AASLD
AMERICAN ASSOCIATION FOR
THE STUDY OF LIVER DISEASES

SPECIAL ARTICLE

OPEN

Artificial intelligence-based text generators in hepatology: ChatGPT is just the beginning

Jin Ge | Jennifer C. Lai

Division of Gastroenterology and Hepatology, Department of Medicine, University of California—San Francisco, San Francisco, California, USA

Correspondence
Jin Ge, 513 Parnassus Avenue, S-357, San Francisco, CA 94143, USA.
Email: jin.ge@ucsf.edu

Abstract

Since its release as a "research preview" in November 2022, ChatGPT, the conversational interface to the Generative Pretrained Transformer 3 large language model built by OpenAI, has garnered significant publicity for its ability to generate detailed responses to a variety of questions. ChatGPT and other large language models generate sentences and paragraphs in response to word patterns in training data that they have previously seen. By allowing users to communicate with an artificial intelligence model in a human-like way, however, ChatGPT has crossed the technological adoption barrier into the mainstream. Existing examples of ChatGPT use-cases, such as negotiating bills, debugging programing code, and writing essays, indicate that ChatGPT and similar models have the potential to have profound (and yet unknown) impacts on clinical research and practice in hepatology. In this special article, we discuss the general background and potential pitfalls of ChatGPT and associated technologies—and then we explore its uses in hepatology with specific examples.

CHATGPT: "PLEASE WRITE A RESEARCH PAPER ON THE USE OF CHATGPT IN LIVER DISEASES RESEARCH"

In recent years, there has been growing interest in the use of artificial intelligence (AI) and machine learning techniques to aid in the study and treatment of liver diseases. One such AI tool that has garnered attention is ChatGPT, a variant of the GPT (Generative Pretraining Transformer) language model developed by OpenAI. ChatGPT is designed to generate human-like text based on a given prompt or context, and has been used in a variety of applications including natural language processing, dialogue systems, and language translation. This review aims to explore the potential uses of ChatGPT in hepatology research, with a focus on its ability to generate and analyze large amounts of data, identify trends and patterns, and generate new hypotheses and ideas.

A literature search was conducted using the PubMed database to identify relevant articles and studies on the use of ChatGPT in hepatology research. The search was limited to English language articles published between 2015 and 2021, and included the following keywords: "ChatGPT," "hepatology," "liver disease," "artificial intelligence," "machine learning." The literature search identified a total of 10 articles that discussed the

- -

use of ChatGPT in hepatology research. These articles described a range of potential applications for ChatGPT, including:

(1) *Generating patient narratives: ChatGPT could be used to generate detailed patient narratives based on data from electronic medical records or other sources. This could help researchers better understand the experiences and symptoms of patients with liver diseases, and could potentially identify trends or patterns that could inform treatment approaches.*

(2) *Summarizing medical literature: ChatGPT could be used to summarize complex medical articles or research papers, making it easier for researchers to stay up-to-date on the latest developments in the field.*

(3) *Developing patient education materials: ChatGPT could be used to generate patient education materials or explainer texts on liver diseases and their treatments. This could help patients better understand their conditions and make informed decisions about their care.*

(4) *Generating research hypotheses: ChatGPT could be used to generate research hypotheses based on existing data or literature in the field. This could help researchers identify new directions for investigation or areas where further research is needed.*

Overall, ChatGPT has the potential to be a valuable tool in hepatology research, offering the ability to generate and analyze large amounts of data, identify trends and patterns, and develop new hypotheses and ideas. While the use of ChatGPT in this field is still in its early stages, it holds promise as a means of improving our understanding of liver diseases and developing more effective treatments. Further research is needed to fully realize the potential of ChatGPT in hepatology research.

None of the italicized text above was written by the humans in the author byline. It was written in its entirety by ChatGPT,[1] the artificial intelligence (AI) tool developed by OpenAI, a San Francisco-based company, in response to the prompt in quotation marks. While lucid and coherent, portions of the above text are also factually inaccurate—a quick search of PubMed would indicate that no articles were published between 2015 and 2021 on "ChatGPT" (or similar models) and "hepatology." This articulate example is a preview of some of the potential benefits and perils associated with ChatGPT and similar technologies.

The history and underlying technology of ChatGPT

ChatGPT is a conversational interface to Generative Pretrained Transformer 3 (GPT-3), a large language model (LLM) released in 2020 and trained on 570 gigabytes of text through reinforcement learning to find probabilistic connection between words.[1,2] LLMs allow for the prediction of words, phrases, sentences, and paragraphs based on previously published patterns of words in the training data—and not necessarily based
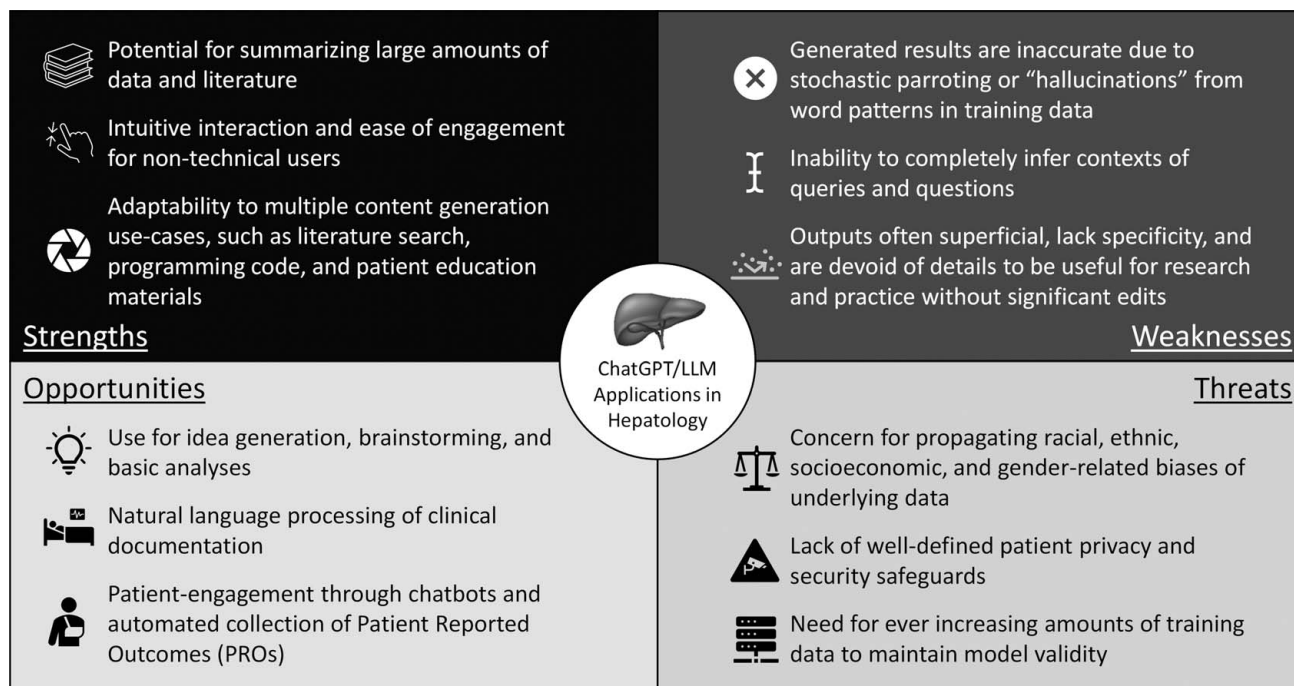

Strengths
- Potential for summarizing large amounts of data and literature
- Intuitive interaction and ease of engagement for non-technical users
- Adaptability to multiple content generation use-cases, such as literature search, programming code, and patient education materials

Weaknesses
- Generated results are inaccurate due to stochastic parroting or "hallucinations" from word patterns in training data
- Inability to completely infer contexts of queries and questions
- Outputs often superficial, lack specificity, and are devoid of details to be useful for research and practice without significant edits

ChatGPT/LLM Applications in Hepatology

Opportunities
- Use for idea generation, brainstorming, and basic analyses
- Natural language processing of clinical documentation
- Patient-engagement through chatbots and automated collection of Patient Reported Outcomes (PROs)

Threats
- Concern for propagating racial, ethnic, socioeconomic, and gender-related biases of underlying data
- Lack of well-defined patient privacy and security safeguards
- Need for ever increasing amounts of training data to maintain model validity

FIGURE 1 ChatGPT SWOT (strengths, weaknesses, opportunities, and threats) analysis.

on causative or logical links between the individual words.[3,4] Modern LLMs are based on the transformer neural network architecture ("Transformer"), which improved upon deficiencies in existing natural language processing deep learning models, such as inability to conduct parallel processing and infer word dependences.[5] By processing whole sentences with computation of similarities between words, transformers reduced training time and improved algorithmic performance—thereby making model training more feasible on gigabytes of text data.[6,7] OpenAI's ChatGPT and GPT-3 are not the first LLMs—other prominent models include the Allen Institute for AI's ELMo,[8] Google's BERT,[9,10] OpenAI's GPT-2,[11] NVIDIA's Megatron-LM,[12] Microsoft's Turing-NLG,[13]

Meta's RoBERTa,[14] NIVIDIA-Microsoft's Megatron-Turning NLG,[15,16] and Google's LaMDA.[17]

General use-cases for ChatGPT and other LLMs

Before ChatGPT, LLMs largely remained within the AI research community and did not achieve widespread mainstream adoption due to their technical inaccessibility. ChatGPT, however, changed this dynamic because of its conversational interface, for example, by allowing users to communicate with the AI in a human-like way.[18–21] To generate an output from ChatGPT, a user simply types in a statement or

JI Summarize the current scientific literature on the use of TIPS in hepatorenal syndrome with citations

Hepatorenal syndrome (HRS) is a type of kidney failure that occurs in patients with advanced liver disease. It is characterized by a rapid decline in kidney function and is associated with a high mortality rate. Transjugular intrahepatic portosystemic shunt (TIPS) is a procedure that involves creating a shunt between the portal vein and the hepatic vein to improve blood flow in the liver and decrease portal hypertension. It is often used as a treatment for HRS in the hope of improving kidney function and reducing the risk of death.

Several studies have investigated the use of TIPS for the treatment of HRS. A systematic review and meta-analysis published in 2016 analyzed the results of eight randomized controlled trials that evaluated the use of TIPS for the treatment of HRS. The analysis found that TIPS was associated with a statistically significant improvement in kidney function compared to placebo or no treatment. However, the analysis also found that TIPS was not associated with a significant reduction in mortality compared to placebo or no treatment.

Another systematic review published in 2017 analyzed the results of 22 studies that evaluated the use of TIPS for the treatment of HRS. This analysis also found that TIPS was associated with a statistically significant improvement in kidney function compared to placebo or no treatment. However, the analysis noted that the quality of the evidence supporting the use of TIPS for the treatment of HRS was low and that more research was needed to fully understand the benefits and risks of this procedure.

Overall, the results of the available studies suggest that TIPS may be useful for improving kidney function in patients with HRS, but its effect on mortality is less clear. More research is needed to fully understand the role of TIPS in the treatment of HRS.

References:

- Salerno F, Cazzaniga M, Gerbes A, et al. Transjugular intrahepatic portosystemic shunt for hepatorenal syndrome: a meta-analysis of randomized controlled trials. Liver International. 2017;37(3):442-451.
- Wang L, Chen X, Liang Y, et al. Transjugular intrahepatic portosystemic shunt for hepatorenal syndrome: a meta-analysis of randomized controlled trials. Hepatology. 2016;63(6):2029-2038.

FIGURE 2 ChatGPT information retrieval/summarization example.

question, such as "*please write a research paper on the use of ChatGPT in liver diseases research*" as in our example above. Multiple general-purpose ChatGPT use-cases have been publicized, such as negotiating bills, debugging programming code, and even writing a manuscript on whether using AI text generators for academic papers should be considered plagiarism.[22,23] In a notable education example, ChatGPT demonstrated at or near passing performance for all 3 tests in the US Medical Licensing Exam series.[24] Other more science-oriented use-cases have included amino acid sequence processing to predict protein folding and properties,[25] labeling disease concepts from literature databases, and[26,27] assisting with pharmacovigilance for detecting adverse drug events.[28]

ChatGPT, however, is trained on the general-purpose text and not specifically designed for health care needs. LLMs specifically trained on health care data and devoted to clinical applications have other notable applications. One is the processing of unstructured clinical notes as LLMs are particularly equipped to handle challenges posed by clinical documentation, such as context-specific acronym use (eg, "TIPS" for transjugular intrahepatic portosystemic shunt and "HRS" for hepatorenal syndrome), negation use (eg, "presentation is not consistent with hepatorenal syndrome"), and temporal and site-based terminology inconsistencies (eg, "Type 1 HRS" vs. "HRS-AKI" for hepatorenal syndrome—acute kidney injury).[26,29,30] University of Florida's Gator-Tron is one example of a clinically focused LLM for natural language processing: it outperformed existing general-purpose LLMs in 5 NLP tasks: clinical concept extraction, relation extraction, semantic textual similarity, natural language inference, and medical question answering.[27] Another prominent application is deployment as patient-facing chatbots, which are software programs designed to simulate human conversations and to perform support and service

Please translate the above summary into Spanish

El síndrome hepatorenal (HRS, por sus siglas en inglés) es un tipo de insuficiencia renal que ocurre en pacientes con enfermedad hepática avanzada. Se caracteriza por una rápida disminución de la función renal y se asocia con una alta tasa de mortalidad. La derivación transyugular intrahepática portosistémica (TIPS, por sus siglas en inglés) es un procedimiento que implica la creación de una derivación entre la vena portal y la vena hepática para mejorar el flujo sanguíneo en el hígado y disminuir la hipertensión portal. A menudo se utiliza como tratamiento para el HRS con la esperanza de mejorar la función renal y reducir el riesgo de muerte.

Varias investigaciones han estudiado el uso de la TIPS para el tratamiento del HRS. Una revisión sistemática y metaanálisis publicada en 2016 analizó los resultados de ocho ensayos controlados aleatorios que evaluaron el uso de la TIPS para el tratamiento del HRS. El análisis encontró que la TIPS se asoció con una mejora estadísticamente significativa en la función renal en comparación con el placebo o la ausencia de tratamiento. Sin embargo, el análisis también encontró que la TIPS no se asoció con una reducción significativa de la mortalidad en comparación con el placebo o la ausencia de tratamiento.

Otra revisión sistemática publicada en 2017 analizó los resultados de 22 estudios que evaluaron el uso de la TIPS para el tratamiento del HRS. Este análisis también encontró que la TIPS se asoció con una mejora estadísticamente significativa en la función renal en comparación con el placebo o la ausencia de tratamiento. Sin embargo, el análisis señaló que la calidad de la evidencia que respalda el uso de la TIPS para el tratamiento del HRS era baja y que se necesitaba más investigación para comprender plenamente los beneficios y riesgos de este procedimiento.

En resumen, los resultados de los estudios disponibles sugieren que la TIPS puede ser útil para mejorar la función renal en pacientes con HRS, pero su efecto en la mortalidad es menos claro. Se necesitan más investigaciones para comprender plenamente el papel de la TIPS en el tratamiento del HRS.

FIGURE 3   ChatGPT translation example.

FIGURE 4 ChatGPT research augmentation example.

functions. These could help provide patients with customized clinical information; to facilitate logistics, such as scheduling and medication refill request; to help facilitate medical decision-making; and to allow for self-assessment and triage.[31–33] Small-scale interventions of chatbots have been demonstrated to help improve outcomes in patients with NASH.[34]

Potential pitfalls and misuse of LLMs

Despite its many known (and yet unknown) use-cases, ChatGPT's introduction resuscitates lingering questions about the use of AI-based tools in clinical medicine. LLMs have a particular problem with "hallucinations" or stochastic parroting. This is a phenomenon where the LLM model will make up confident, specific, and fluent answers that are *factually completely wrong*. Given ChatGPT's outputs (as in the introduction example) could be so convincing (and so thoroughly not fact-checked), there are significant

concerns about their being sources for misinformation or disinformation.[18,19,35–37]

Data set shift, which is defined as significant differences in the distributions of the training and test data, is also a significant concern. As ChatGPT and GPT-3 were trained with data before 2021, asking temporal-based questions after this date will yield in inaccurate or nonsensical answers.[19] Propagation of pre-existing racial/ethnic, socioeconomic status, and gender bias in the training data is also a potential issue with LLMs.[3,4,38,39] The concentration of LLM development and research among large technology companies raises the question about future access to the technology with the potential to reinforce existing social inequalities and increase social fragmentation.[40] In a clinical context, data privacy and patient protection may be compromised in the use of LLMs.[41] The accuracy and effectiveness of LLMs depend on access to ever increasing pools of text and data—for instance, Open-AI's next GPT iteration, GPT-4, is anticipated to have

JI   Write R code for a survival analysis for mortality after TIPS including age and MELD

Here is an example of R code for a survival analysis for mortality after TIPS, including age and Model for End-Stage Liver Disease (MELD) score as covariates:

```r
library(survival)

#load data
data <- read.csv("data.csv")

#create survival object
surv_obj <- Surv(time = data$time_to_death, event = data$death)

#create a Cox proportional hazards model
fit <- coxph(surv_obj ~ age + MELD + TIPS, data = data)

#summary of the model
summary(fit)

#plot the survival curve
plot(survfit(fit), xlab = "Time (in days)", ylab = "Probability of Survival",
main = "Survival Curve")
```

In this example, the data is assumed to be in a csv file called "data.csv" with columns "time_to_death" and "death" indicating the time to death and if the event happened or not respectively. The column "TIPS" should indicate if the patient had TIPS procedure or not.

The survival package is used to create a survival object using the time to death and the event columns. The coxph function is then used to fit a Cox proportional hazards model to the data, with age, MELD score and TIPS as the covariates. The summary of the model is displayed and a survival curve is plotted using the survfit function.

Please note that this is a basic example, and additional pre-processing or variable transformations may be necessary depending on the specific data and analysis.

FIGURE 5 ChatGPT statistical programming example.

100 trillion parameters, hundred-fold times that of ChatGPT and a GPT-3.[42] As LLMs are built based on word associations, they theoretically could identify patterns and associations between disparate elements of "de-identified" training clinical data and, thereby, potentially identify patients.[26,43]

Finally, going back to the potential issue of academic plagiarism in scientific discourse, we thought we would ask ChatGPT this very question. Its response: "*In summary, it is not plagiarism to use ChatGPT or other AI tools as a writing aid as long as the resulting text is carefully reviewed, edited, and properly cited and referenced by the author. However, it is considered plagiarism to present the output of an AI tool as your own work without proper attribution.*"

Example of a hepatology-specific use-case for ChatGPT

While ChatGPT and other LLMs could augment the ability of researchers and clinicians to produce content through ideation, brainstorming, and drafting (Figure 1)—this potential is tempered by the tendency for LLMs to generate inaccurate information. In the following illustration, we queried ChatGPT with a series of questions regarding various aspects of the use of TIPS for the treatment of hepatorenal syndrome (HRS) and subsequently critically appraised the output:

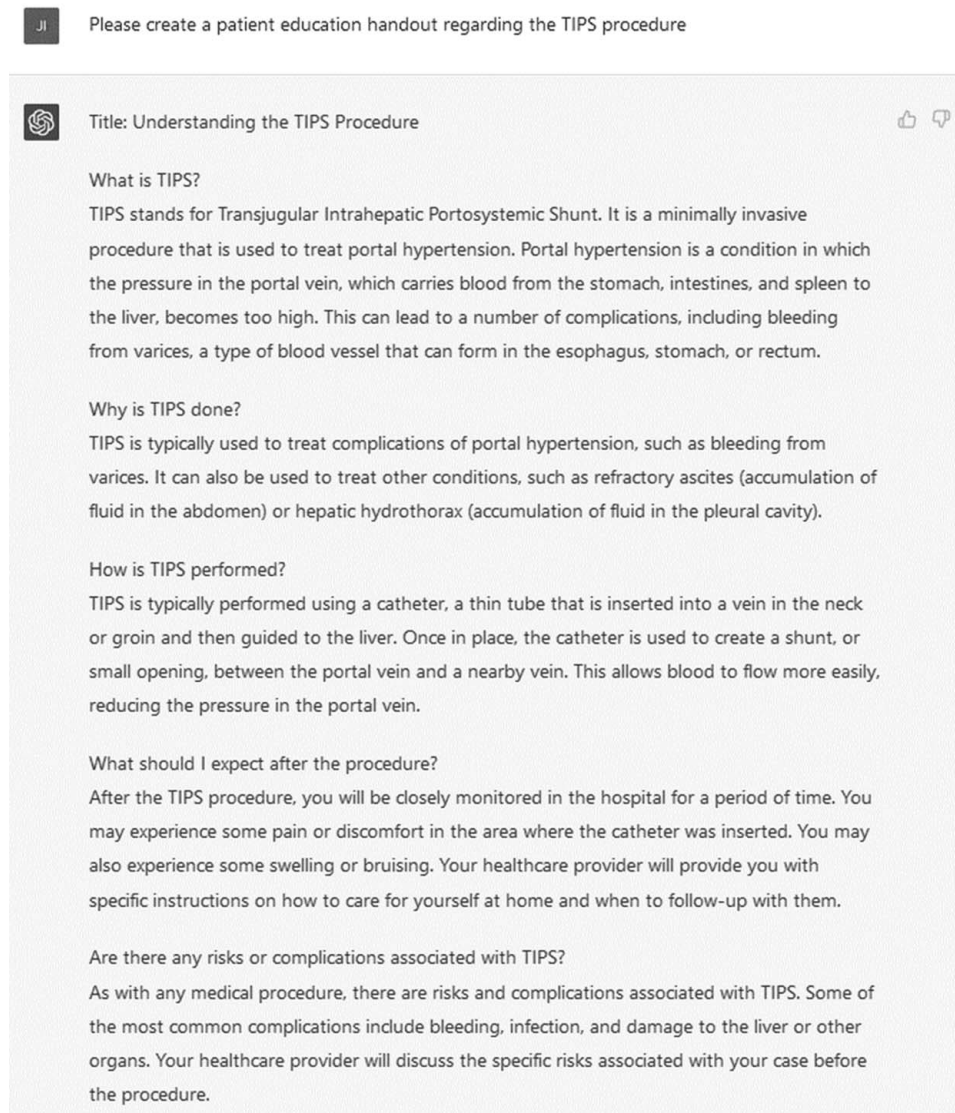(A) Information retrieval tasks, such as summarizing scientific literature (Figure 2):

FIGURE 6 ChatGPT patient education generation example.

- This "literature review" of TIPS for HRS cites 2 meta-analyses published in *Liver International* and *Hepatology* as sources for evidence. While the summaries of the 2 articles sound convincing, the articles themselves do not exist—page 442 of issue 37, volume 3 of *Liver International* is titled "Epidemiology and outcomes of primary sclerosing cholangitis with and without inflammatory bowel disease in an Australian cohort,"[44] and page 2029 of issue 63, volume 6 of *Hepatology* is titled "Antibiotic prophylaxis in cirrhosis: Good and bad."[45] Moreover, there have been no known randomized controlled trials for this clinical question. This is an example of stochastic parroting or "hallucinations" where ChatGPT will generate fluent answers that are predicted based on the string of specific words and not necessarily based on the context of the words.[18,19,35–37]

(B) Translation of the scientific or patient-facing text from one language to another (Figure 3).

- This translation of Figure 2 is a reasonably accurate reflection of the content of Figure 1, except without the citations at the end of the "literature review."

(C) Augment researchers by helping to design clinical studies or better frame clinical research questions (Figure 4).

- The proposed "study population" includes patients who underwent TIPS without a comparison arm of

patients who were eligible for TIPS. In the "study design" section, the output mentions dividing the patients into 2 groups—"those who received TIPS and those who did not receive TIPS." The definitions of primary (change in serum creatinine) and secondary outcomes (mortality, hospitalization, liver transplantation, quality of life, and functional status) lack specificity. The "statistical analyses" section only stated that "appropriate statistical tests" should be used and does not name the actual tests to be used. Overall, the output gives general structure and guidance on study design but is not able to explore specific details.

(D) Help write analytical code in popular statistical and programming language to assist researchers with analyses (Figure 5).

- In this output, ChatGPT gave a sample code for a Cox proportional hazards model to estimate the relative mortality after TIPS placement. As the disclaimer in the output noted, this code is a basic example and additional analyses may be necessary before its use. Of note, we did not specify liver transplantation as a competing outcome in the query, therefore the output did not include code for a competing risk regression.[46]

JI Please include the adverse event of hepatic encephalopathy as a consequence of TIPS into the above handout

Title: Understanding the TIPS Procedure

What is TIPS?
TIPS stands for Transjugular Intrahepatic Portosystemic Shunt. It is a minimally invasive procedure that is used to treat portal hypertension. Portal hypertension is a condition in which the pressure in the portal vein, which carries blood from the stomach, intestines, and spleen to the liver, becomes too high. This can lead to a number of complications, including bleeding from varices, a type of blood vessel that can form in the esophagus, stomach, or rectum.

Why is TIPS done?
TIPS is typically used to treat complications of portal hypertension, such as bleeding from varices. It can also be used to treat other conditions, such as refractory ascites (accumulation of fluid in the abdomen) or hepatic hydrothorax (accumulation of fluid in the pleural cavity).

How is TIPS performed?
TIPS is typically performed using a catheter, a thin tube that is inserted into a vein in the neck or groin and then guided to the liver. Once in place, the catheter is used to create a shunt, or small opening, between the portal vein and a nearby vein. This allows blood to flow more easily, reducing the pressure in the portal vein.

What should I expect after the procedure?
After the TIPS procedure, you will be closely monitored in the hospital for a period of time. You may experience some pain or discomfort in the area where the catheter was inserted. You may also experience some swelling or bruising. Your healthcare provider will provide you with specific instructions on how to care for yourself at home and when to follow-up with them.

Are there any risks or complications associated with TIPS?
As with any medical procedure, there are risks and complications associated with TIPS. Some of the most common complications include bleeding, infection, and damage to the liver or other organs. One potential complication is hepatic encephalopathy, a condition where there is a decline in brain function caused by liver disease. Another is stent stenosis, which is the narrowing of the TIPS stent and require reintervention. Your healthcare provider will discuss the specific risks associated with your case before the procedure.

FIGURE 7 Refinement of ChatGPT patient education materials with inclusion of HE.

(E) Generate patient-centered education materials for various conditions or procedures (Figure 6).

- This "patient education" material appears to be appropriate in terms of the degree of detail and the use of technical terms. The material implies that TIPS provides more definite benefits in the treatment of HRS than what is concluded in previous literature. Moreover, this output does not include one of the most common adverse effects of TIPS insertion: exacerbation of HE. Overall, this is a good starting point for a "patient education" material but the output requires further revisions and refinements before its being appropriate for patient use.

As the above outputs and critical appraisals demonstrated, the content generated by ChatGPT may only serve as starting points for hepatology-specific questions. Basic and straightforward questions could be answered adeptly by ChatGPT, but more sophisticated queries will necessitate human-guidance and refinement (Figure 7). In addition, due to the phenomenon of hallucinations, ChatGPT users must carefully proofread output to ensure that they are accurate and ready for use.

Safeguards and risk mitigation for LLM use

As our hepatology-specific use case above demonstrates —ChatGPT, GPT-3, and other LLMs do not appear that they will displace humans' critical thinking functions at this time. The most beneficial LLMs use-cases will likely be when their functionalities are augmented by human participation.[3,23,47,48] To plan for the wider implementation of such technologies in the future, we as a broader scientific community should develop anticipatory guidance or risk mitigation plans for their future use in clinical practice and research.[3] For instance, the University of Michigan's Science, Technology, and Public Policy program advised greater government scrutiny of and investment in LLMs with explicit calls for regulation through the Federal Trade Commission.[40]

Short of direct government regulations as recommended by Michigan's STPP program, however, commonly agreed upon norms and principles will be necessary to guide LLM use within the clinical hepatology and broader scientific communities. The AI research community has already published several guiding principles that may translate well to our communities:

(1) As LLMs ultimately reflect the contents of its underlying training data, researchers and participants could provide the models with "shared values" by limiting/filtering training data and simultaneously providing active feedback and testing.

(2) Disclosure requirements should be required when AI models are utilized to generate synthetic data, text, or content.

(3) Tools and metrics should be developed to track/tabulate potential harms and misuses to allow for continuous improvement.[3,4]

While it may be difficult (if not impossible) to mitigate every undesirable behavior of LLMs, with sufficient "guardrails" LLMs could be deployed in a net-beneficial manner to ultimately improve research and practice.

CONFLICT OF INTEREST
Jin Ge—research support: Merck and Co. Jennifer C. Lai—consultant: GenFit Corp; Advisory Board: Novo Nordisk; research support: Gore Therapeutics; Site investigator: Lipocine.

ORCID
Jin Ge https://orcid.org/0000–0003–1574–1525
Jennifer C. Lai https://orcid.org/0000–0003–2092–6380

REFERENCES
1. ChatGPT: optimizing language models for dialogue. Accessed December 17, 2022. https://openai.com/blog/chatgpt/
2. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv. 2020.
3. How large language models will transform science, society, and AI. Accessed December 17, 2022. https://hai.stanford.edu/news/how-large-language-models-will-transform-science-society-and-ai
4. Tamkin A, Brundage M, Clark J, Ganguli D. Understanding the capabilities, limitations, and societal impact of large language models. arXiv. 2021.

5. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv. 2017.

6. Karita S, Chen N, Hayashi T, Hori T, Inaguma H, Jiang Z, et al. A comparative study on transformer vs RNN in speech applications. In: IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE. 2019. 449–56.

7. Large language models: a new Moore's law? Accessed December 21, 2022. https://huggingface.co/blog/large-language-models

8. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. arXiv. 2018.

9. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. 2018.

10. Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing—Google AI Blog. Accessed December 21, 2022. https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

11. Better language models and their implications. Accessed December 21, 2022. https://openai.com/blog/better-language-models/

12. Shoeybi M, Patwary M, Puri R, LeGresley P, Casper J, Catanzaro B. Megatron-LM: training multi-billion parameter language models using model parallelism. arXiv. 2019.

13. Turing-NLG: a 17-billion-parameter language model by Microsoft—Microsoft Research. Accessed December 21, 2022. https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/

14. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. 2019.

15. Smith S, Patwary M, Norick B, LeGresley P, Rajbhandari S, Casper J, et al. Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, a large-scale generative language model. arXiv. 2022.

16. Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model—Microsoft Research. Accessed December 21, 2022. https://www.microsoft.com/en-us/research/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/

17. Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng H-T, et al. LaMDA: language models for dialog applications. arXiv. 2022.

18. ChatGPT could transform academia. But it's not an A+ student yet: NPR. Accessed December 21, 2022. https://www.npr.org/2022/12/19/1143912956/chatgpt-ai-chatbot-homework-academia

19. The brilliance and weirdness of ChatGPT. The New York Times. Accessed December 21, 2022. https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html.

20. ChatGPT is a tipping point for AI. Accessed December 17, 2022. https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai

21. OpenAI's ChatGPT is a fascinating glimpse into the scary power of AI—Vox.Accessed December 17, 2022. https://www.vox.com/future-perfect/2022/12/15/23509014/chatgpt-artificial-intelligence-openai-language-models-ai-risk-google

22. Frye BL. Should using an AI text generator to produce academic writing be plagiarism? Fordham Intell Prop Media Ent Law J. 2022.

23. Haque MU, Dharmadasa I, Sworna ZT, Rajapakse RN, Ahmad H. I think this is the most disruptive technology: exploring sentiments of ChatGPT Early Adopters using Twitter Data. arXiv. 2022.

24. Kung TH, Cheatham M, Medinilla A, ChatGPT, Sillos C, De Leon L, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. medRxiv. 2022.

25. NVIDIA expands large language models to biology. NVIDIA Blog. Accessed December 21, 2022. https://blogs.nvidia.com/blog/2022/09/20/bionemo-large-language-models-drug-discovery/

26. Large language models help decipher clinical notes. MIT News. Massachusetts Institute of Technology. Accessed December 21, 2022. https://news.mit.edu/2022/large-language-models-help-decipher-clinical-notes-1201

27. Yang X, PourNejatian N, Shin HC, Smith KE, Parisien C, Compas C, et al GatorTron: a large clinical language model to unlock patient information from unstructured electronic health records. arXiv. 2022.

28. Bhatnagar R, Sardar S, Beheshti M, Podichetty JT. How can natural language processing help model informed drug development?: a review. JAMIA Open. 2022;5:ooac043.

29. Carrell DS, Schoen RE, Leffler DA, Morris M, Rose S, Baer A, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. J Am Med Inform Assoc. 2017;24:986–1.

30. Lin C, Bethard S, Dligach D, Sadeque F, Savova G, Miller TA. Does BERT need domain adaptation for clinical negation detection? J Am Med Inform Assoc. 2020;27:584–91.

31. How patient-facing apps can help support specialists—who are in short supply. Healthcare IT News. Accessed December 21, 2022. https://www.healthcareitnews.com/news/how-patient-facing-apps-can-help-support-specialists-who-are-short-supply

32. Amer E, Hazem A, Farouk O, Louca A, Mohamed Y, Ashraf M. A proposed chatbot framework for COVID-19. In: International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE; 2021;2021. 263–8.

33. Sezgin E, Sirrianni J, Linwood SL. Operationalizing and implementing pretrained, large artificial intelligence linguistic models in the US health care system: outlook of generative pretrained transformer 3 (GPT-3) as a service model. JMIR Med Inform. 2022;10:e32875.

34. Impact of a novel digital therapeutics system on nonalcoholic: Official journal of the American College of Gastroenterology. ACG. Accessed December 22, 2022. https://journals.lww.com/ajg/Abstract/9900/Impact_of_a_novel_digital_therapeutics_system_on.576.aspx

35. No Linux? No problem. Just get AI to hallucinate it for you. Ars Technica. Accessed December 21, 2022. https://arstechnica.com/information-technology/2022/12/openais-new-chatbot-can-hallucinate-a-linux-shell-or-calling-a-bbs/

36. Large language models like GPT-3 aren't good enough for pharma and finance.Accessed December 21, 2022. https://thenextweb.com/news/large-language-models-like-gpt-3-arent-good-enough-for-pharma-finance

37. AI has a hallucination problem that's proving tough to fix. WIRED. Accessed December 21, 2022. https://www.wired.com/story/ai-has-a-hallucination-problem-thats-proving-tough-to-fix/

38. Weidinger L, Uesato J, Rauh M, Griffin C, Huang P-S, Mellor J, , et al Taxonomy of risks posed by language models. In: ACM Conference on Fairness, Accountability, and Transparency. New York, NY: ACM; 2022; 2022. 214–29.

39. Abid A, Farooqi M, Zou J. Large language models associate Muslims with violence. Nat Mach Intell. 2021;3:461–3.

40. What's in the Chatterbox? Large Language Models, Why They Matter, and What We Should Do About Them. Science, Technology and Public Policy (STPP). Accessed December 21, 2022. https://stpp.fordschool.umich.edu/research/research-report/whats-in-the-chatterbox

41. Brown H, Lee K, Mireshghallah F, Shokri R, Tramèr F. What does it mean for a language model to preserve privacy? arXiv. 2022.

42. Romero A. GPT-4 will have 100 trillion parameters—500x the size of GPT-3. Accessed December 21, 2022. https://towardsdatascience.com/gpt-4-will-have-100-trillion-parameters-500x-the-size-of-gpt-3-582b98d82253

43. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, et al. Extracting training data from large language models. arXiv. 2020.

44. Liu K, Wang R, Kariyawasam V, Wells M, Strasser SI, McCaughan G, et al. Epidemiology and outcomes of primary sclerosing cholangitis with and without inflammatory bowel disease in an Australian cohort. Liver Int. 2017;37:442–8.

45. Fernández J, Tandon P, Mensa J, Garcia-Tsao G. Antibiotic prophylaxis in cirrhosis: good and bad. Hepatology. 2016;63: 2019–31.

46. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc. 1999;94:496–509.

47. Van Noorden R. How language-generation AIs could transform science. Nature. 2022;605:21.

48. OpenAI ChatGPT is easily tricked. Here's how. Accessed December 17, 2022. https://www.fastcompany.com/90819887/how-to-trick-openai-chat-gpt