

卫生信息化国际发展动态

北京市卫生健康大数据与政策研究中心

2024. 12. 27

（十二）聊天机器人

1. 标题：人工智能驱动的聊天机器人在医疗保健中的效用和局限性

来源：Cureus.

时间：2024 年 11 月.

链接：<https://doi.org/10.7759/cureus.73127>. PMID: 39650926.

概要：在人工智能（AI）和医疗保健深度融合的重要节点上，临床医生需要掌握聊天机器人的变革能力。目前，已有部分学者和机构开始探索 ChatGPT 等人工智能驱动的聊天机器人的潜在好处。本研究将通过向免费的 AI 聊天机器人 ChatGPT（OpenAI）的 3.5 版本提出一系列录制的提问并得到回复的方式，来全面了解医疗保健领域 AI 聊天机器人的当前知识和实践，探索 ChatGPT（OpenAI）在各种场景中的效用。在临床方面，聊天机器人可以促进医疗信息检索、分诊、患者咨询、诊断培训和行政责任等任务，特别是对于初级医生或接受培训的临床医生，并评估了他们翻译和简化复杂医学主题的能力。另外，出于研究目的，研究还评估了聊天机器人提出的想法、准备协议、协助手稿写作、指导统计分析和推荐合适期刊的能力。在教育领域，对聊天机器人进行了简化复杂主题、审查程序步骤、生成临床场景和制定多项选择题等多项测试。然后，研究还比较了医疗保健领域中其他几个 AI 聊天机器人高级平台——ChatGPT、Google Bard、Microsoft Bing、Ada Health、Babylon Health 等，但是，大多数研究还都集中在易于访问的 ChatGPT。最后得出结论：聊天机器人若得到最佳使用，可以成为医疗保健中临床、研究和教育等不同领域的宝贵资源，其潜力在于提高效率、指导决策以及促进患者护理和教育，但是在应用时还需要多了解一些其局限性，并谨慎使用。

2. 标题：医学研究人员对科研中使用人工智能聊天机器人的态度与看法

来源：Lancet Digit Health.

时间：2024 年 12 月.

链接：[https://doi.org/10.1016/S2589-7500\(24\)00202-4](https://doi.org/10.1016/S2589-7500(24)00202-4).

概要：聊天机器人是计算机或计算机控制的机器人执行通常与人类智能相关的任务并整合认知功能的能力。当前的翻译语言模型通常使用递归或卷积神经网络的复杂系统。在医学和保健领域，人工智能系统可以减少诊断和治疗错误，促进和增加身体活动，还可以降低医疗保健成本。人工智能也成为科学研究中越来越重要的工具，其中需要检索、分析和解释大量数据，可以提高科学研究的效率、准确性和可重复性，然而人工智能聊天机器人在科学研究中的使用仍存在需要解决的挑战和局限性，比如在研究中使用 AI 相关的道德问题，其中包括剽窃、放大偏倚和不准确的风险、研究欺诈、版权问题和内容生成的低透明度。这些挑战可能会导致 AI 聊天机器人传播错误信息，从而产生有害后果。为了更好地了解医学研究人员对在科学过程中使用 AI 聊天机器人的态度和看法，本研究进行了一项大规模的国际横断面调查，旨在调查研究人员对 AI 聊天机器人的熟悉程度、在科学研究中使用这些 AI 聊天机器人的潜在好处和局限性，以及可能影响其采用的因素。研究在线调查在 2023 年 7 月 9 日至 8 月 11 日开放，从 PubMed 索引的 122323 篇文章中确定了 61560 名通讯作者。2452 人（4.0%）提供了回复，2292 人中有 2165 人（94.5%）完成了调查。在 2137 名受访者中，只有 244 名（11.4%）报告了有关人工智能工具的机构培训，而在 2131 名受访者中，有 211 名（9.9%）指出了有关人工智能聊天机器人使用的机构政策。尽管对益处的看法不一，但 2048 人中有 1428 人（69.7%）表示对进一步培训感兴趣。尽管许多人认为人工智能聊天机器人可以减少行政工作量（1941 年的 1299[66.9%]），但对决策过程的理解是不足的（1923 年的 1484[77.2%]）。总的来说，这项研究突出了研究人员对人工智能聊天机器人的浓厚兴趣，为研究人员在科学环境中对 AI 聊天机器人的看法和使用提供了有价值的见解，但也指出存在一些明显的挑战，比如缺乏正式的培训机会、AI 聊天机器人使用管理不足、加强支持结构和教育举措的需求迫切等。通过解决这些挑战，研究人员和机构可以充分利用 AI 聊天机器人的潜力来推进学术工作并推动科学研究的创新。

（徐健编辑）

译文一：

人工智能驱动的聊天机器人在医疗保健中的效用和局限性

Jafar Hayat, Mohammad Lari, Mohammad AlHerz, Ali Lari, 徐健（译）

介绍

在人工智能（AI）和医疗保健的交叉点上，提供商必须掌握聊天机器人当前的变革能力。这些工具可以在广泛的场景中生成特定于上下文的解决方案，从而增强日常工作流程和效率。本文探讨了聊天机器人在各种临床、学术和管理职责中的多种应用，并提出这些 AI 工具如何在适当的提示下演变为医疗保健领域的宝贵资源的观点。

材料和方法

本研究旨在全面了解医疗保健领域 AI 聊天机器人的当前知识和实践，所以提问和问题是向免费的 AI 聊天机器人 ChatGPT（OpenAI）的 3.5 版本提出。从聊天机器人获得的第一个回复被记录下来，并在下面进行讨论。本研究的主要结果是评估和讨论 ChatGPT 在结果图中突出显示的几种临床场景中作用，其次是按几个子标题分层评测和讨论聊天机器人在一般医疗保健各个方面的作用。之后描述性讨论了研究结果，并进行了分类。这项观察性研究没有使用对照研究，但将 ChatGPT 提供的干预措施与现有文献中的当前护理标准进行了比较。

文献综述使用包括“聊天机器人”、“语言模型”、“人工智能”和“ChatGPT”等在内的关键字，结合“研究”、“临床”、“教育”、“分诊”、“检查”、“管理”、“翻译”、“文档”和“医疗保健”等医学术语，对 Medline、Embase 和 Web of Science 在内的相关数据库进行了全面检索，对聊天机器人的应用进行初步调查，以确定语言模型最近应用的共同主题和趋势，例如 OpenAI 的 GPT 系列、Google 的 BERT 和 Claude。因此，选择标准为与人工智能聊天机器人在医疗保健中的使用直接相关的文章，无论是在患者护理、通过患者调度提高管理效率、简化数据收集或解释、计费、资源分配、质量改进还是为医疗专业人员提供教育支持。

结果

AI 驱动的聊天机器人在临床实践和研究中表现出了巨大实用性。在临床环境中，聊天机器人通过提供结构化的病史采集框架（图 1）并根据症状建议鉴别诊断来协助分诊和患者咨询。

```
I am the emergency physician on-call covering the triage; I have a 50-year-old female patient that is hypertensive presenting with 3 days history of progressive shortness of breath. Can you briefly list pertinent points in the history and examination along with an appropriate differential diagnosis. Be concise and use bullet-points.
```

Absolutely, here are key points to cover during your history, examination, and the differential diagnosis.

History:

- Characterize the shortness of breath: Onset, duration, progression, any associated symptoms such as chest pain, cough, wheezing, or fever.
- Cardiac history: History of hypertension, other cardiovascular diseases, dyslipidemia, or diabetes mellitus.
- Respiratory history: History of asthma, COPD, pulmonary fibrosis, or other lung diseases.
- Social history: Smoking status, occupational exposure, recent travel.
- Family history: cardiovascular diseases, asthma, or other respiratory diseases.
- Review of systems: Recent weight gain, leg swelling (heart failure), palpitations (arrhythmia), anxiety or panic attacks.

Physical Examination:

- Vital signs: Blood pressure, heart rate, respiratory rate, oxygen saturation, temperature.
- Respiratory: Inspect for use of accessory muscles, cyanosis, audible wheezes, percuss and auscultate the lungs.
- Cardiovascular: Inspect and palpate for jugular venous distension, palpate the pulse, auscultate the heart (murmurs, extra heart sounds).
- Lower extremities: Look for edema which might suggest heart failure.
- Neurological: Check for any signs of stroke.
- Abdomen: Hepatojugular reflux and hepatic enlargement (congestive heart failure).

Differential Diagnosis:

- Acute decompensated heart failure: This may present with hypertension, especially in a patient with known history of hypertension.
- Acute coronary syndrome: This might present with dyspnea, especially in a hypertensive patient.
- Pulmonary embolism: Particularly if there is a history of recent immobilization, surgery, or hypercoagulable state.
- Pneumonia: Especially if there are signs of infection such as fever or productive cough.
- Chronic obstructive pulmonary disease (COPD) or Asthma exacerbation: If the patient has a history of either.
- Anxiety or panic disorder: May present with shortness of breath, but usually in absence of other significant physical findings.

图 1. 来自 ChatGPT 的提问和回复，建议病史和检查中的相关要点

他们推荐了如图 2 的相关调查，并通过生成清晰、无行话的表格来促进患者同意流程，帮助患者更有效地了解程序、风险和益处（图 3）。

Can you list the important tests, laboratory tests, and radiologic investigations?

Sure, based on your patient's presentation, the following tests and investigations could be considered:

Bedside Tests:

- **ECG:** To assess for ischemia, infarction, arrhythmias, or signs of pulmonary hypertension.
- **Pulse Oximetry:** To check the oxygen saturation level.
- **Spirometry (if available and patient is able):** Can aid in diagnosing obstructive diseases like asthma or COPD.

Laboratory Tests:

- **Complete Blood Count (CBC):** To check for signs of infection, anemia, or polycythemia.
- **Cardiac Enzymes (Troponin I or T):** To rule out myocardial infarction.
- **B-type Natriuretic Peptide (BNP):** This can help differentiate cardiac vs. non-cardiac causes of dyspnea.
- **D-Dimer:** Can be used to rule out pulmonary embolism in low-risk patients.
- **Arterial Blood Gas (ABG):** To assess for hypoxia or hypercapnia and acid-base status.
- **Basic Metabolic Panel:** To assess electrolytes, kidney function, and glucose level.
- **Liver Function Tests (LFTs):** If suspicion of right-sided heart failure (for assessing hepatic congestion).
- **Thyroid Function Tests:** If suspicion of hyperthyroidism.

Radiological Investigations:

- **Chest X-Ray:** To identify signs of heart failure, pneumonia, pneumothorax, or other lung diseases.
- **CT Pulmonary Angiogram:** If high suspicion for pulmonary embolism and patient is stable.
- **Echocardiogram:** To assess cardiac function, wall motion, and valvular disease. Especially useful if suspecting heart failure.
- **V/Q scan:** Alternative to CT pulmonary angiogram in certain cases for diagnosis of pulmonary embolism.

Remember, the choice of investigations should be guided by clinical judgement, and the patient's history and examination findings. It's also important to consider the patient's preferences and overall health status before ordering investigations.

图 2. 询问同一患者的相关检查的随访提问

I am the surgical trainee on-call, a patient of mine has been diagnosed with appendicitis and my consultant has asked me to take consent for a laparoscopic appendectomy. Can you formulate a consent form that includes all possible complications?

Patient Consent

1. **Understanding the procedure:** I understand that a laparoscopic appendectomy involves removing my appendix using small incisions and special surgical instruments. The surgeon will inflate my abdomen with gas and insert a laparoscope to visualize the surgical area.
2. **Benefits:** The main benefit of this procedure is to relieve my symptoms and prevent potential life-threatening complications of appendicitis. Laparoscopic surgery may lead to a faster recovery, less post-operative pain, and a shorter hospital stay compared to open surgery.
3. **Risks and Complications:** I understand that every surgical procedure has potential risks and complications. These can include but are not limited to:
 - General anesthesia complications (<1%).
 - Pain, bruising, and swelling at the incision sites (common).
 - Infection (2-3%).
 - Bleeding, which may require a blood transfusion or additional surgery (<1%).
 - Damage to surrounding structures including bowel, bladder, blood vessels, and nerves (<1%).
 - Development of a hernia at the incision site (1-5%).
 - Conversion to an open procedure if the laparoscopic approach is not safe or feasible (2-4%).
 - Formation of blood clots (deep vein thrombosis) that can migrate to the lungs (pulmonary embolism) (<1%).
 - Post-operative ileus or temporary pause in bowel function (common, but usually resolves within a few days).
 - Risks related to the appendicitis itself, such as abscess or peritonitis, if the appendix has ruptured (10-20% if appendix has ruptured).
 - Complications related to potential undiagnosed conditions, like appendiceal tumors (<1%).
4. **Alternatives:** I have been informed about alternative options, including not having the operation, antibiotic treatment, or open appendectomy, and I understand their potential risks and benefits.
5. **Consent:** I voluntarily give my consent to undergo a laparoscopic appendectomy. I understand the procedure and its risks, and my questions have been satisfactorily answered.

图 3. 提示询问有关患者同意腹腔镜阑尾切除术的详细信息

聊天机器人进一步简化了不同人群的医学语言，帮助患者理解诊断和治疗，尤其是对于不同年龄和教育水平的个体。此外，聊天机器人通过协助记录、听写和确保准确的记录保存来减轻管理负担。

图 4. 是一个提问，产生一个回复，使用隐喻和通俗的术语来轻松地与孩子解释克罗恩病。

I am a practicing doctor, I have to explain to a 5th grader their recent diagnosis of Crohn's disease. Can you explain this disease in a way that they can understand it, using simpler lay terms.

Absolutely! Let's imagine that your body is like a big, long tunnel where the food you eat goes on a journey. This tunnel is called your digestive system. It starts at your mouth, goes down through your stomach and intestines, and ends at the place where you poop.

Now, in some people, parts of this tunnel can sometimes get angry and swollen, kind of like how your skin turns red and puffy if you get a bug bite. This is what's happening inside your body when you have something called Crohn's disease.

Crohn's disease usually gets upset in the last part of your tunnel, the part right before where you poop, but it can happen anywhere along the way. When it's upset, you might feel pain in your tummy, you might not feel like eating, and you might need to use the bathroom a lot. Sometimes, you might even feel very tired.

The tricky thing about Crohn's is that it likes to play hide and seek. It can make you feel yucky for a while and then disappear, making you feel all better. But then, it can come back again. Doctors call these times when you feel sick "flare-ups", and times when you feel good "remission".

The good news is, we have lots of tools to help manage Crohn's disease. You'll have a team of doctors and nurses helping you, and they might give you some special medicine to keep the tunnel calm and less upset. You may also need to eat certain foods that are easier for your tunnel to handle.

And remember, even though Crohn's can be a big nuisance, it doesn't stop you from doing most of the things you love. With the right care and support, you can still go to school, play sports, and spend time with friends. You're just as special and capable as everyone else - you just have a tunnel that needs a little extra care and attention.

图 4. 使用隐喻和通俗的术语来轻松地与孩子解释克罗恩病

在研究中，聊天机器人通过识别潜在的研究领域并指导临床医生就热门主题集思广益进行系统综述来支持想法的产生。他们协助方案准备，使用预定义的模板生成全面的研究方案，并通过改进结构和纠正语言错误来改进手稿写作。聊天机器人还指导研究人员进行统计分析和功效采样，但由于可能存在不准确之处，因此必须谨慎（图 5）。此外，聊天机器人通过将手稿内容与相关期刊相匹配来提供期刊推荐（图 6）。

Can you perform a rough power sample analysis for my study:
An RCT comparing postoperative pain after procedure A and B.
At a 95% confidence level, 0.8 effect size/power at a 0.05 significance level.

To calculate the sample size for a randomized controlled trial (RCT), the effect size, power, and significance level are crucial parameters. You mentioned that the effect size (Cohen's d) is 0.8 (large effect), the power ($1 - \beta$) is 0.80 (80% chance of detecting an effect if there is one), and the significance level (α) is 0.05 (5% chance of falsely rejecting the null hypothesis).

Assuming equal sample sizes in each group, a two-sided test, and using the parameters you provided, the total sample size required for this study is approximately 52 subjects, with 26 in each group (Procedure A and Procedure B).

This estimate is a general guideline. The actual sample size you need might be larger to account for potential dropouts or other sources of variability.

For a precise calculation and to adjust for other factors such as loss to follow-up, it's recommended to consult with a statistician. Also, different statistical software packages might provide slightly different estimates due to variations in the algorithms they use to calculate sample size.

图 5. 聊天机器人执行的简短功效分析

如果信息缺失，聊天机器人通常会指导用户进行填写。

I have completed a randomized controlled trial comparing the outcomes of antibiotic coated nails versus external circular frames for the treatment of infected non-unions of the tibia.

Can you recommend 5 appropriate journals to submit to? Two of them should be open access.

Absolutely, based on the specifics of your study, here are five journals that might be suitable for your submission:

1. **Journal of Bone and Joint Surgery (JBJS):** A highly respected, general orthopedic journal that frequently publishes clinical trials. Its broad readership could give your study a wide impact.
2. **Clinical Orthopaedics and Related Research (CORR):** This journal is a premiere location for high-quality orthopedic research. It's well known for publishing in-depth clinical studies and trials.
3. **Bone & Joint Journal (BJJ):** This is another highly respected journal in the field of orthopedic surgery and is recognized for its high standard of research publication.

Open Access Journals:

1. **Journal of Orthopaedic Surgery and Research (JOSR):** An open-access, peer-reviewed journal that is known for publishing high-quality clinical research in the field of orthopedics.
2. **BMC Musculoskeletal Disorders:** A reputable open-access journal that covers all aspects of the prevention, diagnosis, and management of musculoskeletal and associated disorders.

Remember, each journal has its own submission guidelines and criteria that you will need to thoroughly review before submitting your work. Additionally, be aware that open-access journals typically require an article processing charge (APC).

图 6. 根据提供给聊天机器人的简要信息对日记帐的建议

在教育领域里，聊天机器人有效地总结了药物，包括它们的机制、剂量和潜在的相互作用（图 7）。他们还审查了程序，提供了详细的分步解释，以帮助初学者和有经验的临床医生。聊天机器人还进一步生成了临床场景和多项选择题（MCQ），增强了交互式学习和自我评估。

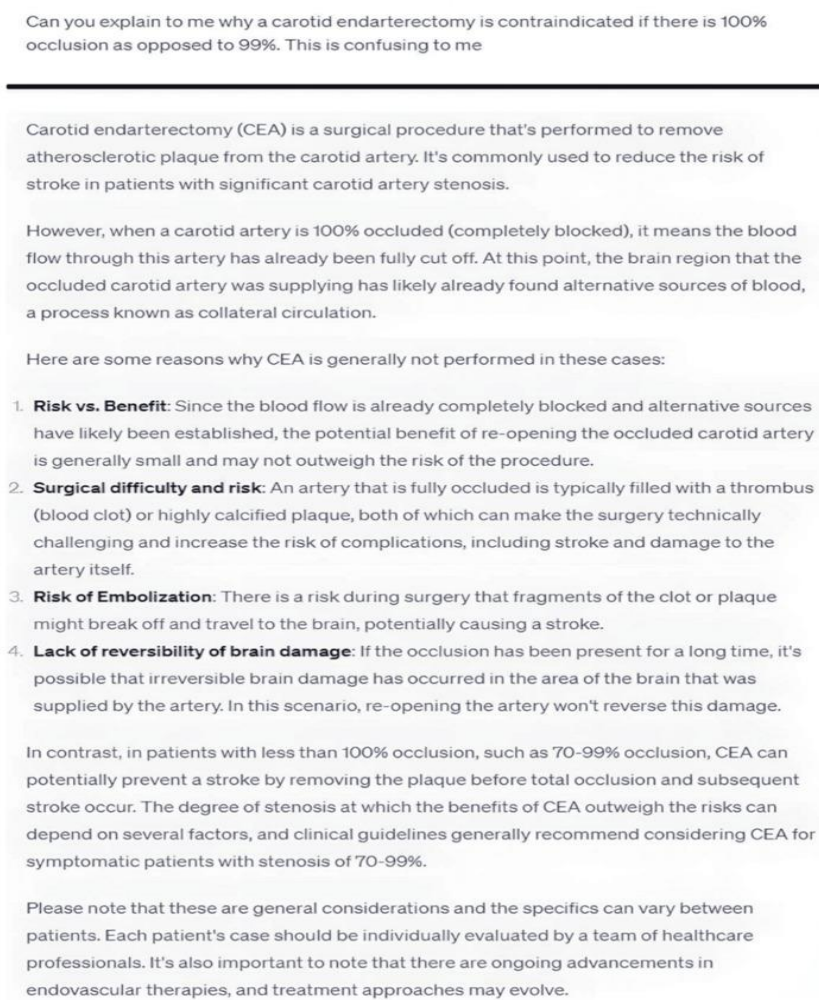


图 7. 聊天机器人解释为何颈动脉内膜切除术在一定程度上的闭塞是禁忌

虽然聊天机器人提供了宝贵的支持，但它们偶尔会生成不准确的信息，因此需要人工监督关键决策。

讨论

什么是聊天机器人？

聊天机器人是模拟类似人类对话的程序。他们利用 AI 技术、自然语言处理（NLP）和机器学习（ML）来评估用户输入并提供适当的输出（响应）。当聊天机器人由 AI

提供支持时，它能够处理复杂的对话，并且可以从不同的用户交互中学习。这使他们能够理解各种类型的非结构化输入。这与早期基于规则的系统形成鲜明对比，后者只有在使用某种结构化输入时才能响应。基于语音的聊天机器人代表 AI 驱动的聊天机器人，具有基于语音的交互功能，无需打字即可进行更自然的对话。

如何工作？

虽然聊天机器人有一定的局限性（诊断能力有限、缺乏同理心和人情味、无法进行和评估身体检查以及技术限制），但用户可以通过理解聊天机器人对话的算法流程来实现最佳响应。用户通过提供输入（称为提问）来启动交互。此提问用作主查询，并为聊天机器人的输出建立上下文。请务必注意，聊天机器人对话是迭代的，涉及多次交互。这意味着通过多个提示构建上下文可以产生所需的响应。这引起了人们对所谓的“提示工程”的关注，它指的是改进指令或提示，以有效地从聊天机器人那里获得各种响应。此外，聊天机器人可以进行微调，以根据特定的数据集和数据域生成响应，例如将医疗数据整合到其数据库中。通过从专业数据中提取响应，聊天机器人确保其响应与当前的医疗信息保持一致。

临床效用

在临床环境中，聊天机器人是非常宝贵的工具，为临床医生提供了大量优势。这些虚拟助手类似于高级搜索引擎，可以无缝检索医疗信息、提出治疗算法、协助患者咨询并减轻管理任务的负担。特别是对于初级医生或接受培训的临床医生，聊天机器人可以创建综合表格，以简洁地概述各种检查的利弊，指导在患者检查中选择最佳的下一步，并提出不同的治疗算法。本文虽然简短，但将探讨聊天机器人的一些基本应用，展示它们在医疗保健领域的变革潜力。

分诊和初始患者咨询，随叫随到的助理

在个人层面，临床医生可以利用聊天机器人来分析患者症状并促进病史采集过程。这些聊天机器人通过使用自然语言处理、上下文分析和症状分类、模式识别和与医学数据库的匹配，以及通过临床指南和最佳实践的整合，智能地检查症状并为临床医生提供一个结构化的框架，用于进行全面的患者病史和检查。通过参考潜在

鉴别诊断列表，聊天机器人会提出相关问题，以确保不会忽视重要细节（图 1）。此交互式过程允许临床医生与聊天机器人进行持续对话，整合新信息并提示聊天机器人提供更详细的响应。这种协作方法将临床医生的专业知识与聊天机器人的分析能力相结合。然而，在某些情况下，它在这个阶段在紧急分诊设置中的应用可能会受到限制。

协助诊断和提供一般检查

医疗保健专业人员可以使用聊天机器人来协助诊断过程和随后的患者检查。通过动态交互，临床医生可以利用聊天机器人对及时信息的访问来得出量身定制的鉴别诊断。此外，临床医生可以提供更新的患者数据，促使聊天机器人建议适当的诊断并推荐相关检查。例如，如果患者出现呼吸道症状，聊天机器人可以提供相关问题和潜在鉴别诊断的列表。根据要求，聊天机器人可以提出调查建议，以帮助缩小诊断范围并排除更严重的情况（图 2）。这个动态过程有助于临床医生构建他们的方法，并确保在初始患者咨询期间进行全面评估。

患者同意大纲

获得患者同意是临床医生日常工作的一个重要方面，尤其是对于初级医疗保健专业人员而言。在这方面，聊天机器人可以通过协助制定使用清晰语言和避免医学术语的同意书来发挥支持作用（图 3）。这些聊天机器人能够搜索与特定手术相关的信息，包括其风险、益处和替代方案，并生成患者易于理解的简洁段落。此功能有两个显著的好处。首先，它帮助临床医生使用通俗易懂的语言有效地传达信息，确保患者理解手术的重要内容。其次，患者自己可以在交互式对话中与聊天机器人互动，使他们能够积极参与同意过程。此外，聊天机器人可以轻松地为不同的程序创建量身定制的同意书，为临床医生和患者节省时间和精力。

行政负担

医疗保健专业人员的日常工作可能要求很高，特别是由于管理任务和临床文档。聊天机器人可以充当虚拟助手，协助听写并减少手动输入数据所需的时间和精力。它们还提供了语法和语言错误检查的好处，确保了准确和结构良好的文档。最近对聊天机器人编写出院小结和手术记录的能力进行了评估，产生了令人鼓舞的结果。此外，可以通过语音命令提示聊天机器人以获取药物剂量等信息或澄清令人困惑的话题。

解释和翻译

聊天机器人能够使用更简单的语言来适应不同的年龄和教育水平。当试图用更简单的术语解释诊断、治疗、药物和预后时，这很有用（图 4）。此外，聊天机器人是高效的翻译者，能够以清晰简洁的方式将不同的文本翻译成许多不同的语言。

研究实用程序

AI 驱动的语言模型的实用性近年来引起了广泛关注。很明显，聊天机器人将在未来的研究中发挥重要作用，尤其是考虑到语言模型的快速发展。然而，研究人员对这些语言模型的可靠性和在研究中的潜在滥用存在许多合理的担忧。虽然聊天机器人可以通过多种方式协助更大规模的研究，但本文将主要侧重于讨论在个人层面采用的方法选择。

提出想法

聊天机器人可以帮助临床医生确定医学感兴趣的领域。通过与聊天机器人互动，临床医生可以集思广益并发现需要进一步研究的潜在研究领域，例如对当前趋势主题进行系统评价。但是，值得注意的是，最新的在线医学数据库的可用性可能会限制此功能的范围，从而可能限制它作为灵感来源。

准备协议

聊天机器人在帮助准备研究方案方面很有价值。通过使用预定义的模板，临床医生可以输入他们的研究思路和方法，从而允许聊天机器人生成全面的协议。此功能对于确保完整性和节省手动数据输入所花费的时间特别有用。

研究写作和语言

聊天机器人可以作为构建和格式化手稿的有用指南。他们可以协助编辑文档的整体布局，同时识别和纠正语言和语法错误。他们在语言处理方面的效率使他们能够就语言相关问题提供及时的反馈。还可以要求聊天机器人以不同的写作风格构建语言。然而，他们在保持上下文的同时准确撰写整篇手稿的能力似乎不足。

统计分析和功效抽样

聊天机器人可以提供有关各种统计分析技术的指导，以更简单的术语提供解释以更好地理解。它们可以帮助研究人员根据研究设计选择合适的分析检验，并提供有关一般统计查询的信息。此外，聊天机器人能够在获得必要信息时进行功效分析，帮助研究人员确定适合其研究的样本量（图 5）。然而，有报道称发电样本不准确。

因此，应该有一位具有统计经验的人参与监督聊天机器人的分析建议。

期刊推荐

聊天机器人可以通过分析各种期刊的范围并将其与文章的标题和前提相匹配，帮助研究人员为他们的文稿找到合适的期刊（图 6）。通过考虑研究的内容和重点，聊天机器人可以提供有关与文稿非常吻合的期刊的建议。此外，他们就可以搜索和识别适合研究主题和方法的期刊的有效策略提供指导。

学习和教育实用程序

聊天机器人可以帮助学习和查看信息，特别是对于正在接受培训的医学生和临床医生。通常，用户依靠标准搜索引擎来收集信息，但这可能会让人不知所措，因为他们需要搜索多个来源才能找到简洁的摘要。相比之下，聊天机器人能够提供简短易懂的个性化回复，使用户的学习过程更加高效和有效。

简化复杂的主题可以节省时间和精力，同时还可能有助于信息保留。聊天机器人在提供各种药物的快速摘要方面发挥着至关重要的作用，包括它们的剂量、作用机制和安全性，即使在考虑与其他药物的潜在相互作用时也是如此。这使得聊天机器人成为快速提取简洁信息的绝佳工具（图 7）。

此外，聊天机器人可以帮助审查程序和手术中涉及的步骤。他们可以详细解释每个步骤及其背后的根本原因。事实证明，此功能对初学者特别有用，因为它可以帮助他们理解基本概念并复习程序。同样，经验丰富的临床医生也可以从聊天机器人中受益，使用它们来查看他们已经熟悉的程序中涉及的步骤。

聊天机器人能够制定广泛的临床问题和患者场景，使其成为临床医生和学生的宝贵资源。通过生成跨各种主题的真实场景，聊天机器人模拟了现实世界的遭遇，并为用户提供了在实际环境中应用其知识的机会。这些场景可以伴随着相关问题，鼓励用户批判性思考并提供准确的回答。

此外，聊天机器人可以生成不同难度级别的多项选择题（MCQ）。此功能使用户能够通过自我测试来评估他们的理解并强化关键概念。通过提供各种临床问题和患者场景，聊天机器人可以作为一种交互式学习工具，可以提高临床推理和知识保留。

局限性

虽然 AI 驱动的聊天机器人提供了便利和令人印象深刻的语言理解能力，但重要

的是要了解它们的局限性。聊天机器人主要用作高级搜索引擎，在复杂级别 1 解释语言。然而，他们偶尔会误解医学查询，缺乏对可能导致不准确回答的细微差别的完全理解。这个值得注意的问题被称为“幻觉”，涉及这些人工智能系统产生的信息可能不基于实际输入数据，可能导致误导性或错误的结论。临床医生在仅依赖聊天机器人生成的答案时应谨慎行事。此外，重要的是要注意，聊天机器人无法实时访问最新的医学文献或复杂的在线医学图书馆。因此，他们的回答可能并不总是反映可用的最新信息。但是，此缺点将来可能会得到解决。本叙述性综述的另一个局限性是本文涵盖的回复范围狭窄，这为 AI 生成的结果与其他临床场景留下了差异的空间。在开发技术时应始终考虑错误风险，重要的是不要只为临床医生和患者依赖此类技术。

为了保持平衡的方法，医疗保健专业人员应该使用聊天机器人作为补充工具，而不是仅仅依赖它们。临床医生的专业知识和临床判断不应被聊天机器人的功能所掩盖。通过认识到这些限制，医疗保健专业人员可以有效地利用聊天机器人来增强他们的实践，而不会影响患者护理。同样，虽然聊天机器人在研究中的效用很有希望，但解决它们的局限性和担忧至关重要。在研究中使用聊天机器人需要考虑伦理问题，虽然目前在这个问题上还没有达成共识，但许多期刊要求聊天机器人的使用透明。为了减少聊天机器人的不受监管的使用，存在聊天机器人检测程序，但其准确性尚不确定。

此外，聊天机器人依赖于某个时间点的预先存在的数据，这意味着如果没有软件扩展，它们就无法访问最新证据。最后，研究人员必须独立验证聊天机器人提取的数据，并避免可能导致不准确的常规不受监管的使用。

Chatbot 扩展和其他聊天机器人

为了解决聊天机器人的当前限制，各种互联网浏览器软件提供了增强其功能的功能。特定的软件扩展可以使 ChatGPT 等聊天机器人访问反映各个领域最新发展的实时数据和最新信息。此外，还有一些可用的扩展可以生成文章或视频的简明摘要，将它们浓缩成更小的段落或要点。此外，通过扩展将聊天机器人与 Web 浏览器集成，允许将标准搜索引擎与 AI 驱动的语言理解集成。这些扩展可以增强聊天机器人的功能和性能。然而，即使没有扩展，较新版本的 ChatGPT 也可能会添加额外的功能来

增强他们的在线推广（GPT-4）。

医疗保健领域的 AI 聊天机器人领域具有多个高级平台。本研究中讨论了 ChatGPT（OpenAI Incorporated, 美国旧金山 Mission District）、Google Bard（Alphabet Inc., 美国加利福尼亚州）和 Microsoft Bing（Microsoft Corporation, 美国华盛顿州）。Google 的 Bard 专为生成上下文丰富的响应而量身定制，可能会提高患者在医疗保健互动中的参与度和理解力。Microsoft 的 Bing 搜索功能提供了广泛的当前医学知识，有助于最新的临床决策。然而，直接比较很少，他们在考试中的表现结果相似。在针对特定患者的医疗保健中，Ada Health 和 Babylon Health 经常被引用。Ada Health 的优势在于它能够使用庞大的医学数据库进行详细的症状分析并指导初步患者评估。另一方面，Babylon Health 专注于虚拟医疗保健服务，提供远程咨询和个性化健康建议，这与远程医疗的趋势相一致。尽管如此，大多数表面研究都集中在 ChatGPT 上，这可能是由于易于访问。

结论

探索像 ChatGPT 这样的 AI 驱动的语言聊天机器人强调了它们在医疗保健和研究方面的潜力。这些数字工具可以协助临床工作、减轻行政负担并增强教育平台。然而，采用这种技术需要一个平衡的观点，承认它们目前的局限性，尤其是在需要创造性方法来解决问题的特定情况下。展望未来，聊天机器人的发展可能会使它们成为个性化医疗保健不可或缺的一部分，其进步可能使这些工具能够提供更细致的、针对患者的见解。因此，使用聊天机器人的知识和熟练程度可能会为医疗保健提供者的日常工作提供优势。AI 语言模型开发的快速速度和轨迹表明，聊天机器人可能很快就会超越目前作为医疗保健辅助工具的角色。尽管人工智能技术在医疗保健领域的广泛实施存在重大障碍，但聊天机器人已经证明以各种方式对医疗保健提供者有用。显然，遵循以前提高医疗保健效率的数字创新的轨迹，聊天机器人前景广阔。尽管存在障碍和一些阻力，但谨慎实施聊天机器人以确保安全并避免潜在陷阱至关重要。

***注：原文和译文版权分属作者和译者所有，若转载、引用或发表，请标明出处。**

译文二：

医学研究人员对科研中使用人工智能聊天机器人的态度与看法：一项国际横断面调查

Jeremy Y Ng, Sharleen G Maduranayagam, Nirekah Suthakar, Amy Li,
Cynthia Lokker, Alfonso Iorio, R Brian Haynes, David Mohe, 徐健 (译)

介绍

人工智能 (AI) 广义上是指计算机或计算机控制的机器人执行通常与人类智能相关的任务并整合认知功能的能力。当前的翻译语言模型通常使用递归或卷积神经网络的复杂系统，该系统由两部分组成：编码器和解码器。最好的模型还使用关注机制来连接这些部分。在医学和保健领域，人工智能系统减少了诊断和治疗错误，促进和增加了身体活动，据估计还降低了医疗保健成本。人工智能也成为科学研究中越来越重要的工具，其中需要检索、分析和解释大量数据。

聊天机器人是 AI 程序，旨在使用文本或语音模拟与人类用户的对话。流行的 AI 聊天机器人包括但不限于 ChatGPT、Bing Chat、YouChat 和 Google Bard。在科学研究中，AI 聊天机器人可以自动执行文献搜索和评论、数据分析和大型数据集解释、实验设计和手稿写作等任务。它们还提高了非母语人士的科学文章的可读性，可能增加公平性并解决促进研究多样性的障碍。还可以训练 AI 聊天机器人通过估计研究的复制可能性来区分可重复和不可重复的研究。总体而言，AI 聊天机器人提高了科学研究的效率、准确性和可重复性。¹³

然而，人工智能聊天机器人在科学研究中的使用存在需要解决的挑战和局限性。一些主要问题包括 AI 聊天机器人在执行或报告科学任务方面的准确性和可靠性，以及与在研究中使用 AI 相关的道德问题，其中包括剽窃、放大偏倚和不准确的风险、研究欺诈、版权问题和内容生成的低透明度。这些挑战可能会导致 AI 聊天机器人传播错误信息，从而产生有害后果。

为了更好地了解医学研究人员对在科学过程中使用 AI 聊天机器人的态度和看法，我们进行了一项大规模的国际横断面调查。这项调查旨在调查研究人员对 AI 聊天机器人的熟悉程度、在科学研究中使用这些 AI 聊天机器人的潜在好处和局限性，

以及可能影响其采用的因素。

方法

检索策略和选择标准

我们获得了 MEDLINE 中索引的所有期刊的完整列表（截至 2023 年 4 月约 5300 种期刊）及其国家医学图书馆（NLM）ID。我们在 Ovid MEDLINE 中开发了这些 NLM 标识符的检索策略，该策略仅限于 2023 年 3 月 1 日至 4 月 7 日编入索引的记录。与所有产生文章相关的 PubMed 标识符号从 Ovid 导出为 .csv 文件并输入到 R 脚本（基于 easy PubMed 包，2019 创建）中，以捕获作者姓名、附属机构和电子邮件地址。检索了 122323 个 PubMed ID 并将其导出到 R 中。通讯作者电子邮件提取于 2023 年 5 月 29 日完成，并在调查管理前清理了潜在错误或重复项。该协议已在开放科学框架上注册，可在线获取。

受访者招募

受访者必须自我认同为医学研究人员，并在各自的研究领域完成至少一个最终学位，或具有至少 5 年以研究为重点的角色经验（例如，研究协调员）。学生（本科生和研究生）没有资格参加这项调查。

包括医学领域的一系列学科（例如，生命科学、公共卫生、生物工程和医学教育）。我们使用便利抽样来招募受访者，针对我们的抽样策略确定的通讯作者。一封电子邮件，其中包含研究伦理委员会批准的招募脚本，详细说明了研究及其目的，并使用 Microsoft Outlook Mail Merge 软件通过电子邮件发送了调查链接于 2023 年 7 月 9 日。当受访者点击调查链接时，他们被定向到一个包含知情同意书的网页。同意该表格的受访者将随后继续进行调查。此调查已关闭，这意味着只有受邀的受访者才能参与。

如果受访者没有回复原始邀请电子邮件，则会在 2023 年 7 月 17 日至 7 月 24 日期间分批发送两次提醒电子邮件，每批 10000 封。在最后一封电子邮件提醒之后有 1 周的等待期。在线调查于 2023 年 7 月 9 日至 8 月 11 日开放。

没有提供经济补偿；参与完全是自愿和匿名的。除筛选问题外，如果参与者不想

回答，可以跳过所有其他问题。在调查期间，可以随时通过关闭浏览器来撤回参与，但是在提交调查后就无法撤回。

数据分析

该调查是使用 Survey Monkey 软件（加拿大安大略省渥太华大学）创建、分发和收集的。该调查的初稿（由 JYN 创建并由团队其他成员审查）是根据 AI 和科学研究领域专家的文献综述和意见制定的。在进行调查之前，调查问题由一组作者和两名具有作者列表之外的调查方法专业知识的受邀研究人员进行了 beta 测试。

该调查包含 29 个封闭式问题，涵盖人口统计信息、AI 聊天机器人的经验、AI 聊天机器人在科学过程中的作用、AI 聊天机器人在科学过程中的感知优势和挑战，以及开放式问题。

使用 Microsoft Excel 导出和分析调查数据。使用计数和百分比来总结收集的数据。对从开放式问题中收集的定性数据进行主题分析，并事先进行试点编码（由 AL 和 NS 进行）。每位作者都独立编码了前 20 名受访者对问题 14（第一个开放式问题）的回答，然后根据他们的结果合作开发了一个共享代码簿。在对代码达成共识后，对所有剩余的开放式响应进行编码，并迭代更新代码本。各个代码被分组为主题并通过共识最终确定。任何冲突都通过 AL 和 NS 之间的讨论解决，如果需要，由第三位作者（SGM）解决。然后为每个主题创建描述性定义。

本研究是根据 CHERRIES 指南进行的。渥太华健康科学网络研究伦理委员会（20230288 - 01H）授予了伦理批准。所有收集的数据均保持机密和匿名，没有收集受访者的身份信息。

结果

在对名单进行重复处理后，共向 61560 名通讯作者发出调查邀请，其中有 2452 名（4.0%）作者提供了回复。2292 名（93.5%）受访者符合资格标准，其中 2165 名（94.5%）受访者完成了调查。由于并非所有受访者都完成了所有调查问题，因此每个问题的回答总数有所不同。原始收集的调查数据可以在网上找到。

人口统计数据见附录（第 3-4 页）。大多数受访者为男性（2149 人中有 1161 人 [54.0%]），不到一半的受访者为女性（959 人 [44.6%]），29 人（1.3%）被确定为非

二元、自我描述或不愿说。调查对象年龄以 36-45 岁居多(2156 人中有 706 人[32.8%])。受访者主要在 95 个国家工作,其中来自美国的代表最多(2137 人中有 601 人[28.1%])、加拿大(163 人[7.6%])和意大利(129 人[6.0%])。大多数是高级研究员(2165 人中有 1136 人[52.5%])、大学或学术机构的教职员工(2158 人中有 1369 人[63.4%])，主要从事临床研究(2159 人中有 1107 人[51.3%])，发表过 21 篇以上的研究论文(2133 人中有 1549 人[72.4%])。

受访者使用 AI 聊天机器人的经历及其所属机构对在科研过程中使用 AI 聊天机器人的看法见表 1。值得注意的是，2138 名受访者中有 1294 名(60.5%)熟悉 AI 聊天机器人的概念，在使用过 AI 聊天机器人的受访者中，大多数使用 ChatGPT(1402 名[65.5%])。不到一半的受访者将 AI 聊天机器人用于与科学过程相关的目的(2125 名受访者中有 945 人[44.5%])，882 人(41.5%)将 ChatGPT 用于这些目的。然而，1180 名(55.5%)受访者从未在这种情况下使用过人工智能聊天机器人。受访者认为有可能(2136 人中有 790 人[37.0%])或非常有可能(451[21.1%])在未来的研究中使用 AI 聊天机器人。

问题	回复数量 (%)
您对 AI 聊天机器人的概念有多熟悉?	
非常熟悉	283 (13.2%)
熟悉	1294 (60.5%)
陌生	357 (16.7%)
很陌生	169 (7.9%)
我不确定	35 (1.6%)
总	2138 (100.0%)
您以前使用过以下哪些 AI 聊天机器人(用于任何目的)? *	
必应聊天	246 (11.5%)
聊天 GPT	1402 (65.6%)
YouChat	26 (1.2%)
谷歌吟游诗人	206 (9.6%)
我从未使用过 AI 聊天机器人	658 (30.8%)
其他	75 (3.5%)
总	2136 (100.0%)
您之前曾专门将以下哪些 AI 聊天机器人用于与科学过程相关的目的? *	
必应聊天	87 (4.1%)
聊天 GPT	882 (41.5%)
YouChat	9 (0.4%)
谷歌吟游诗人	82 (3.9%)
我从未将 AI 聊天机器人用于与科学过程相关的目的	1180 (55.5%)
其他	47 (2.2%)
总	2125 (100.0%)

问题	回复数量 (%)
您将来使用 AI 聊天机器人进行研究的可能性有多大？	
大半	451 (21.1%)
可能	790 (37.0%)
不可能	348 (16.3%)
不太可能	248 (11.6%)
我不确定	299 (14.0%)
总	2136 (100.0%)
您的研究机构是否提供有关如何在科学过程中适当使用 AI 工具的任何培训？	
是的，我已经接受了	64 (3.0%)
是的，但我没有服用	180 (8.4%)
不	1487 (69.6%)
我不确定	406 (19.0%)
总	2137 (100.0%)
您的研究机构是否实施了有关在科学过程中使用 AI 聊天机器人的任何政策？	
是的	211 (9.9%)
不	1178 (55.3%)
我不确定	742 (34.8%)
总计	2131 (100.0%)

表 1 受访者对 AI 聊天机器人的体验

2165 名受访者完成了调查，但由于并非所有受访者都完成了所有调查问题，因此每个问题的回答总数各不相同。AI=人工智能。

*参与者可以选择所有适用的选项。

大多数受访者报告说，他们的研究机构没有提供任何关于使用人工智能工具的培训（2137 人中有 1487 人[69.6%]），尽管 244 人（11.4%）表示他们的机构提供了关于在科学过程中正确使用人工智能工具的培训，244 人中有 64 人（26.2%）完成了此类培训。当被问及他们的机构是否实施了任何关于在科学过程中使用 AI 聊天机器人的政策时，211 名受访者中有 9 名（9.9%）回答是。

表 2 显示了参与者对 AI 聊天机器人在科学过程中的感知作用的回答。略多于一半的受访者回答说有必要进行一些培训（2049 人的 1048[51.2%]），665 人（32.5%）认为需要进行大量培训才能在科学过程中有效使用 AI 聊天机器人。大多数受访者表示明显有兴趣了解更多信息或接受有关在科学过程中使用 AI 聊天机器人的培训（2048 人中有 1428 人[69.7%]）。受访者还认为，人工智能聊天机器人在未来的科学研究中将非常重要（2050 年的 741[36.2%]）或重要（935[45.6%]）。大多数人表示 AI 聊天机器人将产生非常积极或积极的结果（2048 人中有 1236 人[60.4%]），而 380 人（18.6%）认为 AI 聊天机器人会对未来的科学研究产生负面影响或非常消极。

问题	回复数量 (%)
您认为研究人员需要多少培训和教育才能在科学过程中有效地使用 AI 聊天机器人？	
好多	665 (32.5%)
一些	1048 (51.2%)
半	173 (8.4%)
没有	34 (1.7%)
我不确定	129 (6.3%)
总计	2049 (100.0%)
您是否有兴趣了解更多信息或接受有关如何在科学过程中使用 AI 聊天机器人的培训？	
是	1428 (69.7%)
不	189 (9.2%)
或	431 (21.0%)
总计	2048 (100.0%)
您认为 AI 聊天机器人在未来的科学研究中有多重要？	
非常重要	741 (36.2%)
重要	935 (45.6%)
重要	110 (5.4%)
非常不重要	18 (0.9%)
我不确定	246 (12.0%)
总计	2050 (100.0%)
一般来说，您如何看待 AI 聊天机器人对科学研究未来的潜在影响？	
非常积极	282 (13.8%)
积极	954 (46.6%)
负	290 (14.2%)
非常消极	90 (4.4%)
我不确定	432 (21.1%)
总计	2048 (100.0%)

表 2 AI 聊天机器人在科学过程中的作用

受访者将他们对人工智能聊天机器人在科学过程的不同步骤中有多大帮助的说法的同意程度分为五分制，从非常有帮助到非常没有帮助（附录 p 1）。值得注意的是，2044 名受访者中有 1235 人（60.04%）认为人工智能聊天机器人在进行文献搜索方面非常有帮助或有帮助。受访者对人工智能聊天机器人帮助理解或选择研究方法的完整性持不同意见；2033 人中有 820 人（40.3%）认为人工智能聊天机器人没有帮助或非常没有帮助，而 754 人（37.1%）认为它们会非常有帮助或有帮助。在回答这个问题的参与者中，高级研究人员在使用人工智能聊天机器人时选择“总是”、“经常”和“有时”的频率高于其他职业阶段的研究人员。这些发现可以在网上找到：参与者在科学过程中使用人工智能工具的情况。绝大多数受访者认为，人工智能聊天机器人在撰写或编辑手稿（2040 年的 1329 人[65.2%]）、申请研究经费（2032 年的 1183 人[58.2%]）以及将研究材料翻译成另一种语言（2044 年的 1346 人[65.9%]；在 2039 名受访者中，902 名（44.2%）认为人工智能聊天机器人对同行评审或批评其他

研究人员的工作没有帮助或非常没有帮助。2041 名受访者中有 848 名（41.6%）认为人工智能聊天机器人对生成演示材料非常有帮助，2040 名受访者中有 1059 名（51.9%）认为人工智能聊天机器人对研究参与非常有帮助。最后，在 2040 名受访者中，有 1261 人（61.8%）认为人工智能聊天机器人对每个形成一般管理任务非常有帮助或有帮助。

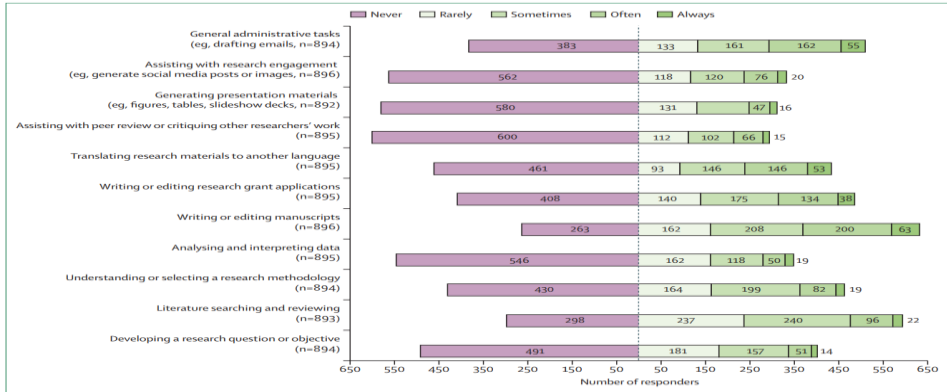


图 1 受访者在科学过程的不同步骤中使用 AI 聊天机器人

受访者对他们在科学过程的不同步骤中与人工智能聊天机器人的好处有关的陈述的同意程度进行了评分，评分范围从非常不同意到非常同意（图 2）。最同意和最同意人工智能聊天机器人的好处包括提高数据分析的效率和速度（1939 年的 1026[52.9%]），减少研究人员的工作量和行政负担（1941 年的 1299[66.9%]）。使科学信息和援助的持续可用性和可及性（1939 年为 1081[55.7%]），比人类研究人员更有效地处理和分析大型数据集（1935 年为 1086[56.1%]），为开展科学研究提供了具有成本效益的解决方案（1933 年为 974[50.4%]），提供了一个更容易获得和包容的研究环境（1936 年为 1120[57.9%]），并提高了科学交流和传播的质量和效率（1932 年为 1002[51.9%]）。

关于人工智能聊天机器人是否可以通过提供标准化和系统化的方法来提高研究的可重复性和透明度的评级好坏参半（图 2）；1936 名参与者中，285 人（14.7%）强烈不同意或不同意（589 人[30.4%]），474 人（24.5%）不确定，456 人（23.6%）同意或强烈同意（132 人[6.8%]）。1937 名参与者中有 565 人（29.2%）不确定 AI 聊天机器人是否可以提高数据分析和实验的准确性和精度，而 243 人（12.6%）强烈反对或不同意（481 人[24.8%]）的说法。对于人工智能聊天机器人是否可以通过提供标准化的数据分析方法来减少人为错误或偏见，受访者的看法不一，1932 年有 530 人（27.4%）不确定，523 人（27.1%）同意，149 人（7.7%）强烈同意；然而，472 名

(24.4%) 受访者表示不同意, 258 名 (13.4%) 强烈反对。总体而言, 大多数受访者 (2050 年的 1676 人[88.1 %]) 认为 AI 聊天机器人在未来的科学研究中将非常重要或重要 (附录 p 2)。



图 2 受访者同意在科学过程中使用 AI 聊天机器人的潜在好处

关于人工智能聊天机器人是否可以通过提供标准化和系统化的方法来提高研究的可重复性和透明度, 评级不一 (图 2); 1936 名参与者中有 285 名 (14.7%) 强烈不同意或不同意 (589[30.4%]), 474 名 (24.5%) 不确定, 456 名 (23.6%) 同意或强烈同意 (132[6.8%]) 该陈述。1937 名参与者中有 565 名 (29.2%) 不确定 AI 聊天机器人是否可以提高数据分析和实验的准确性和精密度, 而 243 名 (12.6%) 强烈不同意或不同意 (481[24.8%]) 的说法。对于 AI 聊天机器人是否可以通过提供标准化的数据分析方法来减少人为错误或偏倚, 受访者的看法不一, 因为 1932 人中有 530 人 (27.4%) 不确定, 523 人 (27.1%) 同意, 149 人 (7.7%) 强烈同意; 然而, 472 (24.4%) 受访者不同意, 258 (13.4%) 强烈不同意该声明。总体而言, 大多数受访者 (2050 年的 1676[81.8%]) 认为 AI 聊天机器人在未来的科学研究中将非常重要或重要 (附录 p2)。

受访者对与在科学过程中使用人工智能聊天机器人的潜在挑战相关的陈述的同意程度进行了评分, 评分范围从非常不同意到非常同意 (图 3)。最同意和最同意人工智能聊天机器人的挑战包括有偏见或扭曲的数据输出 (1924 年的 1384[71.9%]),

道德和法律问题（1927 年的 1480[76.8%]），研究人员的抵制或抵制（1924 年的 1374 [71.4%]），缺乏问责制（1924 年的 1405[73.3%]），决策过程中缺乏透明度和可解释性（1923 年的 1456[75.7%]），对人工智能聊天机器人如何做出决策和生成响应的理解不足（1923 年的 1484[77.2%]），处理依赖于上下文或特定于情境的信息的能力不足（1924 年的 1426[74.1%]），捕捉人类思想和推理的细微差别和复杂性的能力不足（1921 年的 1467 [76.4%]），数据隐私和安全问题（1921 年 1401[72.9%]），以及与用户接受和采用相关的挑战（1924 年 1437[74.7%]）。

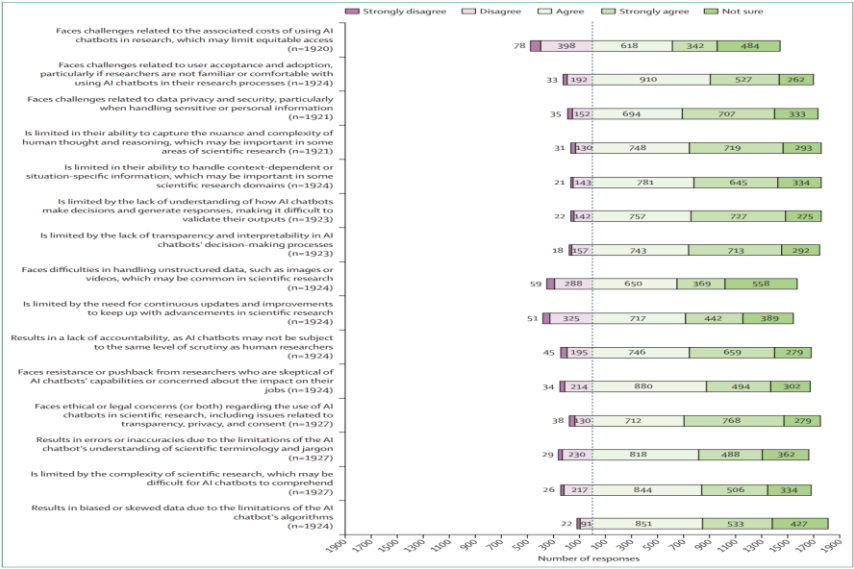


图 3 受访者对在科学过程中使用 AI 聊天机器人的潜在挑战的一致性

开放式问题允许参与者就人工智能聊天机器人在科学研究中的使用提供额外的反馈和评论。开放式问题的指定主题、副主题、个别代码和频率可以在网上找到。专题分析的结果也可以在网上找到。在专题分析中发现的主要主题包括“人工智能聊天机器人训练的性质及其组成”、“研究机构中人工智能聊天机器人政策的性质”、“人工智能聊天机器人在科学过程附加步骤中的帮助”、“人工智能聊天机器人在科学过程附加步骤中的使用程度”、“人工智能聊天机器人在科学过程中的附加好处”和“关于人工智能聊天机器人使用或调查的最终评论”。探索这些主题的更多细节显示在网上。

讨论

据我们所知，就设计和规模而言，这项调查是同类型调查中的第一次，它为医学研究人员在科学过程中对 AI 聊天机器人的知识、使用和看法以及当这些 AI 聊天机

机器人新应用于该领域时对医学研究方式的可能影响提供国际视角。总体而言，受访者对在科学过程中使用 AI 聊天机器人可能的好处表达了不同的意见，而大多数受访者同意其缺点和挑战。许多人同意 AI 聊天机器人可能在科学过程的各个步骤中提供帮助，但由于许多限制和错误，在当前状态下没有帮助。

在我们的调查中，大多数受访者已经熟悉这个概念，并且之前曾将 AI 聊天机器人用于研究目的，其中最常用的 AI 聊天机器人是 ChatGPT。绝大多数受访者认为 AI 聊天机器人在撰写和编辑手稿方面最有帮助。然而，只有少数受访者（11.4%）表示，他们的机构提供了关于在科学过程中适当使用 AI 工具或聊天机器人的培训，一小部分（9.9%）报告说他们的附属机构在这种情况下实施了政策。考虑到许多广泛使用的 AI 聊天机器人存在重大且众所周知的缺点（例如，幻觉定义为做出不正确的预测或模式），这些发现令人惊讶，不幸的是，解决 AI 局限性的宣传活动仅在一些研究机构提供，很少有人对 AI 聊天机器人在研究中的使用实施明确的政策。缺乏可用的培训 and 政策可能是由于过去 2 年在研究中使用了 AI 聊天机器人，或者可能是由于一些出版组织和期刊的研究文章禁止使用聊天机器人等 AI 工具。培训不足并不是因为缺乏兴趣，因为大多数受访者都表达了对学习更多内容和接受有关在科学过程中使用 AI 聊天机器人的培训的兴趣。

缺乏培训和政策也引发了对研究诚信的担忧，因为许多期刊现在要求作者披露所有使用 AI 的行为，但未申报的 ChatGPT 辅助手稿正在同行评审期刊上发表。¹⁸这一事实表明，在同行评议过程中缺乏有效的保护工具或机制来检测人工智能生成的材料，但也表明一种令人担忧的模式，即研究诚信可能已经受到损害。

AI 聊天机器人的主要感知好处包括节省起草手稿、语言翻译或生成图形和表格的时间。AI 聊天机器人以自动化研究过程中耗时的任务并提供写作和编辑帮助而闻名；我们的研究结果也反映了这些好处。

有趣的是，61.8%的受访者认为 AI 聊天机器人对一般管理任务非常有帮助或有帮助，44.5%的受访者以前曾将它们用于研究。大多数受访者报告说从未使用过 AI 聊天机器人，因为当前许多 AI 聊天机器人模型仍然存在很大限制。受访者最担心的是，由于决策和响应生成过程背后的理解或透明度不足，难以验证 AI 聊天机器人的输出。

由于在处理医疗数据时具有已知的模式识别能力，因此人工智能工具的使用已经在医学和医疗保健领域得到了广泛讨论。Ali 及其同事的系统评价发现 AI 在患者

医疗保健中的好处包括早期诊断、患者监测和自动决策。Goodman 及其同事¹⁹指出，纳入支持系统的 AI 工具指导了临床决策，使用患者健康记录在临床不确定性期间进行算法预测。例如，脓毒症预警系统可以使用实时数据在医生检测到任何临床体征之前识别患者的早期脓毒症。最终，医生可以将 AI 驱动的临床决策支持预测纳入临床决策；但是，这些预测需要批判性评估和人工判断，因为 AI 算法包含偏差，也可能产生误报。同样，Ruksakulpiwat 及其同事表示虽然 ChatGPT 可以很好地为放射科医生提供一般和基本的医疗信息并简化医疗报告，但这个 AI 聊天机器人也包含错误，目前需要人工验证。在我们的调查中，许多受访者指出，AI 聊天机器人可以提供实时反馈和见解，在数据分析（尤其是大型数据集）中非常有用，并有助于识别和预测科学趋势和机会。受访者还强调，如果要将 AI 聊天机器人用于科学研究，其输出将需要人工判断和评估。

许多 AI 聊天机器人在初次使用时很容易访问和公开获得，几乎不需要任何费用。YouGov 2023 年的一项民意调查发现，62% 的美国受访者对人工智能的增长有些或大部分担忧。此外，56% 的受访者表示支持联邦对 AI 的监管，而不是科技公司的自我监管，因为 82% 的受访者不相信科技高管会监管 AI。这些发现得到了 Gillespie 及其同事的支持，世卫组织对 17000 多人进行了一项全球调查，发现 61% 的受访者对信任 AI 系统持谨慎态度。受访者最关心 AI 系统的安全性、安保、数据隐私措施和公平性，这是我们的调查受访者在科学研究中使用 AI 聊天机器人方面表达的主要担忧之一。

2023 年的调查由针对用户体验研究人员或设计师和研究运营专家进行的用户访谈发现，77.1% 的受访者已经在至少部分工作中使用人工智能，大约一半（51.1%）为此目的使用 ChatGPT。调查报告称，人工智能被提及最多的好处是效率，因为 40.4% 的受访者使用 AI 进行定性编码，45.5% 的受访者使用 AI 编写报告。另一项对 1600 名研究人员的调查也支持这些感知到的好处，因为受访者认为研究中的 AI 工具将增加数据处理、加快计算速度、自动化数据采集并减少研究人员花费的总时间和资源。此外，人工智能在研究中的负面影响主要被认为是数据中的偏倚和歧视风险（55%）、欺诈（55%）以及在没有完全理解的情况下对模式识别的依赖增加（69%）。同样，Noy 及其同事²⁵目前一项涉及 453 名受过大学教育的专业人士的试验结果表明，使用 ChatGPT 显著提高了生产力、改善了任务质量并缩短了任务完成时间。由于方法略有模糊性，因此需要谨慎解释这些发现，这可能会夸大结果。

大多数研究人员希望了解更多信息并接受有关使用 AI 聊天机器人的培训，这表明需要为生物医学研究人员开发有关在科学过程中使用 AI 聊天机器人的结构化培训。围绕 AI 聊天机器人在研究中的使用制定固定的指南和标准化培训可能有助于保持研究的完整性。由于当前模型存在许多限制或错误以及道德问题，许多研究人员还对 AI 聊天机器人在研究中使用的担忧。在这些工具可用于研究之前，需要解决这些问题。最后，AI 聊天机器人可能通过帮助非母语人士克服语言障碍来提高研究公平性，鉴于它们继续免费或以低成本广泛提供。

这项调查提供了对早期 AI 聊天机器人在医学研究中的应用的见解，并可以纳入其中来衡量这项技术的演变。目前，关于 AI 聊天机器人的使用存在亟待关注的底层问题和挑战。未来的研究应侧重于围绕科学研究中适当使用 AI 聊天机器人制定指南和正式培训，并促进对这些工具的更深入理解。此外，还需要制定关于在学术界和研究界使用 AI 工具的道德准则，考虑到不当使用可能对研究人员、机构、教育和出版业产生的影响，以及适当使用可以为知识传播提供的潜在好处。未来的研究应该调查非通讯作者，以确保更多样化的观点和观点。此外，AI 开发人员也需要解决 AI 聊天机器人模型的当前限制和错误。

因为横断面调查在某个时间点观察一个群体，所以它们进行起来既快速又便宜。我们研究的另一个优势是，在 2023 年 3 月 1 日至 4 月 7 日期间检索了通讯作者的姓名和电子邮件地址，以最大限度地减少退回或不活跃的电子邮件。

我们的研究有几个局限性。由于我们自己的语言和资源限制，非英语研究人员在很大程度上被排除在样本之外。此外，英语部分流利的受访者在完成调查时可能面临挑战，这可能会限制他们充分分享自己态度和观点的能力，而英语不流利的人的观点可能被完全排除在外；因此，我们的研究结果可能不适用于那些主要用英语以外的语言发表研究的人。另一个限制是，我们调查的原始回复率将被低估，因为我们没有考虑退回或不活跃的电子邮件账号和自动回复，也没有验证 61560 名被认定为医学研究人员的作者中有多少人实际上是医学研究人员。尽管我们旨在评估来自不同学科的大量研究人员样本，但许多受邀者表示他们无法参与，因为他们以前从未使用过 AI 聊天机器人。在这种情况下，其中一位作者（JYN）鼓励这些研究人员仍然参加调查。尽管我们尽最大努力鼓励所有受邀者，但我们研究结果的普遍性可能会受到影响，因为那些强烈支持或反对使用 AI 聊天机器人的人更有可能回答我们的调查；因此，那些经验不足或对使用 AI 聊天机器人不确定的人在调查中可

能代表性不足。尽管该池的低响应率会影响泛化性，但参与者的总数仍然代表较大的样本量。大多数调查回复来自高收入国家的个人，这可能会扭曲研究结果，并限制我们的结果对低收入国家研究人员的普遍性，这些国家对 AI 聊天机器人的采用和使用可能大不相同。另一个限制是 AI 聊天机器人在研究中的使用只能在一个时间点观察到，因此，我们的结果提供了研究人员在 AI 聊天机器人使用早期阶段的观点快照。作为横断面调查设计固有的，我们的研究也容易受到回忆和无反应偏倚的影响。

总之，我们的研究为研究人员在科学环境中对 AI 聊天机器人的看法和使用提供了有价值的见解。尽管很大一部分受访者表现出熟悉度，并将 AI 聊天机器人纳入他们的研究过程，但我们的研究结果凸显了明显的挑战，包括缺乏正式的培训机会和管理 AI 聊天机器人使用的机构政策。尽管 AI 聊天机器人在减轻工作量和负担方面具有明显的好处，但人们对这些技术背后的透明度和决策过程仍然存在担忧。我们的研究强调了加强支持结构和教育举措的迫切需求，以促进 AI 聊天机器人有效地集成到科学工作流程中。通过解决这些差距，研究人员和机构可以充分利用 AI 聊天机器人的潜力来推进学术工作并推动科学研究的创新。

***注：原文和译文版权分属作者和译者所有，若转载、引用或发表，请标明出处。**