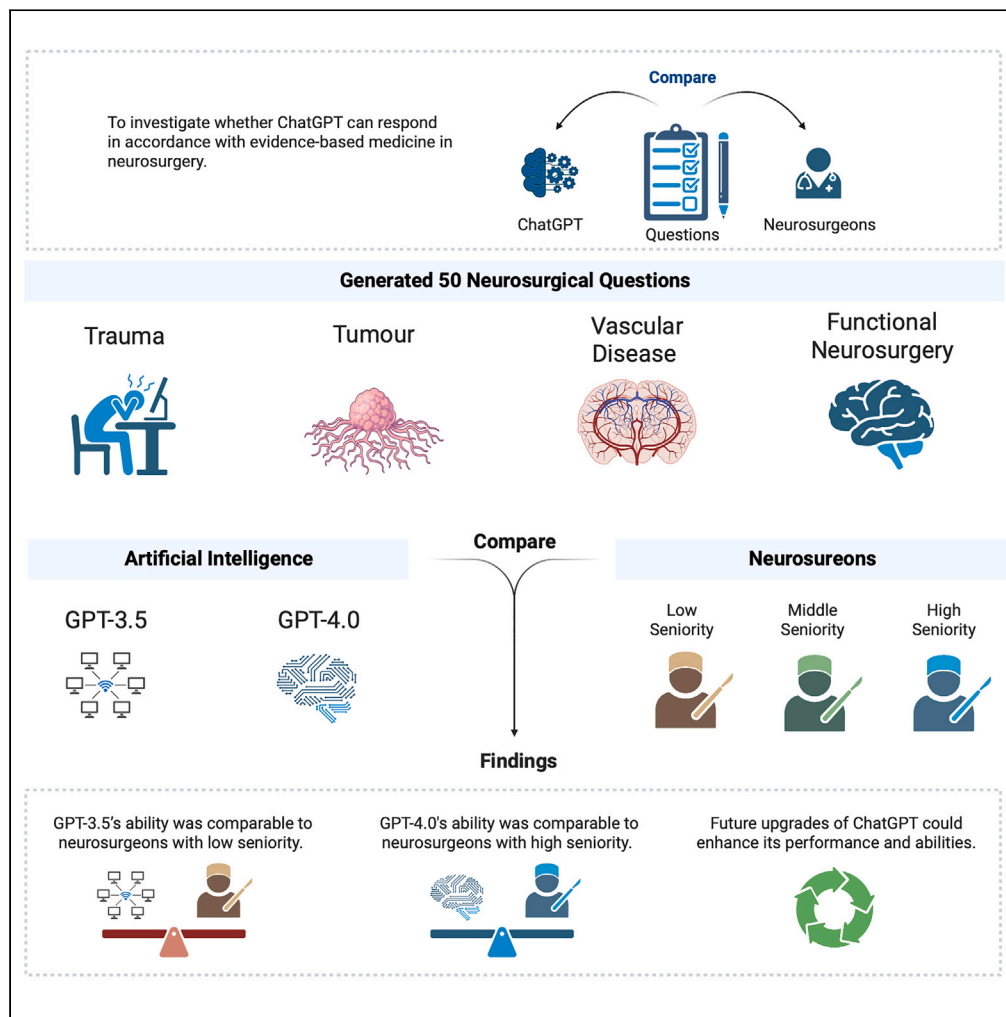


## Article

# A descriptive study based on the comparison of ChatGPT and evidence-based neurosurgeons



Jiayu Liu, Jiqi Zheng, Xintian Cai, Dongdong Wu, Chengliang Yin

604269346@qq.com (D.W.)  
chengliangyin@163.com (C.Y.)

## Highlights

GPT-3.5's ability was comparable to neurosurgeons with low seniority

GPT-4.0's ability was comparable to neurosurgeons with high seniority

Future upgrades of ChatGPT could enhance its performance and abilities

## Article

## A descriptive study based on the comparison of ChatGPT and evidence-based neurosurgeons

Jiayu Liu,<sup>1</sup> Jiqi Zheng,<sup>2</sup> Xintian Cai,<sup>3</sup> Dongdong Wu,<sup>4,\*</sup> and Chengliang Yin<sup>5,6,\*</sup>

## SUMMARY

ChatGPT is an artificial intelligence product developed by OpenAI. This study aims to investigate whether ChatGPT can respond in accordance with evidence-based medicine in neurosurgery. We generated 50 neurosurgical questions covering neurosurgical diseases. Each question was posed three times to GPT-3.5 and GPT-4.0. We also recruited three neurosurgeons with high, middle, and low seniority to respond to questions. The results were analyzed regarding ChatGPT's overall performance score, mean scores by the items' specialty classification, and question type. In conclusion, GPT-3.5's ability to respond in accordance with evidence-based medicine was comparable to that of neurosurgeons with low seniority, and GPT-4.0's ability was comparable to that of neurosurgeons with high seniority. Although ChatGPT is yet to be comparable to a neurosurgeon with high seniority, future upgrades could enhance its performance and abilities.

## INTRODUCTION

ChatGPT, an artificial intelligence (AI) product developed by OpenAI, leverages large language models to generate texts similar to those written by humans, exhibiting immense potential in the medical domain.<sup>1</sup> Facilitating unique narrative responses by developing question-and-answer tasks with the help of a man-machine dialogue interface, ChatGPT presents medical information in a conversational and interactive way.<sup>2,3</sup> With the development of ChatGPT, many researchers have focused on how to apply ChatGPT in medicine. In recent years, researchers have explored the practical application value of ChatGPT in many aspects, such as medical education,<sup>4</sup> clinical management,<sup>5</sup> healthcare practice,<sup>6</sup> scientific writing,<sup>7</sup> clinical decision-making,<sup>8</sup> and medical information acquisition.<sup>9</sup>

Clinical decision support system (CDSS) provides medical information and recommendations for diagnosis and treatments to physicians and patients.<sup>10</sup> CDSS import a large amount of medical knowledge into the computer, which uses algorithms to simulate doctors' clinical diagnosis and treatment ideas and independently or assist doctors to diagnose and treat patients.<sup>11</sup> Researchers have shown that ChatGPT can assist with clinical decision support, which achieves optimization of clinical decision-making.<sup>12</sup> Some researchers have evaluated the value of ChatGPT in the CDSS of many clinical disciplines, such as infection,<sup>13</sup> emergency,<sup>14</sup> surgery,<sup>15</sup> ophthalmology,<sup>16</sup> and psychiatry.<sup>17</sup>

Neurosurgical diseases include trauma, tumors, vascular diseases (VDs), and functional neurosurgical diseases, which are complex and diverse, and their conditions change rapidly, requiring the assistance of CDSS.<sup>18</sup> In addition, neurosurgeons have the characteristics of long training cycles, high training costs, and inconsistent professional levels.<sup>19</sup> Although CDSS is needed more than other clinical specialties, research on ChatGPT-assisted CDSS in neurosurgery needs further exploration. T. H. Kung et al.<sup>20</sup> found that ChatGPT performed well in the United States Medical Licensing Exam (USMLE) without specialized training or reinforcement. Based on this result, we hypothesize that AI will potentially replace doctors, revolutionizing modern clinical medicine.<sup>3,21</sup>

With the development of ChatGPT, it is necessary to ascertain the accuracy of ChatGPT's evidence-based medicine-aligned responses. This analysis of ChatGPT's abilities may provide evidence of whether ChatGPT's ability is comparable to that of an evidence-based doctor in neurosurgery.

## Methods

## Study design

We conducted a descriptive study to investigate whether ChatGPT can provide responses in accordance with evidence-based medicine and compare this ability of ChatGPT with that of neurosurgeons by responding to questions. Specifically, the following were investigated: (1) the responses of ChatGPT compared to recommendations based on the best available evidence in the guidelines; (2) the responses of ChatGPT

<sup>1</sup>Department of Neurosurgery, the First Medical Centre, Chinese PLA General Hospital, Beijing 100853, China<sup>2</sup>School of Health Humanities, Peking University, Beijing 100191, China<sup>3</sup>Department of Graduate School, Xinjiang Medical University, Urumqi 830001, China<sup>4</sup>Department of Information, Daping Hospital, Army Medical University, Chongqing 400042, China<sup>5</sup>Faculty of Medicine, Macau University of Science and Technology, Macau 999078, China<sup>6</sup>Lead contact

\*Correspondence: 604269346@qq.com (D.W.), chengliangyin@163.com (C.Y.)

<https://doi.org/10.1016/j.isci.2023.107590>

**Table 1. Assessment of the responses**

Question	GPT-3.5	GPT-4.0	Low seniority	Middle seniority	High seniority
1. Is it appropriate to perform a lumbar puncture on an acute craniocerebral injury patient suspected of having a craniocerebral hematoma?	2	3	3	3	3
2. Is cranial repair necessary for an acute craniocerebral injury patient with a comminuted skull fracture found during surgery?	3	3	3	1	3
3. Is it appropriate to use sterile gauze to fill the nasal cavity or external ear canal in a patient with cerebrospinal fluid leakage due to skull base fracture?	3	3	2	2	3
4. Is scalp reduction and replantation appropriate for a patient with scalp avulsion more than 12 h after the injury?	3	2	1	2	2
5. A concussion patient whose headache is severe can you use morphine class drugs?	3	3	2	3	3
6. If patients with acute epidural hematoma are satisfied with intraoperative brain decompression, is it still necessary to perform subdural exploration in this situation?	0	2	1	2	2
7. For a patient after intracranial tumor resection, is it appropriate to raise the wound cavity drainage bag after surgery to quickly drain the fluid in the wound cavity?	3	3	3	3	3
8. Is surgical treatment appropriate for a small cavernous hemangioma found deep in the brain stem but asymptomatic?	3	3	3	3	3
9. Is intratumoral mass resection appropriate for patients with hemangioblastoma growing in the brain stem?	2	2	2	2	3
10. Is maximum safe tumor removal beneficial to the prognosis of a patient with high-grade glioma?	3	3	3	3	3
11. Can the total radiation dose be reduced appropriately for patients with high-grade glioma with large tumor volume and growing in vital functional areas?	3	2	2	2	3
12. Should a glioblastoma patient be treated with temozolomide during and after radiotherapy?	3	3	2	3	3
13. Should craniospinal axis irradiation be performed for patients with ependymoma without evidence of intracranial and spinal tumor spread after surgery?	3	3	2	2	3
14. Is radiotherapy plus temozolomide chemotherapy appropriate for a WHO Grade 3 glioma patient without 1p/19q combined deletion found by postoperative pathological examination?	1	3	1	3	3
15. Is total resection appropriate for patients with medulloblastoma found intraoperatively to invade the fourth ventricle floor?	1	1	1	0	2
16. Intraoperatively, meningioma was found in a patient with cavernous sinus lesions involving the inner and outer membranes of the neck. If the surgeon is not overly keen on total resection, is it appropriate to use stereotactic radiotherapy to control tumor growth after surgery?	2	3	3	2	3
17. For patients with meningiomas invading the anterior 1/3 of the sagittal sinus, can they be resected intraoperatively along with the sagittal sinus?	1	3	0	2	3
18. Should facial nerve reconstruction be performed as soon as possible for a patient with acoustic neuroma who accidentally ruptured the facial nerve during surgery?	3	3	2	3	3

(Continued on next page)

Table 1. Continued

Question	GPT-3.5	GPT-4.0	Low seniority	Middle seniority	High seniority
19. What is the preferred drug treatment for a patient with prolactinoma?	3	3	3	3	3
20. Is radiotherapy an option for craniopharyngioma?	3	3	3	2	3
21. In a patient with cystic angio-reticulocyte, is it necessary to resect the nodules during surgery?	0	2	1	3	3
22. What is the standard medicine for primary central nervous system lymphoma at first diagnosis?	3	3	2	2	3
23. Does intracranial melanoma need to be excised locally?	3	3	1	2	2
24. What is the preferred drug for treating delayed cerebral ischemia caused by cerebral vasospasm caused by aneurysmal subarachnoid hemorrhage?	2	3	3	3	3
25. For patients with aneurysmal subarachnoid hemorrhage whose intraoperative aneurysmal neck is relatively wide, it is hard to perform spring coil embolization. Should stent-assisted embolization be selected?	3	3	3	3	3
26. In a patient with a ruptured middle cerebral artery bifurcation aneurysm with parenchymal hematoma, surgical clipping or segmental interventional embolization of the aneurysm is preferred?	3	3	2	3	3
27. Should preoperative hypertension of patients with aneurysmal subarachnoid hemorrhage be raised to ensure cerebral perfusion?	0	3	2	3	3
28. Should patients with giant intracranial arteriovenous malformations be surgically removed quickly?	0	3	0	1	2
29. Should internal carotid endarterectomy be performed in asymptomatic patients with carotid artery stenosis exceeding 70%?	0	3	0	2	3
30. Should antithrombotic drugs be stopped immediately for patients with cerebral hemorrhage after taking antithrombotic drugs?	2	3	2	2	3
31. Can mechanical thrombectomy be performed on a 4-h pre-cranial circulation artery occlusion patient after thrombolytic therapy?	2	3	1	1	3
32. Is the combination of aspirin and clopidogrel recommended for Moyamoya disease patients?	1	2	1	2	3
33. Can Moyamoya disease patients with acute cerebrovascular event be treated with immediate surgery?	1	2	3	1	3
34. Which treatment option can significantly reduce the dose of anti-Parkinson's drugs, subthalamic deep brain stimulation or medial deep brain stimulation of the pallidum?	1	1	3	3	2
35. Is deep brain stimulation recommended for dystonia in patients with severe schizophrenia?	1	3	3	1	2
36. Should spinal spongiform vascular malformations be treated aggressively in patients with mild symptoms caused by repeated bleeding?	0	0	2	2	3
37. During microvascular decompression for patients with trigeminal neuralgia, arteries or veins are the most common vessels that compress the trigeminal nerve?	1	3	3	3	2

(Continued on next page)

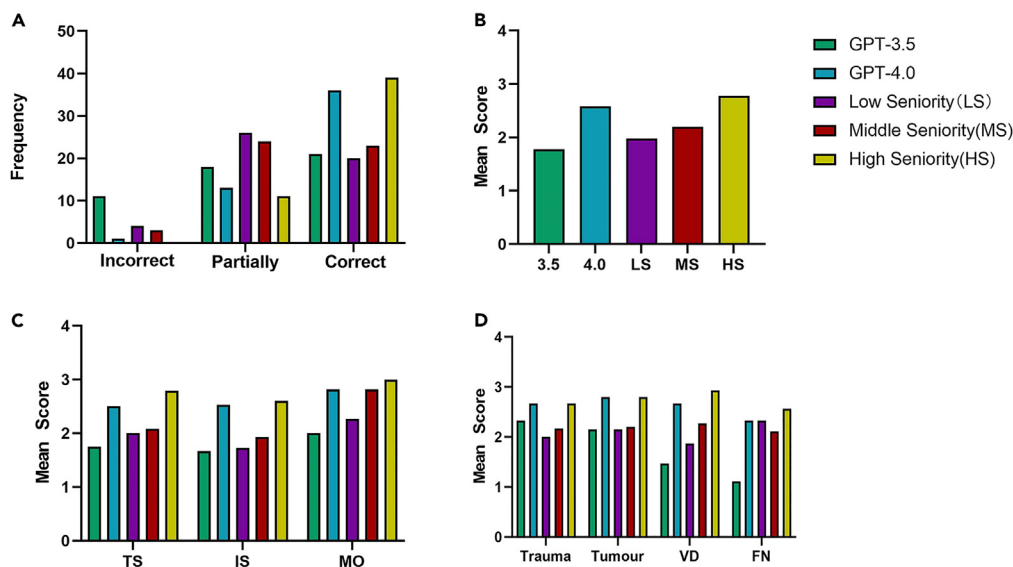
**Table 1. Continued**

Question	GPT-3.5	GPT-4.0	Low seniority	Middle seniority	High seniority
38. Can trigeminal nerve sensory roots be cut off for patients with trigeminal neuralgia without vascular compression during microvascular decompression?	0	1	1	0	2
39. When doctors suspect intracranial infection in patients with intracranial tumors after surgery, is it appropriate to wait for bacterial culture results before applying antibiotics to treat the patient?	3	3	3	3	3
40. Can electric field therapy be used as monotherapy in patients with recurrent glioblastoma?	0	1	0	1	2
41. What is the preferred treatment for patients with brain stem hemorrhage greater than 5 mL and Glasgow score $\leq 7$ ?	0	3	3	2	3
42. What is the most effective symptomatic medication for Parkinson's disease?	1	3	3	3	3
43. What is the first-line chemotherapy regimen for WHO Grade 3 oligodendroglioma with 1p/19q combined deletion?	2	3	1	0	2
44. What is the preferred therapeutic target for deep brain stimulation to control parkinsonism, associated cognitive or emotional problems, and psychiatric symptoms in patients with Parkinson's disease?	0	1	1	1	3
45. What is the preferred lipid-lowering drug for non-cardiogenic ischemic stroke patients with low-density lipoprotein $\geq 2.6$ mmol/L?	2	3	2	3	3
46. What is the first therapeutic agent to relieve acromegaly in patients with growth hormone adenoma?	1	2	3	3	3
47. What is the preferred treatment for patients with subarachnoid hemorrhage and acute hydrocephalus with increased intracranial pressure and impaired consciousness?	3	3	1	3	3
48. What is the preferred treatment drug for patients with trigeminal neuralgia?	3	3	3	3	3
49. What is the preferred treatment for focal epilepsy patients over 65 years old?	0	3	1	3	3
50. What are the treatment options for patients with primary systemic dystonia who do not improve after oral medication or botulinum toxin treatment?	3	3	3	2	3

compared to those of neurosurgeons; (3) the responses of ChatGPT and neurosurgeons according to items' specialty classification; and (4) the responses of ChatGPT and neurosurgeons according to items' question types. Because this was not a study of human subjects, neither receiving approval from the institutional review board nor obtaining informed consent was required.

### Study data

We performed this exploratory research on May 9, 2023, and generated a total of 50 questions on neurosurgery according to the guidelines for neurosurgery of central nervous system (CNS) of trauma, tumor, VD, and functional neurosurgery (FN), including National Comprehensive Cancer Network (NCCN) guidelines for CNS cancers, European Academy of Neurology, American Stroke Association, Movement Disorder Society, American Association of Neurological Surgeons, Congress of Neurological Surgeons, and Chinese Medical Association (Table 1). We generated the questions, such as answering yes or no, or asking questions about drugs, with simple answers that facilitated us to count the final answers. We posed each question three times to GPT-3.5 and GPT-4.0. We also recruited three neurosurgeons, each with high, middle, and low seniority, to write response questions, and the responses were recorded. In this study, we classified neurosurgeons according to the guidelines set by the National Health Commission of the People's Republic of China. The classification was based on their experience and qualifications: (1) low: attending physician (has gotten the doctor qualification and has worked in neurosurgery for more than three years); (2) middle: associate chief physician (has worked in neurosurgery as attending physician for more than three years); (3) high: chief physician (has worked in neurosurgery as an associate chief physician for more than three years).



**Figure 1. The comparison of ChatGPT and evidence-based neurosurgeons**

(A–D) Statistical results (A) the accuracy in different groups; (B) comparison of mean scores in different groups; (C) comparison of mean scores according to question type; (D) comparison of mean scores according to specialty classification.

Three senior neurosurgeons not involved in this study assessed the responses as “consistent” or “inconsistent” with solid recommendations based on the best available evidence in the above guidelines. Senior neurosurgeons are the highest level of physicians, with more experience and expertise than high-, middle-, and low-seniority neurosurgeons. They make a career of researching medical science and technology, saving lives, and treating diseases. In our study, we recruited three neurosurgeons, with high, middle, and low seniority, to respond to questions, and three senior neurosurgeons assessed the responses as “consistent” or “inconsistent.” “Inconsistent” means that the answer did not match the answer to the question determined by the three senior neurosurgeons based on guidelines and personal experience. The total score for each question is rated on a scale from 0 to 3 according to the number of “consistent” responses. Divergent opinions were resolved via discussion.

#### Data measurement

The response data of ChatGPT and neurosurgeons were compared. We divided questions into trauma, tumor, VD, and FN according to neurosurgery subspecialties. The mean scores were analyzed based on specialty classification and question type, which were further classified into treatment strategy (TS), intraoperative strategy (IS), and medicine option (MO). The mean scores were analyzed according to the questions’ specialty classification and question type.

#### Statistical analysis

SPSS statistical software 25.0 (IBM Corp., Armonk, New York, USA) was used for data analysis. Numerical variables are expressed as the mean  $\pm$  SD. Qualitative variables are described as the absolute value of cases in the distinctive group. Student’s t test was performed to evaluate the data with a normal distribution. Repeated measure analysis of variance was used for statistical assessment. Significant differences between groups were indicated when  $p < 0.05$ .

## RESULTS

### Baseline characteristics

GPT-3.5 had 42.00% (21/50) of responses in accordance with guidelines, with 22.00% (11/50) inconsistent, and 36.00% questions (11/50) for which one response or two responses were inappropriate. GPT-4.0 had 72.00% (36/50) of responses in accordance with guidelines, with 2.00% (1/50) inconsistent and 26.00% questions (13/50) for which one response or two responses were inappropriate.

Three neurosurgeons with low seniority had 40.00% (20/50) of responses in accordance with guidelines, with 8.00% (4/50) inconsistent, and 52.00% questions (26/50) for which one response or two responses were inappropriate. Three neurosurgeons with middle seniority had 46.00% (23/50) of responses in accordance with guidelines, with 1.20% (3/50) inconsistent, and 48.00% questions (24/50) for which one response or two responses were inappropriate. Three neurosurgeons with high seniority had 78.00% (39/50) of responses in accordance with guidelines, with no inconsistency, and 22.00% of questions (11/50) for which one response or two responses were inappropriate. (Figure 1A).

**Table 2. Comparison of mean scores according to question type**

Characteristic	GPT-3.5	GPT-4.0	low seniority			middle seniority			high seniority		
			Mean scores	p value (with GPT-3.5)	p value (with GPT-4.0)	Mean scores	p value (with GPT-3.5)	p value (with GPT-4.0)	Mean scores	p value (with GPT-3.5)	p value (with GPT-4.0)
treatment strategy	1.750 ± 1.260	2.500 ± 0.885	2.000 ± 1.022	0.377	0.056	2.080 ± 0.881	0.224	0.076	2.790 ± 0.415	0.001	0.110
intraoperative strategy	1.670 ± 1.291	2.530 ± 0.743	1.730 ± 1.100	0.806	0.013	1.930 ± 1.033	0.452	0.014	2.600 ± 0.507	0.005	0.719
medicine option	2.000 ± 1.095	2.820 ± 0.405	2.270 ± 0.786	0.432	0.052	2.820 ± 0.405	0.042	1.000	3.000 ± 0.000	0.013	0.167

### Comparison of mean scores

The mean score was  $1.780 \pm 1.217$  in GPT-3.5 and  $2.580 \pm 0.758$  in GPT-4.0 ( $p < 0.001$ ). The mean score was  $1.980 \pm 1.000$  in the low-seniority group,  $2.200 \pm 0.904$  in the middle-seniority group, and  $2.780 \pm 0.418$  in the high-seniority group. The mean score significantly differed between the group of GPT-3.5 and middle seniority ( $p = 0.024$ ) and GPT-3.5 and high seniority ( $p < 0.001$ ). The mean score also significantly differed between the group of GPT-4.0 and low seniority ( $p < 0.001$ ) and GPT-4.0 and middle seniority ( $p = 0.005$ ). We recorded no between-group differences in the mean score in group GPT-3.5 and low seniority ( $p = 0.242$ ) and GPT-4.0 and high seniority ( $p = 0.058$ ). (Figure 1B).

### Comparison of mean scores according to the question type

In TS, the mean score was  $1.750 \pm 1.260$  in GPT-3.5 and  $2.500 \pm 0.885$  in GPT-4.0 ( $p = 0.004$ ). The mean score was  $2.000 \pm 1.022$  in the low-seniority group,  $2.080 \pm 0.881$  in the middle-seniority group, and  $2.790 \pm 0.415$  in the high-seniority group. The mean score of GPT-3.5 in TS was significantly lower than that of the high-seniority group ( $p = 0.001$ ). We recorded no between-group differences in other groups.

In IS, the mean score was  $1.670 \pm 1.291$  in GPT-3.5 and  $2.530 \pm 0.743$  in GPT-4.0 ( $p = 0.007$ ). The mean score was  $1.730 \pm 1.100$  in the low-seniority group,  $1.930 \pm 1.033$  in the middle-seniority group, and  $2.600 \pm 0.507$  in the high-seniority group. The mean score of GPT-3.5 was significantly lower than that of the high-seniority group ( $p = 0.005$ ). The mean score of GPT-4.0 was significantly higher than that of the low- and middle-seniority group ( $p = 0.013$ ,  $p = 0.014$ , respectively). We recorded no between-group differences in other groups.

In MO, the mean score was  $2.000 \pm 1.095$  in GPT-3.5 and  $2.820 \pm 0.405$  in GPT-4.0 ( $p = 0.020$ ). The mean score was  $2.270 \pm 0.786$  in the low-seniority group,  $2.820 \pm 0.405$  in the middle-seniority group, and  $3.000 \pm 0.000$  in the high-seniority group. The mean score of GPT-3.5 was significantly lower than that of the middle- and high-seniority groups ( $p = 0.042$ ,  $p = 0.013$ , respectively). We recorded no between-group differences in other groups. (Table 2 and Figure 1C).

### Comparison of mean scores according to specialty classification

In trauma, the mean score was  $2.330 \pm 1.211$  in GPT-3.5 and  $2.670 \pm 0.516$  in GPT-4.0 ( $p = 0.465$ ). The mean score was  $2.000 \pm 0.894$  in the low-seniority group,  $2.170 \pm 0.753$  in the middle-seniority group, and  $2.670 \pm 0.516$  in the high-seniority group. The mean score of GPT-4.0 was significantly higher than that of the low-seniority group ( $p = 0.025$ ). We recorded no between-group differences in other groups.

In tumor, the mean score was  $2.150 \pm 1.089$  in GPT-3.5 and  $2.600 \pm 0.681$  in GPT-4.0 ( $p = 0.025$ ). The mean score was  $1.900 \pm 1.021$  in the low-seniority group,  $2.200 \pm 0.951$  in the middle-seniority group, and  $2.800 \pm 0.417$  in the high-seniority group. The mean score of GPT-3.5 was significantly lower than that of the high-seniority group ( $p = 0.012$ ). The mean score of GPT-4.0 was significantly higher than that of the low-seniority group ( $p = 0.005$ ). We recorded no between-group differences in other groups.

In VD, the mean score was  $1.470 \pm 1.246$  in GPT-3.5 and  $2.670 \pm 0.816$  in GPT-4.0 ( $p = 0.002$ ). The mean score was  $1.870 \pm 1.060$  in the low-seniority group,  $2.270 \pm 0.799$  in the middle-seniority group, and  $2.930 \pm 0.258$  in the high-seniority group. The mean score of GPT-3.5 was significantly lower than that of the middle- and high-seniority groups ( $p = 0.013$ ,  $p < 0.001$ , respectively). The mean score of GPT-4.0 was significantly higher than that of the low-seniority group ( $p = 0.041$ ). We recorded no between-group differences in other groups.

In FN, the mean score was  $1.110 \pm 1.167$  in GPT-3.5 and  $2.330 \pm 1.000$  in GPT-4.0 ( $p = 0.010$ ). The mean score was  $2.330 \pm 1.000$  in the low-seniority group,  $2.110 \pm 1.167$  in the middle-seniority group, and  $2.560 \pm 0.527$  in the high-seniority group. The mean score of GPT-3.5 was significantly lower than that of the low- and high-seniority groups ( $p = 0.002$ ,  $p = 0.005$ , respectively). We recorded no between-group differences in other groups. (Table 3 and Figure 1D).

## DISCUSSION

With the development of ChatGPT, many researchers have evaluated its application value in the medical field. Sallam<sup>22</sup> indicated that ChatGPT was an efficient and promising tool for scientific research, healthcare practice, and education. ChatGPT can conduct comprehensive

**Table 3. Comparison of mean scores according to specialty classification**

Characteristic	GPT-3.5	GPT-4.0	low seniority			middle seniority			high seniority		
			Mean scores	p value (with GPT-3.5)	p value (with GPT-4.0)	Mean scores	p value (with GPT-3.5)	p value (with GPT-4.0)	Mean scores	p value (with GPT-3.5)	p value (with GPT-4.0)
trauma	2.330 ± 1.211	2.670 ± 0.516	2.000 ± 0.894	0.530	0.025	2.170 ± 0.753	0.793	0.203	2.670 ± 0.516	0.465	–
Tumor	2.150 ± 1.089	2.600 ± 0.681	1.900 ± 1.021	0.234	0.005	2.200 ± 0.951	0.858	0.057	2.800 ± 0.417	0.012	0.163
vascular disease	1.470 ± 1.246	2.670 ± 0.816	1.870 ± 1.060	0.271	0.041	2.270 ± 0.799	0.013	0.138	2.930 ± 0.258	<0.001	0.262
functional neurosurgery	1.110 ± 1.167	2.330 ± 1.000	2.330 ± 1.000	0.002	1.000	2.110 ± 1.167	0.053	0.559	2.560 ± 0.527	0.005	0.512

literature reviews and involve improved language and the ability to express and communicate research ideas and results.<sup>7</sup> Regarding medical education, ChatGPT can provide professional academic guidance in the qualification examination for physicians,<sup>20</sup> medical student education,<sup>23</sup> and communication skills.<sup>4</sup> In addition, ChatGPT can help streamline the clinical workflow with possible cost savings and increased efficiency.<sup>24,25</sup>

The application of ChatGPT in neurosurgery mainly includes medical and surgical education, administrative assistance, and clinical decision-making. Sevgi et al.<sup>26</sup> found that although using ChatGPT for neurosurgery training might revolutionize learning for medical students, residents, and neurosurgeons, it should not be considered a dependable source of information. Its reliability should be further evaluated and improved. Besides, Singh et al.<sup>27</sup> presented the future applications and drawbacks of ChatGPT integration in medical and surgical education, clinical decision-making, and administrative assistance. They indicated that ChatGPT might be a helpful tool for neurosurgical medical education, planning and decision-making, and the productivity and effectiveness of clinicians soon.

Some researchers have found the capability of ChatGPT in CDSS. ChatGPT can provide general basic answers regarding medical knowledge, lifestyle, treatment, and radiology.<sup>15,28</sup> Haemmerli et al.<sup>29</sup> found ChatGPT performed well for adjuvant treatment recommendations and may become a promising assisted tool for physicians. Based on this, ChatGPT could also be considered in helping create policies related to concussion and repetitive brain trauma associated with neurodegenerative disease risk. To further explore the application of ChatGPT in neurosurgery, we first analyzed ChatGPT's ability to be comparable to the evidence-based neurosurgeon's ability.

### Is ChatGPT's ability comparable to that of evidence-based neurosurgeons?

We conducted tests on GPT-3.5 and GPT-4 to evaluate their abilities to respond in accordance with evidence-based medicine. In this study, although GPT-3.5 had poor performance to some extent, GPT-4.0 had a correct response rate of 72.00%, meaning that GPT-4.0 could answer consistently with recommendations from evidence-based medicine in most cases. This finding underscores that most medical responses provided by GPT-4.0 were based on the latest medical evidence and were consistent with experts' opinions and guidelines, even without explicit training on a medical database. However, GPT-3.5 and GPT-4.0 produced some responses needed to meet evidence-based medicine. For example, ChatGPT provided an answer that contradicts guidelines regarding whether electric field therapy can be used as monotherapy in patients with recurrent glioblastoma<sup>30</sup> (Figure 2). ChatGPT's advice is based on the latest scientific evidence and adheres to expert guidelines and treatment protocols, but its dataset dates from 2021 and is not updated in real time.<sup>31</sup> Life science and medicine develop rapidly, and new knowledge is produced constantly. Even human medical experts should keep learning new things and continually update their knowledge base. The neurosurgical field involves highly complex, specialized, and in-depth content. The GPT-4.0 responses are highly consistent with evidence-based medicine, despite not being trained explicitly on medical databases. ChatGPT is as good as the dataset it is trained on, which is the same as all AI programs.<sup>32</sup> We believe that if ChatGPT is trained on medical databases, the professional response standards might be further improved, revolutionizing modern clinical medicine.<sup>33,34</sup>

In this study, we compared the responses of ChatGPT to those of neurosurgeons. Statistical results show that GPT-3.5's ability to respond in accordance with evidence-based medicine was comparable to that of neurosurgeons with low seniority, and GPT-4.0's ability was comparable to that of neurosurgeons with high seniority. It seems that ChatGPT, and specifically GPT-4.0, responded better than most neurosurgeons. We believe that as ChatGPT continues to upgrade, it will add a new dimension to healthcare.<sup>35</sup> With the exponential growth of the amount of information accumulated by AI, it provides better assistance for doctors' clinical diagnosis and treatment at the individual level, such as optimal treatment strategies, IS, targeted therapy, and personalized treatment. However, it should be noted that biomedical research and clinical knowledge results cannot be equated. There is a complex transformation relationship between the two, and this transformation process has difficulty reflecting the open data sources. In the face of highly specialized and complex medical knowledge, ChatGPT operates based on statistics and association. It is good at simple thinking activities such as word processing and common dialogue. However, it is difficult to form objective knowledge reflecting human physiological or pathological activities, which often requires a rigorous logical reasoning





Can electric field therapy be used as monotherapy in patients with recurrent glioblastoma?



Electric field therapy, also known as tumor-treating fields (TTF) therapy, is a treatment modality that involves the use of low-intensity alternating electric fields to disrupt the division of cancer cells. It has been approved for the treatment of recurrent glioblastoma in combination with standard chemotherapy.

As of my knowledge cutoff in September 2021, electric field therapy is not typically used as monotherapy for recurrent glioblastoma. The standard treatment for recurrent glioblastoma usually involves a combination of surgery, radiation therapy, and chemotherapy. Electric field therapy, specifically Optune (NovoTTF-100A) system, is often added to this regimen.

The FDA-approved indication for Optune is in combination with temozolomide chemotherapy for the treatment of adult patients with newly diagnosed glioblastoma (GBM) and as monotherapy for recurrent GBM after surgical and radiation options have been exhausted. The evidence supporting the use of electric field therapy as monotherapy for recurrent glioblastoma is limited.

It's important to consult with a neuro-oncologist or a medical professional who specializes in the treatment of glioblastoma for the most up-to-date information on treatment options and recommendations. They can provide guidance based on the latest research and individual patient factors to determine the most appropriate treatment plan.

**Figure 2. Screenshot of ChatGPT's answer to a question**

process, and sometimes personal experience and even intuition. Therefore, clinicians do not need to worry about this.<sup>32</sup> The profession of doctors continues to show its value because of the complexity of various factors and the variability of individual differences in patient diagnosis and treatment in the office and the operating room.<sup>32</sup> At this time, only human doctors can play the role of comprehensive judgment and clinical treatment.

### Comparison of mean scores according to classification

Performance comparison between GPT-3.5 and GPT-4.0 across neurosurgery subspecialties and question types revealed that GPT-4.0 generally outperformed GPT-3.5, despite no significant difference in the mean scores for most items. According to question type, we found that the deficiencies of GPT-3.5 were improved in GPT-4.0, both in treatment or IS and MOs. In particular, GPT-4.0 performed better than neurosurgeons with low and middle seniority in IS.<sup>36,37</sup> This result indicates that ChatGPT can better assist the clinical medical practice in the development of ChatGPT, which can better integrate the latest achievements of contemporary science and technology to improve the level of human medical treatment further.

In addition, the initial version of GPT was trained on a 40 GB text dataset using a modified transformer architecture and had a model size of 1.5 billion parameters. Later on, OpenAI introduced GPT-3 in 2020. To enhance the output of GPT-3.0, OpenAI introduced a new model called InstructGPT, also known as GPT-3.5, particularly for conversational models. It is based on GPT-3.0 but improves its language capabilities and ability to follow provided instructions. The models of GPT-3.5 introduce a significant element: human training. The GPT-4.0 dialogue model is a fine-tuned version of GPT-3.5. However, it works within even stricter guardrails, an early prototype of AI alignment with human values by forcing it to comply with many rules. This enhancement enables the model to understand the context better and distinguish nuances, resulting in more accurate and coherent responses. Furthermore, GPT-4.0 has a maximum token limit of 25,000 words, significantly increasing from GPT-3.5's 3,125 words (<https://openai.com>). Our study also suggests that ChatGPT's knowledge in specific fields (e.g., MO) remains insufficient, but these deficiencies can be optimized by upgrading ChatGPT.

According to question type, we also found similar results in which the performance of GPT-4.0 is better than neurosurgeons with low and middle seniority. Although ChatGPT has been used in the medical field, the results of this study show that ChatGPT can provide detailed, specific, and personalized responses to medical problems in various subspecialties of neurosurgery, demonstrating its potential for medical diagnosis and treatment.<sup>38,39</sup>

### Ethical consideration

Firstly, many studies have found that AI-assisted CDSS can optimize procedures and improve clinical practice efficiency; most are theoretical and retrospective studies and lack high-quality clinical trials.<sup>24,25,40</sup> Secondly, the clinical application of AI requires regulation. Researchers must perform clinical studies according to relevant regulations. Only products approved by medical regulatory authorities can be used in

clinical practice in compliance.<sup>33,41</sup> Thirdly, using conversational AI for specialized research will likely introduce inaccuracies, bias, and plagiarism.<sup>23</sup> Physicians who use ChatGPT risk being misled by false or biased information. As an AI, ChatGPT lacks autonomous consciousness, does not take a position on its output, and does not bear any legal or ethical responsibility. Therefore, humans should always remain accountable for scientific practice. In the future clinical application of ChatGPT, it is crucial to fully consider relevant ethical issues to ensure that AI products meet the corresponding ethical requirements in clinical practice.

In addition, private entities own and control many AI technologies, raising privacy and data security issues. The nature of the use of AI could mean institutions such as corporations, hospitals, and the public might have convenience in obtaining and utilizing patient health information. Regarding the “black box,” AI sometimes cannot easily be regulated by human medical experts.<sup>42,43</sup> The learning algorithms’ methods and “reasoning” used for reaching their conclusions can be partially or entirely opaque to human observers.<sup>44</sup> This opacity may also threaten the security of personal health information.

Although ChatGPT shares some similarities with the human mind regarding knowledge generation, the differences between the two are striking. ChatGPT is a machine learning model algorithm, and the essence of its output language is still derived from the understanding of semantic rules. At the same time, human beings can make differentiated expressions according to different situations and scenes, and human beings have a broader process of individual cognitive modeling. The ability to have a more profound sense of the social environment functions that machines do not have for the time being also makes human intelligence full of possibilities. We believe that no matter how ChatGPT develops in the future, it will not be able to replace the role of doctors in clinical diagnosis and treatment but will be able to assist clinical practice better.<sup>45</sup> Neurosurgeons should exercise caution when using ChatGPT and consider it a supplement to their clinical knowledge and experience.<sup>28</sup>

In summary, GPT-3.5’s ability to respond in accordance with evidence-based medicine was comparable to that of neurosurgeons with low seniority, and GPT-4.0’s ability was comparable to that of neurosurgeons with high seniority. Although ChatGPT is not yet comparable to neurosurgeons with high seniority, future upgrades could enhance its performance and abilities. Neurosurgeons and neuroscientists should be aware of the progress of this AI chatbot and contemplate its potential integration into clinical practice.

### Limitations of the study

This study has several limitations. First, the number of questions needed to be increased. The questions created have not been divided into questions that directly question neurosurgical information and require the ability to interpret based on this information. Second, the sample size in this study was insufficient, and the group of neurosurgeons recruited for this study was limited. We focus on the high, middle, and low seniority of neurosurgeons, and the grouping still needs to be improved. In the future, further inclusion in the research subjects can be based on specific doctor titles and levels to obtain more robust results. Third, this study did not include neurosurgeons from hospitals of different levels and lacked comparative data between ChatGPT and neurosurgeons in different hospitals. It still needs to be further improved in the future.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

### ACKNOWLEDGMENTS

Not applicable.

### AUTHOR CONTRIBUTIONS

Jiayu Liu: literature search, study design, data collection, data interpretation, and writing;

Jiqi Zheng: All use of ChatGPT in this study was done solely by Jiqi Zheng;

Jiqi Zheng and Xintian Cai: provided feedback on manuscript texts;

Chengliang Yin and Dongdong Wu: study design and provided feedback on all manuscript texts.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: May 31, 2023

Revised: June 21, 2023

Accepted: August 4, 2023

Published: August 9, 2023

## REFERENCES

- Howard, A., Hope, W., and Gerada, A. (2023). ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect. Dis.* 23, 405–406. [https://doi.org/10.1016/S1473-3099\(23\)00113-5](https://doi.org/10.1016/S1473-3099(23)00113-5).
- Mann, D.L. (2023). Artificial Intelligence Discusses the Role of Artificial Intelligence in Translational Medicine: A JACC: Basic to Translational Science Interview With ChatGPT. *JACC. Basic Transl. Sci.* 8, 221–223. <https://doi.org/10.1016/j.jacbts.2023.01.001>.
- Graber-Stiehl, I. (2023). Is the world ready for ChatGPT therapists? *Nature* 617, 22–24. <https://doi.org/10.1038/d41586-023-01473-4>.
- Khan, R.A., Jawaid, M., Khan, A.R., and Sajjad, M. (2023). ChatGPT-Reshaping medical education and clinical management. *Pakistan J. Med. Sci.* 39, 605–607.
- Galido, P.V., Butala, S., Chakerian, M., and Agustines, D. (2023). A Case Study Demonstrating Applications of ChatGPT in the Clinical Management of Treatment-Resistant Schizophrenia. *Cureus* 15, e38166.
- Patel, S.B., and Lam, K. (2023). ChatGPT: the future of discharge summaries? *Lancet Digit. Health* 5, e107–e108.
- Chen, T.-J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *J. Chin. Med. Assoc.* 86, 351–353.
- Kao, H.J., Chien, T.W., Wang, W.C., Chou, W., and Chow, J.C. (2023). Assessing ChatGPT's capacity for clinical decision support in pediatrics: A comparative study with pediatricians using KIDMAP of Rasch analysis. *Medicine (Baltim.)* 102, e34068. <https://doi.org/10.1097/MD.00000000000034068>.
- Johnson, D., Goodman, R., Patrinely, J., Stone, C., Zimmerman, E., Donald, R., Chang, S., Berkowitz, S., Finn, A., and Jahangir, E. (2023). Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the Chat-GPT model.
- Kawamoto, K., Houlihan, C.A., Balas, E.A., and Lobach, D.F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 330, 765.
- Ash, J.S., McCormack, J.L., Sittig, D.F., Wright, A., McMullen, C., and Bates, D.W. (2012). Standard practices for computerized clinical decision support in community hospitals: a national survey. *J. Am. Med. Inf. Assoc.* 19, 980–987. <https://doi.org/10.1136/amiajnl-2011-000705>.
- Liu, S., Wright, A.P., Patterson, B.L., Wanderer, J.P., Turer, R.W., Nelson, S.D., McCoy, A.B., Sittig, D.F., and Wright, A. (2023). Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *J. Am. Med. Inf. Assoc.* 30, 1237–1245.
- Howard, A., Hope, W., and Gerada, A. (2023). ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect. Dis.* 23, 405–406.
- Bradshaw, J.C. (2023). The ChatGPT Era: Artificial Intelligence in Emergency Medicine. *Ann. Emerg. Med.* 81, 764–765. <https://doi.org/10.1016/j.annemergmed.2023.01.022>.
- Lukac, S., Dayan, D., Fink, V., Leinert, E., Hartkopf, A., Veselinovic, K., Janni, W., Rack, B., Pfister, K., and Heitmeir, B. (2023). Evaluating ChatGPT as an Adjunct for the Multidisciplinary Tumor Board Decision-Making in Primary Breast Cancer Cases.
- Mihalache, A., Huang, R.S., Popovic, M.M., and Muni, R.H. (2023). Performance of an Upgraded Artificial Intelligence Chatbot for Ophthalmic Knowledge Assessment. *JAMA Ophthalmol.* e232754. <https://doi.org/10.1001/jamaophthalmol.2023.2754>.
- Praveen, S.V., and Lohia, R. (2023). What do psychiatry researchers feel about ChatGPT? A study based on Natural Language Processing techniques. *Asian J. Psychiatr.* 85, 103626. <https://doi.org/10.1016/j.ajp.2023.103626>.
- Warnke, P. (2020). Neurosurgery, neurological surgery or surgical neurology: semantics or real impact? *J. Neurol. Neurosurg. Psychiatry* 91, 1024–1026. <https://doi.org/10.1136/jnnp-2020-324008>.
- Usachev, D., Konovalov, A.N., Likhтерman, L.B., Konovalov, N.A., and Matuev, K.B. (2022). [Neurosurgeon training: current problems and modern approaches]. *Vopr. Neurokhir.* 86, 5–16. <https://doi.org/10.17116/neiro2022860115>.
- Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., and Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit. Health* 2, e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.
- Gordijn, B., and Have, H.T. (2023). ChatGPT: evolution or revolution? *Med. Health Care Philos.* 26, 1–2. <https://doi.org/10.1007/s11019-023-10136-0>.
- Patel, S.B., Lam, K., and Liebreiz, M. (2023). ChatGPT: friend or foe. *Lancet Digit. Health* 5, e102.
- van Dis, E.A.M., Bollen, J., Zuidema, W., van Rooij, R., and Bockting, C.L. (2023). ChatGPT: five priorities for research. *Nature* 614, 224–226.
- Jeblick, K., Schachtner, B., Dextl, J., Mittermeier, A., Stüber, A.T., Topalis, J., Weber, T., Wesp, P., Sabel, B., and Ricke, J. (2022). ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2212.14882>.
- Gunawan, J. (2023). Exploring the future of nursing: Insights from the ChatGPT model. *Belitung Nurs. J.* 9, 1–5.
- Sevgi, U.T., Erol, G., Doğruel, Y., Sönmez, O.F., Tubbs, R.S., and Güngör, A. (2023). The role of an open artificial intelligence platform in modern neurosurgical education: a preliminary study. *Neurosurg. Rev.* 46, 86.
- Singh, R., Reardon, T., Srinivasan, V.M., Gottfried, O., Bydon, M., and Lawton, M.T. (2023). Implications and future directions of ChatGPT utilization in neurosurgery. *J. Neurosurg.* 1, 1–3.
- He, Y., Tang, H., Wang, D., Gu, S., Ni, G., and Wu, H. (2023). Will ChatGPT/GPT-4 be a Lighthouse to Guide Spinal Surgeons? *Ann. Biomed. Eng.* 51, 1362–1365. <https://doi.org/10.1007/s10439-023-03206-0>.
- Haemmerli, J., Sveikata, L., Nouri, A., May, A., Egervari, K., Freyschlag, C., Lobrinus, J.A., Migliorini, D., Momjian, S., and Sanda, N. (2023). ChatGPT in glioma patient adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board? Preprint at medRxiv. <https://doi.org/10.1101/2023.03.19.23287452>.
- Kim, Y.Z., Kim, C.Y., and Lim, D.H. (2022). The Overview of Practical Guidelines for Gliomas by KSNO, NCCN, and EANO. *Brain Tumor Res. Treat.* 10, 83–93. <https://doi.org/10.14791/btrt.2022.0001>.
- Zhou, Z., Wang, X., Li, X., and Liao, L. (2023). Is ChatGPT an Evidence-based Doctor? *Eur. Urol.* 84, 355–356. <https://doi.org/10.1016/j.eururo.2023.03.037>.
- Bernstein, J. (2023). Not the Last Word: ChatGPT Can't Perform Orthopaedic Surgery. *Clin. Orthop. Relat. Res.* 481, 651–655. <https://doi.org/10.1097/CORR.0000000000002619>.
- Kleebayoon, A., and Wiwanitkit, V. (2023). ChatGPT/GPT-4 and Spinal Surgeons. *Ann. Biomed. Eng.* 51, 1657. <https://doi.org/10.1007/s10439-023-03223-z>.
- Javan, R., Kim, T., Mostaghni, N., and Sarin, S. (2023). ChatGPT's Potential Role in Interventional Radiology. *Cardiovasc. Intervent. Radiol.* 46, 821–822. <https://doi.org/10.1007/s00270-023-03448-4>.

35. Parsa, A., and Ebrahimpzadeh, M.H. (2023). ChatGPT in Medicine; a Disruptive Innovation or Just One Step Forward? *Arch. Bone Jt. Surg.* 11, 225–226. <https://doi.org/10.22038/abjs.2023.22042>.
36. Weidman, A.A., Valentine, L., Chung, K.C., and Lin, S.J. (2023). OpenAI's ChatGPT and Its Role in Plastic Surgery Research. *Plast. Reconstr. Surg.* 151, 1111–1113. <https://doi.org/10.1097/PRS.00000000000010342>.
37. Oh, N., Choi, G.S., and Lee, W.Y. (2023). ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Ann. Surg. Treat. Res.* 104, 269–273. <https://doi.org/10.4174/astr.2023.104.5.269>.
38. Chen, W., Liu, X., Zhang, S., and Chen, S. (2023). Artificial intelligence for drug discovery: Resources, methods, and applications. *Mol. Ther. Nucleic Acids* 31, 691–702. <https://doi.org/10.1016/j.omtn.2023.02.019>.
39. Scerri, A., and Morin, K.H. (2023). Using chatbots like ChatGPT to support nursing practice. *J. Clin. Nurs.* 32, 4211–4213. <https://doi.org/10.1111/jocn.16677>.
40. Varghese, J. (2020). Artificial intelligence in medicine: chances and challenges for wide clinical adoption. *Visc. Med.* 36, 443–449.
41. Shumway, D.O., Kriege, K., and Wood, S.T. (2023). Discordance of the Urinary and Pleural Fluid Antigen Test and False Positive for *Streptococcus pneumoniae* in Empyema Secondary to Necrotizing Bacterial Pneumonia. *Cureus* 15, e37458. <https://doi.org/10.7759/cureus.37458>.
42. Quinn, T.P., Jacobs, S., Senadeera, M., Le, V., and Coghlan, S. (2022). The three ghosts of medical AI: Can the black-box present deliver? *Artif. Intell. Med.* 124, 102158.
43. Durán, J.M., and Jongsma, K.R. (2021). Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *J. Med. Ethics* 47, 329–335.
44. Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med. Ethics* 22, 122–125.
45. Uprety, D., Zhu, D., and West, H.J. (2023). ChatGPT-A promising generative AI tool and its implications for cancer care. *Cancer* 129, 2284–2289. <https://doi.org/10.1002/cnrc.34827>.

## STAR★METHODS

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chengliang Yin ([chengliangyin@163.com](mailto:chengliangyin@163.com)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- All data reported in this paper will be shared by [lead contact](#) upon reasonable request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Not applicable.

### METHOD DETAILS

We performed this exploratory research on May 9, 2023, and generated a total of 50 questions on neurosurgery according to the guidelines for neurosurgery of central nervous system (CNS) of trauma, tumour, vascular disease (VD) and functional neurosurgery (FN), including National Comprehensive Cancer Network (NCCN) guidelines for CNS cancers, European Academy of Neurology, American Stroke Association, Movement Disorder Society, American Association of Neurological Surgeons, Congress of Neurological Surgeons. (Table 1) We generated the questions, such as answering yes or no, or asking questions about drugs, with simple answers that facilitated us to count the final answers. We posed each question three times to GPT-3.5 and GPT-4.0. We also recruited three neurosurgeons, each with high, middle, and low seniority, to write response questions, and the responses were recorded. Three senior neurosurgeons not involved in this study assessed the responses as "consistent" or "inconsistent" with solid recommendations based on the best available evidence in the above guidelines. In our study, we recruited three neurosurgeons, each with high, middle, and low seniority, to respond to questions, and three senior neurosurgeons assessed the responses as "consistent" or "inconsistent." "Inconsistent" means that the answer did not match the answer to the question determined by three senior neurosurgeons based on guidelines and personal experience. The total score for each question is rated on a scale from 0 to 3 according to the number of "consistent" responses. Divergent opinions were resolved via discussion.

### QUANTIFICATION AND STATISTICAL ANALYSIS

SPSS statistical software 25.0 (IBM Corp., Armonk, New York, USA) was used for data analysis. Numerical variables are expressed as the mean  $\pm$  SD. Qualitative variables are described as the absolute value of cases in the distinctive group. Student's t-test was performed to evaluate the data with a normal distribution. Repeated measure analysis of variance was used for statistical assessment. Significant differences between groups were indicated when  $p < 0.05$ .

### ADDITIONAL RESOURCES

Not applicable.