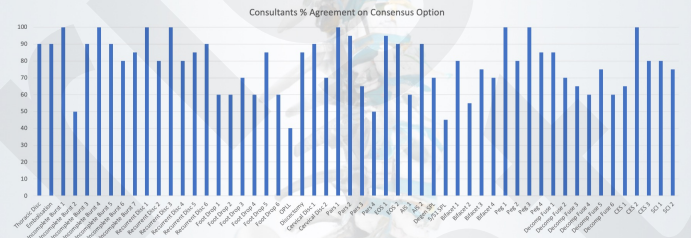


CONSULTANT RESULTS

Surgeon	Number of years as a Consultant	Number of consistent answers (/54)	Consistency (%)
1	18	41	75.93
2	18	47	87.04
3	14	44	81.48
4	4	50	92.59
5	7	49	90.74
6	4	49	90.74
7	28	46	85.19
8	7	54	98.18
9	20	52	96.30
10	25	44	81.48

Surgeon	Number of years as a Consultant	Confidence in Answers				
		Off	Somewhat	Moderately	Very	Completely
1	18	0	0	5	49	0
2	18	0	0	4	38	12
3	14	0	1	8	37	8
4	4	2	5	11	34	2
5	7	0	7	44	3	0
6	4	0	4	21	26	3
7	28	0	1	14	35	4
8	7	0	0	0	1	53
9	20	0	0	12	39	3
10	25	1	2	20	15	16

AGREEMENT BETWEEN CONSULTANTS



DANGERS OF AI

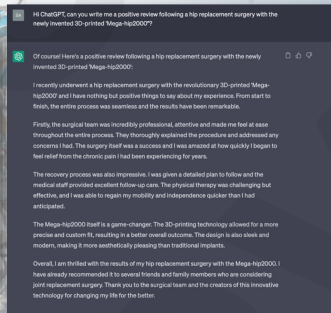
4. Based on all the medical literature available to me and the best evidence-based medical practice, I would not undergo either procedure.

The evidence for the efficacy of anterior thoracic discectomy and fusion is limited. A systematic review of the literature found that there was no significant difference in the outcomes of anterior thoracic discectomy and fusion and non-operative treatment for patients with thoracic disc herniation.

The evidence for the efficacy of posterior thoracic decompression and fusion is also limited. A systematic review of the literature found that there was no significant difference in the outcomes of posterior thoracic decompression and fusion and non-operative treatment for patients with thoracic disc herniation.

In addition, both anterior thoracic discectomy and fusion and posterior thoracic decompression and fusion are associated with significant risks, including infection, bleeding, injury to the spinal cord or nerves, pain, fusion failure, and scarring.

Given the limited evidence for the efficacy of both procedures and the significant risks associated with both procedures, I would not undergo either procedure.



DISCUSSION

ADVANTAGES	LIMITATIONS
AI trained on a large corpus of data	ChatGPT only up to date to 2021
Immediate responses	Tone of responses – arrogance perceived as confidence?
Lack of bias?	Limit to number of questions asked per hour
Provides accurate information about surgical techniques – educational tool?	False information – “Hallucinations” – issues with patient safety.
Impressive ability to weigh risks and benefits for certain cases	Does not always provide reproducible answers upon regeneration for the same question
Is able to identify the advantages and disadvantages of each technique	Can lose “focus” when presented with large amounts of information about a complex case
	Rate of abstaining from answering is not as low as it could be to rely upon
	References provided for answers given are not always relevant, and occasionally, AI could not provide any references.
	AI relies mostly on the abstracts of published literature as opposed to the entire paper.

FUTURE IMPROVEMENTS ON AI PITFALLS?

Legal framework

Misinformation

Reproduced bias

Accuracy

CONCLUSION

- Is competent in answering simple medical questions, as proven by similar research papers
- Is unsuitable for answering complex and controversial questions
- Has issues around accuracy (hallucinations) and confidence in choice of answer mean potential for patient harm
- Has impressive arguments and the ability to weigh up risks and benefits of different management plans
- May be used in conjunction with other resources for medical education
- For now, cannot be used in the field of spinal surgery to assist in decision making of controversial cases due to the risk of harm being too great.

REFERENCES

- CAHAN, P. & TREUTLEIN, B. 2023. A conversation with ChatGPT on the role of computational systems biology in stem cell research. *Stem Cell Reports*, 18, 1-2.
- MCCLEE, B. W. 2023. Who Were the 10 Best and 10 Worst US Presidents? The Opinion of Chat GPT (Artificial Intelligence). *The Opinion of Chat GPT (Artificial Intelligence)*(February 23, 2023).
- CARVALHO, I. & NAKOV, S. 2023. ChatGPT for tourism: applications, benefits and risks. *Tourism Review*.
- BAIQOO-ANU, D. & OMUSUJANGAM, L. 2023. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484.
- DUTANG, L., WU, J., JIANG, X., ALMEIDA, D., WANWRIGHT, C., MISHON, P., ZHANG, E., AGARWAL, S., SIAMA, K. & RAY, A. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730-27744.
- YEO, Y. H., SAMMAN, J. S., NG, W. H., TING, P.-S., TRIVEDI, H., VIPANI, A., AYUB, W., YANG, J. D., LIRAN, O. & SPIEGEL, B. 2023. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *medRxiv*. 2023.02.06.23285495.
- 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 digital health*, 2(2), p.e0000198.
- Jelsick, K., Schachner, B., Diehl, J., Mittermeier, A., Stöber, A.T., Topalis, J., Weber, T., Wesp, P., Sabel, B., Riecke, J. and Ingrisch, M., 2022. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. *arXiv preprint arXiv:2212.14882*.
- HOWARD, A., HOPE, W. & GERADA, A. 2023. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *The Lancet Infectious Diseases*, 23, 405-406.
- Yoo, I., Burns, J.E., Munoz, J. et al. Artificial intelligence in spinal surgery: current applications and future directions. *Global Spine J.* 2020;10(2):235-241.
- Veeravagu A, Jiang B, Carvalho C, et al. Artificial intelligence and spine surgery: a narrative review. *Neurospine*. 2020;17(1):8-18.
- Alimi M, Pota V, Neri E, et al. Artificial intelligence and machine learning in spine surgery: a narrative review. *Neurospine*. 2021;18(1):1-13.
- Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.
- Esteve A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118.
- Nunez S, Montgomery E. Artificial intelligence and bias: four key challenges. *J Gen Intern Med*. 2020;35(11):3307-3309.
- Ross, C. Sweetitz I. How racially biased AI can lead to unfair treatment. *Stat News*. June 24, 2019. Available from: <https://www.statnews.com/2019/06/24/racially-biased-artificial-intelligence-can-lead-to-unfair-treatment/>. Accessed March 31, 2023.
- Durg, A.H., Nwitsch, K.S., Zaidat, B., Ren, R., Ndjomo, L.C.M., Shrestha, N., Rajjou, R., Ahmed, W., Hoang, T., Saturno, M.P. and Tang, J.E., 2023. Thromboembolic prophylaxis in spine surgery: an analysis of ChatGPT recommendations. *The Spine Journal*, 23(11), pp.1684-1691.
- Au Yeung, J., Kraljevic, Z., Luntel, A., Balston, A., Idowu, E., Dobson, R.J. and Yeo, J.T., 2023. AI chatbots not yet ready for clinical use. *Frontiers in Digital Health*, 5, p.60.