Assessing the Potential Role of ChatGPT in Spine Surgery Research

Ari Ruben Berg¹, Dhruv Mendiratta, Isabel Herzog¹, Ashok Para, Neil Kaushal, Michael Vives ¹Rutgers New Jersey Medical School

INTRODUCTION: In less than two months, Chat Generative Pre-trained Transformer (ChatGPT), a complex machine learning model, has garnered more than 100 million users worldwide. ChatGPT's recent role as an author on scientific literature has academicians calling its validity into question. While it is generally able to maintain factual basis in its writing, the scholarly writing capabilities still have ways to go. Nevertheless, ChatGPT's ability to identify gaps of knowledge from the existing literature may allow it to function in novel research generation. Systematic reviews have gained popularity in surgical literature due to their ability to consolidate available data into one comprehensive analytical study. Therefore, ChatGPT may have a niche in identifying areas that need to be addressed via a systematic review. The aim of this study is to assess ChatGPT's ability to generate novel systematic ideas on topics within spine surgery.

METHODS: ChatGPT was instructed to give ten novel systematic review ideas for five popular topics in spine surgery literature: microdiscectomy, laminectomy, spinal fusion, kyphoplasty, and disc replacement. Therefore, a total of 50 systematic review ideas were generated. ChatGPT's ability to generate novel systematic review ideas was evaluated by conducting a literature search in PubMed, CINAHL, EMBASE, and Cochrane. The number of non-systematic review articles and number of systematic review papers that had been published on each ChatGPT-generated idea were recorded.

RESULTS:

Overall, ChatGPT had a 66% accuracy rate in creating novel systematic review ideas. These 50 research ideas and their associated novelty assessments are shown in **Table 1**. More specifically, the accuracy rates were 80%, 80%, 30%, 70%, and 70% for microdiscectomy, laminectomy, spinal fusion, kyphoplasty, and disc replacement, respectively. However, there was a 32% rate of ChatGPT generating ideas for which there were 0 non-systematic review articles published. There was a 71.4%, 50%, 22.2%, 50%, 62.5%, and 51.2% success rate of generating novel systematic review ideas, for which there were also non-systematic reviews published, for microdiscectomy, laminectomy, spinal fusion, kyphoplasty, and disc replacement, and overall, respectively.

DISCUSSION AND CONCLUSION: Despite being limited in information past the end of 2021, ChatGPT still successfully generated novel ideas at an overall rate of 66%. As this technology improves through reinforcement learning and database expansion, there is a role for ChatGPT as a research tool. Its utility in generating new ideas can save spine surgeons time in having to parse through hundreds to thousands of articles by themselves. With more specific instructions and user-driven modulation, the accuracy of ChatGPT can be increased; future studies should investigate this.

Moralizationy				875
Comparison of microdinaceionsy automothetimous addedy and youngar putients.	0		Y	
A systematic series of the incidence of received hemistion after microdiscustersy.			Y .	
The impact of patient factors such as obesity and emoking on microdiscochemy eacorney.	7		Y .	
A review of the effectiveness of different neglical approaches for hasher microdecochemy.	7	3	N	
The impact of prosperative imaging stadies on the maxim of microdiscubury.			Y	
A systematic review of the success rates and complication rates of minimally invaries microdiscustomy compared to	1		Y.	
anana ina mananany.			v	
A maintain review of the cost officiences of niceducetowy commend to alternative trainants for homized				
tanbar diras.	,			
The role of physical therapy and reliabilitation in importing extremes after microdecochemy.	*		N	
A review of patient-reported autosmos after microdiscubency, including quality of life and whate to work.			Y.	
Latinetwy				875
Comparison of subcomes between minimally invarive and traditional spea laminectomy for spinal stanosis.		3	N	
Analysis of the effort of print age and consorbidities on mecomes following laminocomy for spind recovir.			Y	
Assument of the long-term seconter of laminochemy for degenerative disc disease.	0		Y	
Boview of the use of nerve monitoring todatiques during laminocomy to minimize the tick of nerve injury.			Y	
Analysis of the toke of purepensive physical therapy in optimizing recovery following laminocurrent.	2		N	
Comparison of lamino tump with other surgical options for the treatment of Domianol discs.	10	4	Y	
Assument of the innext of different acceled annuality on nation satisfaction after lanine-term			Y	
Analysis of the risk of respection following laminocharry for spinal atmosis.			Y	
Comparison of automas harmon laninostany and laninostany for the tradment of prival density.	3		Y	
framination of the impact of adjacent treatments such as quident injustions and nows blocks an automou following				
lanihertung				
Spinal Penins				385
The long term automos of cernical spinal facion for the incoment of degenerative disc discuse	>190	3	N	
The impact of smalling on spinal flation outcomes	>198	2	N	
The comparative efficacy of minimally invasive vs. traditional optical factors for handwar degenerative conditions.	3		Y	
The safety and effectiveness of allegenth to autografi for spinal factors			N	
The rule of biologies in promoting spinal lasion macross	>198	4	N	
The impact of patient characteristics (ope, obesity, consorbidities) on spind faston extremes	>198		N	
The effectiveness of spind fasion in the treatment of spind tumors	0		Y	
The impact of revision mayary on the recome of spind fusion	>199		Y	
The cost effectiveness of spind fusion compared to non-magical transmoms for choose back pain.	19		N	
The effectiveness of different going factors turbuleness for the treatment of lamber spinal response	>100		N	
Exhedrary				3%
Comparison of pain and functional concornes after kephophaty on renderspharts.	15	3	N	
Embodiary oncome is network with amounter comboil for turns three with transmic for turns			Y	
The effective of the second			N N	
conservation of the second s			× ×	
neg war neg ne rannekilder er kylegeny wenne pieten.			Y Y	
ran militar su hannar commonanter su eliferation à annuary.				
annengen antennen er rennen soge mennen an oppere age somsten atte spedenet.				
ranne we another recourts or refler to anno new stillings.				
Adjunctive therapies for improvement of Siggloplasty entromes, such as physical therapy or pharma-scheropy		-	Y II	
Kyphoplasty for sortubral compression fluoranes in patients with spind tumors.	-198		N	
Xyphophaty in the management of vertebral fractions in patients with Askyloning Spondyläti.		•	v	
Die Keplarenent				~
Classed outcomes of total data replacement in the carvical spine	>199	18	N	
A comparison of lotal data replacement with listion surgery in the lambar spine			Y	
Long-turns datability of hotal disc replacement in the humber and serviced spins	0		Y .	
Economic evaluation of total disc replacement compared to fusion surgary	9		Y	
Patient satisfaction following total disc explanation surgery	0		Y .	
Revision and superation rates following total disc replacement supery	22		Y	
A symmetic review of the radiological outcomes of total disc replacement surgery	>199		N	
The impact of age, gender, and body mass index on the opportunit of total disc replacement surgery	4		Y	
Clinical occurres of total date replacement in patients with degrammine dire dataset	>100	24	N	
Tatal disc replacement for the warment of homized nucleus pulpous and radioslopathy.	3		Y	