

数智健康国际动态

北京市卫生健康大数据与政策研究中心

2025. 6. 26

（六）数智药学

人工智能（AI）的飞速发展正给医疗健康发展带来前所未有的机遇，已在辅助诊断、药物研发、健康管理、疾病预测和医院管理等多方面展现出巨大潜力。本期专题将聚焦 AI 在药物研发与应用中展现的特有价值。

众所周知，传统药物研发耗资巨大、周期漫长，动辄需要十年和数十亿美元投入。AI 的引入则可有效克服这些困境，在高效识别药物靶点、筛选优化分子、助力老药新用，并在数据管理、药物发现、疾病治疗等各方面都表现出独特优势，甚至引发整个行业的变革。下面我们通过两篇代表性论文一窥 AI 为药物研发和应用带来的机遇与挑战。

第一篇文章系统梳理了 AI 在药物开发中的最新进展及其对药物化学和医疗保健的影响。文章首先清晰地介绍了 AI 在制药领域的核心概念、应用类别和工作原理，并展示了其在药物设计、递送系统、疾病治疗等方面的具体价值；接着详细探讨了 AI 在 3D 打印药物、纳米技术和预测疗法中的前沿应用，并客观分析了行业采用 AI 面临的挑战；最后得出核心结论：AI 已在多个药物发现环节证明其价值，能显著提升研发效率和提供新视角，但需要注意的是，它并非万能药，其研发与应用也需要科学管理和风险控制。

第二篇文章则通过一个探索性案例来揭示 AI 在实际应用中的挑战。研究利用最新 AI 工具 ChatGPT 评估其向患者提供药物说明的准确性。研究选取 4 名药学专家和 33 名药学博士实习生，采用问卷调查的方式，评估了 ChatGPT 对三种常见药物提供说明的准确性，并与权威数据库进行了对照。研究发现：虽然 ChatGPT 能生成基础的药物说明信息，但其回答仍存在一些信息缺失和可能的误导风险，尤其对患者自主用药而言可靠性不足。专家与实习生的评估存在显著的差异，也凸显了经验对判断的重要性。因此，研究强调 ChatGPT 目前还不能作为患者独立的药物信息来源，

必须经过专家验证，未来仍需持续改进模型和加强监管。

这两篇文献，从药品研发到药物使用，生动展现了 AI 赋能药物研发与应用的双重面貌：一方面，它带来令人振奋的机遇——有望实现更高效、低成本、精准化的药物研发；另一方面，它也面临严峻的挑战——即使像提供用药说明这样看似简单的任务，也暴露了在准确性、完整性、安全性上的不足，更不必说复杂的药物研发全流程。而克服这些挑战还绝非易事，需要全社会甚至全球产学研医管各界开放协作、携手创新实现：共同构建安全共享的数据治理环境、突破核心技术瓶颈、建立敏捷可靠的监管框架和大力培养跨学科复合型人才。唯有如此，我们才能驾驭 AI 的强大引擎，突破人类医药服务边界，将这场深刻的医药革命真正转化为普惠大众、提升生命质量的力量源泉。

（徐健编辑）

译文一：

人工智能技术在制药领域的兴起综述

Ayesha Sultana, Rahath Maseera, Abdul Rahamanulla, Alima Misiriya, 徐健 (译)

来源: Futur J Pharm Sci.

时间: 2023 年 8 月.

链接: <https://doi.org/10.1186/s43094-023-00517-w>.

背景

智能的一个关键组成部分是逻辑推理能力，这长期以来一直是 AI 研究的主要领域。1955-1956 年，兰德公司和卡内基梅隆大学的 Allen Newell、J. Clifford Shaw 和 Herbert Simon 创建了定理证明程序，被认为是该领域的一个重要里程碑。数字计算机或由计算机控制的机器人执行通常由智能生物完成的的任务的能力称为 AI。该短语经常用于开发具有类似于人类认知能力的 AI 系统，例如推理、寻找意义、概括和从经验中学习的能力。自 1940 年代第一台数字计算机被开发出来以来，人们已经确定，计算机可以被编程以执行极具挑战性的工作。

当今制药市场，是在经历漫长而昂贵的药物开发过程后才能接受药物。大多数药物耗资数十亿美元，需要 10 年或更长时间才能进入药品市场。这需要更多的资金和更多的时间进行药物开发。AI 的概念似乎更有希望克服这些缺点，并且应该会带来可功的药物开发计划。AI 正在用于药物、假肢以及不断发展和先进的机器人等新技术。AI 在药物开发过程中的其他好处包括识别药物靶点、从数据库中提取具有化学修饰的分子，以及有时重新利用药物。

正文

很明显，人工智能的当前发展导致了許多可用于创造药品的技术发展。那些能够将这项技术用作战术工具的人将成为竞争“赢家”。将这种潜力转化为行动是困难的，因为成功应用程序的地方有机会提高生产力，并提高产品一致性和质量(图 1)。

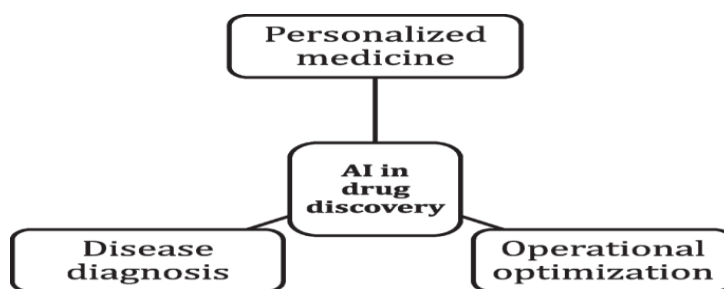


图 1 AI 趋势

AI 目标

- **专家系统开发:** 涉及创建智能运行的自动化系统，并为人们提供如何开展工作的建议。
- **机器人工智能:** 它将帮助计算机中类似认知模式的发展，使它们能够像人一样行事并采取必要措施来解决具有挑战性的情况。
- **多领域应用:** AI 将有助于实现多个领域，例如心理学、医学、伦理学、自然科学、医疗保健等。

文献综述

在一项研究中，模型树（AI 的一种形式）被用于对速释片剂的配方数据库进行建模。人工神经网络是众所周知的，经常用于药品配方学科，用于将模型的性能与这些网络进行比较。相关系数（ R^2 ）用于评估衍生模型的可预测性。从模型树研究中得出的多元线性方程在质量上与神经网络模型相似，能够预测片剂拉伸强度、崩解时间和药物溶出曲线。然而，这些方程式发现了在配方数据库中保密的其他重要数据。模型树作为一种透明的技术，被认为是公式设计师的有用工具。

通过分析激酶的化学结构，作者能够使用 KinomeX（一个基于深度神经网络的基于 AI 的在线平台）确定它们的多药理学。该平台使用的 DNN 是从 300 多种激酶中获得的 14,000 个生物活性数据点上训练的。因此，它在确定药物对激酶家族和特定激酶亚家族的一般选择性方面具有实际意义，有助于各种修饰剂的开发。一个值得注意的例子是 Cyclica 基于云的蛋白质组筛选 AI 技术的配体表达。使用这种方法可以寻找可以与某种小化学物质同时具有脱靶和脱靶相互作用的受体。这有助于解释药物的潜在负面影响。

AI 预测药物与靶点相互作用的能力也被用于避免多药理学并帮助重新利用药物。已重新利用的药物有资格立即进行 II 期临床试验。这也降低了支出，因为重新推出现有药物比引入全新的药理学实体（4130 万美元）更便宜（840 万美元）。可以使用“guilt by association”技术来预测药物与疾病之间的新关系，该技术是基于知识的或计算驱动的。机器学习（ML）方法利用支持向量机（SVM）、神经网络（NN）、逻辑回归和深度学习（DL）等方法，在计算驱动的网络中很受欢迎。在重新利用药物时，PREDICT、SPACE 和 logistic 回归平台等 ML 技术会考虑药物的基因表达谱。

制药行业 AI 的基本概念

AI 开发分为两大类。第一类包括技术方法和软件，例如专家系统，它们模仿人类经验并从一组规则中得出结论。第二类是人工神经网络（ANN），模拟大脑运作方式的设备，例如人工神经网络。人工神经网络的泛化能力是其重要特征，非常适合处理药品开发中与配方优化相关的问题（表 1）。

表 1 基于 AI 的药物递送计算工具列表

工具	描述	网站 URL	引用
Alpha fold	蛋白质 3D 结构预测	https://deepmind.com/blog/alphafold	[10]
Chemputer	用于报告化学合成程序的更标准化格式	https://zenodo.org/record/1481731	[11]
DeepChem	基于 python 的 AI 工具，用于各种药物输送任务预测	https://github.com/deepchem/deepchem	[12]
Deep neural net-QSAR	分子活性预测	https://github.com/Merck/DeepNeuralNet-QSAR	[13]
Deep tox	毒性预测	www.bioinf.jku.at/research/DeepTox	[14]
Delta vina	用于重新评分蛋白质-配体结合亲和力的评分函数	https://github.com/chengwang88/deltavia	[15]
Hit dexter	用于预测可能对生化分析有反应的分子的 ML 模型	http://hitdexter2.zbh.uni-hamburg.de	[16]
Neural graph fingerprints	新分子的性质预测	https://github.com/HIPS/neural-fingerprint	[17]

工具	描述	网站 URL	引用
NNScore	基于神经网络的蛋白质-配体相互作用评分函数	https://pubs.acs.org/doi/full/10.1021/acscen-tsci.8b00507	[18]
ODDT	用于化学信息学和分子建模的综合工具包	https://github.com/aspuru-guzik-group/ORGANIC	[19]

AI 原理

构建 AI 系统涉及在机器中仔细复制人类的特征和技能，并利用机器的处理能力来超越我们的才能。要了解 AI 的不同子领域以及它们如何应用于行业的不同领域，需要深入研究该主题。

- **机器学习 (ML)**: ML 教会计算机如何根据先验知识得出结论并做出决策。在不依赖人类经验的情况下，它识别模式并检查历史数据，以推断这些数据点的重要性并得出可能的结论。

- **深度学习 (DL)**: 深度学习是一种机器学习技术。DL 在 2000 年代重振了神经网络，它训练了更深的网络。为了对结果进行分类、推断和预测，它教机器通过层来解释输入。例如，它有助于了解复杂的内部表示，这对于理解困难的语言或分析对象是必要的，方法是通过深入的活动向量层，并通过了解随机梯度来找到激励这些向量的连接强度。

- **神经网络 (NN)** 这些系统以类似于人类神经细胞的方式运作。它们是一组算法，通过捕获众多潜在变量之间的关系来模拟人脑的工作方式。

- **机器阅读、理解和解释语言的能力称为自然语言处理 (NLP)**。一旦计算机识别出用户的预期消息，它就会做出适当的反应。

- **计算机视觉**: 通过剖析图像并检查物品的各个方面，计算机视觉算法试图理解图像。这有助于机器对一组照片进行分类和学习，使其能够根据先前的观察结果产生卓越的结果。

- **认知计算**: 通过以与人类相同的方式分析文本、音频、图像和其他输入，认知计算算法试图模拟人脑的功能并提供所需的结果。此外，报名参加有关 AI 应用的免费课程（图 2）。

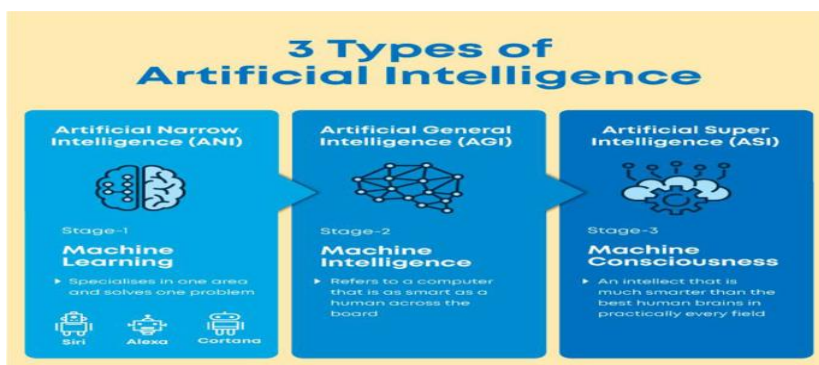


图 2 AI 的三种类型

AI 在制药和药物输送中的应用

当 AI 系统用于监管制造或临床试验等流程时，长期学习的力量通常会在训练后被消除。然而，自从最近采用质量源于设计（QbD）方法以来，制药行业已经有所改善，最新的工业 4.0 举措似乎描绘了一个快速发展的行业。因此，如果开发出早期的 AI 应用程序，它很有可能被投入使用。与其他科学领域相比，制药科学可能会导致数据编纂和标准化的延迟。数据积累和标准化对于前者中有效训练 AI 至关重要（图 3）。

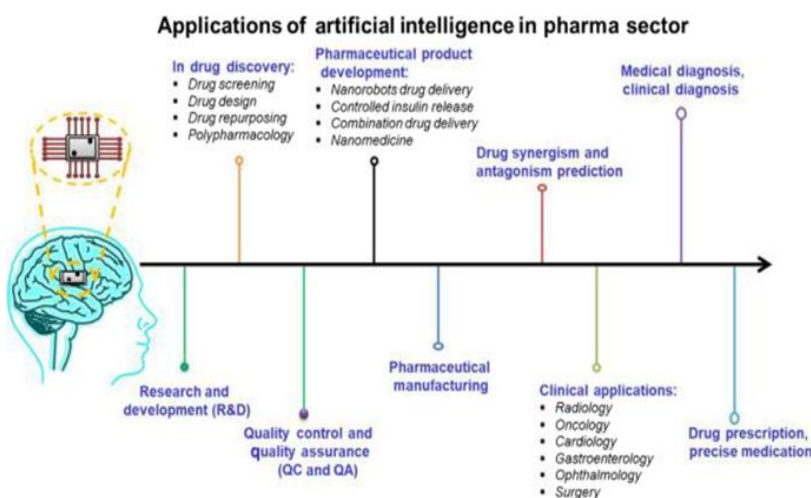


图 3 AI 在制药领域的应用

AI 应用

以下是如何在数据处理中使用 AI 的一些示例：

1. 产生最相关结果的数据搜索和搜索引擎优化。
2. 可用于执行一系列依赖于参数的指令的 If - then 的逻辑链。

3. 模式检测，在大量数据集中发现值得关注的模式，以获得原始见解
4. 使用概率模型预测未来结果。

医疗保健领域 AI

- **管理:** AI 系统正在协助日常管理活动，以最大限度地减少人为错误并最大限度地提高生产力。医疗记录的自然语言处理（NLP）转录有助于组织患者信息，以便临床医生可以更轻松地阅读这些信息。
- **远程医疗:** 在非紧急情况下，患者可以联系医院的 AI 系统来分析他们的症状，输入他们的生命体征，并确定他们是否需要医疗帮助。只给他们最紧急的情况可以减轻医学专家的工作量。
- **辅助诊断:** 得益于计算机视觉和卷积神经网络，AI 现在能够读取 MRI 扫描以比放射科医生快得多的速度寻找肿瘤和其他恶性生长，误差幅度明显缩小。
- **机器人手术:** 机器人手术的误差幅度非常小，可以每天 24 小时可靠地进行手术而不会感到疲劳。由于精度高，它们比传统技术侵入性更小，可以节省患者在医院康复所需的时间。
- **重要统计数据监测:** 对一个人的健康状况进行持续评估对于评估他们的表现是必要的。可穿戴技术的应用越来越广泛;但是，这些数据不容易获得，需要对其进行分析以提供有用的见解。许多应用可以挽救生命，因为生命体征甚至可以在患者意识到之前预测健康变化（图 4）。



图 4 AI 的优缺点

下一代 3D 打印药物的 AI

制药 3D 打印（3DP）管道和 AI 可以协同工作。个体化药物的给药必须取代医学中长期存在的“一刀切”概念。制药 3DP 可以在诊所提供定制药物，但现在它需要合格的 3DP 从业者的存在和技能。有许多标准的工艺优化工具，包括有限元分析（FEA）和机理建模，但没有一个能够彻底优化制药 3DP 的各个阶段。相比之下，ML 可以对创建 3DP 药物的每个步骤进行智能优化。这最终将消除对 3DP 药物开发的持续专家投入的要求，从而消除该技术临床实施的障碍（表 2）。

表 2 人工智能在制药行业使用的期望和策略

公司	使用 AI 的策略
Expectation 1	新药和生物制造商的开发
Novartis	为了对用新化学物质处理的细胞的数字图像进行分类，并将具有相似效果的化合物分组，诺华正在采用机器学习
Berg	在 1000 多个癌症和健康细胞样本中，创建了一个 AI 模型来揭示以前未被发现的癌症通路
Wuxi NextCODE	使用强大的模式识别在临床试验中开发和评估的药物
Expectations 2	解决以前难以治疗的疾病
Mission therapeutics	用于抗阿尔茨海默病和帕金森病的去泛素酶（DUB）抑制剂的 AI 辅助药物开发
Helax	利用 HealNet，这是一个人工智能平台，可帮助科学家识别更多治疗疾病的药物，同时最大限度地减少痛苦、费用和危险
Expectations 3	临床试验期间志愿者的药物依从性
CURATE	通过在特定模式水平上优化治疗剂量，开发了一个基于 AI 的平台来阻止晚期癌症疾病的进展
Expectations 4	更好地分析和利用临床数据
BenevolentBio	知识图谱是使用 AI 平台创建的，该平台从学术论文、专利、临床试验和患者记录等来源接收数据
Expectations 5	寻找合适的临床患者并缩短周转时间

公司	使用 AI 的策略
Santen and twoXAR	使用 AI 开发了一个用于发现、筛选和优先排序眼部疾病新候选药物的平台

采用纳米技术的 AI

由于当前分子商品的生产时间更长、成本更高、生产率降低，人工智能在制药业务、制药和药物输送中的重要性越来越大。然而，即使是当前制剂的开发也是基于耗时、昂贵且不可预测的错误研究。大数据、人工智能和多尺度建模方法正在通过一种称为“计算药学”的新系统集成到药学中，该系统正在对药物递送范式提出重大的潜在变化。该系统的出现是为了应对过去十年计算能力和算法的指数级增长。体内药代动力学参数、药物分布、物理稳定性、体外相关性以及物理和化学性质的预先制定以及活性预测都是将人工智能技术应用于药品开发的活动的例子(图. 5)。

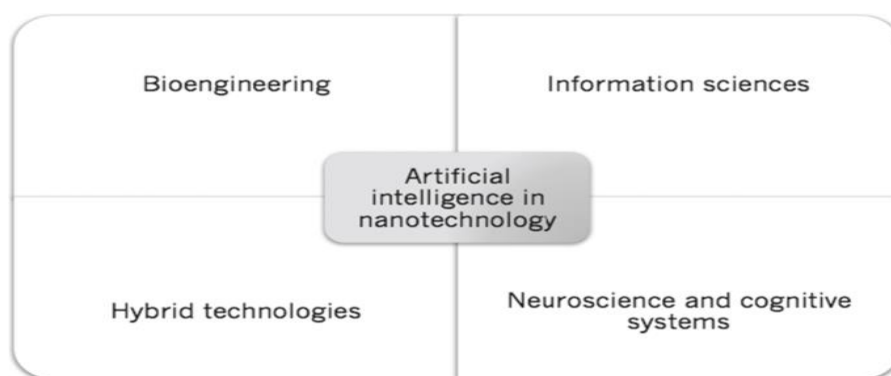


图 5 纳米技术 AI

AI 预测新疗法

AI 的改进重新激发了人们对罕见病治疗的兴趣。目前，全球有超过 3.5 亿人患有 7000 多种罕见疾病。为了研究治疗罕见疾病的新型药物，总部位于英国（英国）的生物技术公司 Heal 已获得 1000 万美元的资金。另一家瑞士生物技术公司 Thera chon 已获得 6000 万美元，用于开发治疗罕见遗传疾病的药物。

制药行业采用 AI

- 许多制药公司与科技公司和 AI 初创公司合作或收购，与致力于 AI 驱动的药物发现的专业企业和初创公司取得联系。这使他们能够利用自己的知识和工具，基

于已知的理论和经验构建有前途的候选治疗药物。

- 学术界的互动：随着制药公司开始接受人工智能，预计企业和学术界之间的合作伙伴关系将扩大。
- 建立内部专业知识并为员工提供他们需要的工具。
- 开放的科学倡议和研发挑战：这种用于药物研究的有用人工智能采用方法比以前的方法具有更低的财务风险（表 3）。

表 3 AI 开发

聚合和合成信息	将新的核糖核酸（RNA）测序技术与专有机机器学习相结合
了解疾病机制	全基因组筛选分析 提供蛋白质的详细结构 3D 结构 身份识别参与调节细胞周期的蛋白 发现下一代癌症疗法 训练计算机视觉和医学学习模型
开发新型候选药物	基于结构的深度（卷积神经网络）CNN 筛选化合物库对疾病的疗效 基于网络的机器学习方法 预测小分子的生物活性 确定生物靶标 预测每个代谢物质量的身份

AI 挑战

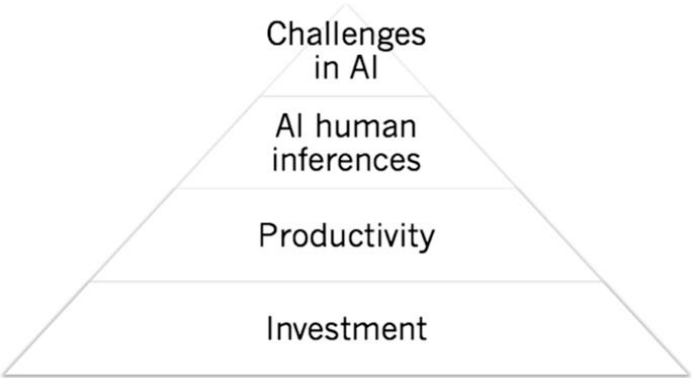


图 6 AI 面临的挑战

- 鉴于该技术的新且深奥，不为人知，许多制药企业仍将 AI 视为“黑匣子”。
- IT 基础设施不足是因为当今使用的大多数 IT 程序和系统在创建或规划时都没有考虑到人工智能。更糟糕的是，制药公司必须投入大量资金来改善其 IT 基础设施。

- 由于大部分数据是自由文本格式，因此制药公司必须采取额外的步骤，以分析友好的方式收集和排列数据。尽管存在这些限制，但毫无疑问，人工智能已经在改变生物技术和制药行业（图 6）。

结论

AI 已在各种药物发现领域证明了其实用性。AI 可以帮助科学家进行药物开发和交付的设计、规划、质量管理、维护和质量控制。它不是万能的，不会在一夜之间带来翻天覆地的变化，但它有可能提高效率，提供有用的见解，并突出药物发现过程中的新视角。制药公司目前正在经历一场革命性的转变，在开发新的科学和实践时，风险还需得到谨慎管理。AI 在创新药研发过程中的成功将通过整合许多不熟悉的和新的领域来衡量。数据管理、药物发现、糖尿病治疗、数字咨询和 AI 的其他用途可用于该领域。有大量证据表明，医疗 AI 可以帮助医生和患者在 21 世纪更有效地提供医疗保健服务。

***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：

ChatGPT 向患者提供药物说明的准确性评估： 一个探索性案例研究

Norah Othman Abanmy, Nadia Al-Ghreimil, Jawza F Alsabhan

Heyam Al-Baity, Rana Aljadeed, 徐健（译）

来源：Front Artif Intell.

时间：2025 年 4 月.

链接：<https://doi.org/10.3389/frai.2025.1550591>.

1. 引言

自 2022 年 11 月向公众推出 ChatGPT 以来，拥有超过 1300 万用户，生成的输出或响应数量可观。但是，无法保证生成的响应的正确性，可能有时不准确和不可靠。

ChatGPT 在医疗保健领域的几种潜在应用已被确定，包括自动诊断、个性化推荐和虚拟患者咨询。

然而，ChatGPT 在医疗保健环境中的应用有几个局限性和伦理考量。ChatGPT 在医疗保健环境中的一个相当大的缺点就是它缺乏特定领域的经验。尽管它可以产生与个人类似的反应，但它可能不具备提供可靠医疗建议所需的知识和经验深度。不正确的诊断、对患者护理的不当建议或对现有事实的误解可能是由于信息不足或不准确造成的。此外，严重的隐私问题与医疗专业人员的 ChatGPT 应用程序有关。如果底层人工智能（AI）模型分析或保留交互期间提供的患者数据，则敏感信息可能会泄露。

本研究旨在评估 ChatGPT 在药房应用程序上下文中提供准确信息的能力。由于 ChatGPT 有可能提高患者参与度、提供与药物相关的信息并支持医疗保健服务，因此评估其准确性和有效性至关重要。本研究考察了 ChatGPT 对三种常见药物 tirzepatide, citalopram 和 apixaban 的反应，并将它们与 Micromedex 的患者指导报告进行了比较。

2. 材料和方法

这是一个案例研究，旨在调查 ChatGPT 生成的患者指示报告的准确性，其中 CareNotes 被选为比较参考。这是通过对 37 名参与者（4 名药学专家和 33 名药学博士实习生）的在线调查完成的，如下所述。

2.1. 数据采集

在线调查开始时，参与者被明确告知第一份报告来自 CareNotes，第二份报告是使用“大型语言模型 ChatGPT”生成，并收到问卷设置的描述和如何回答的说明。所有参与者都提供了参与该研究的书面知情同意书，并且已获得沙特国王大学机构审查委员会对本研究的 IRB 批准（编号：KSU-HE-23-1266）。

以下药物 tirzepatide, citalopram 和 apixaban 是由两名研究人员（NOA 和 JFA）在咨询了同一部门的临床药剂师后选择的。人们一致认为患者对这些药物的指导很重要，并且它们通常开给当地社区的患者。

使用了可公开访问的免费 ChatGPT3.5 版（于 2023 年 12 月 21 日发布）。针对 ChatGPT 的问题由两名研究人员（NA 和 JFA）执行，通过提出以下问题来模拟患者在现实生活中的担忧：“我必须服用[药物名称]。我应该如何使用它，我应该注意什么？研究人员用[药物名称]代替了上述三种药物中的一种。

ChatGPT 被提示五次，为三种药物中的每一种产生五种不同的反应。这样做是为了解释 ChatGPT 文本输出的可变性，并实现对其生成能力的良好覆盖。每次聊天会话都会重新启动，以确保公正的响应；也就是说，我们为每种药物生成了五个不同的报告版本。所有 15 份 ChatGPT 生成的报告均可应要求提供。

2.2. 问卷

通过在查阅已发布的问卷后回答为此目的开发的问卷，测试了 ChatGPT 生成的报告与 CareNotes 报告的相似性和信息准确性。来自沙特阿拉伯利雅得沙特国王大学药学院的 4 名药学专家和 33 名药学博士实习生参与了评估。参与者被要求通过独立回答问卷来对 ChatGPT 报告进行评分。每位专家针对所有 5 个 ChatGPT 反应评估

了三种药物，即每个专家进行了 15 次评估。每个实习生对每种选定的药物评估一个版本，即每个实习生进行三次评估。

问卷包括 4 个问题，用于测试报告的正确性、完整性、危害性和药物依从性差。这四个术语的定义如下。

- 正确性：正确的信息与 CareNotes 报告中的信息类似。
- 完整性：与药物说明相关的所有关键医疗信息。
- 有害性：患者可能错误地解释信息，这可能导致身体或心理伤害或滥用药物。
- 对服药依从性的负面影响：患者可能会错误地解释信息，这可能会阻碍他/她遵守医疗保健提供者的建议。

每个问卷都包含三个块：（i）CareNotes 报告，（ii）ChatGPT 生成的报告的单一版本，以及（iii）一系列评估 ChatGPT 报告准确性的问题。我们要求参与者在五点李克特量表（制定为陈述）上他们对每个标准的同意程度进行评分。此外，每个问题都附有后续问题，我们要求参与者为他们的评估提供证据。

1. 事实正确性：“ChatGPT 报告的事实正确。”

跟进：“复制 ChatGPT 报告的所有不正确的文本段落（如果适用）。”

2. 完整性：“患者的相关医疗信息包含在 ChatGPT 报告中。”

跟进：“在 ChatGPT 报告中列出所有缺失的医疗信息（如果适用）。”

3. 潜在危害：“ChatGPT 报告导致患者得出错误的结论，这可能会导致身体、心理伤害和/或滥用药物。”

跟进：“列出所有可能有害的结论，这些结论可能来自 ChatGPT 报告（如果适用）。”

4. 依从性差的可能性：“ChatGPT 报告导致患者得出错误的结论，这可能会导致对药物的依从性差。”

跟进：“列出所有可能错误的结论，这些结论可能来自 ChatGPT 报告，导致依从性差（如果适用）。”

2.3. 数据分析

收集问卷并检查完整性。针对三种药物（tirzepatide、citaropram 和 apixaban）中的每一种，评估参与者在李克特量表上的事实正确性、完整性、潜在危

害和依从性差的可能性评分，并以文本中的百分比和数字中的权重报告。

以下方程用于对药物（ d ）的单个版本（ v ）的单个测量（ m ）进行加权平均评分：

$$weight_{d,v,m} = \frac{1}{responses} \sum_{i=1}^5 count_{likert_i} * i \quad (1)$$

其中，响应数是针对该版本收集的响应总数，计数 $_{likert_i}$ 是特定李克特 $_i$ 评级的响应数，因此，李克特 $_1$ 对应于“强烈不同意”评级，李克特 $_5$ 对应于“强烈同意”评级。

对“正确性”和“完整性”措施的“强烈同意”和“同意”回答被视为积极，而“潜在危害”或“可能较差的依从性”措施被视为负面。因此，我们计算了补码。对于后两个（负测量），分数计算如下：

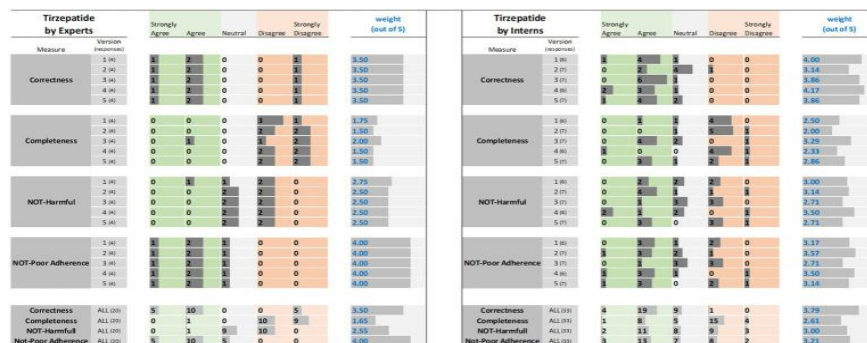
$$weight_{d,v,m} = \frac{1}{responses} \sum_{i=1}^5 count_{likert_i} * (5 - i + 1) \quad (2)$$

方程 1 适用于“完整性”和“正确性”的测量，而方程 2 适用于“潜在危害”和“依从性差的可能性”，因此在图中显示为“无害”和“依从性不差”。

3. 结果

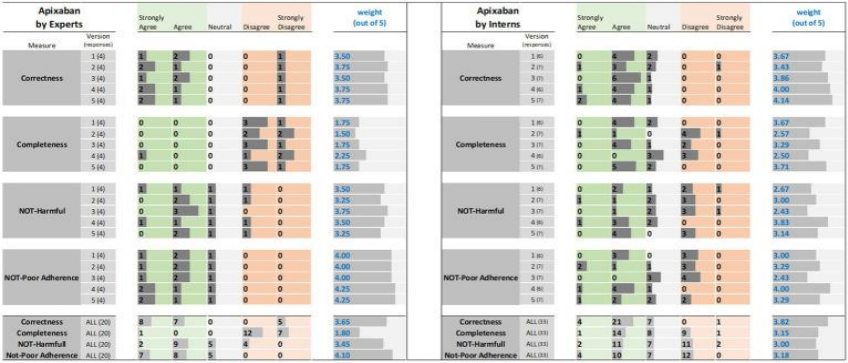
4 名专家和 33 名药学博士实习生参与了对三种常用处方药（tirzepatide、citaropram 和 apixaban）患者咨询的 ChatGPT 报告的评估。每个实习生只评估了三种药物中的每一种的一个版本。因此，版本 1 和 4 由 6 名实习生评估，而版本 2、3 和 5 由 7 名实习生评估。作为第二层验证，药房专家评估了 ChatGPT 为每种药物生成的所有五个版本的报告。

图 1.



Tirzepatide 评估结果：（左）专家和（右）实习生。结果（即单个评分和权重）用水平数据条编码，并按度量、版本划分，然后在图底部显示所有五个版本的总数。每个版本号旁边的括号内是收集的响应数。“强烈同意”和“同意”列为绿色，以反映正面评级。请注意，在绿色区域内具有更多和更长的深灰色条对应于较高的权重。

图 3.



apixaban 评估结果：（左）专家和（右）实习生。结果（即单个评分和权重）用水平数据条编码，并按度量、版本划分，然后在图底部显示所有五个版本的总数。每个版本号旁边的括号内是收集的响应数。“强烈同意”和“同意”列为绿色，以反映正面评级。请注意，在绿色区域内具有更多和更长的深灰色条对应于较高的权重。

专家和实习生对 tirzepatide、citalopram 和 apixaban 的 ChatGPT 报告的评估结果分别显示在图 1-3 中。

3.1. 专家评估

以下各节描述了对三种药物的 ChatGPT 报告的专家评估结果（图 1-3）。

3.1.1. 正确性测量

对于 tirzepatide 和 apixaban，四分之三的专家同意 ChatGPT 报告（包括五个版本）在事实上是正确的。ChatGPT 报告中第四位专家指出的错误被 tirzepatide 和 apixaban 注意到。这些包括某些版本中 tirzepatide 和 apixaban 的给药频率不正确、使用 tirzepatide 进行低血糖管理的说明不正确、监测策略以及如何处理错过的 apixaban 剂量。对于 citalopram，两位专家不同意版本 1、2 和 3 的正确性。他们认为，尽管 CareNotes 报告中没有提及这些信息，但与食物相关的给药时间说明是不正确的。然而，citalopram 的第 2 版和第 4 版被两位不同的专家认为是错误的。

3.1.2. 完整性度量

所有专家都认为，ChatGPT 生成的报告不完整，除了 citalopram 第 4 版、tirzepatide 第 3 版和 apixaban 第 4 版，据报道每份报告都只由一名专家完成。ChatGPT 未涵盖的领域是给药说明中的步骤、储存条件、禁忌症或警告、药物相互作用、严重副作用（甲状腺癌、眼睛或视力问题）、错过剂量的指导、怀孕期间的预防措施和母乳喂养的母亲。对于 citalopram，缺失的信息包括关于漏服、药物相互作用、副作用、储存条件和药物作用发作的指导。对于 apixaban，缺失的信息包括药物相互作用、漏服剂量、怀孕和哺乳期间的预防措施、警告和过敏反应。

3.1.3. 潜在危害措施

还评估了对 ChatGPT 报告的潜在危害。经证实，tirzepatide 报告可能会导致错误的结论，除了一位关于第 1 版的专家外，所有专家都同意这一点。据报道，tirzepatide 的 1 版是有害的，因为没有提及：（1）有关药物不良反应的任何信息，主要是甲状腺癌、胰腺炎和眼睛或视力问题；（2）药物相互作用和储存条件；或（3）如何处理漏服。就 citalopram 而言，只有一位专家报告了 ChatGPT 的第 1、2 和 4 版对患者造成潜在伤害，而第 3 版则由两位专家报告具有潜在危害，因为没有提到关于心律问题的预防措施，血清素综合征（与某些其他药物一起使用时可能危及生命），以及出血风险增加。关于 apixaban，四位专家之间没有达成共识。一位专家认为版本 2、4 和 5 可能会造成伤害，因为 ChatGPT 报告中没有关于突然停药 apixaban 的信息，这可能会导致中风和血栓。另一位专家认为版本 1 可能会造成伤害，因为没有关于药物与食物相互作用的信息。一位专家的回应是中立的，没有评论，而另一位专家不同意 ChatGPT 报告可能会对患者造成潜在伤害。

3.1.4. 依从性差的可能性

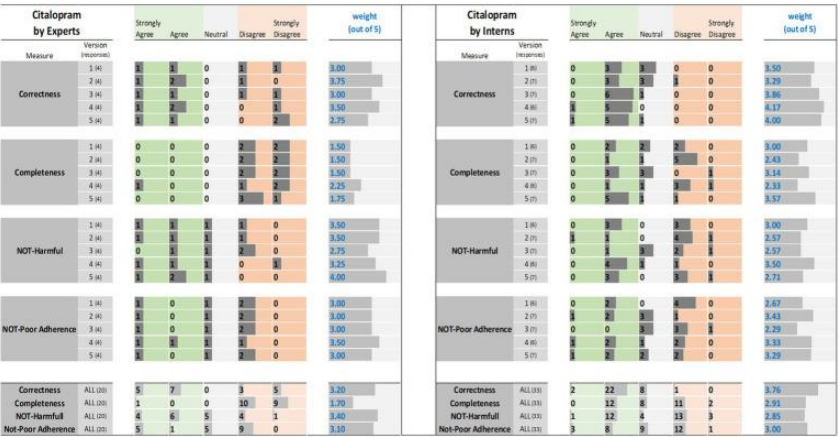
三位专家不同意 tirzepatide 和 apixaban ChatGPT 报告可能导致依从性差，一位专家持中立态度，没有发表评论。然而，对于 citalopram，两位专家一致认为它可能会导致依从性差，因为 ChatGPT 报告提到睡前服用该药物，这会导致失眠，从而导致依从性差，而不吃药可能会导致因胃不适而导致依从性差，此外还应明确解释（4 周内）药物作用的发生问题，以帮助患者继续服药。

3.2. 药学博士实习生评估

3.2.1. 正确性测量

实习生对 tirzepatide 的评估表明，第 3 版对 ChatGPT 报告的正确性一致性最高（85.7%），其次是第 1 版和第 4 版（83.3%）（[图 1](#)）。除了一名实习生外，没有对 tirzepatide 第 2 版的分歧。citalopram 第 4 版的一致性最高（100%），其次是第 3 版和第 5 版（均为 85.7%），而一名实习生对第 2 版仅报告了 1 个不一致（[图 2](#)）。对于 apixaban，版本 3 和 5 的一致性最高（85.7%），其次是版本 4（83.3%）。然而，一名实习生不同意 apixaban 的第 2 版（[图 3](#)）。ChatGPT 报告中提到的 apixaban 错误点是建议在服药时限制或避免饮酒，因为酒精会增加出血的风险。然而，尚无关于 citalopram 或 tirzepatide 的报道。

图 2.



citalopram 评估结果：（左）专家和（右）实习生。结果（即单个评分和权重）用水平数据条编码，并按度量、版本划分，然后在图底部显示所有五个版本的总数。每个版本号旁边的括号内是收集的响应数。“强烈同意”和“同意”列为绿色，以反映正面评级。请注意，在绿色区域内具有更多和更长的深灰色条对应于较高的权重。

3.2.2. 完整性度量

只有 57% 的实习生同意 tirzepatide 的第 3 版已经完成，其次是第 5 版（42.9%）（[图 1](#)）。citalopram 的第 5 版完整性得分最高（71.4%），其次是第 3 版（42.9%）（[图 2](#)）。对于 apixaban，版本 5 的得分最高（71.4%），其次是版本 1（66.7%）（[图 3](#)）。

根据实习生的说法, ChatGPT 未涵盖的领域是有关 tirzepatide 品牌名称、主要禁忌症、警告/预防措施、给药技术、错过剂量说明、治疗期间的监测参数、储存条件、药物相互作用的一些细节, 以及一些副作用也被遗漏, 包括甲状腺癌、视力问题、低血糖和肾功能障碍。

ChatGPT 未涵盖的关于 citalopram 的领域包括品牌名称、可用剂型、错过剂量说明、主要与单胺氧化酶抑制剂的药物相互作用以及非处方药;禁忌症和注意事项;严重的副作用, 包括心律问题和血清素综合征;特殊人群的安全性, 例如孕妇和哺乳期母亲;药物停药说明;以及预期的药物作用发作。

对于 apixaban, 如果患者无法吞咽片剂, 则提供品牌名称、剂型和给药说明信息;错过剂量说明;可能的副作用, 例如过敏反应;患者预防措施, 尤其是在脊柱问题或背部手术的情况下;药物相互作用;禁忌;怀孕和哺乳期;警告;和治疗期间所需的监测实验室检查报告为缺失。

3.2.3. 潜在危害措施

57.1% 的实习生报告称 ChatGPTtirzepatide 第 5 版响应可能造成危害, 其次是第 3 版 (42.9%) ([图 1](#))。citalopram 第 2 版的有害得分最高 (71.4%), 其次是第 5 版 (57.1%) ([图 2](#))。apixaban 的第 1 版和第 2 版的有害得分最高, 分别为 57.1% 和 50% ([图 3](#))。实习生报告的原因如下: 有关禁忌症、怀孕和母乳喂养期间的预防措施、错过剂量说明、严重副作用、过敏反应、给药技巧和储存说明的信息可能会导致患者受伤。对于 citalopram, 没有发现明显的副作用, 例如心律问题、血清素综合征 (与某些其他药物一起使用时可能危及生命)、出血副作用风险增加、驾驶过程中的药物预防措施以及突然停药, 如果不提及, 所有这都会导致患者受伤。此外, 一位实习生提到, 在 ChatGPT 中报告这一说法后, “通常, 初始剂量很低, 可能会根据需要逐渐增加”, 当患者认为 citalopram 对他们不起作用时, 他们可能会尝试自行增加剂量。对于 apixaban, 没有关于漏服、药物-食物相互作用、硬膜外手术期间的禁忌症、妊娠期药物的安全性以及治疗期间所需的监测参数的信息, 包括出血的体征和症状, 所有这些都可能导致对患者造成伤害。

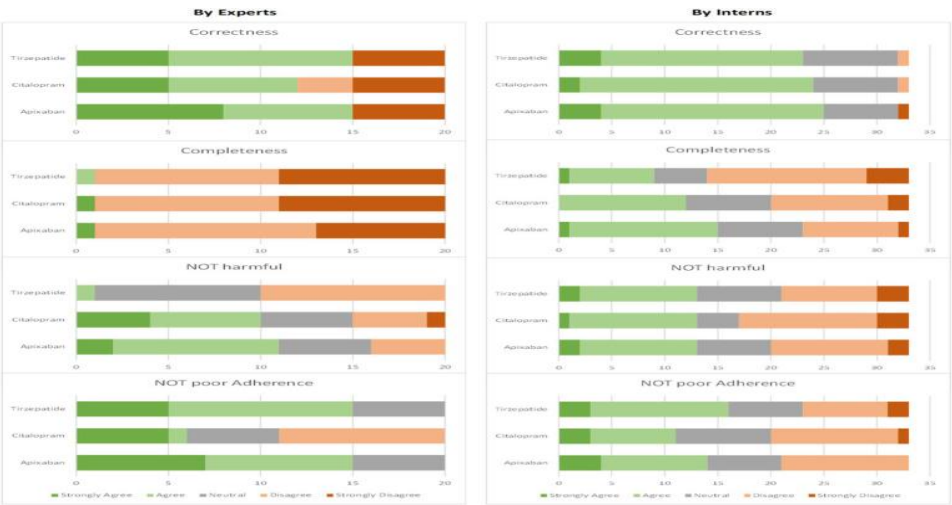
3.2.4. 依从性差的可能性

第 3 版和第 5 版的 tirzepatide 的 ChatGPT 报告有 42.9% 报告, 第 1 版报告

为 33.3% (图 1)。实习生对 citalopramChatGPT 报告进行评分的报告显示,这将导致第 1 版的依从性差为 66.7%,第 3 版的依从性为 57.1%(图 2)。在 apixaban 的情况下,57.1% 的实习生认为版本 3 最有可能导致依从性差,其次是版本 1 (50%) (图 3)。虽然在 CareNotes® 中没有提到,但 tirzepatide 的自由文本评估表明,实习生担心没有提到药物需要时间才能产生理想的效果,这可能会阻止患者继续服用。此外,未报告缺失的剂量说明。对于 citalopram,提及药物的自杀作用,而不是表示逐渐停药、不完全的副作用以及 citalopram 的延迟作用可能会导致依从性差。出血、不完全副作用和突然停药的风险也可能导致依从性差。

对专家和实习生评估的总体结果的全面看法表明,在实习生评估的情况下,有一种趋势是比专家评估更重视正确性和完整性的衡量标准。关于潜在的危害和依从性差的可能性,他们给出的权重低于专家的权重。专家和实习生的总体测量结果如图 4 所示。

图 4.



所有版本的每项措施的评估结果共同: (左) 专家 (4 名专家 × 5 个版本 = 20 个回复) 和 (右) 实习生 (33 个回复)。

4. 讨论

ChatGPT 的日益普及使 AI 能够对包括医疗保健在内的多个行业产生积极影响。然而, ChatGPT 输出的可靠性程度仍不确定。已发表的研究中强调了对从 AI 输出中收集的医疗信息的准确性和效用的担忧, 强调了仔细评估和评价的必要性。

对三种药物进行了 ChatGPT 和 CareNotes 报告之间的评估和比较。为了帮助评

估，该团队利用了 Micromedex 数据库。Micromedex 是一种广泛认可的在线参考工具，可为医疗保健专业人员和患者提供全面的药物信息。药学专家和 PharmD 实习生评估了 ChatGPT 生成的每份报告。他们采用结构化的评估流程，重点关注四个关键领域：正确性、完整性、潜在危害和对依从性的负面影响。

结果表明，虽然 ChatGPT 有望成为生成基本药物信息的工具，但它对于患者的自力更生来说还不够可靠。所有三种药物的报告完整性都存在重大差距，这可能会误导患者并导致潜在伤害或依从性差。

药学专家和 PharmD 实习生之间的评估分数存在显著差异。专家们更为挑剔，尤其是在 tirzepatide 等领域，不正确的给药频率是所有评估版本中的一致问题，一位专家强调了 ChatGPT 如何推荐与标准做法不同的频率。这可以防止患者错误地给药。此外，apixaban 报告缺乏管理漏服的具体说明，这是专家发现的一个严重遗漏，但实习生的评价较轻。例如，apixaban 的第 2 版被一位专家评为不完整，而实习生则认为它可以接受患者指导。

相比之下，实习生表现出更宽松的评价，尤其是 citalopram，尽管专家指出在使用的最初几周内遗漏了关于效果延迟发作和潜在副作用的遗漏，但其中一个版本从实习生那里获得了 100% 的正确性评分。

这种差异可归因于两组之间的经验和知识差距。具有广泛临床实践的专家更擅长识别报告中的潜在风险或缺失元素。实习生可能不完全了解药物信息准确性所涉及的复杂性，使他们更有可能忽视 ChatGPT 响应中的小问题。

此外，研究结果表明，ChatGPT 在提供全面的药物相关信息方面尚不完全可靠，尤其是对于患者指导。例如，在 tirzepatide 中，在所有版本中，专家和实习生都发现了缺失的关键安全性信息，包括缺乏糖尿病患者低血糖管理的指导，这是接受降糖治疗的患者的关键考虑因素。此外，apixaban 报告一直错过有关储存和特定预防措施的关键说明，如果患者仅依赖 AI 生成的内容做出医疗决策，这将带来重大风险。一位专家强调，如果不提供这些细节可能会导致危险的滥用，尤其是对于有潜在疾病的患者。

此外，citalopram 的第 1 版被专家评为特别不完整，其中一位专家指出没有关于血清素综合症的警告，血清素综合症是一种与选择性血清素再摄取抑制剂药物（如 citalopram）相关的可能危及生命的疾病。然而，实习生不太可能注意到这些

差距，因此正确性和完整性得分更高。

因此，虽然 ChatGPT 可以作为医疗保健环境中的补充工具，但未经专家验证，患者无法独立依赖其当前形式。完整性方面的差距和有害不准确的可能性意味着医疗保健专业人员必须在将信息提供给患者之前对其进行审查和验证。

在五个版本的 ChatGPT 药物报告之间观察到明显差异，尤其是在正确性和完整性方面。因此，该系统无法生成全面准确的药物信息。

我们的研究结果与之前的研究结果一致，这些研究调查了 ChatGPT 回答在回答药物信息问题方面的准确性和实用性。Morath 等人发现，在输入到 ChatGPT 的 50 个药物信息查询的回复中，只有 13 个被认为是正确的。其余的回答要么不正确，要么仅部分正确。在 27 个回复中发现了包含可能对患者健康不利的信息的回复；14 例被认为是“对患者造成伤害的低风险”，13 例被认为是“对患者造成伤害的高风险”。另一项评估 ChatGPT 作为药物相关问题有效且可靠资源的适当性的研究发现，10 个回复中只有 39 个通过了评估。其余的回答要么不准确、不完整，要么缺乏直接回答。

这项研究的主要优势在于，我们在患者信息的四个关键领域检查了 ChatGPT：正确性、完整性、潜在危害和对依从性的负面影响。药房实习生和专家完成了审核，以确保最终评估的稳健性和准确性。我们还使用了免费版本的 ChatGPT。这确保了我们的评估公众容易获得的药物信息的准确性。这项研究最显著的局限性之一是团队仅评估了三种药物。因此，这些评估的结果可能无法捕捉到与不同药物相关的其他领域。此外，不能排除专家和实习生在评估过程中存在的潜在偏见。此外，尽管研究人员试图模仿患者的问题，但值得注意的是，ChatGPT 生成的报告响应了专家要求的提示。不能保证患者会问类似的问题。缺乏 ChatGPT 知识的患者可能会根据初始查询和后续问题生成不同的报告。这限制了可用药物信息的数量。因此，需要进一步的研究来评估患者请求的 ChatGPT 报告和生成的药物信息的可读性。

未来检查 ChatGPT 在提高患者服药依从性方面的作用的工作也是强制性的，因为可以调查患者的素养、对 AI 生成内容的信任以及不完整或误导性信息。

总体而言，研究结果表明，ChatGPT 在完整性和安全性细节方面存在不一致，这表明在可以自信地将其用作患者药物信息的主要来源之前，还需要进一步改进。

5. 总结

ChatGPT 还不能被视为可靠的独立患者药物信息来源。需要持续的模型改进，例如将 AI 驱动的工具与专家验证集成，以达到能够始终如一地提供高质量、安全和全面的药物相关建议的水平。此外，为了提高 ChatGPT 的可靠性，监管指南是强制性的。

***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。