake, neuropsychological and neurological tests, neuroimaging, and biomarkers. This comprehensive approach allows for the classification of patients into specific stages of the disease, including its preclinical phases.

Given the complexity of AD diagnosis, we explored the potential of artificial intelligence, particularly Chat-Generative Pre-Trained Transformer (ChatGPT), to aid in the complex process of diagnosis. In the ever-evolving landscape of artificial intelligence, ChatGPT stands as a remarkable achievement in the ongoing quest to enhance human-machine communication. ChatGPT operates on the principles of deep learning, employing an extensive neural network model with the capacity to comprehend and produce text with a nuanced understanding of context, tone, and intent [4]. ChatGPT is increasingly finding applications in medicine. A growing body of research emphasizes the utility of ChatGPT in various medical contexts, such as patient education, interaction, and the dissemination of healthcare information [5–10]. ChatGPT has even demonstrated some levels of accuracy in diagnosing medical conditions [11–16]. This underscores the potential for ChatGPT to play a valuable role in Medicine, including aiding in the complex process of AD diagnosis.

While the potential of ChatGPT in medical diagnosis has been explored in various medical conditions, little research has been conducted on its application to AD. A study has attempted to predict AD from spontaneous speech using ChatGPT [17]. This study has shown that text embedding, as generated by ChatGPT, can be used to not only detect patients with AD from healthy controls but also to infer the patients' cognitive scores. This study was however solely based on speech data. Using neu-

ropsychological data, a study [18] has investigated the ability of ChatGPT to assist in AD screening. This study has assessed the ability of ChatGPT to provide general interpretations of neuropsychological test scores for a patient with mild AD. Results demonstrated that ChatGPT could accurately interpret the scores obtained on neuropsychological tests. However, ChatGPT has showed several limitations as it did not utilize standardized scores and did not specify the affected cognitive domains. It is important to note that the preliminary study by El Haj et al. [18] was concerned with the interpretation of neuropsychological data and did not directly address the diagnosis of AD. The present study expands upon the study by El Haj et al. [18]. The present study goes beyond the interpretation of neuropsychological data and aims to assess whether ChatGPT can effectively integrate neuropsychological, medical, neurological, neuroimaging, and biomarker information to contribute to the diagnostic process of AD. Importantly, the present study incorporates multiple case scenarios to evaluate the ability of ChatGPT to classify patients into specific stages of AD, including its preclinical phases. This broader assessment seeks to address the complex and multifaceted nature of AD diagnosis.

In summary, while research has shown the potential of ChatGPT in the field of medical diagnosis for various conditions [11–16], there is a dearth of knowledge about its application in the context of AD and neurological conditions in general. While a prior study by El Haj et al. [18] explored the ability of ChatGPT to interpret neuropsychological data in a single patient with AD, it did not directly address its broader potential in assisting with the diagnosis of AD. The present study thus assesses whether ChatGPT can effectively synthesize information from various medical domains to contribute to the intricate diagnostic process of AD. Furthermore, the study assesses the ability of ChatGPT to classify patients into distinct stages of AD, including its preclinical stages. To achieve this, we provided ChatGPT with a comprehensive dataset comprising patient histories, complaints, physical examinations, as well as biomarker, neuroimaging, and neuropsychological data. The patients presented varying degrees of AD severity, with one patient at the MCI stage. To validate the accuracy of ChatGPT in AD diagnosis, we compared its assessments with those made by neurologists and geriatricians who were blinded to the aims of the study. Additionally, we enlisted other blinded physicians to determine whether the diagnoses were formulated by their human colleagues or by artificial

intelligence, akin to a Turing Test. Our expectations were that ChatGPT would demonstrate a high level of accuracy in AD diagnosis, potentially providing valuable insights into the diagnostic process.

METHODS

The four cases

In this study, four typical cases were generated, one representing the mild stage of AD, one representing the moderate stage of AD, one representing the advanced stage of AD, and one representing MCI. A comprehensive overview of the demographic, professional, and health characteristics of each case is provided below. The cases were constructed for the sole purpose of this study. Genuine patient data typically contains sensitive and confidential information related to their health, medical history, and personal particulars and, for ethical reasons concerning data security and patient confidentiality, could not be shared with an artificial intelligence system like ChatGPT. The current version of ChatGPT does not seem to have robust safeguards in place to protect patient data from potential unauthorized access, breaches, or misuse. Additionally, the matter of data ownership is a matter of debate when considering the use of ChatGPT. It remains unclear who has ownership of the data and how it may be employed beyond immediate healthcare contexts. These considerations hold paramount importance in the context of patient privacy and data protection.

Procedures

We provided ChatGPT, as well as two blinded specialists, with a comprehensive clinical picture of the four cases as may be provided to specialists.

ChatGPT diagnosis

See the Supplementary Material for a transcript of the ChatGPT diagnoses.

Physicians' diagnosis

We provided a geriatrician and a neurologist with the same information as ChatGPT and asked them to provide a diagnosis. These physicians were blind to the study's aims and hypotheses. To compare their diagnoses with those of ChatGPT, we requested the physicians to describe the main symptoms of each case in approximately half a page and provide the

diagnosis. Both physicians provided the appropriate diagnosis for each case (i.e., Case1: mild AD, Case2: moderate AD, Case3: severe AD, Case4: MCI). The agreement between their diagnoses and those made by ChatGPT was 100%. However, unlike ChatGPT, they specified the stage of Case1 without being prompted to do so.

ChatGPT versus physicians' diagnosis

We provided another geriatrician and another neurologist, who were blinded to the aims of the study, with the diagnoses of the four cases as made by ChatGPT and the other two physicians. we informed them that the diagnoses were made either by colleagues or by artificial intelligence. We asked these physicians to decide whether each diagnosis was made by: 1) a colleague, 2) ChatGPT, or 3) "do not know". The physicians answered "do not know" for all four cases.

DISCUSSION

Within the constantly evolving landscape of artificial intelligence, we investigated whether AD diagnosis can be aided or supplemented using ChatGPT. We provided ChatGPT with descriptions of four cases: one representing the mild stage of AD, one representing the moderate stage of AD, one representing the advanced stage of AD, and representing MCI. ChatGPT accurately diagnosed the four cases. Similar diagnoses were made by two blinded professionals. Interestingly, we invited other blinded professionals to decide whether these diagnoses were made by colleagues or artificial intelligence, and they could not distinguish between the two, thereby, ChatGPT passing the Turing Test. These findings suggest that ChatGPT can be a useful tool for symptom assessment and may aid with the clinical diagnosis of AD.

AD is a complex disorder as multiple mechanisms (e.g., cognitive, neurobiological, and even environmental) are involved in its pathogenesis and progression, and it is stills unclear how each mechanism contributes to the disease and specific features of its clinical presentation. AD diagnosis is a challenging clinical task as it requires considering a wide range of parameters. Moreover, there is not a single test that can definitively diagnose AD. The diagnosis is typically based on a combination of medical history, clinical evaluation, and various biomarker assessments. The AD diagnosis is also challenging because clinicians need to consider that symptoms may overlap with those of other medical conditions, especially

at the mild stage of AD where memory problems and other cognitive manifestations can overlap with normal age-related changes and other dementing disorders. Also, AD progresses somewhat differently in each individual, making it challenging to establish a standard diagnosis for each stage of the disease. Our study, however, provides initial evidence for the potential value of ChatGPT in the diagnosis of AD. As evident by ChatGPT's responses, its diagnostic process is built upon the successful analysis of a variety of cognitive, clinical, neurological and biomarker information, in addition to the patients' histories and complaints. Thus, ChatGPT can perform certain basic diagnostic tasks for AD. ChatGPT can even situate patients within a precise stage of AD. Clinicians have the option to utilize ChatGPT in a supplementary capacity. For instance, ChatGPT can be furnished with comprehensive case descriptions to help identify and delineate primary symptoms. Subsequently, clinicians can assess and verify this analysis to expedite the diagnostic procedure. Furthermore, the diagnostic approach of ChatGPT can serve as a valuable model for trainees, offering them a basis for symptom recognition and, in turn, aiding in the development of their capacity to evaluate and address any irregularities.

While ChatGPT may prove to be a valuable tool in aiding the diagnosis of AD, it is imperative to emphasize that, in its current developmental stage, it falls far short of replacing clinicians. Physicians are trained to deliver comprehensive diagnoses by considering the myriad mechanisms underlying the onset and progression of AD. This was, somehow, evidenced in our study as the blinded physicians automatically diagnosed Case 1 as mild stage AD, unlike ChatGPT who only did so after prompting. However, ChatGPT automatically situated the other cases into their respective stage of AD, possibly because it built upon our previous prompt. We note with interest and surprise that ChatGPT demonstrated the ability to learn from its previous interaction with a user. Another concern with the use of ChatGPT in diagnosis is that the clinical interview, an integral component of the diagnostic process, should be administered by physicians. Hence, it is vital to view ChatGPT as a complementary tool to clinical judgment, and it should always be employed by physicians who possess the training and expertise to detect and interpret AD symptoms.

Our study can be situated within the growing body of research on the application of ChatGPT in Healthcare [5–10, 19–24] and, specifically, in the area of medical diagnosis [11–16]. This research shows how ChatGPT may demonstrate high levels of accu-racy in diagnosing medical conditions. Our study also extends a previous study showing that ChatGPT could accurately interpret scores of neuropsychological tests [18]. However, the present study is original in that it provides evidence of the ability of ChatGPT to diagnose AD and, critically, its ability to classify patients into specific stages of AD, including its preclinical phases.

One potential limitation of our study may be the use of generated cases. These cases were what one may consider typical "textbook" examples. The consideration of including data related to real cases could further strengthen the generalizability of our findings, although ethical aspects must be carefully considered. Adherence to ethical and privacy considerations is paramount when integrating ChatGPT into the diagnostic process. For example, the matter of data ownership is a subject of debate, as ChatGPT may use data to enhance its learning process. This is why we chose to deliberately create the four cases rather than use actual patient data. Privacy and professional confidentiality pose core ethical challenges when employing Large Language Models in the medical domain. The deployment of personal assistants powered by these models raises significant concerns regarding patient privacy and practitioner confidentiality. The sensitivity of data transmitted to these "personal assistants" requires scrutiny by regulatory bodies to mitigate potential intrusions in the medical domain. Another ethical challenge lies in the potential misuse of language models. The unregulated application of language models poses a prominent risk in the medical domain. Therefore, it is crucial to implement control measures aligned with legal frameworks to prevent any deleterious misuse of these potent linguistic tools.

In conclusion, while the use of ChatGPT in AD diagnosis is promising, it should be seen as an adjunct to clinical judgment rather than a replacement. The findings of this study open new avenues for the application of artificial intelligence in medical diagnosis, ultimately benefiting patients and contributing to the advancement of the field. It is a step toward harnessing the potential of Artificial Intelligence to improve diagnosis as well as the lives of individuals affected by AD and similar conditions.

AUTHOR CONTRIBTUIONS

Mohamad El Haj (Conceptualization; Investigation; Methodology; Writing – original draft); Claire

Boutoleau Bretonnière (Resources; Supervision; Validation); Karim Gallouj (Methodology; Supervision; Validation); Nathalie Wagmann (Resources; Supervision; Validation); Pascal Antoine (Project administration; Resources); Dimitrios Kapogiannis (Supervision; Validation; Writing – review & editing); Guillaume Chapelet (Investigation; Methodology; Supervision; Validation; Writing –review & editing).

ACKNOWLEDGMENTS

FUNDING

CONFLICT OF INTEREST

MEH is an Editorial Board Member of this journal but was not involved in the peer-review process of this article nor had access to any information regarding its peer-review.

DATA AVAILABILITY

The data supporting the findings of this study are available within the article.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/ADR-230191.

REFERENCES

[1] Dubois B, Villain N, Frisoni GB, Rabinovici GD, Sabbagh M, Cappa S, Bejanin A, Bombois S, Epelbaum S, Teichmann M, Habert M-O, Nordberg A, Blennow K, Galasko D, Stern Y, Rowe CC, Salloway S, Schneider LS, Cummings JL, Feldman HH (2021) Clinical diagnosis of Alzheimer's disease: Recommendations of the International Working Group. *Lancet Neurol* **20**, 484-496.

[2] Santos CY, Snyder PJ, Wu W-C, Zhang M, Echeverria A, Alber J (2017) Pathophysiologic relationship between Alzheimer's disease, cerebrovascular disease, and cardiovascular risk: A review and synthesis. *Alzheimers Dement (Amst)* **7**, 69-87.

[3] Anderson ND (2019) State of the science on mild cognitive impairment (MCI). *CNS Spectrums* **24**, 78-87.

[4] Bhatia P (2023) ChatGPT for academic writing: A game changer or a disruptive tool? *J Anaesthesiol Clin Pharmacol* **39**, 1-2.

[5] Khan RA, Jawaid M, Khan AR, Sajjad M (2023) ChatGPT – Reshaping medical education and clinical management. *Pak J Med Sci* **39**, 605-607.

[6] Lee H (2023) The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ*, doi: 10.1002/ase.2270.

[7] Dave T, Athaluri SA, Singh S (2023) ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* **6**, 1169595.

[8] Xue VW, Lei P, Cho WC (2023) The potential impact of ChatGPT in clinical and translational medicine. *Clin Transl Med* **13**, e1216.

[9] Jeblick K, Schachtner B, Dexl J, Mittermeier A, Stüber AT, Topalis J, Weber T, Wesp P, Sabel BO, Ricke J, Ingrisch M (2023) ChatGPT makes medicine easy to swallow: An exploratory case study on simplified radiology reports. *Eur Radiol*, doi: 10.1007/s00330-023-10213-1

[10] Baumgartner C (2023) The potential impact of ChatGPT in clinical and translational medicine. *Clin Transl Med* **13**, e1206.

[11] Delsoz M, Raja H, Madadi Y, Tang AA, Wirostko BM, Kahook MY, Yousefi S (2023) The use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports. *Ophthalmol Ther* **12**, 3121-3132.

[12] Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, Landman A, Dreyer K, Succi MD (2023) Assessing the utility of ChatGPT throughout the entire clinical workflow: Development and usability study. *J Med Internet Res* **25**, e48659.

[13] Balas M, Ing EB (2023) Conversational AI models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel Pro Differential Diagnosis Generator. *JFO Open Ophthalmol* **1**, 100005.

[14] Cascella M, Montomoli J, Bellini V, Bignami E (2023) Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios. *J Med Syst* **47**, 33.

[15] Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T (2023) Diagnostic accuracy of differential-diagnosis lists generated by generative pre-trained Transformer 3 Chatbot for clinical vignettes with common chief complaints: A pilot study. *Int J Environ Res Public Health* **20**, 3378.

[16] Xv Y, Peng C, Wei Z, Liao F, Xiao M (2023) Can Chat-GPT a substitute for urological resident physician in diagnosing diseases?: A preliminary conclusion from an exploratory investigation. *World J Urol* **41**, 2569-2571.

[17] Agbavor F, Liang H (2022) Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health* **1**, e0000168.

[18] El Haj M, Boutoleau-Bretonnière C, Chapelet G (2023) ChatGPT's dance with neuropsychological data: A case study in Alzheimer's disease. *Ageing Res Rev* **92**, 102117.

[19] Li W, Zhang Y, Chen F (2023) ChatGPT in colorectal surgery: A promising tool or a passing fad? *Ann Biomed Eng* **51**, 1892-1897.

[20] Cheng K, Li Z, Guo Q, Sun Z, Wu H, Li C (2023) Emergency surgery in the era of artificial intelligence: ChatGPT could be the doctor's right-hand man. *Int J Surg* **109**, 1816-1818.

[21] Cheng K, Li Z, Li C, Xie R, Guo Q, He Y, Wu H (2023) The potential of GPT-4 as an AI-powered virtual assistant for surgeons specialized in joint arthroplasty. *Ann Biomed Eng* **51**, 1366-1370.

[22] Ahn C (2023) Exploring ChatGPT for information of cardiopulmonary resuscitation. *Resuscitation* **185**,109729.

[23] Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM (2023) Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med* **183**, 589-596.

[24] Haver HL, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, Yi PH (2023) Appropriateness of breast cancer prevention and screening recommendations provided by ChatGPT. *Radiology* **307**, e230424.