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 2023.7.29

（六）大语言模型

1. **标题：** ChatGPT 和学术界的大语言模型：机遇与挑战

来源： BioData Min.

时间： 2023 年 7 月.

链接： https://doi.org/10.1186/s13040-023-00339-9.

概要： 自 2022 年发布 ChatGPT 以来，学术界就对该技术的应用表达出不同意见。ChatGPT 是由近两年才出现的深度学习模型——大语言模型（LLM）提供支持。这些模型在大量公开可用的文本数据上进行训练，而后在人机对话中生成类似人类的响应。ChatGPT 和 LLM 在教育和学术研究上表现出了令人兴奋的潜能，同时也存在着一些令人不安的伦理问题。本研究将围绕这一议题，对该技术在学术写作、教育和编程等领域中的应用进行了深入研究，并对其可能产生的偏见、可扩展性和可访问性进行了充分讨论，认为 ChatGPT 和 LLM 的应用及搜索能力已远超 Instagram、Facebook、TikTok、google 和 Microsoft 等科技巨头，但用于训练 AI 模型的计算硬件在训练和部署期间的制造和能耗都很昂贵，虽然该技术目前还是一个非营利项目，但是有巨大获利动能，人们在大规模应用前还需要认真考虑其训练数据和初创作品的归属、荣誉等伦理问题；另外，随着公司利用 ChatGPT 和 LLM 竞相创新时，很可能会部署一些不道德和不负责任的解决方案，加剧和延续边缘群体的系统性偏见和差异，所以还需要相关部门对算法公平性、意外伤害和边缘人群适用性进行严格审查。最后，文章总结了在学术界使用基于 LLM 的聊天机器人的五点建议：（1）我们必须找到它们有效使用方法，（2）ChatGPT 和 LLM 使用不能构成抄袭（尽管它们可能会产生抄袭文本），（3）我们必须能量化其中的偏见，（4）用户必须谨慎对待它们的准确性，以及（5）它们应用于研究和作为学术工具的未来是光明的。

2. **标题：医疗领域应用大语言模型（或生成式 AI）的监管必要性**

来源： NPJ Digit Med.

时间： 2023 年 7 月.

链接： https://doi.org/10.1038/s41746-023-00873-0.

概要： 人工智能(AI)的快速发展创生了像 GPT-4 和 Bard 等复杂的大语言模型(LLM)。LLM 在医疗领域的潜在应用现在倍受关注。这是因为其应用非常广泛，不仅可以促进临床文献、创建出院摘要、生成临床、操作和程序说明、获得保险预授权、总结研究论文、或作为聊天机器人为患者回答问题并提供特定数据和关注点，还可以帮助医生根据医疗记录、图像、实验室结果来诊断病情，并建议治疗方案或计划，同时还能提高患者的自主搜索能力、改进科学写作、增强研究公平性、简化医疗保健工作流程、节省成本、改进医学教育中的个性化学习等。但是,在医疗领域应用 LLM 时需要格外谨慎，因为 LLM 在规模和复杂性、硬件要求、广泛的适用性、实时适应、社会影响、数据隐私和安全等多方面都与已受监管的基于人工智能的医疗技术有很大不同。虽然美国 FDA 已发布了一份讨论文件，概述了针对医疗设备中使用的 AI 和 ML 技术量身定制的潜在监管方法，但截止目前，FDA 还没有专门针对基于 AI 技术的特定分类，而是沿用医疗设备监管框架进行风险评估,也不能解决调节式自适应算法和自我深度学习的监管。而且，截止目前 LLM 还没接受过医学信息语料库或医疗大数据的预训练，有时还会“产生幻觉”结果，或带有潜在的医学偏见等，这些都可能将 LLM 引入后风险和挑战放大，而 GPT-4 则更将潜能和风险提升到另一新高度。因此，监管机构需要对 LLM 从其部署就开始监管，而且最好以与当前市场上 AI 技术不同的方式进行监管。之后，文章还列出 GPT-4 发布后大语言模型将在医疗领域可能发展出的潜能与监管所面临的挑战，认为 LLM 监管应确保医疗专业人员和患者可以在不对人造成伤害和不对数据或隐私造成损害的情况下使用 LLM。最后，文章总结了监管机构在将 LLM 引入医疗领域时需要做的五件事，使这一愿景得以实现。

（徐健编辑）

译文一：

ChatGPT 和学术界的大语言模型：机遇与挑战

Jesse G. Meyer, Ryan J. Urbanowicz, Patrick C. N. Martin, Karen O' Connor, Ruowang Li, Pei-Chen Peng, Tiffani J. Bright, Nicholas Tatonetti, Kyoung Jae Won, Graciela Gonzalez-Hernandez, and Jason H. Moore, 徐健（译）

1. 前言

自 2022 年 1 月发布 ChatGPT 以来，学术界就对这项技术的使用表达了不同意见。这种基于人工智能（AI）的聊天机器人以对话方式与用户互动，使用类似人类的语言来回答问题并生成内容。它还被训练用来创建计算机代码。ChatGPT 跟踪以前的提示和响应，根据输入和输出的顺序更正和调整后续答案。ChatGPT 由大语言模型（LLM）提供支持，LLM 是一种近 2 年才出现的深度学习模型。这些模型在大量公开可用的文本数据（如书籍、文章和网页）上进行训练，以在对话中生成类似人类的响应。

学术界面临着由 ChatGPT 和 LLM 驱动的技术发展。一方面，ChatGPT 和 LLM 在教育和研究方面的潜力令人兴奋。它可以用作课堂辅助工具，以快速回答问题，或用作辅助文献综述和文章大纲的学习工具。另一方面，也有一些令人不安的伦理问题需要考虑。例如，LLM 可能会采用偏见，并在训练数据集中延续刻板印象，将虚假信息呈现为真相。在本文中，我们将讨论围绕这种 AI 技术的优势和担忧。我们研究了像 ChatGPT 这样的 LLM 在教育、编程和学术写作中的使用。我们还评论了 ChatGPT 等 AI 模型的偏见、可扩展性和可访问性。我们希望这篇文章能引起人们对学术界纳入基于 LLM 的聊天机器人的兴趣和进一步讨论，同时仔细考虑伦理问题。最后，我们总结了对学术界使用基于 LLM 的聊天机器人的立场。

2. 背景

要理解本社论中讨论的 ChatGPT 和其他 LLM，需要几点背景知识。首先，这项技术的一个关键进步是它的迭代能力，其中先前的响应和输出调整后续输出。其次，ChatGPT 有两个版本：使用模型版本 3.5 的免费版本和当前使用模型版本 4.0 的付

费版本。OpenAI 很可能使用免费版本的反馈来推动改进，使其成为付费版本。版本 4.0 的另一个重要功能是它可以接受图像输入。例如，它可以使用网站设想图来生成构建该网站所需的代码。还有其他平台，例如来自 Google 的 Bard，以及其他 LLM，例如 BLOOM，这是一种开放访问的 176B 参数多语言 LLM，可用于为类似聊天的界面提供支持并用在部署其他应用程序。最后，一个重要的概念是“提示工程”和充当“提示工程师”，即由于 ChatGPT 或 Bard 的迭代性质以及对文本提示的确切选择的敏感性，成功诱导所需输出的能力并不总是微不足道的。最近，一个及时工程师的招聘启事在互联网上广泛流传，其中模糊的要求资格提供每年高达 300 万美元的薪水>突出了用于聊天的 LLM 的炒作。

基于 LLM 的聊天机器人的一个基本挑战是它们可以将虚假信息呈现为真相，也称为“幻觉”。事实上，ChatGPT 和 Bard 可以歪曲他们的能力。例如，当被问及它是否可以帮助找到相关出版物在评论文章中引用时，ChatGPT 确认它可以，但随后继续列出五个完全虚构的出版物。在重复试验中，有时它列出了一两篇真实的论文，但有时，一篇论文是编造的，或者论文是真实的，但与调查无关，除了可能有相同的作者。因此，由大语言模型（LLM）生成的答案可以正确格式化，但不一定是事实。人们需要记住，它不是人类：它没有被训练以准确反映其自身能力和有限的方式做出反应，而是被训练成构建文本语句，模仿在给定提示的训练文本中发现的类似语句。这些问题限制了 LLM 聊天机器人在某些领域的效用，例如医疗领域，事实对于确保最佳健康结果至关重要。例如，至少在当前状态下，根据 ChatGPT 的输出制定治疗计划是不明智的，因为这会导致大量误诊和无效治疗，从而对患者造成伤害。由于这种限制，LLM 的任何使用都应仔细检查输出的准确性。

3. 用于学术写作的聊天 GPT

3.1 使用来自 ChatGPT 的文本的伦理和规则

LLM 可以有益于科学研究和写作，但应该如何记录它们的使用？使用 ChatGPT 和其他 LLM 生成的文本是一个伦理灰色地带。将别人的作品作为自己的作品呈现被认为是抄袭，但如果该文本是由人工智能模型生成的呢？此外，如果该文本是通过作者引导 AI 的迭代过程生成的呢？当涉及到人工智能生成的文本时，目前还不清

楚个人贡献在哪里结束，抄袭在哪里开始。在某些情况下，需要确定写作来源，例如手稿作者身份。事实上，我们已经看到了 ChatGPT 被列为作者的手稿。在一个案例中，ChatGPT 在预印本中被列为作者，但在最终出版物中却没有被列为作者。一些期刊已经发布了关于 ChatGPT 是否可以成为作者以及在最终出版物中允许 LLM 使用哪些文本的指南。*科学*已经宣布，他们发表的论文中不允许使用 AI 生成的文本或图形，并且 ChatGPT 不能在其期刊中被命名为作者。*《自然》*杂志发布了类似的指南，指出：（1）任何 LLM 工具都不会被接受为研究论文的作者，（2）在研究中使用 LLM 工具的研究人员应在方法和/或致谢部分记录它们的使用方式。

3.2 ChatGPT 作为写作和编辑工具

根据 2020 年 NSF 科学与工程指标，在美国培养的博士后中有 49%出生在海外。在工程，数学和计算机科学领域，60%的博士学位授予国际学生。有效的学术写作在学术出版物的成功和质量中起着至关重要的作用，然而，它是国际博士后和学生面临的主要挑战之一。

LMM 可以为来自不同背景的研究人员提供好处，特别是对于非英语母语人士，作为一种写作和编辑工具，以提高学术写作的质量。随着时间的推移，基于软件的编辑工具取得了重大进步。从文档编辑器（例如 Microsoft Word）实施的基本拼写检查到 Grammarly，Scribbr 和 Quillbot 等商业在线服务，这些基于软件的编辑工具具有各种功能，可以纠正语法错误并提高写作清晰度。但是，这些工具通常具有一组预定义的写作样本评估，并根据这些评估提供相应的报告。相比之下，LMM 通过使用自定义提示在评估方面提供了更大的灵活性。例如，用户可以使用诸如“您能否解释您检测到的语法错误？”之类的提示来请求 LMM 为已识别的语法错误提供解释。这些解释为用户提供了额外的信息，以评估报告的准确性。已经报道了设计创意提示以改善语法、可读性、语气、意图等的例子。

需要强调的是，使用 LMM 作为编辑工具不同于基于提示的文本生成。作为一种编辑工具，LMM 利用人类创作的文本作为修改和建议的基础。这个过程类似于要求朋友或同事校对写作样本并提供建议。因此，使用带有预先编写的文本的 LMM 不太可能引起诸如抄袭之类的伦理问题，这在使用 LMM 基于提示的文本生成时可能会出现。

3.3 如何使用 ChatGPT 来帮助撰写论文？

在一些例子中，手稿报告了 ChatGPT 在某些测试或任务上的表现。在另一个例子中，ChatGPT 的回应被用来推测为什么服用雷帕霉素可能有益于特定的哲学论证。作为对简短提示的回应，ChatGPT 对 Pascal's Wager 和 Rapamycin 进行了合乎逻辑的介绍，并均衡地讨论了如何将该论点应用于服用该药物。在第三个例子中，研究人员展示了与 ChatGPT 围绕人工智能、使用 LLM 的聊天机器人以及高等教育中的抄袭的对话结果。在这个例子中，ChatGPT 很好地总结了 ChatGPT、抄袭、ChatGPT 如何用于抄袭以及大学教授如何结合不同类型的评估，如演讲和活动，以避免仅仅依赖潜在的抄袭论文来获得成绩。

研究人员可能希望避免使用 LLM。例如，研究人员可能会试图使用 ChatGPT 来汇总字段以获得概述，但是，由于 ChatGPT 通常会将虚假信息呈现为事实，这是不可取的。最好使用维基百科等资源来广泛了解某个主题。然而，有一些专门为提供准确的科学信息而开发的 LLM 的新例子，例如 perplexity.ai。在询问我们最近工作的提示中，该工具找到了我们的预印本，并将其放在其他相关手稿的上下文中，并具有相关但不全面的引用。

3.4 ChatGPT 和 LLM 用于资助提案撰写

在资助提案中使用 LLM 可能需要不同的规则，其中已经写成的通常是由帮助首席研究员（PI）提出建议的工作人员制作的，如博士后研究员或研究生通常会帮助 PI 撰写与其项目相关的资助提案，该提案随后由 PI 提交。使用专业写作人员或自由资助提案作者也很常见，在某些情况下，他们会制作 PI 提交的提案的全文。与员工甚至承包商合作起草提案并不违反 NIH 的政策。尽管提案与效绩指标相关，但它是由该机构正式提交的，所以不能假定为唯一作者。根据 LLM 使用资助和论文的趋势，联邦资助机构可能需要发布关于是否许可的指南。

毫无疑问，ChatGPT 已被用于撰写资助提案。在一位作者尝试 ChatGPT 如何帮助撰写资助提案的经验中，ChatGPT 可以为感兴趣的方向提供一些想法，甚至提出合理的逻辑理由，我们需要进行研究，这些研究可以找到进入 NIH 资助的重要性部分。ChatGPT 甚至可以提供可能合乎逻辑地用于部分资助的通用文本，例如目标页

面。它可以建议可能测试感兴趣的假设的特定实验。对于这篇社论的一位作者来说，一个令人惊讶的效果是，它有助于克服作家的障碍，因为只要在页面上有一些文本就足以继续进行编辑。正如 Jodi Picoult 所说："你总在编辑一个糟糕页面。您无法编辑空白页。但是，资助必须在短时间内表达许多复杂想法，因此，由于 ChatGPT 的大部分文本都是通用的且经常重复，根据我们的经验，ChatGPT 生成的文本都没有进入最终的资助提案。

3.5 LLM 的未来：作为资助提案审查者？

每年，科学家都会审查数以千计的资助提案，花费数千小时的精力和旅行时间。尽管投入了大量资金，但目前的审查制度还是主观判断，比如，发到一个研究部的一些提案得分很低，但另一个研究部几乎相同的提案则获得资助。可以想象，LLM 最终可以用来创建较少主观的评价，但前面还有很长的路要走，需要多层情报。这将使审查时间更快，浪费的人力时间更少，然后可以用于更多的研究。预计 LLM 可能会用于个人评审员的资助提案审查，美国国立卫生研究院于 2023 年 6 月 23 日发布了 NOT-OD-23-149 指南，规定"美国国立卫生研究院禁止 NIH 科学同行审稿人使用自然语言处理器、LLM 或其他生成式人工智能(AI)技术来分析和制定资助申请和研发合同提案的同行评审批评。"在不久的将来，使用 LLM 进行资助提案审查也似乎不太可能被接受。

4. LLM 的可扩展性和可访问性

ChatGPT 不是一个独一无二的工具，大语言模型已经存在了几年，具有生成有意义和有用的文本的不同能力。但是，其中许多工具都可以作为命令行界面（CLI）免费获得，这可能解释了 ChatGPT 迅速成名的原因。OpenAI 巧妙地设计了 GPT-3 模型，以便与直观的用户界面一起大规模部署。任何拥有电子邮件地址的人都可以开始生成文本。公众兴趣飙升，ChatGPT 在 100 个月内获得了 2 亿活跃用户，这一壮举甚至无法与 Instagram，Facebook 或 TikTok 等公司相提并论。

据估计，ChatGPT 每天有 13 万次查询，有时难以满足需求，查询无法处理。相比之下，谷歌搜索每天收到大约 8 亿次查询。大规模部署人工智能模型是未来人工智能技术必须面对的挑战，即使对于谷歌（LaMDA 的 Bard power）和 Microsoft

（人工智能驱动的 Bing）等科技巨头来说也是如此，它们已经宣布在各自的搜索引擎中集成人工智能。

用于训练 AI 模型的计算硬件在训练和部署期间的制造和能耗方面都很昂贵。对 Open AI 可以使用的计算能力（10，000 台 Nvidia V100）的猜测使硬件成本接近 5 万美元，而不考虑研究、开发和能源使用。最近的一项研究估计，BERT（6 亿参数 LLM）的训练完成将产生 21 至 78 公吨的一氧化碳。提醒一下，ChatGPT 是一个 175 亿个参数的 LLM，这只考虑了训练成本，而不是与生成文本以响应潜在的数十亿日常查询相关的任何成本。与大规模人工智能模型相关的能量和环境压力可以通过选择可持续能源、提高算法效率和开发低能耗人工智能专用硬件来减少。

OpenAI 可能最初是一个非营利项目，但面对大规模的 AI 应用成本，不可否认的是，它有获得财务盈利的动力。ChatGPT 提供免费套餐以及每月 20 美元的高级订阅计划，即使在高峰时段也能优先有效地访问该模型。虽然人工智能有可能彻底改变教育和学术研究，但付费访问可能会进一步扩大贫富之间的差距。如果人工智能服务保持免费，我们必须想知道隐藏的成本是什么，以及用户信息将如何用于盈利——这是免费使用在线服务的常见做法。

数据使用的问题还揭示了一个伦理难题，这个难题一直困扰着艺术界，包括 DALL-E 和 Mid-Journey 等其他生成式人工智能模型:谁拥有训练数据?艺术家、作家和音乐家的作品构成了这些模式成功的基本要素，他们可能会发现自己被作品的货币化所轻视，而没有得到任何荣誉。

5. 教育

ChatGPT 在教育的多个层面都有希望，有可能在基础教育、本科和研究生教育中担任各种支持角色。关于 ChatGPT 的使用或滥用的可能性，已经出现了各种评论。在这里，我们从学生和教育者的角度考虑 ChatGPT 的潜在作用。

LLM 通过协助研究和学术写作或将其用作交互式学习指南，其中可能包括生成练习考试并提供即时反馈，为学生提供教育方面的益处和机会。各级学生的机会包括使用 ChatGPT 作为创意灵感的来源、快速直接回答特定问题的方法以及内容生成器（例如草稿、格式、总结和/或编辑）。在高等教育中，基于 LLM 的聊天机器人可用于提高学生的参与度，促进小组活动，创建交互式学习工具，并提供即时反馈和

评估。

对于学习者和教育者来说，使用 ChatGPT 并非没有挑战和限制。过度依赖可能会阻碍学生发展关键技能的能力，如写作。人们担心，学生可能会轻易地不诚实使用 ChatGPT、在完成作业或考试时作弊或者在没有引用的情况下写报告/论文，这可能被视为抄袭。ChatGPT 产生的数据可能不准确或有偏差，用户需要意识到这一点，以确保这些缺陷不会在他们的工作中传播。

尽管存在这些担忧，但应该指出的是，学生可以根据教育机构的道德或荣誉准则，以建设性的方式参与 ChatGPT。最终，问题不应该是学生“是否”使用 ChatGPT，而是“如何”使用，这与父母如何帮助孩子完成作业的问题没有什么不同。例如，一个学生在回复一个学生写论文的请求时，逐字提交由 ChatGPT 生成的文本，这显然是不可取的。相反，如果一个学生在写文章时扮演了提示工程师、事实核查员和编辑的角色，就应该得到积极的评价。

虽然对教育者来说，发现作弊者仍然是一个未解决的挑战，但如果我们从允许(如果不是鼓励的话)学生使用 ChatGPT 的假设开始，可以采取一些策略来解释这一点。例如，学生可以被要求(1)提交他们与 ChatGPT 互动的成绩单，(2)将他们的最终提交与 ChatGPT 生成的记录进行对比，并跟踪变化，或者(3)指出 ChatGPT 所犯的错误，以及如何在学生提交的作业中解决这些错误。正如 OpenAI 首席执行官萨姆·奥特曼所说:“生成文本是我们都需要适应的东西。我想，我们适应了计算器，改变了数学课的测试内容。”毫无疑问，这是一个更极端的版本，但它的好处也更极端。也许 ChatGPT 的到来只是进一步突出了我们的教育系统通常评估学生的众多缺点。

关于学生从 LLM 学习的一个更重要的问题是，无论提供的信息是否准确，它通常都以自信且似乎知道的方式做出回应。学生应该意识到，该系统并不像看起来那么聪明，并且可能会不可预测地歪曲或发明信息。归根结底，ChatGPT 与将互联网用于教育目的没有什么不同，在教育目的中，大多数信息都应该以一定程度的怀疑态度来消化，需要跨多个（理想情况下是主要）来源进行确认。

对于教育工作者来说，ChatGPT 为课程开发和设计、课程规划、评估和评价提供了许多新的机会。由于学生使用 ChatGPT 和其他 LLM 课程是不可避免的，教育工作者需要改变教学方法和教育评估，例如将 ChatGPT 的使用整合到课程中，并且更多地关注过程而不是最终产品进行评估。教授学生正确使用基于 LLM 的聊天机器人所需的技能，包括这些工具的局限性，对于确保学生负责任地使用它们并披露它们

的使用情况，以及确保所有学生都能使用这些工具，以避免在学习中造成差异，都是非常重要的。

6. 协助编程

ChatGPT 还承诺在以下方面促进计算机编程:(1)学习编码或使用特定的包、库或框架，(2)编写新代码，(3)解释现有代码，(4)调试现有代码，(5)增加代码的紧凑性或效率，或(6)将代码从一种编程语言翻译为另一种编程语言。根据询问，ChatGPT 声称能够帮助各种编程语言，包括 Python, Java, JavaScript, c++, C, Ruby, Julia, PHP, SQL, HTML, CSS 等。

作为一种教育编程资源，ChatGPT 可以作为交互式讲师，回答有关使用哪种语言、代码语法和语义、最佳实践、可用库或包、替代方法、集成开发环境(ide)和编程环境的问题。ChatGPT 还可以生成简单/清晰的示例代码，每行都有注释，以及代码功能的自然语言摘要(突出显示关键变量、方法或包含的包的底层功能)。与使用谷歌搜索或 stackoverflow、geeksforgeeks 等网站搜索“如何”编码问题相比，ChatGPT 似乎能够直接为学习者提供一个简单、可解释、相关的解决方案，以解决特定的编码问题。教育工作者还可以让 ChatGPT 生成编码教程，以及与特定编程主题相关的作业和考试问题。

对于任何技能水平的程序员，ChatGPT 都可以自动编写新代码。值得注意的是，已经有一些例子将 ChatGPT 直接集成到 ide 中以提高代码编写效率(例如 JetBrains 或 Visual Studio 的相应插件)，以及其他相关的 AI 应用程序来促进代码编写(例如 Github Copilot)。然而，ChatGPT 在编写代码时的效用和适用性最终取决于(1)用户在描述所需代码的功能和参数时提供的详细程度，以及(2)所请求代码的规模和复杂性。目前，ChatGPT 更有可能成功地准确编写较小的代码块，而它在编写更大/更复杂的程序(例如软件包)时的可靠性则值得怀疑。

除了编写新代码之外，ChatGPT 还可以提供许多辅助功能来改进现有代码。像 GitHub 这样的平台上的代码共享为加速计算领域的研究和开发提供了巨大的机会。然而，解释别人写的代码(例如，学生甚至是我们自己多年前写的代码)可能

会很困难，因为通常有许多有效的方法来编程一个特定的任务。当代码文档不完善，注释有限、不准确或含糊不清时尤其如此。用户可以询问 ChatGPT 一行或一段代码是做什么的，ChatGPT 将尝试将其分解成单独的部分，解释变量、命令和步骤，并包括它认为代码正在做什么的一般摘要。在相关的任务中，可以要求 ChatGPT 在代码中添加或纠正注释，作为自动化代码文档的一种手段。ChatGPT 还可以促进代码调试，尽管与手工编码的情况一样，识别影响预期代码功能或性能的错误可能比识别那些简单地阻止代码成功运行的错误要困难得多。此外，用户可以要求它尝试和简化现有代码，以使其更紧凑、可解释或计算效率更高，或者将代码从一种编程语言翻译成另一种编程语言。

与普通应用的注意事项一样，ChatGPT 可能在某种程度上不可预测地将不正确的代码表示为正确的。此外，它可能不知道或无法预测在特殊情况下可能破坏代码功能的边界情况，并且它可能不会在默认情况下提供最佳或最有效的编码解决方案。最终，由 ChatGPT 编写的代码可能会提供一个有用的起点，但是它生成的代码和注释都应该由程序员和用户进行检查和验证，以确认它完全满足预期的目的。对于任何高风险的应用程序，直接依赖 ChatGPT 生成的代码是不明智的，因为安全性、责任、隐私和信任是至关重要的。因此，至少目前看来，有经验的人类程序员远没有被淘汰。

7. 偏见

随着人工智能/机器学习解决方案的增长和扩散，对算法公平性、意外伤害和与边缘化群体使用这些解决方案相关的公平性的审查越来越多。在医疗保健环境中部署不道德和不负责任的设计 AI/ML 解决方案可能会加剧和延续边缘群体的系统性偏见和差异。随着公司竞相与 ChatGPT 等 LLM 进行创新，根深蒂固的偏见将被放大，危害将显现出来。在 LLM 编码偏差的根本原因得到解决之前，用于临床应用的 LLM 将遭受同样的命运，传播偏差。解决算法偏见的根本原因的变革工作始于在项目设计中提出基本问题，例如，是否需要人工智能解决方案，如果需要，目的是什么;减少偏见的策略是什么?如何避免利用资料;团队的跨学科程度如何(例如，是否有伦理学家、法律专家等)?重要的是，开发能够减少算法偏见和系统性种族主义的 LLM

需要采取行动;评估公平性的技术方法必须以透明的方式成为模型评估过程的一部分，包括对方法和度量选择的披露。减轻数据和模型开发管道中的偏见的多管齐下的方法可以包括预处理算法，如重称重，异构影响去除器或学习公平表示;在处理技术，如偏见消除，对抗性去偏见或歧视意识;和后处理，如拒绝选项分类或均等赔率后处理。当临床应用 LLM 在设计时有意识地、透明地考虑到这些社会因素，它们就可以成为促进公平和改善护理机会的额外工具。

8. 结论

总结一下我们对 LLM 使用的立场:

1） LLM 必受欢迎

这篇社论提供了几个关于 LLM 如何提高学术写作、教育和编程效率的例子和观点。因此，在我们看来，毫无疑问，我们是否应该在所有可能的应用中采用这些工具。LLM 必被接受，以提高所有学科的教学和研究效率。

2）必需量化 LLM 偏差指标

ChatGPT 的输出来自互联网上最公开的文本，直到 2021 年。因此，ChatGPT 和其他 LLM 的性能模仿可用文本，并且它们与训练数据一样有偏见。例如，众所周知，ChatGPT 会使诸如护士是女性，医生是男性之类的刻板印象永久化，这种偏见来自于培训数据。很难定义评估训练数据和模型输出中的偏差水平的指标。当在文本分析或生成研究中利用 LLM 的输出时，我们必须讨论固有偏见作为限制。

3）使用 LLM 的学位论文不构成抄袭

从某种意义上说，使用 ChatGPT 和 LLM 的输出可能仍然看起来像一个道德灰色地带。通过设计，LLM 将人类文本“编码”为以后使用的统计模型。ChatGPT 的输出将匹配互联网上现有的文本，特别是如果样本足够小的话。因此，从严格意义上讲，像 ChatGPT 这样的 LLM 确实会产生抄袭。但是，如果没有提示，ChatGPT

不会做任何事情。缺乏及时的工程和迭代，感觉更像是抄袭。经过几轮快速工程制作输出后，它感觉更像是原创作品。严格来说，所有的工作都是由相关人员精心制作/设计/推动的。从这个意义上说，模型输出就是那个人的工作。ChatGPT 只是一个类似拼写检查或 Grammarly 的工具。就像从盒子里烤蛋糕一样:蛋糕是面包师的作品，而不是销售蛋糕粉的公司。为了安全起见，我们建议由 LLM 或任何人类生成的任何文本都要通过剽窃检测软件进行评估，以确定是否与已发表的文本意外接近。

4）LLM 可以产生虚假或不准确的陈述

在许多情况下，LLM 的准确性很低。该模型可以生成格式正确的信息，例如不存在的引用。ChatGPT 会歪曲它的知识，它不会回答“我不知道”，而是会很容易地提供看似自信的捏造信息。用户有责任引用适当的来源，并确保内容是真实的。无论 LLM 创造了什么，在与他人分享之前，必须由敏感的工程师“采纳”，因此，LLM 的输出属于提示的作者，作者对其有效性和真实性负责。

5）未来

LLM 应用于研究和作为学术工具的前景是光明的。虽然我们已经回顾了一些关于 LLM 的当前应用和讨论，但仍有许多创造性的应用有待开发和评估。此外，我们看到了将 LLM 与其他有用的工具(如深度学习和自动化机器学习)结合起来的机会。未来将迅速展开，我们将在未来几个月看到许多有用的进展，但挑战也将随行永无止境的开源和商业工具。

***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：

医疗领域应用大语言模型(或生成式AI)的监管必要性

Bertalan Mesk, Eric J. Topol, 徐健（译）

1. 简介

人工智能(AI)的快速发展导致了复杂的大语言模型(LLM)的发展，如OpenAI的GPT-4和谷歌的Bard。ChatGPT是GPT-4的前身，于2022年11月发布，其空前受欢迎程度反映出其是新技术用户接受速度最快的一个——2个月内达到1亿。

这种快速增长引发了关于这种对话聊天机器人在医疗保健和医学实践中可能发挥的作用的全球讨论。LLMs的各种应用已出现，包括促进临床文献、创建出院摘要、生成临床、操作和程序说明、获得保险预授权、总结研究论文、或作为聊天机器人为患者回答问题，并提供他们的特定数据和关注点。LLM还可以帮助医生根据医疗记录、图像、实验室结果来诊断病情，并建议治疗方案或计划。同时，通过对患者的数据、症状和关注点进行个性化评估，患者搜索可能会变得比以前更自主。

系统评价还强调了其他潜在好处，例如改进科学写作、增强研究公平性、简化医疗保健工作流程、节省成本以及改进医学教育中的个性化学习。

鉴于对患者预后和公共卫生的潜在影响，必须考虑如何监管这些新的基于人工智能的工具。在不损害其有希望的进展的情况下对这些医学和医疗保健LLM的监管是确保安全、保持伦理标准、预防不公平和偏见以及保护患者隐私的及时而关键的挑战。以前人工智能认识到的任何担忧现在都随着LLM的多能性而显著放大。

本文探讨在医疗领域中应用LLM的潜在风险和益处，并认为有必要对LLM进行不同于市场上已有的基于AI的医疗技术的监管，以减轻潜在危害并保持公众对这些突破性技术的信任。

2. LLM与已受到监管的基于AI的技术不同

LLM 在规模、功能和潜在影响方面与以前的深度学习方法有很大不同。在这里，我们概述了 LLM 的关键特征，这些特征使它们与传统的深度学习技术区分开来。

2.1 规模和复杂性

LLM 在海量数据集上进行训练，并利用数十亿个参数，从而产生前所未有的复杂性。这种复杂程度需要考虑与可解释性、公平性和意外后果相关的挑战的监管。此外，LLM 使用的标记可以是单词、子单词甚至字符，作为在训练和生成过程中用于表示和处理语言的最小文本单位。标记化是自然语言处理(NLP)中的关键步骤，它允许 LLM 有效地分析和生成文本，因为这些模型旨在处理标记序列而不是整个句子或段落。目前，词语切分还不在医疗监管机构的管辖范围内。

2.2 硬件要求

与以前的深度学习模型相比，LLM 在每秒浮点运算（FLOP）和图形处理单元（GPU）使用方面需要大量计算资源，因为它们的规模大、训练数据广泛、专为 NLP 任务设计的神经网络模型 Transformer 架构，并需要微调。

2.3 广泛的适用性

与为解决特定的医疗问题或临床需求而训练的专业深度学习模型不同，LLM 具有跨越各领域的多功能，例如医疗保健、金融和教育。因此，一刀切的监管框架不适合 LLM，监管必须适应各种行业特定问题。

2.4 实时适应

LLM 可以根据用户输入和不断变化的环境实时调整其响应。这种动态行为要求监管纳入持续的监测和评估机制，以确保负责任的使用和遵守伦理准则。这类似于基于人工智能的自适应医疗技术对监管机构的要求。

2.5 社会影响

LLM 的广泛采用有可能从根本上改变社会的各个方面。因此，监管不仅必须解决 LLM 的技术方面，还必须解决其更广泛的伦理、社会和经济影响。

2.6 数据隐私和安全

LLM 对大量培训数据的依赖引发了与数据隐私和安全有关的担忧。监管应建立强大的框架来保护敏感信息，并防止未经授权访问或滥用这些强大的模型。

LLM 这些独特的特征需要量身定制的监管方法。这种方法必须是适应性的、整体性的，并认识到 LLM 带来的各种挑战和潜在后果，确保其在各个领域的负责任和合乎伦理的使用。

3. FDA 对 AI 的 Pre-LLM 监管

美国食品和药物管理局（FDA）一直引领着全球关于监管的讨论，一个突出的例子就是提供了从 3D 打印药物到基于人工智能的医疗工具的新兴技术法规。

随着数字健康技术的日益普及，FDA 开始监管软件即医疗设备（SaMD），SaMD 是指执行医疗功能并用于预防，诊断、治疗或监测各种疾病或病症的软件解决方案。

作为这种方法的延续，FDA 一直在调整其监管框架，以专门针对医疗设备中的 AI 和机器学习（ML）技术。FDA 发布了一份讨论文件，概述了他们针对医疗设备中使用的 AI 和 ML 技术量身定制的潜在监管方法。该讨论文件提出了一种产品全生命周期（TPLC）方法来规范基于 AI/ML 的 SaMD，该方法侧重于在这些技术的整个生命周期中持续监控和改进。拟议的框架还强调了透明度、实际性能监控以及对 AI/ML 算法修改和更新的明确期望的重要性。

目前，FDA 没有专门针对基于 AI 的技术的特定类别，而是在现有的医疗设备监管框架内对其进行评估。他们根据风险水平将此类设备分为三大类：

- I 类（低风险）：这些设备的风险最小，并受到一般控制，例如注册、列表、标记和良好生产规范。I 类设备的例子包括无动力手术器械和牙线。一些基于人工智能的低风险医疗技术可能属于这一类，具体取决于它们的预期用途。

- II 类（中风险）：这些设备比 I 类设备具有更高的风险水平，并且受到一般控制和特殊控制，例如性能标准，上市后监督或特定标记要求。II 类设备的示例包括输液泵、手术帘和电动轮椅。许多基于人工智能的医疗技术，如诊断成像系统可能属于这一类。
- III 类（高风险）：这些设备的风险最高，并受一般控制、特殊控制和上市前批准（PMA）的约束。III 类设备通常支持或维持人类生命，在防止人类健康损害方面具有重要意义，或存在潜在的不合理的疾病或伤害风险。III 类设备的例子包括植入式起搏器、人工心脏瓣膜和一些用于关键医疗决策的基于 AI 的技术。

基于人工智能的医疗技术也可能受到 FDA 的数字健康软件预认证（Pre-Cert）计划的约束，该计划旨在简化 SaMD 的监管流程，包括基于人工智能的技术。

这一过程中一个里程碑是，他们在 2019 年发布了专门基于人工智能的医疗技术数据库，并获得了监管机构的批准。截至 2023 年 4 月，该数据库中包含 521 台设备。最受欢迎的类别是放射学、心血管学和血液学，分别有 392 台、57 台和 15 台。绝大多数(96%)获得 510(k)批准，18(3.5%)获得 de novo 通路批准，3(0.5%)获得上市前批准(PMA)批准。

正如其他论文所指出的那样，这些设备中只有少数在随机对照试验（RCTs）中进行了测试，只有有限数量的研究使用外部验证、前瞻性评估和各种指标来探索人工智能在实际临床环境中的全面影响、评估用例的范围相对较窄，没有或透明度很低。

综上所述，虽然 FDA 在监管人工智能方面取得了进展，但还不能解决两个相关但不同的先进技术问题的监管。一个是关于调整式自适应算法，该算法可以根据输入数据或其在特定任务上的表现调整其参数或行为。这种适应性使算法能够随着时间的推移提高其性能或响应不断变化的条件。

另一个与深度学习中所谓的自学功能有关。它指的是系统在没有直接监督的情况下自学能力，这种方法通常需要无监督或自监督学习，其中模型从输入数据中学习模式和表示，而不依赖于标记的示例。这种自学深度学习模型可以通过优化其内部表示来发现数据中的底层结构和关系，而无需明确的指导。

4. 医学实践中的 LLM 时代

到目前为止，还没有 LLM 接受过医学信息语料库或数百万患者记录、图像、实验室数据、办公室访问或床边对话的预培训。关于 GPT-4 的培训细节尚未公布，GPT-4 是 2023 年 3 月发布的最先进的 LLM。然而，LLM 具有变革潜力，其用例范围可以从临床文档到提供个性化健康计划。图 11 描述了 10 个医疗专业人员用例和 10 个患者用例。

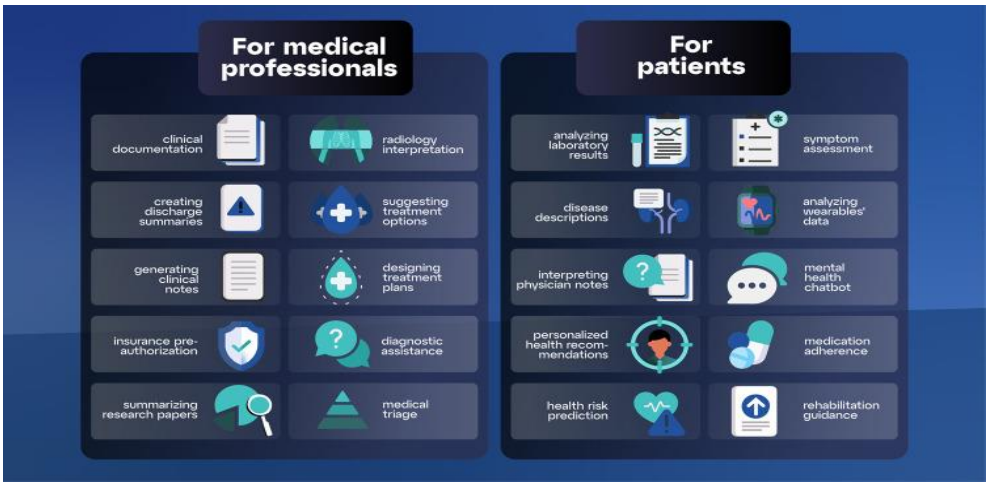

图 1 医疗专业人员 LLM 用例的 10 个示例和 10 个患者例子

同时，将这些模式引入医疗领域会导致风险和挑战的放大。

它开始给医生带来新的挑战，因为患者到达会议时不仅在谷歌搜索他们的症状后收到回复，而且还来自类似 ChatGPT 的聊天机器人。关于 ChatGPT 在多大程度上可以用于医学研究和总结同行评审论文的讨论已经存在，当它只提供在明确要求后做出回应的来源时。此外，据报道，其中一些来源是编造的。

LLM 有时会“产生幻觉”结果，这是指生成了不是基于输入数据或事实信息的输出。此类错误信息可能与诊断、治疗或推荐的测试有关。对于外行来说，这样的输出以高度的信心传达，并且很容易被提示者接受为真理——这有可能是危险的。无论是由于不完整或有偏见的训练数据，其统计性还缺乏上下文。它带来了在医疗环境中提供不可靠或完全错误的答案的重大风险，这可能会产生严重后果。

另一个问题是，使用 LLM 时的医学偏见会影响临床决策、患者结果和医疗保健公平性。如果训练数据包含偏差，例如某些人口群体的代表性不足，过分强调特定治疗或过时的医疗实践，LLM 可能会无意中在其输出中学习和传播这些偏差。GPT-4 的偏倚输出可能导致不正确的诊断或次优的治疗建议，从而可能对患者造成伤害或延迟适当的护理。

GPT-4 将潜力和风险提升到一个新的高度。它将能够阅读图像上的文本（包括医生的手写笔记），并分析图像的内容和上下文。表 1 总结了以前版本和新版本在医疗保健相关和医疗提示方面的主要差异。它表明 GPT-3 可以处理带有一般查询的简单提示，而 GPT-4 能够分析复杂的多级提示，并提供更复杂的结果，例如案例描述或研究论文摘要。

表 1　ChatGPT 和 GPT-4 提示的深度和细节的差异

提示	ChatGPT	GPT-4
提示 1－诊断症状不明确的患者	患者表现为疲乏、体重减轻和偶尔头晕。这些症状的可能原因是什么？	一名 45 岁男性患者因持续性疲乏、意外体重减轻 3 磅和头晕发作 15 个月，来就诊。请提供鉴别诊断并建议相关诊断试验。
提示 2－治疗建议	二型糖尿病有哪些常见的治疗方法？	一名最近诊断为 2 型糖尿病的 55 岁女性，HbA1c 水平为 8.5%。概述一个全面的治疗计划，包括生活方式改变、药物选择和随访监测。
提示 3－患者教育	用简单的术语解释高血压。	创建患者友好的高血压教育讲义，包括病情、危险因素、症状、潜在并发症和管理策略的概述。
提示 4－回顾医学研究	告诉我运动对心理健康的好处。	总结最近关于身体活动与心理健康结果之间关系的研究结果，包括潜在机制、运动类型以及对不同人群的建议。
提示 5－临床病例场景	描述肺炎患者。	创建涉及社区获得性肺炎的 65 岁患者的详细临床病例场景，包括当前病史、相关既往病史、体格检查结果、诊断测试结果和治疗计划。

GPT-4 在医疗保健中的应用引发了伦理问题，需要制定监管框架。需要解决透明度、问责制和公平性等问题，以防止潜在的伦理失误。例如，应让医疗保健专业人员和患者了解人工智能参与决策过程，并向其提供人工智能建议的解释。

此外，监管可以帮助确保人工智能驱动的模型不会延续或加剧现有的医疗保健差异。通过强制使用多样化和具有代表性的数据来源，监管机构可以抵消人工智能训练数据中的潜在偏见，从而促进医疗服务交付的公平性。

在此类环境中使用 GPT-4 和 ChatGPT 需要强有力的法规来确保患者信息的机密性和安全性。这可能包括数据匿名化、加密和安全存储的具体准则，以及防止第三方未经授权访问或滥用数据的措施。

作为广泛实施的标志，医疗公司、数字健康服务和医疗保健组织已经开始将 ChatGPT 实施到其核心业务中。例子包括 Microsoft 拥有的 Nuance，因为他们决定将 GPT-4 AI 添加到其医疗笔记工具中；还有一家名为 Nabla 的法国初创公司，该公司声称是第一家使用 GPT-3 构建工具来帮助医生完成文书工作的初创公司。

所有这些例子和挑战促使监管机构不仅在部署 LLM 模型时开始监管 LLM，而且以与目前市场上的 AI 技术不同的方式对其进行监管。

5. LLM 的监管挑战

大多数 LLM 已经在全球范围内发布，没有针对特定国家的迭代需要监管机构采取全球方法。从监管角度来看，也不清楚 LLM 将属于哪些技术类别。然而，基于 LLM 和以前的深度学习方法之间的差异，可能需要一个新的监管类型来解决 LLM 特定的挑战和风险。

监管机构只有在 LLM 的开发人员声称其 LLM 可用于医疗目的的情况下，才必须为 LLM 设计法规，或者如果 LLM 是为特定的医疗目的而开发、改编、修改或针对特定的医疗目的。即使目前广泛使用的 LLM 不属于任何一类，专门接受过医疗数据和数据库培训的 LLM 的医疗替代方案也可能属于任何一类。

一个突出的例子是 DeepMind 和谷歌研究人员发表的 Med-PaLM。在该研究中，作者提出了一个框架，用多个维度对模型答案进行人工评估，包括事实性、准确性、可能的危害和偏见。此外，使用提示策略的组合，他们的模型在美国医疗执照考试题目中达到了 67.6%的准确率，比之前最先进的技术高出 17%以上。由于人类评估揭示了 LLM 提供的响应中的关键差距，他们引入了指令提示调整，由此产生的模型 Med-PaLM 表现令人兴奋，但仍然不如临床医生。从那时起，GPT-4 在同一次考试中可以达到超过 85%的准确率。

随着不仅可以分析文本而且可以分析图像的 GPT-4 的发布，可以预期该模型将在不久的将来发展到分析上传的文档、研究论文、手写笔记、声音和视频（表 2）。

表 2　LLM 现在可以分析的内容形式类型列表及将来可能的新版本

内容类型	潜在应用	可用性
文本/对话	聊天机器人、文本分析、文档	是的
图像分析	检测图像的内容和上下文	2023 年
文档/PDF 分析	分析研究论文并创建文档摘要	不适用
声音	语音转文本应用程序和基于声音的交互	不适用
视频	分析视频内容并创建深度伪造	不适用

这强调了这样一个观念，即规范当前的 LLM 模型是不够的，因为可以期望具

有那些高级功能的新迭代以与先前迭代相似的速率实现。如果不考虑这些未来的补充，那么当这些更新变得广泛可用时，只关注语言模型的规则可能会错过重要的更新。

拥有批准设备的公司决定在其服务中实施 LLMs，面临着额外的挑战。也就是说，如果该技术已经被批准用于医疗用途，FDA 将如何监管最近注入 LLM 的基于人工智能的医疗技术？表 3 总结了这些监管挑战。

表 3 与 LLM 引起的相关监管挑战清单

监管挑战	简短描述
患者数据隐私	确保用于训练大语言模型的患者数据完全匿名并防止潜在的违规行为。这带来了重大的监管挑战，因为根据美国的 HIPAA 等隐私法，任何违规行为都可能导致严重后果。
知识产权	如果 LLM 生成类似于专有医学研究或文献的内容，则可能导致有关知识产权的问题。
医疗事故责任	确定当人工智能的建议导致患者受到伤害时，谁应该负责。是人工智能开发人员、使用它的医疗保健专业人员，还是采用它的机构？
质量控制与标准化	需要监管以确保人工智能生成的医疗建议的可靠性和一致性，这些建议可能会根据用于训练人工智能的数据而有所不同。
知情同意	当人工智能工具用于他们的医疗保健管理时，需要通知患者并给予同意。这是具有挑战性的，因为患者可能很难完全理解人工智能使用的含义。
可解释性和透明度	法规需要确保人工智能如何做出决策的透明度。这对于人工智能模型来说尤其具有挑战性，由于其复杂的算法，这些模型通常被称为“黑匣子”。
公平与偏见	需要监管来防止人工智能模型中的偏差，这些偏差可以在训练过程中使用患者数据引入。这可能导致医疗保健结果的差异。
数据所有权	定义和规范谁拥有大语言模型学习的数据可能具有挑战性，尤其是在涉及患者数据时。
过度依赖人工智能模型	过度依赖人工智能可能会导致人类专业知识下降，如果人工智能出现故障或提供不正确的信息，则可能导致潜在的错误。需要制定法规来平衡人工智能和人类专业知识的使用。
持续监控和验证	确保人工智能工具在不同人群中的持续性能、准确性和有效性是一项关键的监管挑战。

已经有关于监管 LLM 的建议，尽管这些建议来自医疗保健之外。在一篇工作论文中，Hacker 等人提出了一个新的术语，通过区分开发人员、部署人员、专业和非专业用户以及 LLM 输出的接收者来捕获 AI 价值链。作者还提出了四种策略，以确保这些模型值得信赖，并为整个社会的利益而部署。详细地说，监管应侧重于具体的高风险应用，而不是预先训练的模型本身，并应包括（i）有关透明度的义务，（ii）风险管理，（iii）非歧视条款，以及（iv）内容审核规则。

Mökander 指出，现有的审计程序未能解决 LLM 带来的治理挑战，并提出了三项贡献来填补这一空白：1）确定需要制定新的审计程序，以捕捉 LLM 带来的风险;2）通过借鉴 IT 治理和系统工程的最佳实践，概述以可行和有效的方式审计 LLM 的蓝图;3）讨论审计 LLM 前景的局限性。

这些潜在的解决方案可以作为医疗保健新法规的基准。无论哪种情况，监管机构和立法者都需要迅速采取行动，以跟踪 LLM 前所未有的演变和进步的动态。

作为监管机构面临越来越大压力的一个迹象，2023 年 3 月，埃隆·马斯克(Elon Musk)和史蒂夫·沃兹尼亚克(Steve Wozniak)等一批杰出的计算机科学家和科技行业高管呼吁“所有人工智能实验室立即暂停对比 GPT-4 更强大的人工智能系统的训练，至少暂停 6 个月”。他们在信中提到，“最近几个月，人工智能实验室陷入了一场失控的竞赛，开发和部署越来越强大的数字思维，没有人——甚至包括它们的创造者——能够理解、预测或可靠地控制这些思维。这种暂停应该是公开和可以核查的，并包括所有关键行为体。如果这种暂停不能迅速实施，政府应该介入并暂停。”

著名的人工智能专家，如吴恩达（Andrew Ng）反对这个想法，而是呼吁在人工智能创造的巨大价值与现实风险之间寻求平衡。我们同意，除非政府介入，否则暂停实际上无法实施，因此，暂停是无法实施的。“让政府暂停他们不了解的新兴技术是反竞争的，开创了一个可怕的先例，也是可怕的创新政策”。

值得一提的是，由于隐私问题和缺乏适当的监管，意大利于 2023 年 4 月成为第一个暂时屏蔽 ChatGPT 的西方国家。

6. 结论

LLM 为医疗保健的未来提供了巨大希望，但它们的使用也带来了风险和伦理挑战。通过采取积极主动的监管方法，可以利用 LLM 等人工智能驱动技术的潜力，同时最大限度地减少潜在危害，并保持患者和医疗保健提供者的信任。

此外，LLM 也可能成为第一类通过实施患者设计来监管的基于人工智能的医疗技术，这意味着，监管机构最终将让患者参与最高级别的决策，确保这些以令人难以置信的速度发展的人工智能工具得到监管，以满足现实生活中的临床和患者需求。

我们在这里总结了可以期望监管机构在将 LLM 引入医学实践时所做的事。

-为 LLM 创建一个新的监管类，因为 LLM 与已经通过监管的基于人工智能的医疗技术明显不同。

-为公司和医疗保健组织提供关于如何将 LLM 部署到其现有产品和服务中的监管指导。

-创建一个监管框架，不仅涵盖基于文本的交互，还包括可能的未来迭代，如分析声音或视频。

-提供一个框架，用于区分专门针对医疗数据进行培训的 LLM 和针对非医疗目的进行培训的 LLM。

-与 FDA 的数字健康预证书计划类似，对开发 LLM 的公司进行监管，而不是对每个 LLM 迭代进行监管。

***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。**