将人工智能应用于心理健康领域的临床决策支持：我们学到了什么？

作者： Grace Golden[1,2], BASc；Christina Popescu[2,3], BSc, MSc；Sonia Israel[2], BSc；Kelly Perlman[2,3], BSc；Caitrin Armstrong[2], BASc；Robert Fratila[2], BSc；Myriam Tanguay-Sela[2],BASc，& David Benrimoh[2,3,4], MD.CM., MSc., MSc., FRCPC

所属机构： [1]西安大略大学，加拿大安大略省伦敦；[2]Aifred Health Inc.，加拿大魁北克省蒙特利尔；[3]麦吉尔大学，加拿大魁北克省蒙特利尔；[4]斯坦福大学，美国加利福尼亚州帕洛阿尔托

作者贡献：

GG 和 DB 构思了本稿件，并参与了稿件撰写和审阅。CP、SI、KP、CA、RF 和 MTS 参与了稿件撰写和审阅。

摘要：采用人工智能（人工智能）模型增强的临床决策支持系统（临床决策支持系统）正成为医疗保健领域具有潜在价值的工具。尽管前景可期，这些系统的开发和实施通常会遇到若干障碍，阻碍其广泛采用的潜力。本文介绍了一项近期开发的人工智能临床决策支持系统 Aifred Health 的案例研究，该系统旨在支持重度抑郁障碍治疗方案的选择和管理。我们同时考察了在该人工智能临床决策支持系统的开发和测试过程中所遵循的原则，以及为促进实施而制定的实际解决方案。我们还提出了在人工智能临床决策支持系统的构建、验证、训练和实施过程中应考虑的建议。这些建议包括：识别关键问题；基于该问题选择机器学习方法的类型；确定所需的数据类型；确定临床决策支持系统为提供临床效用所需的形式；收集医生和患者反馈；以及在多种环境中验证该工具。最后，我们探讨了广泛采用这些系统的潜在益处，同时权衡了实施挑战，例如确保系统不会扰乱临床工作流程，以及以能够赢得最终用户信任的方式设计系统。

引言

临床决策支持系统是一类计算机化的医疗服务现场系统，旨在支持临床决策。[1] 这些系统整合大量临床信息，旨在通过改善临床医生的决策来提升医疗服务提供。[2,3] 临床决策支持系统是新兴工具，有潜力影响身体和心理医疗[4]，，但其在临床实践中的实施仍然有限。本文探讨临床决策支持系统开发与实施相关的障碍和挑战，提出在整个过程中应考虑的建议，并讨论广泛采用可能带来的益处，同时权衡这一过程中面临的挑战。本文将重点阐述的一个主题是：临床决策支持系统的开发和设计与其临床实施之间需要保持连续性。

临床决策支持系统的类型

临床决策支持系统历史悠久，已针对各种躯体及精神疾病开发（详情请参阅补充材料）。[5,6,7] 临床决策支持系统分为两类：基于知识的系统和非基于知识的系统。[1,3] 基于知识的临床决策支持系统建立在静态的循证规则和临床知识之上，可提供诊断或治疗建议，也可发出提醒。[1,3] 与此同时，非基于知识的系统会采用某种形式的统计模型，根据观测数据生成输出。[3] 其中一部分系统采用复杂的统计学习模型，即人工智能。非基于知识的临床决策支持系统也可以包含基于知识的临床决策支持系统（例如，提供预测的临床决策支持系统也可能包含指南信息）。本文将重点讨论集成了人工智能模型的非基于知识的临床决策支持系统（称为人工智能临床决策支持系统），这类系统也包含基于知识的系统的某些要素。

人工智能临床决策支持系统：障碍与挑战将人工智能整合到临床决策支持系统中

人工智能在医学预测中的应用研究日益受到关注。[8,9,10,11]迄今为止，这些人工智能模型大多被开发为独立工具，未集成到临床决策支持系统中，这可能会降低其在更广泛临床界中的应用率。一个可能的原因是，大多数模型源于基础研究，其主要目标是构建模型，而非在优先开发临床实用工具的转化研究背景下开展。[12] 在基础研究背景下开发的模型可能因多种原因而无法转化为临床决策支持系统。例如，模型本身可能存在问题（如所建模型无法推广至目标人群），或者未能将模型集成到临床工作流程中（如模型需要将临床实践中通常不会收集的复杂检测结果作为输入）。[13]

数据源选择注意事项

要将人工智能模型成功整合进临床决策支持系统，模型必须达到特定领域的准确性标准，而这一标准取决于不同领域对可接受风险和不确定性的不同要求。要生成准确的模型，必须有数量充足且质量足够高的数据，这些数据应涵盖目标人群，并对应于与待做决策相关的时间点。数据质量始于获取大量数据，通常来自电子病历（EMR）或临床试验资料库；数据集过小则可能导致输出存在偏差，或难以具备良好的泛化能力。[3,11] 即使模型能够准确、可靠地预测某个终点，选择该终点也至关重要，因为这决定了模型是否具有临床效用。例如，设想一个模型能够准确预测某种疾病的治疗结局，而该疾病只有一种有效治疗方案。在这种情况下，了解预测结果可能不会改变临床管理，因为临床医生不太可能剥夺患者尝试该治疗方案的机会。此外，如果没有临床理由选择某种治疗方案而非另一种，那么能够在多种治疗方案之间进行预测的模型，可能比在指南或临床经验规定了治疗方案尝试顺序时更有用。

临床实践中的采纳

在讨论临床决策支持系统的成败时，应根据特定领域以及系统旨在辅助的具体决策，采用相应指标进行评估。一个可能普遍适用的指标是使用率（即定期使用临床决策支持系统的临床医生人数）；虽然较高的使用率并不能保证系统产生影响，却是其大规模产生影响的先决条件。在常规临床实践中，临床决策支持系统的缺乏，尤其是人工智能临床决策支持系统的缺乏，十分引人注目。这表明，即便存在有用的临床决策支持系统，它们目前也因使用率低而未能产生显著的临床影响。为什么会这样？临床决策支持系统因耗时过多而受到批评[14]，临床医生对其接受度不高，原因是他们对人工智能生成结果缺乏信心，并且普遍不信任该系统[15,16,17]，医学领域一直存在抵制接受人工智能的现象，这至少部分源于机器学习模型缺乏透明度。[18,19] 鉴于医学的特点在一定程度上依赖机制信息，如果系统不解释为何或如何得出预测结果，医生可能不会信任模型预测背后的推理和决策。[18,19] 缺乏信任正是人工智能的“黑箱”问题，而人工智能可解释性方法的研究如今已成为该领域的重点。[20]

一项重要的批评意见是，将临床决策支持系统纳入临床实践会扰乱或干扰工作流程，并且常常导致医师挫败感和职业倦怠，这可能是由于缺乏训练或支持所致。[14,21] 这些例子表明，尽管临床决策支持系统具有改善临床护理和患者结局的潜力，但不当的设计或实施可能引入新的错误或问题，降低工具留存率，或完全阻碍其采纳。[22]

识别相关问题

最后，必须考虑临床决策支持系统设计的一项基本前提：开发临床决策支持系统不应只是为了采用新技术，而应着眼于解决临床医生需要协助的问题，重点解决那些尚无更简单或更有效方案的现存问题。[18]

重点必须放在支持医生需要协助作出的决策上，这也为本文接下来将探讨的案例研究提供了一个有益的切入点。以重度抑郁障碍（MDD）的治疗选择与管理为例。尽管目前有多种药物、心理治疗和神经刺激技术可用于治疗抑郁症[23],，但仍有三分之二的患者在首个疗程后未能达到缓解。[24] 患者及其医生在选择治疗方案时，往往不得不采取试错法。[25] 此外，医生通常会习惯于开具同一批有限的治疗方案（可能是因为熟悉这些方案的副作用和典型结局），结果是患者的初始治疗往往局限于少数几种方案，且通常是药物。[25,26] 除了治疗选择之外，治疗管理也面临挑战：尽管研究表明，使用算法能改善患者结局，但许多临床医生并未常规遵循指南。[27,28] 因此，对于重度抑郁障碍而言，似乎可以通过（1）基于知识的临床决策支持系统来解决治疗管理问题，以及（2）非基于知识的临床决策支持系统来解决治疗选择问题。要让临床决策支持系统解决这一问题，就必须将其设计为协助临床医生实现治疗选择个体化，并进一步协助临床医生管理治疗。[18,25] 在创建 Aifred Health 临床决策支持系统时，我们为自己设定的挑战，正是治疗选择与管理。

构建临床决策支持系统

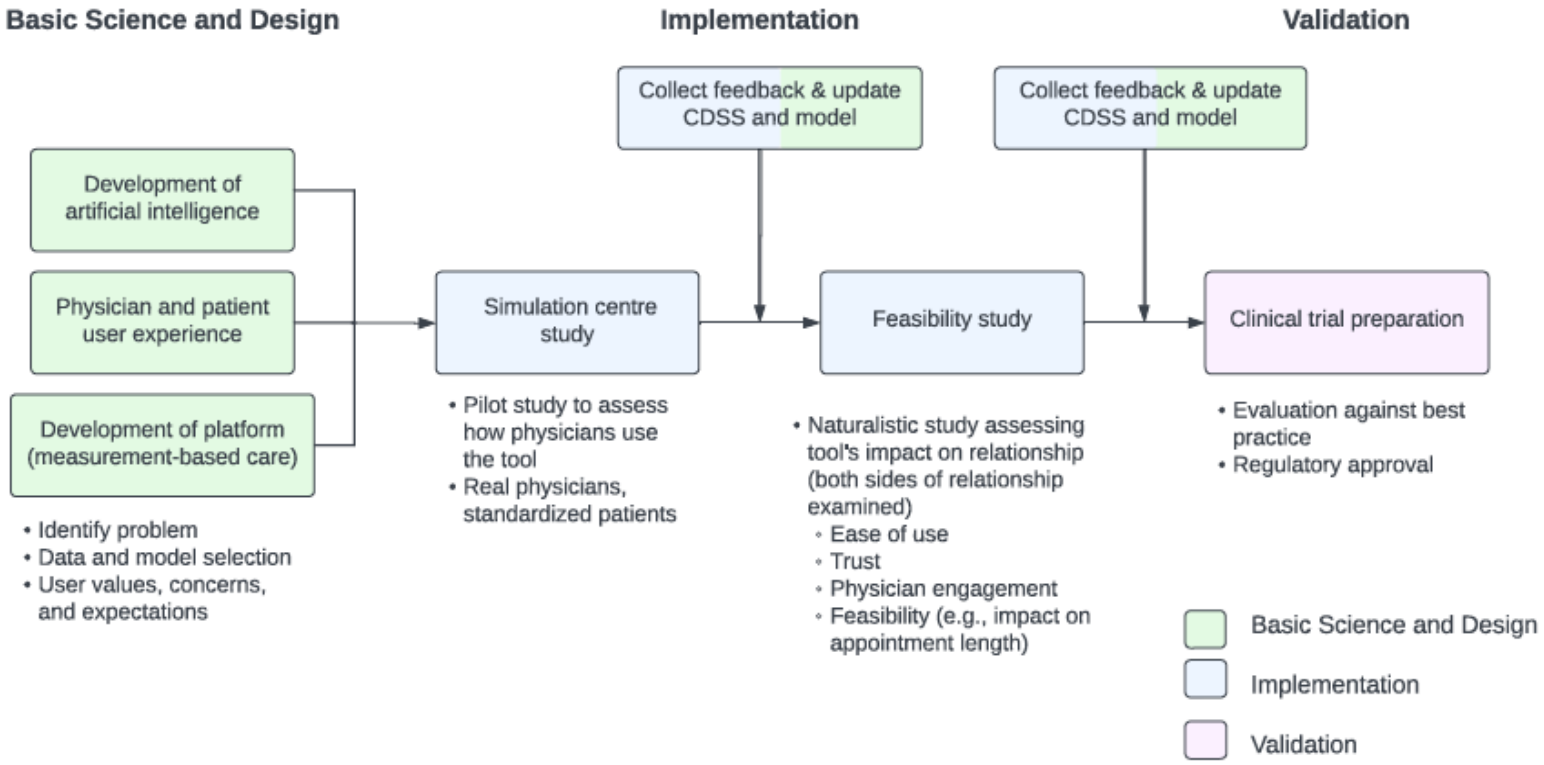
基础科学与设计

构建我们的临床决策支持系统，第一步是通过与临床医生和患者利益相关者进行讨论，并调查循证抑郁症治疗文献，了解临床医生和患者需要哪些功能。我们决定使用人工智能来解决治疗选择问题，因为采用经典统计学的现有方法未能生成可靠的工具来帮助实现个性化治疗。[25,29,30] 我们认为，改善治疗管理并不需要采用人工智能方法，因为现有指南已提供明确的建议，而且基于测量、算法指导的照护已被证明能够改善患者结局，但在临床实践中，这种照护并未得到充分落实。（见表1）。[23,31]

人工智能组件的设计与患者和临床医生用户体验的设计同步进行（见图1）。最终形成的应用程序以临床医生和患者所述的需求为依据；人工智能生成的输出经过设计，以融入临床工作流程，同时对临床医生有用且更易于解读。此外，为避免增加临床医生和患者的负担，并降低患者和临床医生的中断率，人工智能所需的输入尽可能少。[32,33]

此外，早期讨论很明确地表明，临床医生并不只是希望有一个工具来提供患者的总体缓解概率；相反，他们希望该工具能够帮助其在不同治疗方案之间进行选择。正是由于临床医生的意见，模型才将重点放在实现差异化治疗获益预测上——也就是说，构建能够比较多个个体治疗方案的模型。[30] 根据临床医生的反馈，在模型的后续迭代中，我们专注于创建一个模型，使其能够针对每位患者，在预测疗效最佳与最差的治疗方案之间产生具有临床意义的差异。[34]

图 1临床决策支持系统开发、实施与验证流程



人工智能与临床决策支持系统的集成

在了解人工智能需要解决的问题后，我们还需要确定如何将其纳入系统。为协助治疗管理，临床决策支持系统包含加拿大情绪与焦虑治疗网络（CANMAT）2016年指南的操作化版本。[23] 这些操作化指南根据患者当前状态提供信息，并包含一个治疗选择模块，临床医生可在其中获得指南所列治疗方案的信息。在此步骤中，我们将人工智能预测结果 置于指南模块中，但在视觉上将人工智能信息与源自指南的信息区分开来。为尽量减少系统对医生自主权的影响，人工智能会在按顺序排列的列表中为每种治疗方案生成缓解概率，而不是提供单一建议，从而促使临床医生与患者讨论治疗方案，以鼓励共同决策。[34]

如上所述，医生采纳人工智能工具的一大障碍，是由于缺乏可解释性而对这些系统缺乏信任。为解决这一问题，我们创建了“可解释性报告”（详见 Mehltretter[30]），该报告会针对每种药物，列出与预测缓解概率关联最强的五项患者变量。在一项模拟中心研究中，医生认为该报告对模拟患者的相关性与医生信任度相关；医生信任度与患者严重程度相互作用，可预测医生是否会开具人工智能预测的排名靠前的治疗方案之一。[34,35]

在设计用于集成人工智能的临床决策支持系统时，我们也同步开发了人工智能本身。首先，我们全面审查了哪些预测因子可能影响重度抑郁障碍患者的治疗结局，以确定模型应纳入哪些内容，并建立一个参照框架，用于比较模型学到的关联。最初确定的预测因子包括症状、人口统计学信息以及各种生物标志物。[36] 然而，在审查现有数据后，我们发现，许多临床实践中常用的个体治疗方案缺乏生物标志物数据，而且医生不太可能在使用该工具时常规收集患者的这类数据。[36] 因此，我们在模型训练中仅采用易于收集的临床和人口统计学特征。开发过程的一部分是确定采用哪种机器学习方法。最终，我们选择了深度学习，因为它能够发现数据中的复杂非线性模式，而经典统计模型难以做到这一点。[37] 此外，随着新的预测因子类型变得更易于获取，将其整合到核心模型中也会更加容易（更多信息请参见 Mehltretter[29] ）。

数据源选择注意事项

创建模型时，一个重要的决策是选择数据源。数据来源必须涵盖相关预测因子的范围，包含接受不同药物治疗的各类患者，以及能够近似反映真实人群的临床特征（例如合并症）。临床试验数据可能存在偏差，原因是其纳入与排除标准严格（例如，参与者并不能代表诊所接诊的患者）。另一方面，尽管电子病历系统拥有大量可用数据，医疗记录可能缺少严格的结局测量指标[38], 数据可能不完整、缺失、未结构化，或包含

错误[39], 或者数据可能因对连续就诊或就诊不规律的误解而导致错误推断。[40] 电子病历数据也可能受到时间偏倚的影响：随着时间推移，治疗标准发生变化，不同时期的患者可能不具可比性。[41]还存在倾向性偏倚的风险，即临床医生可能更倾向于以特定方式治疗某些患者。[42] 重要的是，倾向性模型等策略有助于克服此类问题，在使用电子病历数据时应予以考虑。[42]

最终，我们确定，与电子病历数据相比，临床试验数据中存在的偏差在实践中更容易识别和调查；此外，纳入/排除标准较宽松的试验能够提供充足的数据，从而构建出一个近似代表真实临床人群的数据集。临床试验数据还有一个明显优势，即包含明确无歧义的结局，可用作训练目标。[29,30,35]

实施

临床实践中的采纳

将该工具应用于临床实践，需要在临床环境中对其进行验证，这对于建立医生与人工智能驱动工具之间的信任至关重要。[25] 试点和质量保证测试也提供了一个机会，让我们能在开展更大规模的试验之前发现并解决软件错误，或工具设计与临床工作流程之间的不匹配问题，从而有望降低不良事件的发生率，并减少设计欠佳对临床试验中临床决策支持系统性能的影响。

鉴于在实施阶段进行充分且反复测试的重要性，我们最早开展的研究之一是在模拟中心收集医生对产品效用的使用体验反馈，然后再将其纳入实际临床实践（更多信息请参阅 Benrimoh[35] 和 Tanguay-Sela[43]）。我们发现，临床医生认为该工具有助于与患者进行共同决策。[35] 值得注意的是，鉴于早期与临床医生的交流主要聚焦于使用该工具所需的时间，

研究发现，该工具可在与标准化病人进行的10分钟会话中成功使用，这表明该工具已准备好在诊所进行可行性测试。

在开展以有效性为重点的临床试验之前，我们的下一步是在真实临床环境中对临床决策支持系统进行可行性研究。本研究的主要关注点是每次预约所花费的时间，因为临床医生利益相关者提到，时间损失将成为他们在实践中使用临床决策支持系统的一大障碍（更多信息请参见Popescu[34]）。我们发现，在引入该工具前后，预约时长并无差异。研究还表明，医生和患者对该工具的参与度持续超过两周（根据Arean等人，2016年的结果，采用两周基准来表示留存成功）[44]，这可能是因为该应用程序与临床护理相结合。[34] 至关重要的是，我们发现，对于任何患者而言，该工具均未对临床医生与患者关系产生负面影响——并且在约一半的病例中（即46%），甚至有报告称使用临床决策支持系统改善了这种关系。[34]这些发现十分重要，因为它们表明，在临床医生与患者关系中加入“第三方”（人工智能临床决策支持系统）不会对其产生负面影响，进而提高该工具的可行性以及整体治疗的疗效。[45]

临床试验准备

该工具进行临床验证的下一步，是开展随机对照试验（RCT），以评估该工具的有效性和安全性。收集的次要终点也将有助于持续评估其可行性以及对照护流程的影响。这是重要的一步，因为许多临床决策支持系统并未通过随机对照试验（RCT）评估其对患者结局的影响。[12]

这项正在进行的试验（NCT04655924）是一项以医生为单位进行集群随机分组、对患者和评估者设盲的临床决策支持系统阳性对照试验。我们将研究设计得尽可能贴近实际并具有实用性，这意味着医生可以自由选择是否采用人工智能建议，或忽略其提供的信息。同样，医生会接受最佳实践方面的训练，并获得问卷结果，但不必严格遵循指南。选择这种实用性设计，是为了得出与真实世界实施相关的结果，因为

研究结束后，不能合理地假设临床医生会按照预先制定的研究方案使用临床决策支持系统。

讨论

采用临床决策支持系统的潜在益处

临床决策支持系统有望提高医疗效率并降低长期成本。[46]借助循证支持来增强医疗服务提供者决策能力的工具，有助于满足心理健康服务需求，并可能赋能更多基层执业人员（如执业护士和家庭医生）提供高质量的医疗服务。此外，人工智能驱动的临床决策支持系统可通过提高个性化水平、改善治疗管理质量，减少需要尝试的治疗方案数量，从而有望缩短患者患病时间。这不仅能显著减轻患者及其家人的痛苦，也能大幅缓解不堪重负的医疗卫生系统和经济体所承受的压力。[25]

应对人工智能临床决策支持系统的局限性

尽管实施临床决策支持系统（CDSS）可能带来诸多益处，但许多临床决策支持系统往往在实施阶段遭遇失败，原因包括缺乏透明度、增加日常实践所需时间、对证据存在不确定性且不信任系统，或扰乱临床工作流程。[14,15,16,17,18] 如果工具的开发旨在创造新颖事物，却不适合当前需求，那么这些挫折的根源往往在最初的概念构思阶段就已埋下。[18] 我们认为，避免这些陷阱的第一步是明确要解决的具体问题，并以此指导系统设计——既包括用户体验，也包括人工智能模型的设计。此外，我们认为，成功的临床决策支持系统应经过严格的测试与验证流程。[47] 此外，这一流程必须是迭代式的，并在每个阶段都提供纳入用户反馈的机会。人们或许会倾向于从大规模临床试验着手，因为这样可以缩短从系统初步开发到临床实施的时间。然而，临床试验成本高昂、资源

密集，且会涉及大量患者，因此只有在确信其很可能成功且不会造成伤害时，才应开展临床试验。

结论

本文讨论了限制临床决策支持系统在医疗保健领域广泛采用的当前障碍，并提出了在构建、验证、训练和实施过程中应考虑的建议。临床决策支持系统是创新且高效的工具，有望在患者和系统层面改善医疗服务提供，但其实施仍面临重大障碍，必须加以解决才能成功实施。

表1. 构建人工智能临床决策支持系统的独特挑战及 Aifred 的应对方法。

挑战	Aifred 的方法
确保临床决策支持系统能够解决正确的问题 。[18]	将治疗选择和管理确定为关键问题；治疗选择需要开发人工智能模型，而管理则需要对指南进行数字化和操作化。
数据需要具备较高的质量。[3,38,39,40]	由于质量不一致，未纳入电子病历数据。结局的可用性不一致，因此未纳入电子病历数据。使用抗抑郁药临床试验。
必须验证一个机器学习模型。[29]	开发了不同类型的机器学习模型，并与现有已发表模型进行了比较。采用临床医生熟悉的指标，以增强信任。

功能必须具备 临床效用。[18,25]	鉴于重度抑郁障碍有多种药物选择，且 临床医生难以从中做出选择，我们开发了一种 人工智能驱动的临床决策支持系统，用于实现差异化治疗 获益预测，以协助在不同治疗方案之间进行选择。 该临床决策支持系统还纳入了 CANMAT 指南的操作化版本，以协助进行治疗管理， 从而比在标准护理中更有效地将其整合进临床工作流程。 在标准护理中所能实现的程度。
建立医生对 人工智能的信任。 [15,16,17,18]	开发了可解释性报告，其中最多提供五项 在决定.....方面最重要的患者变量 特定药物的缓解概率最重要的患者变量；开展了 利益相关者需求评估，以了解医生 希望从该工具中获得什么以及会如何使用它；进行了 严格的临床验证流程，并报告了结果。.
临床决策支持系统不能干预 这种关系，例如过度 劫持决策 过程。[18,25,34]	研究工具对医患关系的影响 在医生与 标准化病人的模拟临床互动中，以及在临床可行性研究期间。

资助：本研究由 Aifred Health Inc；加拿大国家研究委员会创新研究援助计划；支持 IMADAPT 的 ERA-Permed Vision 2020；魁北克省政府 Nova Science；以及 MEDTEQ 新冠纾困拨款资助。**利益冲突：**GG 和 CP 曾受雇于 Aifred Health 或目前受雇于该公司，或曾从该公司获得经济报酬。MTS 受雇于 Aifred Health，并持有期权。SI、KP、CA、RF 和 DB 是 Aifred Health 的股东、员工、董事或创始人。**数据可用性声明：**本文为观点文章，因此没有相关数据。

参考文献

1. Berner, E. S., & La Lande, T. J. (2007). 临床决策支持系统概述。载于 E. S. Berner (编), 临床决策支持系统（第 3–22 页）。Springer New York. https://doi.org/10.1007/978-0-387-38319-4_1

2. Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). 运用临床决策支持系统改善临床实践：旨在识别成功关键特征的试验系统综述。*BMJ*, *330*(7494), 765. https://doi.org/10.1136/bmj.38398.500764.8F

3. Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020年). 临床决策支持系统概述：益处、风险与成功策略。*npj*数字医学, *3*(1), 17. https://doi.org/10.1038/s41746-020-0221-y

4. Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., & Tang, P. C. (2001). 循证医学实践的临床决策支持系统。美国医学信息学会杂志, *8*(6), 527–534.https://doi.org/10.1136/jamia.2001.0080527

5. Jia, P., Zhang, L., Chen, J., Zhao, P. 和 Zhang, M.（2016年）。临床决策支持系统对用药安全的影响：综述。*PLOS ONE*，*11*（12），e0167683。https://doi.org/10.1371/journal.pone.0167683

6. Kwan, J. L., Lo, L., Ferguson, J., Goldberg, H., Diaz-Martinez, J. P., Tomlinson, G., Grimshaw, J. M. 和 Shojania, K. G.（2020年）。计算机化临床决策支持系统与医疗服务的绝对改善：对照临床试验的荟萃分析。*BMJ*，m3216。https://doi.org/10.1136/bmj.m3216

7. Roshanov, P. S., Misra, S., Gerstein, H. C., Garg, A. X., Sebaldt, R. J., Mackay, J. A., Weise-Kelly, L., Navarro, T., Wilczynski, N. L., Haynes, R. B., 和 CCDSS系统综述团队。（2011）。

慢性病管理的计算机化临床决策支持系统：决策者与研究人员合作开展的系统综述。*Implementation Science*，*6*(1)，92。https://doi.org/10.1186/1748-5908-6-92

8. Cohen, J. P., Bertin, P., & Frappier, V.（2020年）。Chester：基于网络交付的本地计算胸部X线疾病预测系统。*ArXiv:1901.11210 [Cs, q-Bio]*。http://arxiv.org/abs/1901.11210

9. Lakhani, P., & Sundaram, B.（2017）。胸部放射摄影中的深度学习：使用卷积神经网络自动分类肺结核。*Radiology*，*284*(2), 574–582。https://doi.org/10.1148/radiol.2017162326

10. Larson, D. B., Chen, M. C., Lungren, M. P., Halabi, S. S., Stence, N. V., & Langlotz, C. P.（2018）。深度学习神经网络模型评估儿童手部X线片骨骼成熟度的性能。*Radiology*，*287*(1), 313–322。https://doi.org/10.1148/radiol.2017170236

11. Squarcina, L., Villa, F. M., Nobile, M., Grisan, E., & Brambilla, P.（2021年）。深度学习在抑郁症治疗反应预测中的应用。*Journal of Affective Disorders*，*281*，618–622。https://doi.org/10.1016/j.jad.2020.11.104

12. Tulk Jesso, S., Kelliher, A., Sanghavi, H., Martin, T., & Henrickson Parker, S.（2022）。临床医生参与临床人工智能工具的开发与评估：系统性文献综述。*Frontiers in Psychology*，*13*，830345。https://doi.org/10.3389/fpsyg.2022.830345

13. Waring, J., Lindvall, C., & Umeton, R.（2020年）。自动化机器学习：医疗保健领域的最新进展综述与机遇。*Artificial Intelligence in Medicine*，*104*，101822。https://doi.org/10.1016/j.artmed.2020.101822

14. Muhiyaddin, R., Abd-Alrazaq, A. A., Househ, M., Alam, T., & Shah, Z. (2020年). 临床决策支持系统对医生的影响：范围综述。*Studies in HealthTechnology and Informatics*, *272*, 470–473. https://doi.org/10.3233/SHTI200597

15. Curry, L., & Reed, M. H. (2011). 初级保健场景中用于诊断成像的电子决策支持。美国医学信息学会杂志：*JAMIA*, *18*(3), 267–270. https://doi.org/10.1136/amiajnl-2011-000049

16. Shibl, R., Lawley, M., & Debuse, J. (2013). 影响决策支持系统接受度的因素。决策支持系统，*54*(2), 953–961. https://doi.org/10.1016/j.dss.2012.09.018

17. Sousa, V. E. C., Lopez, K. D., Febretti, A., Stifter, J., Yao, Y., Johnson, A., Wilkie, D. J., & Keenan, G. M. (2015). 运用模拟研究护士对临床决策支持建议的接受与不接受。*Computers, Informatics, Nursing: CIN*，*33*(10), 465–472. https://doi.org/10.1097/CIN.0000000000000185

18. Shortliffe, E. H., & Sepúlveda, M. J. (2018). 人工智能时代的临床决策支持。*JAMA*, *320*(21), 2199–2200. https://doi.org/10.1001/jama.2018.17163

19. Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., & Cilar, L.（2020年）。医疗保健领域基于机器学习的预测模型的可解释性。*WIREs Data Mining and KnowledgeDiscovery*，*10*（5）。https://doi.org/10.1002/widm.1379

20. Kleinerman, A., Rosenfeld, A., Benrimoh, D., Fratila, R., Armstrong, C., Mehltretter, J., Shneider, E., Yaniv-Rosenfeld, A., Karp, J., Reynolds, C. F., Turecki, G., & Kapelner, A.（2021 年）。利用潜在空间中的原型设计进行治疗选择，并将其应用于抑郁症治疗。*PLOS ONE*，*16*（11），e0258400。https://doi.org/10.1371/journal.pone.0258400

21. Jankovic, I., & Chen, J. H. (2020年). 临床决策支持与临床医生倦怠危机的影响。医学信息学年鉴，*29*(01), 145–154。https://doi.org/10.1055/s-0040-1701986

22. Graham, T. A. D., Kushniruk, A. W., Bullard, M. J., Holroyd, B. R., Meurer, D. P., & Rowe, B. H. (2008). 基于网络的临床决策支持系统的可用性如何可能导致不良医疗事件。*AMIA ... 年度研讨会论文集。AMIA研讨会*，257–261。

23. Kennedy, S. H., Lam, R. W., McIntyre, R. S., Tourjman, S. V., Bhat, V., Blier, P., Hasnain, M., Jollant, F., Levitt, A. J., MacQueen, G. M., McInerney, S. J., McIntosh, D., Milev, R. V., Müller, D. J., Parikh, S. V., Pearson, N. L., Ravindran, A. V., Uher, R., 与CANMAT抑郁症工作组。（2016年）。加拿大情绪与焦虑治疗网络（CANMAT）2016年临床指南：成年人重度抑郁障碍的管理：第3节。药物治疗。《加拿大精神病学杂志》，*61*(9), 540–560.https://doi.org/10.1177/0706743716659417

24. Warden, D., Rush, A. J., Trivedi, M. H., Fava, M., 与Wisniewski, S. R.（2007年）。STAR*D项目结果：研究发现的综合综述。《当代精神病学报告》，*9*(6), 449–459.https://doi.org/10.1007/s11920-007-0061-3

25. Benrimoh, D., Fratila, R., Israel, S., Perlman, K., Mirchi, N., Desai, S., Rosenfeld, A., Knappe, S., Behrmann, J., Rollins, C., & You, R. P. (2018). Aifred Health，一种由深度学习驱动的心理健康临床决策支持系统。载于《NIPS '17竞赛：构建智能系统》（第 251–287 页）。Spinger International Publishing。

26. Frank, R. G., & Zeckhauser, R. J. (2007). 定制治疗与现成治疗：医生选择中的行为倾向。《健康经济学杂志》，*26*(6), 1101–1127.https://doi.org/10.1016/j.jhealeco.2007.08.002

27. Adli, M., Wiethoff, K., Baghai, T. C., Fisher, R., Seemüller, F., Laakmann, G., Brieger, P., Cordes, J., Malevani, J., Laux, G., Hauth, I., Möller, H.-J., Kronmüller, K.-T., Smolka, M. N., Schlattmann, P., Berger, M., Ricken, R., Stamm, T. J., Heinz, A., & Bauer, M. (2017). 如何有效

算法指导治疗是否适用于抑郁症住院患者？德国算法项目3随机对照多中心试验的结果。《国际神经精神药理学杂志》，*20*(9)，721–730。https://doi.org/10.1093/ijnp/pyx043

28. Trivedi, M. H., Rush, A. J., Crismon, M. L., Kashner, T. M., Toprac, M. G., Carmody, T. J., Key, T., Biggs, M. M., Shores-Wilson, K., Witte, B., Suppes, T., Miller, A. L., Altshuler, K. Z., & Shon, S. P. (2004). 德克萨斯州药物算法项目中重度抑郁障碍患者的临床结果。《普通精神病学档案》，*61*(7), 669. https://doi.org/10.1001/archpsyc.61.7.669

29. Mehltretter, J., Rollins, C., Benrimoh, D., Fratila, R., Perlman, K., Israel, S., Miresco, M., Wakid, M., & Turecki, G. (2020年). 对用于抑郁症差异化治疗选择的深度学习模型所选特征的分析。《人工智能前沿》，*2*, 31. https://doi.org/10.3389/frai.2019.00031

30. Mehltretter, J., Fratila, R., Benrimoh, D. A., Kapelner, A., Perlman, K., Snook, E., Israel, S., Armstrong, C., Miresco, M., & Turecki, G. (2020年). 抑郁症治疗选择中的差异化治疗获益预测：对 STAR*D 和 CO-MED 数据的深度学习分析。计算精神病学, *4*(0), 61. https://doi.org/10.1162/cpsy_a_00029

31. Kozicky, J. M., Schaffer, A., Beaulieu, S., McIntosh, D., & Yatham, L. N. (2022年). 使用一种基于网络的即时应用程序，以提高对 CANMAT 和 ISBD 2018年指南中双相障碍管理相关建议的遵循度。《双相障碍》, 24(4), 392-399. 10.1111/bdi.13136

32. Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). 心理健康应用程序的客观用户参与度：系统检索与基于样本组的使用情况分析。《医学互联网研究杂志》，21(9)，e14567. https://doi.org/10.2196/14567

33. Torous, J., Lipschitz, J., Ng, M., & Firth, J. (2020). 用于缓解抑郁症状的智能手机应用程序临床试验脱落率：系统综述与荟萃分析。《Journal of Affective Disorders》， 263, 413–419. https://doi.org/10.1016/j.jad.2019.11.167

34. Popescu, C., Golden, G., Benrimoh, D., Tanguay-Sela, M., Slowey, D., Lundrigan, E., Williams, J., Desormeau, B., Kardani, D., Perez, T., Rollins, C., Israel, S., Perlman, K., Armstrong, C., Baxter, J., Whitmore, K., Fradette, M.-J., Felcarek-Hope, K., Soufi, G., ... Turecki, G. (2021年). 评估人工智能驱动、基于网络的临床决策支持系统在成年人抑郁症治疗中的临床可行性：纵向可行性研究。JMIR Formative Research, 5(10), e31862. https://doi.org/10.2196/31862

35. Benrimoh, D., Tanguay-Sela, M., Perlman, K., Israel, S., Mehltretter, J., Armstrong, C., Fratila, R., Parikh, S. V., Karp, J. F., Heller, K., Vahia, I. V., Blumberger, D. M., Karama, S., Vigod, S. N., Myhr, G., Martins, R., Rollins, C., Popescu, C., Lundrigan, E., ... Margolese, H. C. (2021年). 利用模拟中心评估人工智能驱动的临床决策支持系统在抑郁症治疗中的初步可接受性及其对医患互动的影响。BJPsych Open, 7(1), e22. https://doi.org/10.1192/bjo.2020.127

36. Perlman, K., Benrimoh, D., Israel, S., Rollins, C., Brown, E., Tunteng, J.-F., You, R., You, E., Tanguay-Sela, M., Snook, E., Miresco, M., & Berlim, M. T. (2019). 重度抑郁障碍中抗抑郁治疗结局预测因子的系统性荟萃综述。Journal of Affective Disorders, 243, 503–515. https://doi.org/10.1016/j.jad.2018.09.067

37. Rashidi, H. H., Tran, N., Albahra, S., & Dang, L. T. (2021年). 医疗保健与检验医学中的机器学习：监督学习与自动化机器学习概述. 国际实验室血液学杂志, 43(S1), 15–22. https://doi.org/10.1111/ijlh.13537

38. Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). 英国的精神科医生不使用结局测量指标：全国调查. 英国精神病学杂志, 180(2), 101–103. https://doi.org/10.1192/bjp.180.2.101

39. Kristianson, K. J., Ljunggren, H., & Gustafsson, L. L.（2009）。从半结构化电子病历系统中提取门诊患者数据：促进数据用于质量控制和研究的访问与使用模型。《健康信息学杂志》，15(4), 305–319。https://doi.org/10.1177/1460458209345889

40. Zheng, K., Gao, J., Ngiam, K. Y., Ooi, B. C., & Yip, W. L. J.（2017）。消除电子病历中的偏差。第23届ACM SIGKDD知识发现与数据挖掘国际会议论文集，2171–2180。https://doi.org/10.1145/3097983.3098149

41. Yuan, W., Beaulieu-Jones, B. K., Yu, K.-H., Lipnick, S. L., Palmer, N., Loscalzo, J., Cai, T., & Kohane, I. S.（2021年）。病例对照设计中的时间偏差：阻碍对未来进行可靠预测。《自然·通讯》，12(1), 1107。https://doi.org/10.1038/s41467-021-21390-2

42. Afzal, Z.、Masclee, G. M. C.、Sturkenboom, M. C. J. M.、Kors, J. A. 和 Schuemie, M. J.（2019年）。利用电子病历中的文本特征生成并评估倾向性模型。PLOS ONE, 14(3)，e0212999。https://doi.org/10.1371/journal.pone.0212999

43. Tanguay-Sela, M., Benrimoh, D., Popescu, C., Perez, T., Rollins, C., Snook, E., Lundrigan, E., Armstrong, C., Perlman, K., Fratila, R., Mehltretter, J., Israel, S., Champagne, M., Williams, J., Simard, J., Parikh, S. V., Karp, J. F., Heller, K., Linnaranta, O., ... Margolese, H. C.（2022年）。评估人工智能驱动的临床决策支持系统在借助模拟中心进行抑郁症治疗时的感知效用。《精神病学研究》，308，114336。https://doi.org/10.1016/j.psychres.2021.114336

44. Arean, P. A.、Hallgren, K. A.、Jordan, J. T.、Gazzaley, A.、Atkins, D. C.、Heagerty, P. J. 和 Anguera, J. A.（2016年）。移动应用对抑郁症的使用及有效性：一项全远程临床试验的结果。《医学互联网研究杂志》，18(12)，e330。https://doi.org/10.2196/jmir.6482

45. Nolan, P. 与 Badger, F.（2005）。提升初级保健满意度的医生与抑郁症患者关系的相关方面。《精神病学与心理健康护理杂志》，12（2），146–153。
https://doi.org/10.1111/j.1365-2850.2004.00806.x

46. Castillo, R. S. 与 Kelemen, A.（2013）。成功实施临床决策支持系统的相关考量。CIN：计算机、信息学与护理，31（7），319–326。
https://doi.org/10.1097/NXN.0b013e3182997a9c

47. Day、S.、Shah、V.、Kaganoff、S.、Powelson、S. 和 Mathews、S. C.（2022）。评估数字健康初创企业的临床稳健性：横断面观察性分析。《医学互联网研究杂志》，24(6)，e37677。
https://doi.org/10.2196/37677

补充材料

引言

历史背景

首先，我们将介绍医疗保健领域临床决策支持系统的开发和使用背景。美国《国家医疗保健法》支持临床决策支持系统的开发，因为此类系统具有成本效益，并能够集成到某些电子病历系统中（Sutton 等，2020年）。临床决策支持系统具有多种功能和优势，例如能够减少处方错误、减轻不良事件、降低成本（例如，减少重复检查和重复医嘱）、提供诊断和决策支持、改善临床工作流程、发送提醒和警报，以及改善患者结局（Khairat 等，2018年；Sutton 等，2020年）。

自20世纪70年代以来（Shortliffe & Buchanan，1975），研究者已针对多种躯体及精神疾病开发了临床决策支持系统（Shortliffe & Buchanan，1975；Jia等人，2016年；Kwan等人，2020年；Roshanov等人，2011年）。尽管临床算法或数据驱动算法及模型日益复杂，研究和电子病历中的数据量不断增加，计算存储能力和算力也有所提升，但临床决策支持系统的有效性因工具而异，差异显著（Kwan等人，2020年；Roshanov等人，2011年）。例如，Kwan等人（2020年）开展的一项系统综述和荟萃分析报告称，在纳入的108项研究中，临床决策支持系统使接受符合系统预设标准的医疗服务的患者比例提高了5.8%。然而，作者指出，排名前四分之一的研究之间存在显著异质性，其流程依从性的改善幅度从10%到62%不等，这表明临床决策支持系统干预可能产生的影响范围很广。Roshanov等人（2011年）的另一项系统综述考察了临床决策支持系统是否能够改善慢性病管理流程及相关患者结局，发现针对糖尿病和血脂异常的系统有望改善患者结局。尽管如此，针对高血压、哮喘、慢性阻塞性肺疾病、心脏疾病及其他疾病的系统，在改善患者结局方面很少显示出益处；不过，这些疾病领域的研究表明，临床决策支持系统可能有助于提升患者所获医疗服务的质量（Roshanov等人，2011年）。照护流程的改善可能源于

临床决策支持系统的推送功能（例如，在移动设备上弹出的通知消息）、复杂决策支持（例如，基于指南的反馈，以及有关筛查、诊断和处方的建议），或文档选项（Kwan等人，2020年）。另一项评估临床决策支持系统对用药安全影响的综述也证实了这一观察：75%的试验显示其对医疗流程有积极影响，但仅有20%的试验显示其对患者结局有积极影响（Jia等人，2016年）。Roshanov等人（2011年）指出，只有56%的试验将患者结局作为主要终点进行考察，这表明临床决策支持系统，或旨在考察其影响的研究，并不一定以改善患者结局为设计目的。相反，人们常常认为改善照护流程就能改善结局或其他重要指标。例如，临床决策支持系统的设计目的可能是提高临床医生的效率或接诊能力，使其能够为更多患者提供服务，并期望借此改善总体结局，因为接受医疗服务的患者人数增加了。或者，研究设计者可能不确定临床决策支持系统对患者结局的预期影响，因此会将评估系统的可行性作为主要终点，而不再进一步研究其对结局的影响。这一区别很重要，因为将照护流程改善转化为患者结局改善可能面临多种障碍。这些障碍可能包括：最终用户不确信所测量结局的适当性，或不认可所提供的决策支持；在常规实践中缺乏足够的时间来训练或使用临床决策支持系统；或者系统功能有限，无法满足临床需求。此外，许多临床决策支持系统都包含面向患者的组件，因此，在设计过程中考虑患者对系统的看法至关重要（Kawamoto等人，2005年）；患者对系统的信任程度或对其效用的看法，可能影响其采纳意愿，进而影响系统的整体有效性。

实施

临床实践中的采纳

鉴于在实施阶段进行充分且反复测试的重要性，我们最早开展的研究之一是在模拟中心收集医生对产品效用的体验反馈，以便在将其纳入真实临床实践之前进行评估（参见

Benrimoh 等，2021 和 Tanguay-Sela 等，2022，了解更多信息）。Aifred 的模拟中心研究评估了 20 名参与者使用临床决策支持系统的情况，其中 11 名为精神科医生，9 名为家庭医生。研究通过观察医生使用该系统与三名标准化病人（分别患有轻度、中度和重度抑郁症）进行诊疗，帮助我们了解医生在与患者的会诊中会如何使用该系统。研究为医生提供了使用该工具的基础训练，并指示他们在会诊过程中按照自己认为合适的方式使用该工具（Benrimoh 等，2021）。模拟研究揭示了医生对该工具的若干看法。其中一个重要发现是，临床医生指出，该工具有助于与患者进行共同决策；当临床医生与患者分享应用程序屏幕，将他们“邀请”参与其中时，标准化病人对该工具的评价最高（Benrimoh 等，2021）。50% 的医生表示，他们会对所有重度抑郁障碍患者使用该工具；85% 的医生认为，他们会对复杂病例或难治性患者使用该工具（Tanguay-Sela 等，2022）。研究还发现，在与标准化病人进行会诊时，使用该工具的时间不到 5 分钟，这表明该工具已准备好在诊所开展可行性测试。

这项准备度评估促成了我们在真实临床环境中开展临床决策支持系统的可行性研究，可视为在开展以有效性为重点的临床试验之前进行的一次“压力测试”。在这项可行性研究中，我们主要关注每次预约所花费的时间，因为临床医生在项目早期的利益相关者讨论中指出，时间损失是他们在实践中使用临床决策支持系统的最大障碍。7名临床医生为14名确诊重度抑郁障碍的患者提供了治疗（Popescu 等，2021）。将未使用临床决策支持系统的基线预约与后续使用该系统的就诊进行比较后发现，预约时长没有显著差异（Popescu 等，2021）。我们并未要求医生在患者预约之外使用临床决策支持系统。这些结果消除了人们对该系统可能增加医生工作负荷或占用额外时间的担忧。此外，62%的患者和71%的医生表示信任该工具，92%的患者和71%的医生认为该工具易于使用（Popescu 等，2021）。最后，尽管我们告知医生可以自行决定使用或忽略该应用程序，医生仍在96%的预约中使用了该应用程序（更多信息请参见 Popescu 等，2021）。值得注意的是，医生和患者双方

都在使用该工具超过两周后仍保持了持续参与度（根据 Arean 等人（2016）的研究结果，我们采用两周基准来表示留存成功；该研究作者指出，应用程序使用量在两周后显著下降），这可能归因于该应用程序与临床护理相结合（Popescu 等，2021）。至关重要的是，我们发现，该工具没有对任何患者的临床医生与患者关系产生负面影响；在大约一半的案例中，甚至有报告称使用临床决策支持系统改善了这种关系（Popescu 等，2021）。这一点十分重要，因为此前有人担心，在临床医生与患者关系中加入一个“第三方”（人工智能临床决策支持系统）可能会对关系产生负面影响，进而降低该工具的可行性以及整体治疗效果（Nolan & Badger，2005）。

参考文献

Arean, P. A., Hallgren, K. A., Jordan, J. T., Gazzaley, A., Atkins, D. C., Heagerty, P. J., & Anguera, J. A.（2016年）。移动应用程序在抑郁症治疗中的使用情况和有效性：一项全远程临床试验的结果。《医学互联网研究杂志》，*18*（12），e330。https://doi.org/10.2196/jmir.6482

Benrimoh, D., Tanguay-Sela, M., Perlman, K., Israel, S., Mehltretter, J., Armstrong, C., Fratila, R., Parikh, S. V., Karp, J. F., Heller, K., Vahia, I. V., Blumberger, D. M., Karama, S., Vigod, S. N., Myhr, G., Martins, R., Rollins, C., Popescu, C., Lundrigan, E., ... Margolese, H. C.（2021年）。利用模拟中心评估人工智能驱动的临床决策支持系统在抑郁症治疗中的初步可接受性及其对医患互动的影响。*BJPsych Open*，*7*（1），文章1。https://doi.org/10.1192/bjo.2020.127

Jia, P., Zhang, L., Chen, J., Zhao, P., & Zhang, M. (2016年). 临床决策支持系统对用药安全的影响：综述。*PLOS ONE*, *11*(12), e0167683. https://doi.org/10.1371/journal.pone.0167683

Khairat, S., Marc, D., Crosby, W., & Al Sanousi, A. (2018). 医生不采用临床决策支持系统的原因：批判性分析。*JMIR*医学信息学, *6*(2), e24. https://doi.org/10.2196/medinform.8912

Kwan, J. L., Lo, L., Ferguson, J., Goldberg, H., Diaz-Martinez, J. P., Tomlinson, G., Grimshaw, J. M., & Shojania, K. G. (2020年). 计算机化临床决策支持系统与医疗服务的绝对改善：对照临床试验的荟萃分析。*BMJ*, m3216. https://doi.org/10.1136/bmj.m3216

Nolan, P. 和 Badger, F.（2005）。提升初级保健满意度的医生与抑郁症患者关系要素。《精神病学与心理健康护理杂志》，*12*(2)，146–153。https://doi.org/10.1111/j.1365-2850.2004.00806.x

Popescu, C., Golden, G., Benrimoh, D., Tanguay-Sela, M., Slowey, D., Lundrigan, E., Williams, J., Desormeau, B., Kardani, D., Perez, T., Rollins, C., Israel, S., Perlman, K., Armstrong, C., Baxter, J., Whitmore, K., Fradette, M.-J., Felcarek-Hope, K., Soufi, G., ... Turecki, G.（2021年）。评估一种人工智能驱动的网络

纵向可行性研究。《JMIR Formative Research》，5(10)，e31862。https://doi.org/10.2196/31862 Roshanov, P. S., Misra, S., Gerstein, H. C., Garg, A. X., Sebaldt, R. J., Mackay, J. A., Weise- Kelly, L., Navarro, T., Wilczynski, N. L., Haynes, R. B., & CCDSS系统综述团队。（2011）。用于慢性病管理的计算机化临床决策支持系统：决策者与研究者合作系统综述。《Implementation Science》，6(1)，92。https://doi.org/10.1186/1748-5908-6-92 Shortliffe, E. H., & Buchanan, B. G.（1975）。医学中的不精确推理模型。《数学生物科学》，23(3–4)，351–379。https://doi.org/10.1016/0025- 5564(75)90047-4 Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I.（2020）。临床决策支持系统概述：益处、风险及成功策略。《npj数字医学》，3(1)，17。https://doi.org/10.1038/s41746-020-0221-y Tanguay-Sela, M., Benrimoh, D., Popescu, C., Perez, T., Rollins, C., Snook, E., Lundrigan, E., Armstrong, C., Perlman, K., Fratila, R., Mehltretter, J., Israel, S., Champagne, M., Williams, J., Simard, J., Parikh, S. V., Karp, J. F., Heller, K., Linnaranta, O., ... Margolese, H. C.（2022）。利用模拟中心评估人们对人工智能驱动的抑郁症治疗临床决策支持系统的感知效用。《精神病学研究》，308，114336。https://doi.org/10.1016/j.psychres.2021.114336