Case Report

Patients With Acute Stroke Symptoms Consulting Large Language Model Chatbots Before Presenting to the Emergency Department: Case Series

Sara Dierickx[1,2*], MD; Paolo Speziali[3*], MSc; Helena Gossye[1,2*], MD, PhD; Sofia Maldonado Slootjes[1,2*], MD; Sylvie De Raedt[1,2*], MD, PHD; Anaïs Missiaen[1,2*], MD; Michel Van Lint[4*], MD; Edwin Van Dellen[5*], MD, PHD; Fenne Vandervorst[1,2*], MD

[1]NEUR Research Group, Center for Neurosciences (C4N), Vrije Universiteit Brussel, Brussels, Belgium
[2]Department of Neurology, Universitair Ziekenhuis Brussel, Brussels, Belgium
[3]AI Research Group, Vrije Universiteit Brussel, Brussels, Belgium
[4]Department of Ophthalmology, Universitair Ziekenhuis Brussel, Brussels, Belgium
[5]Department of Psychiatry, Brain Centre, University Medical Center Utrecht, Utrecht, The Netherlands
*all authors contributed equally

Corresponding Author:

Fenne Vandervorst, MD
Department of Neurology
Universitair Ziekenhuis Brussel
Laarbeeklaan 101
Brussels 1090
Belgium
Phone: 32 24776801
Email: fenne.vandervorst@uzbrussel.be

Abstract

Background: Chatbots are being increasingly consulted for health-related questions. Understanding their potential to recognize serious conditions such as stroke is crucial for assessing their role in clinical decision-making and public health.

Objective: This case series describes 3 patients who consulted large language model (LLM)–powered chatbots before seeking care at the emergency department and explores the potential benefits and limitations of such chatbot use in the context of acute stroke symptoms.

Methods: Cases were identified when patients or their family members spontaneously reported during history-taking that they had consulted a chatbot at symptom onset and the corresponding chat history was available for review. We describe 3 patients who consulted LLM-powered chatbots before presenting to the emergency department with symptoms subsequently diagnosed as stroke. The chatbot interactions were reviewed for recognition of stroke as a possible diagnosis and for recommendations to seek medical care.

Results: In 2 cases, the LLM-powered chatbots correctly recognized the likelihood of stroke and the need for urgent medical assessment. In the third case, the symptoms were not initially recognized as signs of stroke, and medical help was advised only once they persisted.

Conclusions: LLM-powered chatbots have the potential to recognize stroke symptoms and promote timely access to medical care. However, their responses may not reliably identify stroke symptoms or provide sufficiently urgent advice. Deployment of such chatbots in the context of acute neurological symptoms requires clinical oversight and patient education regarding their limitations.

Interact J Med Res 2026;15:e107318; doi: 10.2196/107318

Keywords: artificial intelligence; AI; ischemic stroke; transient; large language models; public health; stroke

Background

Stroke symptoms are highly variable and depend largely on the location of the vascular lesion within the central nervous system. A defining feature of symptoms caused by cerebrovascular events is their sudden onset, which helps distinguish them from other central nervous system disorders, such as demyelinating diseases. The FAST (face drooping, arm weakness, speech difficulties, time to call emergency services) mnemonic summarizes key stroke symptoms, with the "time" component emphasizing the urgency of immediate medical attention [1]. Timely treatment—including thrombolysis and thrombectomy—remains the most critical determinant of outcome, with every minute being decisive. Mass media campaigns focusing on individual components of the FAST acronym have been shown to improve public awareness of stroke symptoms and increase thrombolysis rates [2]. However, their effects tend to diminish after approximately 6 months, highlighting the need for repeated initiatives. Additionally, the FAST acronym is less sensitive in detecting posterior circulation stroke symptoms, often presenting with ataxia and visual changes. More extensive acronyms have been developed and tested; however, they were associated with lower retention rates [3].

AI systems such as ChatGPT 4.1 (OpenAI) [4] and Grok 3 (SpaceXSI) [5] are becoming increasingly accessible worldwide. Large language model (LLM)–based systems can extract information from conversations, integrate relevant context, and apply advanced communication skills. In the setting of acute stroke, history-taking can be exceptionally challenging. Patients may experience anosognosia, visual or language disturbances such as hemianopsia and aphasia, or impaired use of an upper limb due to paresis, and/or neglect, all of which can limit their ability to provide accurate information or interact effectively.

At our hospital, 3 patients experiencing acute ischemic stroke presented at the emergency department after consulting an LLM-powered AI chatbot. Atypical stroke symptoms were correctly identified in one conversation, despite—or possibly due to—the presence of spelling errors. The original and translated conversations are provided in Multimedia Appendices 1–3.

Of 3 patients, 2 were appropriately prompted to seek emergency medical care. The third patient was initially advised to rest; however, after 13 hours of persistent symptoms, he was referred to seek medical attention. Each case is described separately, followed by a brief interpretation and commentary by medical specialists.

These cases exemplify an emerging phenomenon increasingly encountered in clinical practice: patients using AI chatbots to seek medical advice regarding symptoms they experience. Given the growing accessibility and use of these technologies, it is important to understand their potential role in the diagnostic pathway. The aim of this article is to evaluate the benefits and risks associated with chatbot-based consultations by linking chatbot interactions with clinical and radiological findings. In the Discussion section, we further explore the practical implications of these findings and consider future perspectives on the integration of chatbots into health care systems.

Case Presentation

Ethical Considerations

Ethics approval was not required for this case report. Written informed consent was obtained from all patients included in this case report.

Case 1

Description

A 79-year-old male patient with a medical history of diabetes and hypertension experienced sudden blurred vision lasting approximately 4 to 5 minutes. One hour later, he experienced transient word-finding difficulties for 3 to 4 minutes. He subsequently drove a 54-year-old female friend to a dental appointment and mentioned these episodes to her. Curious about the potential causes of such speech difficulties, she consulted ChatGPT 4.1 for insights.

ChatGPT 4.1 generated a list of differential diagnoses, including transient ischemic attack (TIA), migraine with aura, focal seizures, multiple sclerosis, stress or fatigue, hypoglycemia, medication side effects, alcohol or drug use, and dry mouth and jaw or dental issues. It advised that if the symptoms were isolated and the individual felt well afterward, the episode could be considered benign. However, if the symptoms recurred and were accompanied by weakness, vision problems, confusion, or numbness, it recommended seeking medical attention. The chatbot then asked whether she wanted to continue the discussion and offered to ask additional questions, to which she agreed. ChatGPT 4.1 then asked about the duration of the symptoms and any contextual factors (eg, stress, fatigue, alcohol consumption, or the possibility of hypoglycemia). It also inquired about associated symptoms, previous episodes, and the patient's medical history. The patient's friend responded that the symptoms lasted a few minutes and occurred without any specific triggers. She mentioned that the patient had experienced blurred vision 1 hour prior, indicating it was the first such episode and that he was being treated for diabetes with metformin and dulaglutide.

On the basis of this exchange, ChatGPT 4.1 suggested that a TIA was the most likely diagnosis and warned her about its association with an elevated stroke risk. It recommended seeking urgent medical attention by going to the emergency department. Furthermore, it provided guidance on how to clearly summarize the events for the physician and explained what to expect during further medical examinations (Multimedia Appendix 1).

The patient subsequently presented to our emergency department and was admitted for a clinical suspicion of recurrent TIA. Computed tomography (CT) scan demonstrated luminal narrowing of the left internal carotid artery. Magnetic resonance imaging (MRI) revealed no evidence of

recent cerebral ischemia but confirmed the presence of left internal carotid artery narrowing attributable to an underlying dissection. A diagnosis of retinal and MRI-negative cerebral TIA was considered.

Interpretation

On the basis of a brief but accurate description of symptoms, the chatbot concluded transient word-finding difficulties. A broad but largely accurate differential diagnosis was given (only focal seizures seem less probable based on the description of dysarthria instead of aphasia). Initially, a more benign phenomenon was suggested; however, no further explanation was provided. The question regarding other neurological deficits in this case appears to have prompted the recommendation to seek urgent medical attention, even though isolated, acute, and transient neurological deficits alone should always warrant medical evaluation.

Case 2

Description

A 59-year-old patient with a history of obstructive sleep apnea experienced sudden onset of visual disturbances affecting the right visual field, along with left-sided neck pain; lightheadedness; and paresthesia in the right arm, leg, and face.

Before seeking medical attention, he consulted the AI chatbot Grok 3, which suggested both urgent and nonurgent explanations, including migraine with aura, TIA, stroke, peripheral neuropathy, nerve compression, hypoglycemia, hyperventilation, and ocular disorders. It advised calling the emergency services (112) if the symptoms were sudden in onset; one-sided; or associated with confusion, weakness, or speech difficulties (Multimedia Appendix 2). As the patient's symptoms were both acute and lateralized, he promptly presented to the emergency department.

A perfusion brain CT scan showed hypoperfusion in the territory of the left posterior cerebral artery, while the noncontrast CT showed no abnormalities, and an occlusion of the P1 segment was found on CT angiography. We administered alteplase and conducted intra-arterial digital subtraction angiography, which revealed that the clot had already dissolved. The patient's symptoms improved within minutes to hours, except for a persistent right-sided inferior quadrantanopia. He was admitted to the neurology ward for further stroke workup. Brain MRI confirmed an acute infarction in the left occipital lobe; the image can be found in Figure S1 in Multimedia Appendix 4.

Interpretation

In the original description of symptoms, some spelling errors were made, possibly due to the stroke itself causing sensory disturbances and hemianopsia. Nevertheless, symptoms were recognized correctly. Once again, a broad differential diagnosis was listed, with, at the end, instructions about the degree of urgency. Symptoms with an acute onset; unilateral distribution; and the co-occurrence of speech difficulties, confusion, or weakness were correctly identified as requiring urgent medical advice. However, a milder and more fluctuating or recurrent course can also occur in the context of stroke and should not always be considered as nonurgent.

Case 3

Description

A 25-year-old male patient, without relevant medical history, experienced weakness in the right arm and right leg, accompanied by speech difficulties and peripheral visual loss. The symptoms started while he was washing his hands after using the toilet. He consulted ChatGPT 4.1 for advice and explained that his head felt "weird" after going to the toilet, while his eyes appeared "crooked." ChatGPT suggested several possible explanations, including a drop in blood pressure, dehydration, hypoglycemia, and excessive straining. It advised seeking medical attention if symptoms such as visual disturbances, weakness, confusion, nausea, vomiting, or loss of balance occurred. The chatbot subsequently asked several follow-up questions. The patient reported he remained dizzy and that his visual disturbances were unchanged; no other symptoms were present. On the basis of this response, ChatGPT concluded that a temporary drop in blood pressure most likely caused the symptoms and advised lying down, drinking water, and breathing slowly. It stated that the symptoms should improve within 5 to 10 minutes and recommended seeking urgent medical attention if they persisted. When subsequently asked whether the symptoms had improved, the patient replied that they had not. ChatGPT then asked whether he experienced tingling or weakness in an arm or leg, which the patient denied. After several additional questions, the patient suggested going to sleep for a while. ChatGPT indicated that this was reasonable but advised informing someone beforehand. The patient also mentioned that he normally had mild strabismus but that it had worsened during this episode. ChatGPT responded that this could be associated with a drop in blood pressure. Later, the patient reported feeling better when lying down, which ChatGPT interpreted as further supporting this explanation. When the patient asked whether the symptoms could be related to a very spicy sauce he had eaten for the first time, ChatGPT indicated that this could potentially contribute to a drop in blood pressure. Finally, when the patient reported that the symptoms persisted after 13 hours, ChatGPT advised him to seek medical attention at the emergency department (Multimedia Appendix 3). At the emergency department, neurological examination revealed dysarthria and pronation of the right arm without drift. No objective motor deficit of the leg could be demonstrated, although the patient reported a subjective difference while walking. A contrast-enhanced CT scan of the head showed no abnormalities. The patient was hospitalized at the Stroke Unit. Subsequently, the brain MRI confirmed an acute ischemic stroke located in the left thalamus. An etiological workup did not reveal an underlying cause.

Interpretation

In this case, symptoms were described concisely with attention to the circumstances in which they occurred. Stroke

was unfortunately not mentioned in the list of possible causes of dizziness and altered vision. The following questions focused on persistence of symptoms as well as associated symptoms. Subsequently, recommendations regarding time frames were not applied consistently by the chatbot, resulting in a delayed presentation to the emergency department. Alternative explanations proposed by the patient, including strabismus and the intake of spicy harissa sauce, were endorsed by the chatbot, which may reflect a tendency of LLMs to reinforce user satisfaction.

Discussion

Principal Findings

On the basis of these 3 examples of patients experiencing acute ischemic stroke consulting a LLM-based AI chatbot for medical advice, the following observations were made: first, atypical stroke symptoms can be detected by AI chatbots with little interference from writing or language errors, often occurring in the setting of stroke, due to neurological deficits such as aphasia, hemianopia, and motor impairment; second, all patients described followed the advice given by the chatbot; and third, the way in which certain symptoms are described, including contextual detail and more elaborate phrasing, inherently shaped by the patient's own interpretation of their symptoms, seems to influence how the symptoms are interpreted by the chatbot.

Opportunities of AI Chatbots

Stroke symptoms remain poorly recognized in the general population, and LLM-powered AI chatbots offer the advantage of being available at all times. In 2 of 3 described cases, they were able to recognize stroke symptoms even when described in lay terms with spelling errors in the second case. Language disorders are a characteristic feature of numerous neuropsychiatric conditions, and LLMs have the capability to detect these disturbances, thereby not only preserving comprehension of the communicated message but also possibly contributing to both diagnosis and prognosis [6,7]. In a more general medical context, the potential of chatbots for safe and effective use has been widely reported [8]. For instance, ChatGPT 4.1 achieved passing grades on the United States Medical Licensing Examination [9]. Moreover, in a cross-sectional study of patient questions in an online forum, the quality and empathy of ChatGPT 4.1 responses were rated significantly higher for both quality and empathy, even when compared with the longest physician-authored responses [10]. More recently, specialized LLM-based AI bots have been created with improved capabilities for clinical reasoning, history-taking, and high-quality conversation [11]. Although the versatility of widely used AI LLMs may yield impressive results, challenges arise in highly specialized domains, such as neurology.

Risks of AI Chatbots

Concerns have been raised regarding the limitations of medical advice provided by AI chatbots. For example, these systems can make errors due to discrimination or stereotyping [12]. Because the LLMs are created based on unchecked, potentially biased data, these biases may be adopted in their output [13]. Aside from general reliability issues, chatbots that are optimized directly for user satisfaction can develop manipulative strategies aimed at securing positive feedback (also called "sycophancy") [14] and user retention. As a result, they may be more prone to reinforcing user-suggested explanations and prolonging patient interactions rather than directing users toward appropriate clinical evaluation, as seen in case 3. However, the phenomenon of sycophancy also implies a sensitivity of the chatbot to factors such as the patient's experience of certain symptoms, concerns about illness or symptoms, and general well-being—elements that are important in human patient interaction. Nonetheless, this attentiveness becomes problematic when it causes the chatbot to prioritize the patient's own framing over clinically relevant warning symptoms, as appeared to occur in the third case. These findings reinforce the need for human oversight and robust evaluation prior to clinical deployment.

In addition, evaluating a chatbot's robustness may prove difficult because they are highly sensitive to small changes in user phrasing and to the language used. For instance, during summary generation, keywords such as "not" may be dropped, which can completely reverse the intended meaning of the advice.

Another reported limitation is the occurrence of errors under the form of completely erroneous generations that are often referred to as "perseveration" [13]. They occur when the required information is not available in the training dataset, and the model therefore fills in a plausible-sounding and coherently phrased fabrication. Unfortunately, these are indistinguishable from clinically valid information for a lay person. Overreliance on automated models should thus be avoided.

Finally, there is a lack of "privacy," as user-provided data are not protected under medical confidentiality regulations [12], and questions regarding liability and accountability are likely to arise.

Conclusions and Future Perspectives

Patients increasingly consult AI chatbots for medical advice regarding their diagnosis and treatment. Phenomena observed in these interactions, including stereotyping, sycophancy, and perseveration, underscore the need for human oversight, the development of purpose-built chatbots for medical advice, and rigorous validation prior to deployment in clinical settings.

It is important to recognize, however, that some of these phenomena, such as sycophancy, are not unique to chatbot interactions and may equally occur in human communication. Moreover, in clinical practice, history-taking constitutes only 1 component of the diagnostic workup, complemented by findings from clinical neurological examination, medical imaging, and other investigations. Establishing their position within the existing acute care landscape is an essential

first step, allowing future research to focus on meaningful comparisons.

Beyond their integration into existing care pathways, AI chatbots and their successors have the potential to transform medical practice as we know it. Caution must be taken, as blindly trusting them may lead to overlooked errors with important consequences, while overdiagnosis may lead to medical overconsumption. Given the ongoing need for human expertise and legal concerns regarding privacy and liability, we propose that health care professionals as well as local policymakers should be involved in the development and testing of LLMs' underlying AI chatbots. Furthermore, we should inform patients about risks and limitations.

Curiously but cautiously integrating these new tools into decision-making, information storage, and administrative support may pave the way for AI chatbots supporting clinicians. Assisting lay users to navigate this evolving landscape provides a novel challenge.

Generative AI Use

ChatGPT 4.1 (OpenAI) and Grok 3 (xAI) were used to preserve spelling errors and nuances when translating the original chats.

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

Conceptualization: EVD, FV
Methodology: EVD, FV
Investigation: SD, AM
Writing—original draft: SD, FV
Writing—review & editing: PS, EVD, HG, SMS, SDR, MVL

Conflicts of Interest

None declared.

Multimedia Appendix 1

Chat history for Case 1.
[DOCX File (Microsoft Word File), 17 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Chat history for Case 2.
[DOCX File (Microsoft Word File), 14 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Chat history for Case 3.
[DOCX File (Microsoft Word File), 22 KB-Multimedia Appendix 3]

Multimedia Appendix 4

MRI scan for Case 2.
[DOCX File (Microsoft Word File), 150 KB-Multimedia Appendix 4]

Checklist 1

CARE checklist.
[DOCX File (Microsoft Word File), 18 KB-Checklist 1]

References

1. Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati Prehospital Stroke Scale: reproducibility and validity. Ann Emerg Med. Apr 1999;33(4):373-378. [doi: 10.1016/s0196-0644(99)70299-4] [Medline: 10092713]
2. Advani R, Naess H, Kurz M. Mass media intervention in Western Norway aimed at improving public recognition of stroke, emergency response, and acute treatment. J Stroke Cerebrovasc Dis. Jun 2016;25(6):1467-1472. [doi: 10.1016/j.jstrokecerebrovasdis.2016.02.026] [Medline: 27019989]
3. Darkhabani MZ, Homa-Bonell JK, Thoreson L, Bobholz JA, Spaulding D, Engebose M. BE FAST versus FAST: a randomized pilot trial comparing retention of stroke symptoms between 2 mnemonics. J Am Heart Assoc. Oct 2024;13(19):e035696. [doi: 10.1161/JAHA.123.035696] [Medline: 39291473]
4. ChatGPT. URL: https://chat.openai.com [Accessed 2026-04-13]
5. SpaceXSI. URL: https://x.ai [Accessed 2025-09-22]

6. de Boer JN, Voppel AE, Begemann MJ, Schnack HG, Wijnen F, Sommer IE. Clinical use of semantic space models in psychiatry and neurology: a systematic review and meta-analysis. Neurosci Biobehav Rev. Oct 2018;93:85-92. [doi: 10.1016/j.neubiorev.2018.06.008] [Medline: 29890179]
7. Ciampelli S, Voppel AE, de Boer JN, Koops S, Sommer IE. Combining automatic speech recognition with semantic natural language processing in schizophrenia. Psychiatry Res. Jul 2023;325:115252. [doi: 10.1016/j.psychres.2023.115252] [Medline: 37236098]
8. Luykx JJ, Gerritse F, Habets PC, Vinkers CH. The performance of ChatGPT in generating answers to clinical questions in psychiatry: a two-layer assessment. World Psychiatry. Oct 2023;22(3):479-480. [doi: 10.1002/wps.21145] [Medline: 37713576]
9. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. Feb 8, 2023;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
10. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. Jun 1, 2023;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
11. Tu T, Schaekermann M, Palepu A, et al. Towards conversational diagnostic artificial intelligence. Nature. Jun 2025;642(8067):442-450. [doi: 10.1038/s41586-025-08866-7] [Medline: 40205050]
12. Fournier-Tombs E, McHardy J. A medical ethics framework for conversational artificial intelligence. J Med Internet Res. Jul 26, 2023;25:e43068. [doi: 10.2196/43068] [Medline: 37224277]
13. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med. Aug 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
14. Williams M, Carroll M, Narang A, Weisser C, Murphy B, Dragan A. On targeted manipulation and deception when optimizing LLMs for user feedback [Poster]. Presented at: International Conference on Learning Representations 2025; Apr 24-28, 2025; Singapore, Singapore. URL: https://openreview.net/pdf?id=Wf2ndb8nhf [Accessed 2026-11-06]

Abbreviations

CT: computed tomography
FAST: face drooping, arm weakness, speech difficulties, time to call emergency services
LLM: large language model
MRI: magnetic resonance imaging
TIA: transient ischemic attack

Edited by Taiane de Azevedo Cardoso; peer-reviewed by Milan Toma, Shiyang Zhuang; submitted 17.Jul.2026; initial comments to author 01.Sep.2026; final revised version received 19.Oct.2026; accepted 23.Oct.2026; published 27.Nov.2026

Please cite as:
Dierickx S, Speziali P, Gossye H, Slootjes SM, De Raedt S, Missiaen A, Van Lint M, Van Dellen E, Vandervorst F
Patients With Acute Stroke Symptoms Consulting Large Language Model Chatbots Before Presenting to the Emergency Department: Case Series
Interact J Med Res 2026;15:e107318
URL: https://www.i-jmr.org/2026/1/e107318
doi: 10.2196/107318

© Sara Dierickx, Paolo Speziali, Helena Gossye, Sofia Maldonado Slootjes, Sylvie De Raedt, Anaïs Missiaen, Michel Van Lint, Edwin Van Dellen, Fenne Vandervorst. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.