MEDTECH INSIGHT
CITELINE COMMERCIAL

01 Apr 2024 | **Interviews**

HSS Spine Surgeon Sees Potential For AI Chatbots, But Not To 'Practice Medicine'

by Marion Webb

HSS Spine co-chair Sheeraz Qureshi can envision a future where HSS will use AI chatbots, like ChatGPT, to help educate patients, but it will never replace shared clinical decision-making between doctors and patients.

The Hospital for Special Surgery (HSS) ranks among US health care systems evaluating how AI tools, such as ChatGPT, could be integrated to create efficiencies and ease the rising burden of documentation on health providers.

"I think it's impossible to have enough time to take care of all of the questions and concerns of every patient," Sheeraz Qureshi, co-chief of Spine at HSS, told *Medtech Insight*.

"So I think the advantage of having some sort of a chatbot that can answer some questions with relative accuracy just allows for a patient to be more informed in their decision-making without that being a further time burden on a clinician," Qureshi said.

His remarks followed a study in which Open AI's ChatGPT 3.5 version provided more accurate answers than researchers expected on medical questions about cervical spine surgery.

In light of the findings, Qureshi said he can envision a future where AI tools may be integrated at the academic medical center to enhance interaction and augment communication between providers and patients.

However, there are limitations to what AI can provide. (Also see "*'We Are Definitely In Uncharted Territory': AI Innovation, Regulation And Health Equity Prospects*" - Medtech Insight, 20 Dec, 2023.)

MEDTECH INSIGHT
CITELINE COMMERCIAL
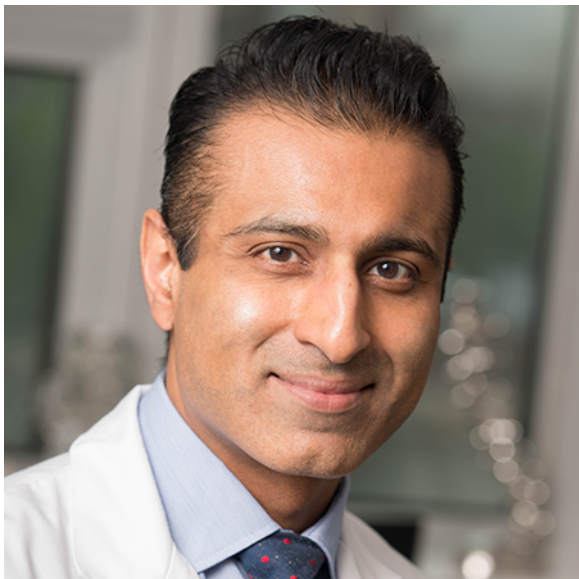
DR. SHEERAZ QURESHI, CO-CHAIR OF SPINE AT HSS *Source: HSS*

"AI is a predictive tool that has no emotion, and I think the decision-making in health care is a shared decision and is much more tied to the personal relationship between the patient and the doctor than we may give it credit for," he said. "It is a practice of medicine and AI will never be able to replace that," Qureshi said.

"AI is a predictive tool that has no emotion, and I think the decision-making in health care is a shared decision and is much more tied to the personal relationship between the patient and the doctor than we may give it credit for ... It is a practice of medicine and AI will never be able to replace that." -- Sheeraz Qureshi

Qureshi believes that large language models, such as ChatGPT – provided they are used by HSS patients, which will be investigated further – could be an added tool to help patients make treatment decisions by answering questions such as, "What is the overall risk of paralysis in patients who undergo cervical spine surgery for spinal cord compression?"

The tools could also be trained using data from HSS to give patients comparative information such as, "What is the average infection rate?" and "What is the infection rate at HSS?"

Given that doctors only have so much time available to answer questions, giving patients access to a bot 24/7 would enable them to become more informed, which he believes could lead to better outcomes by aligning patient-physician expectations.

"There are some bad outcomes that happen because something went wrong, but most of the time I find that there's a sort of dissatisfaction that a patient may have with the outcome, because their expectation wasn't correct or what they interpreted the doctor had told them wasn't really what the doctor meant," he explained.

He noted that, especially in spine surgery, surgeons expect to make the patient better, but that

MEDTECH INSIGHT
CITELINE COMMERCIAL

doesn't necessarily mean 100% better. The more informed the patient is upfront, the greater the likelihood that both parties are aligned relative to potential outcomes.

He also sees value for physicians in using chatbots to obtain authorizations for certain types of treatments from insurers, which he said would save doctors time spent on the phone.

First, Qureshi said, HSS researchers will need to understand how many patients are actually using AI tools to access medical information, because if not enough patients are using AI tools, then integrating them may not be worth the investment.

"My expectation is it will probably depend on the age of patients and their comfort level using things like ChatGPT in other parts of their life," he said.

After HSS' analysis, he expects the health system will train its own algorithms to ensure the most accurate information is returned based on a set of curated data.

ChatGPT 3.5 Scores 8.1 Out Of 10

HSS researchers recently presented findings from a study that evaluated whether ChatGPT 3.5 can answer questions about spine surgery.

For the study, HSS investigators identified nine frequently asked questions about cervical spine surgery, with queries ranging from the benefits and drawbacks of different surgical approaches to side effects and recovery after surgery. The questions were inputted one at a time into ChatGPT version 3.5, released in late 2022, including, "What are the different types of cervical spine surgery?", "What is the difference between ACDF or CDR?" and "How long is the recovery following cervical spine surgery?"

Two independent reviewers scored the responses on accuracy, appropriateness in giving a specific answer to specific questions asked, and readability by a layperson. On average, the responses received a score of 8.1 out of 10 with a 3.9 out of 5 for accuracy and 2.2 out of 3 for appropriateness.

The reviewers said while accuracy was impressive, they observed that the bot failed to provide comprehensive responses and often omitted important factors. As an example, ChatGPT would describe a particular procedure as being more challenging without mentioning other considerations, such as patient indications or surgical training.

"Appropriateness and readability were the two significant strong points for ChatGPT," the reviewers said. The responses were easier for people to understand than research literature. They also praised the bot for prefacing answers with a statement to consult an expert for medical advice. The findings were presented at the annual American Academy of Orthopedic Surgeons

MEDTECH INSIGHT
CITELINE COMMERCIAL

meeting in February.

One of Qureshi's major concerns with using large language models is that, compared with doing a web search, they often fail to provide citations. This can be problematic given AI's known susceptibility to "hallucinations."

Another concern he shares with many doctors in using generative AI is around litigation. (Also see "*A More Cautious Approach To AI, And Other Policy And Regulation Trends From HIMSS 24*" - Medtech Insight, 19 Mar, 2024.)

"I would definitely want a disclaimer," Qureshi said. "If it was something that we 'recommended,' that is, for informational and educational purposes only, [it] should not be taken as absolute advice any one person's specific scenario and does not replace consultation with an expert."

Rather, generative AI models should be seen as tools to bridge information gaps between patients and physicians.

"If the chatbot can be used to answer additional questions for a patient, that's a great use of a chatbot, but at the end of the day that chatbot should not be something that a patient is using to decide whether or not they should have surgery done," Qureshi said.