

医学人工智能中的因果推理与可解释性：问题、进展与展望

摘要

本报告旨在深入探讨当前医学人工智能（AI）领域中因果推理与可解释性（XAI）的核心议题。随着 AI 技术在医疗健康领域的广泛应用，仅仅依赖模型的预测准确性已不足以满足临床实践的复杂需求。因果推理致力于揭示变量间的真实因果联系，超越表象相关性，为理解疾病机理、评估干预效果提供深层认知。可解释性则关注打开 AI 模型的“黑箱”，使其决策过程透明化，从而增强用户信任、确保医疗安全并满足伦理与法规要求。报告将首先界定医学 AI 中因果推理与可解释性的核心概念及其内在联系与区别。随后，将阐述这些技术在提升临床决策质量、保障医疗安全、促进药物研发等方面的关键应用价值。报告还将梳理当前学术界在因果推理和 XAI 方法上的研究现状，包括结构因果模型、倾向性评分、LIME、SHAP 等主流技术，并剖析其面临的瓶颈，如混杂偏倚的处理、模型验证的困难、解释的可靠性与评估标准缺乏等。此外，本报告将回顾该领域过往研究的重要里程碑，特别是 Judea Pearl 等学者在因果科学理论上的奠基性工作，以及关键推断方法和解释技术的发展。最后，通过分析 Aetion、Verge Genomics、Causaly 等企业的产业化案例，展示因果 AI 与 XAI 在实际医疗场景中的应用与商业潜力。报告总结了当前面临的挑战，并展望了未来研究方向，强调了跨学科合作、稳健评估框架构建以及因果与解释方法深度融合对推动医学 AI 向更智能、更可靠、更符合伦理的未来发展的重要性。

1. 引言：医学人工智能中的因果推理与可解释性

近年来，人工智能（AI）在医学领域的应用取得了显著进展，从医学影像分析到疾病风险预测，再到个性化治疗方案的制定，AI 技术正深刻改变着医疗健康的面貌。然而，随着模型复杂性的增加，尤其是深度学习等“黑箱”模型的广泛应用，其决策过程的不透明性引发了广泛关注。在性命攸关的医疗领域，仅仅知道“是什么”预测结果，远不如理解“为什么”以及“如何”得出该结果更为重要。因此，因果推理（Causal Inference）和可解释性 AI（Explainable AI, XAI）成为推动医学 AI 向更深层次、更可信赖方向发展的两大关键支

柱。

- 1.1. 问题定义与核心概念

- 1.1.1. 医学 AI 中的因果推理

因果 AI 旨在识别和理解数据中变量间的因果关系，这超越了传统机器学习模型仅能发现相关性的能力¹。它通过运用因果推断方法，探究事件发生的根本原因，从而对潜在机制获得更深层次的理解¹。在医疗领域，理解治疗措施与临床结局之间的因果效应至关重要¹。传统的 AI 模型，如许多机器学习算法，主要通过学习数据中的统计模式来进行预测。然而，这些模式往往反映的是相关性而非因果性。例如，某种药物的使用可能与患者康复相关，但这并不一定意味着药物直接导致了康复，可能存在其他混杂因素（如患者的总体健康状况、生活习惯等）同时影响了药物选择和康复结果。因果推理的目标正是要从复杂的观测数据中剥离出真实的因果效应，回答诸如“如果对患者 X 采取治疗 A，其结局 Y 会如何变化？”这类反事实问题。

- 1.1.2. 医学 AI 中的可解释性 (XAI)

可解释性 AI (XAI) 专注于创建人类能够理解其操作和输出的 AI 模型³。在医疗这样高风险领域，AI 的可解释性对于建立信任、明确责任以及满足监管要求至关重要³。正如 Doshi-Velez 和 Kim 所定义，XAI 中的可解释性是“向人类解释或以可理解的方式呈现的能力”⁴。随着 AI 模型（尤其是深度学习模型）日益复杂化，它们往往被视为“黑箱”，即其内部工作机制难以被人类直观理解。这种不透明性使得临床医生在采纳 AI 给出的建议时犹豫不决，尤其当这些建议可能对患者生命健康产生重大影响时。XAI 的目标就是打开这个“黑箱”，提供关于模型如何做出特定预测或决策的洞察。这不仅有助于提升用户对 AI 系统的信任，还能促进模型的调试、发现潜在偏见，并确保其应用符合伦理和法规标准，例如 HIPAA（健康保险流通与责任法案）和 FDA（美国食品药品监督管理局）的相关规定⁵。

- 1.1.3. 两者的内在联系与差异

因果 AI 因其模型构建基于对因果机制的理解，通常被认为是内在地可解释的

（即“白箱”模型），它在模型构建完成之前（*ante hoc*）就能提供解释，允许领域专家介入并施加约束³。相比之下，许多传统的 XAI 方法，如 LIME 和 SHAP，则是在模型已经构建完成后（*post hoc*），对“黑箱”模型的行为进行解释，这种事后解释有时会引发对其忠实性和可靠性的疑虑³。因果 AI 致力于通过识别根本原因来提供更深层的解释力，而不仅仅是揭示相关性¹。

虽然因果推理和可解释性都旨在增强对 AI 模型的理解，但它们的侧重点和实现路径有所不同。因果 AI 更侧重于回答“为什么会发生某事？”这类关于机制的问题，其解释根植于对因果结构的理解。而传统 XAI 方法更多地是解释“模型是如何得出某个决策的？”，例如通过识别哪些输入特征对模型的输出贡献最大。因此，基于因果关系的解释往往更具有鲁棒性和行动指导意义。例如，一个 XAI 工具可能指出患者的高血压是心脏病风险预测模型中的一个重要特征，而因果 AI 则会尝试阐明高血压是如何通过一系列生理机制（如血管损伤、心脏负荷增加）导致心脏病风险增加的。

下表总结了医学 AI 中因果推理与可解释性的主要区别：

表 1：医学 AI 中因果推理与可解释性（XAI）的比较

特征	因果 AI (Causal AI)	可解释性 AI (XAI)
核心目标	理解现象背后的“为什么”（因果机制）	理解模型决策的“如何”（模型行为）
解释生成	通常为事前（ <i>ante hoc</i> ），模型内在地可解释	多为事后（ <i>post hoc</i> ），为已构建模型提供解释
模型透明度	通常为白箱模型，结构明确	可应用于白箱或黑箱模型，后者通过解释层增加透明度
对信任的影响	潜力更高，因提供机制性洞	提升信任，但事后解释的忠实

	察，解释更根本	性可能受质疑
典型方法/焦点	结构因果模型、反事实推断、干预效果估计、因果发现	LIME、SHAP、特征重要性分析、规则提取、注意力机制
数据依赖	强调识别和处理混杂因素，可能需要干预数据或强假设	主要依赖模型输入输出数据，对数据生成过程的假设较少
行动指导性	更强，可指导干预措施以改变结果	可能较弱，解释可能不直接对应可操作的干预

医学 AI 领域从最初主要关注预测准确性，逐渐演进到对理解、信任和负责任部署的全面需求，这反映了 AI 技术应用的成熟过程。在医疗这一高风险、高影响的领域，准确性固然重要，但若缺乏透明度和因果理解，AI 技术的潜力将难以充分发挥，甚至可能带来负面影响 [2, 5]。因此，因果推理和可解释性成为当前医学 AI 研究与应用的前沿和热点。

值得注意的是，尽管因果 AI 因其“事前”和机制性的解释被认为能提供更根本的理解和更高的可信度 [3]，但在当前的学术研究和应用实践中，诸如 LIME、SHAP 等“事后”XAI 技术似乎被更广泛地讨论和采用 [6]。这一现象可能暗示，虽然从理论上讲，构建真正的因果模型是实现深度理解的理想途径，但在实际操作中，为已有的复杂预测模型（尤其是深度学习模型）添加 XAI 解释层，可能在技术成熟度、实施便捷性或特定应用场景的即时需求方面具有一定优势。构建准确且全面的因果模型本身是一项极具挑战性的任务，需要深厚的领域知识、高质量的数据，并可能涉及复杂的因果发现和验证过程。

2. 背景及应用价值

医学 AI 的发展不仅追求技术上的突破，更需关注其在实际应用中所承载的社会价值和伦理责任。因果推理和可解释性正是连接 AI 技术与其在医疗领域安全、有效、公正应用的关键桥梁。

- **2.1. 医学 AI 对透明度、信任和伦理的需求**

在医疗实践中，透明度和信任是医患关系以及新技术采纳的基石。AI 系统做出的决策，尤其是那些直接影响患者诊断、治疗和预后的决策，如果其过程不透明，将难以获得临床医生和患者的信任³。缺乏对模型内部工作机制的理解和信任，已经成为 XAI 在临床实践中整合缓慢和受限的主要原因之一⁴。当 AI 系统出现错误或产生非预期结果时，如果无法追溯其决策逻辑，就难以进行有效的审计、明确责任归属，并从中吸取教训以改进系统⁴。

这种对透明度和信任的需求，并非仅仅是技术层面的考量，而是深刻嵌入在医学伦理的基本原则之中。患者的自主权要求他们有权了解自身病情及治疗方案的相关信息，并在此基础上做出知情同意。如果 AI 的决策过程是一个无法解释的“黑箱”，那么基于 AI 建议的医疗决策在何种程度上能够真正体现患者的知情同意，便成了一个疑问。同样，医生的专业责任要求他们对医疗决策负责，而对一个不理解其工作原理的 AI 工具的盲目依赖，显然与此相悖。此外，AI 算法中可能存在的偏见，如果不能被揭示和理解，就可能导致医疗资源分配不公或对特定人群产生歧视，违背了医疗的公正性原则。因此，推动医学 AI 的透明化和可解释化，是确保 AI 技术应用符合患者最佳利益、尊重个体权利、促进医疗公平的伦理要求。

- **2.2. 因果推理在提升临床决策质量中的核心价值**

循证医学（EBM）强调临床决策应基于当前可获得的最佳证据²。对于治疗决策而言，最佳证据是指能够确定治疗措施对临床结局的因果效应的证据。如果推荐的治疗方案并不具备有益的因果效应，那么它将无法改善患者的健康状况²。传统医学研究（如观察性研究）常常只能揭示变量间的关联性，而这些关联性可能是由混杂因素导致的非因果联系，若基于此类证据进行治疗决策，可能产生偏倚，甚至对患者造成伤害²。

因果推理的核心价值在于它提供了一套系统性的方法论，帮助研究人员和临床医生从

相关性中甄别出因果性，从而更准确地评估治疗干预的真实效果⁷。例如，通过构建因果图（如 DAGs）明确变量间的假设因果路径，识别潜在的混杂因素，并采用倾向性评分匹配、工具变量分析等方法来调整或控制这些混杂因素的影响，从而得到更接近真实因果效应的估计值¹。这种能力使得 AI 系统不仅能预测“可能会发生什么”，还能回答“如果采取某种干预，将会发生什么？”这类反事实问题，为临床医生在众多治疗方案中进行选择提供了更为坚实的科学依据¹。

除了个体临床决策，因果推理在公共卫生政策制定和监管科学领域也展现出巨大潜力。利用日益增长的真实世界数据（RWD），如电子健康记录（EHR）、医保数据等，结合因果推断方法，可以在药品或医疗器械上市后快速评估其在真实人群中的有效性和安全性⁷。这相较于传统的随机对照试验（RCTs）周期长、成本高、普适性有限的特点，具有显著的补充优势，能够为卫生政策的调整和监管决策提供更及时、更广泛的证据支持⁷。例如，当出现新的公共卫生威胁（如疫情）或需要评估某项广泛实施的健康政策（如疫苗接种计划）时，基于 RWD 的因果分析能够更快地生成证据，指导应对策略。

- **2.3. 可解释性 AI 在保障医疗安全与合规中的关键作用**

医疗安全是医疗质量的核心。XAI 通过使 AI 的预测与临床实际情况相符，有助于提升医疗安全水平⁵。当 AI 模型的决策过程清晰易懂时，临床医生能够更好地判断其输出结果的合理性，识别潜在的错误或风险，从而避免盲从 AI 建议可能导致的不良后果。例如，如果一个 AI 系统预测某位患者有较高的术后并发症风险，XAI 能够解释是哪些因素（如患者的特定生理指标、合并症、生活习惯等）导致了这一高风险评估，临床医生可以结合自身的专业知识对这些因素进行核实和权衡，制定更具针对性的预防措施。

同时，清晰的决策过程也有助于满足日益严格的医疗法规要求，如 HIPAA 和 FDA 的相关规定⁵。在某些情况下，提供决策解释甚至是法律上的强制要求³。例如，如果 AI 系统被用于辅助诊断或治疗方案推荐，一旦发生医疗纠纷，能够提供 AI 决策依据的解释将对于责任认定和事件调查至关重要。此外，对于 AI 医疗器械的审批，监管机构也越来越关注其安全性和可靠性，而 XAI 提供的透明度是证明这些特性的重要

途径。通过确保 AI 决策过程的可追溯性和可审查性，XAI 为构建负责任的医疗 AI 系统提供了技术保障。

3. 学术研究现状及瓶颈

医学 AI 领域的因果推理和可解释性研究在过去十年中取得了显著进展，涌现出多种理论框架和技术方法。然而，在将这些研究成果有效应用于复杂的临床实践中，仍面临诸多挑战。

● 3.1. 因果推理研究进展与主要方法

因果推理的研究旨在从数据中发现和量化因果关系。其核心思想根植于统计学、计算机科学、流行病学和哲学等多个学科。关键方法包括：

- **结构因果模型 (Structural Causal Models, SCMs)**：由 Judea Pearl 等人发展的 SCMs 提供了一个用数学语言描述因果假设的形式化框架。SCMs 通常包含一组变量和一系列方程，这些方程描述了变量之间直接的因果机制¹。
- **因果有向无环图 (Causal Directed Acyclic Graphs, DAGs)**：DAGs 是 SCMs 的图形化表示，用节点代表变量，用有向边代表直接因果关系²。DAGs 能够清晰地展示变量间的因果假设，帮助识别混杂因素（confounders）、对撞因子（colliders）和中介变量（mediators），从而指导研究设计和数据分析策略，以估计无偏的因果效应²。
- **倾向性评分方法 (Propensity Score Methods, PSM)**：在观察性研究中，用于处理选择偏倚和混杂问题。通过估计每个个体接受特定干预的概率（即倾向性评分），然后在评分相似的个体间进行匹配、分层或加权，以平衡干预组和对照组的基线协变量分布，从而模拟随机对照试验的效果³。
- **工具变量分析 (Instrumental Variable Analysis, IV)**：当存在未观测到的混杂因素时，IV 分析是一种有力的工具。一个有效的工具变量与干预（暴露）相关，但与结局仅通过干预发生联系，且不与任何未观测混杂因素相关⁴。孟德尔随机化（Mendelian Randomization）是 IV 分析在遗传流行病学中的一个重要应用，利用随机分配的基因变异作为工具变量来推断暴露因素（如生活方式、生物标志

物) 与疾病结局间的因果关系。

- **因果机器学习 (Causal Machine Learning)**：结合了机器学习的强大预测能力和因果推理的框架，旨在从高维复杂数据中估计异质性因果效应。例如，因果森林 (Causal Forests) 和基于深度学习的反事实估计器等方法，能够识别不同亚组人群对干预的不同反应，为个性化医疗提供支持⁸。
- **因果发现算法 (Causal Discovery Algorithms)**：这类算法尝试从观测数据中自动学习因果结构 (如 DAGs)，而无需预先指定所有因果假设。它们通常基于条件独立性检验或利用数据中的非高斯性、非线性等特征¹。

下表总结了医学 AI 中常用的关键因果推理技术：**表 2：医学 AI 中的关键因果推理技术**

技术名称	简介	常见医学应用	主要假设/局限性
因果有向无环图 (DAGs)	图形化表示变量间因果假设，识别偏倚路径	研究设计、混杂控制、中介分析	依赖领域知识正确构建图结构，难以处理未观测混杂
结构因果模型 (SCMs)	用方程描述因果机制，是 DAGs 的数学基础	复杂系统建模、反事实推断	需要明确函数形式，参数估计可能复杂
倾向性评分方法 (PSM)	通过平衡协变量分布估计平均干预效应	观察性研究中的治疗效果评估	强不可知性 (所有混杂因素均已测量)，倾向分模型设定正确，重叠假设
工具变量分析 (IV)	利用工具变量处理未观测混杂，估计因果效应	药物效果评估 (如孟德尔随机化)、政策评估	存在有效工具变量 (相关性、独立性、排他性)，线性假设 (部分方法)

孟德尔随机化 (MR)	以基因变异为工具变量研究暴露与结局的因果关系	风险因素的病因推断、药物靶点验证	基因变异满足 IV 假设（如无水平多效性），群体分层等偏倚
因果森林/因果 ML	利用机器学习估计异质性因果效应，识别亚组差异	个性化医疗、治疗推荐	依赖于底层因果识别策略的有效性，模型本身可能复杂

● 3.2. 可解释性 AI 研究进展与主流技术

XAI 的研究旨在开发能够使其决策过程透明化的方法和技术。根据解释的范围、模型依赖性以及生成时间，XAI 技术可以分为多种类型⁶。

- **局部解释方法 (Local Explanations)**：关注解释模型对单个样本的特定预测。
 - **LIME (Local Interpretable Model-agnostic Explanations)**：通过在待解释样本附近生成扰动数据，并用一个简单的、可解释的模型（如线性模型）来拟合原始复杂模型在这些扰动样本上的预测，从而得到对该样本预测的局部线性近似解释³。LIME 是模型无关的，可以应用于任何黑箱模型。
 - **SHAP (SHapley Additive exPlanations)**：基于博弈论中的 Shapley 值概念，将预测结果公平地分配给每个输入特征，从而量化每个特征对特定预测的贡献度³。SHAP 值具有良好的理论性质，如一致性和可加性，能够提供比 LIME 更稳定的特征重要性度量。
- **全局解释方法 (Global Explanations)**：旨在理解模型的整体行为和学习到的模式。SHAP 也可以通过聚合大量样本的 SHAP 值来提供全局特征重要性的概览。其他方法包括部分依赖图（Partial Dependence Plots, PDPs）和累积局部效应图（Accumulated Local Effects, ALE plots），它们展示了一个或两个特征如何边际地影响模型的预测。
- **基于模型内部结构的方法**：
 - **注意力机制 (Attention Maps)**：主要用于深度学习模型，尤其是在自然语言处理和计算机视觉领域。注意力机制能够高亮显示模型在做出决策时重点关

注的输入部分（如图像区域或文本序列中的词语），从而提供一种直观的解释⁵。

■ **梯度方法 (Gradient-based Methods)**：如 Saliency Maps、Integrated Gradients 等，通过计算输出相对于输入的梯度来评估特征的重要性。这些方法通常用于解释基于梯度的模型，如神经网络⁶。

- **模型特定解释 (Model-specific Explanations)**：这类方法依赖于特定模型的内部结构，例如决策树的规则路径、线性模型的系数等。
- **事前解释 (Ante-hoc Explanations)**：指模型本身就是可解释的，如逻辑回归、决策树或因果模型。这类模型的结构和参数直接反映了决策逻辑³。

下表总结了医学 AI 中常用的主流 XAI 技术：表 3：医学 AI 中的主流 XAI 技术

技术名称	简介	类型 (局部/全局, 模型无关/特定, 事后/事前)	医学应用	优势/劣势
LIME	通过局部代理模型解释单个预测	局部, 模型无关, 事后	图像诊断解释、风险评分因子识别	易于理解和实现；解释可能不稳定，对扰动敏感
SHAP	基于 Shapley 值量化特征贡献	局部/全局, 模型无关 (理论上), 事后	基因表达分析、疾病预测因子排序、临床决策支持解释	理论基础坚实，提供一致性保证；计算量可能较大，对某些模型近似计算
注意力机制	高亮模型在输入数据中关注的区域	局部 (通常), 模型特定 (DL), 事中/事后	医学影像病灶定位、文本报告关键信息提取	直观易懂，与人类认知相似；可能不完全反映决策逻辑，“注意力不等于解释”

梯度方法	基于梯度评估特征重要性	局部, 模型特定 (可微模型), 事后	神经网络特征可视化	计算高效; 梯度饱和问题, 解释可能不鲁棒
决策树/规则列表	模型本身即为可解释的规则集	全局, 模型特定, 事前	临床指南生成、简单诊断规则提取	透明直观; 对于复杂关系表达能力有限, 可能牺牲预测性能

● 3.3. 当前面临的核心瓶颈与技术挑战

○ 3.3.1. 因果推理的挑战

尽管因果推理方法为医学研究提供了强有力的工具，但在实际应用中仍面临诸多挑战：

- **混杂偏倚 (Confounding Bias)**：检测和度量混杂效应，尤其是未观测到的混杂因素，是因果推理中的核心难题⁹。因果充分性假设（即所有相关的混杂变量都已被观测到）在现实中往往难以满足且无法通过经验数据进行检验⁹。如果存在未被控制的混杂因素，那么观察到的暴露与结局之间的关联可能是虚假的，从而导致错误的因果推断。
- **数据局限性**：虽然随机对照试验（RCTs）是建立因果关系的“金标准”，但在许多医学情境下，RCTs 因成本高昂、伦理限制或实际操作困难而不可行⁹。因此，研究者常常依赖观察性数据（如 EHR 数据）。然而，EHR 数据本身也存在诸多问题，例如：缺乏真正的阴性记录（许多诊断只有阳性记录或缺失状态）、差异性监测（患者与医疗系统交互频率不同导致非随机缺失）、信息性缺失、模型误设等多种偏倚来源¹¹。此外，EHR 数据样本量巨大，虽然能提高统计精度（置信区间变窄），但也可能放大潜在偏倚的影响，导致错误地拒绝零假设¹²。
- **模型验证困难**：因果模型的核心在于估计反事实结局（即如果采取了与实际不同的干预，结局会怎样），但反事实结局在定义上就是不可观测的¹³。这

使得因果模型的验证极具挑战性。目前，因果方法的评估主要依赖于手工制作的模拟数据，但这些模拟数据生成过程往往是现实的简化模型，缺乏真实世界数据的复杂性，其评估结果的外部有效性有限¹³。

- **时间动态性与复杂干预**：许多医学问题涉及随时间变化的暴露、混杂因素和结局，以及复杂的、多组分的干预措施。传统的因果推理方法在处理这类动态和复杂情境时能力有限，需要更高级的模型和分析技术。

○ 3.3.2. XAI 的挑战

XAI 技术虽然取得了不少进展，但在临床应用中仍面临以下瓶颈：

- **解释的忠实性与可靠性**：事后解释方法（如 LIME、SHAP）虽然应用广泛，但其解释的忠实性（即解释是否真实反映了模型的决策逻辑）和可靠性（即解释是否在相似情况下保持一致）有时会受到质疑³。解释本身也可能被“操控”或产生误导。
- **计算成本与实时性**：为复杂的深度学习模型生成高质量的解释，尤其是实时解释，可能需要大量的计算资源，这在时间敏感的医疗场景（如急诊）中可能成为一个主要障碍⁵。
- **临床整合与工作流程**：将 XAI 工具无缝集成到现有的临床工作流程和电子健康记录（EHR）系统中，并对医护人员进行有效培训，使其能够正确理解和使用解释结果，是一项复杂的系统工程⁵。
- **评估标准的缺乏**：当前 XAI 领域一个“关键的空白”是缺乏稳健的、量化的解释评估指标和标准化的评估框架⁶。大多数研究依赖于轶事证据、案例研究或专家意见来评估解释的质量，这使得不同 XAI 方法之间的比较变得困难，也难以客观衡量解释对临床决策的实际效用。
- **性能与可解释性的权衡**：通常情况下，模型的预测性能（如准确率）与可解释性之间存在一种权衡关系¹⁵。高度复杂的模型（如深度神经网络）往往具有更强的预测能力，但其可解释性较差；而简单的、本质可解释的模型（如线性回归、决策树）虽然易于理解，但可能无法捕捉复杂数据中的非线性关系，导致性能下降。如何在两者之间找到适合特定医疗任务的最佳平衡点，

是一个持续存在的挑战¹⁵。

这些挑战之间并非孤立存在，而是相互交织，共同构成了当前医学 AI 领域因果推理与可解释性研究的核心难题。一个根本性的、贯穿始终的瓶颈可以被视为“表征与现实的差距”。无论是因果模型（如 DAGs）还是 XAI 方法，它们都是基于对现实世界的某种表征（即模型和数据）进行操作的。因果模型依赖于我们对因果结构的假设²，而 XAI 方法解释的是模型的行为，而非必然是现实世界现象本身。如果模型或其赖以训练的数据未能准确、完整地表征复杂的医学现实（例如，由于存在未测量的混杂因素¹⁰，或 EHR 数据中固有的偏见和缺失¹²），那么基于此产生的因果推断和解释，无论其方法论多么精妙，都可能在临床实践中产生误导，甚至带来危害。更进一步地，因果推理中的挑战（如未观测混杂）会直接加剧 XAI 面临的困境。如果一个模型的预测受到了由混杂因素引起的虚假相关性的影响，那么 XAI 方法（如 SHAP 或 LIME）可能会将这些具有误导性的相关因素识别为“重要特征”，从而提供一些在技术上忠实于模型行为、但在临床上不正确甚至有害的解释。例如，如果模型错误地学习到某个与疾病结局相关的非致病因素（因为它与真正的致病因素共现），XAI 可能会突出这个非致病因素，引导临床医生关注错误的方向。这恰恰凸显了因果 AI 通过尝试建模真实机制，其内生解释在理论上更为稳健的原因¹。最后，XAI 评估的困境⁶在医学领域尤为突出，因为“好的”解释不仅仅关乎人类是否容易理解，更重要的是其临床效用和安全性。一个简单易懂的解释如果忽略了关键的临床背景，或者强调了具有误导性的因素，那么它可能是无用的，甚至是危险的。缺乏标准化的、与临床相关的评估指标，意味着我们很难严格比较不同 XAI 方法在特定医疗任务中的优劣，也阻碍了 XAI 研究成果向可靠临床工具的转化⁵。

4. 过往研究的里程碑

因果推理和可解释性 AI 的发展并非一蹴而就，而是建立在一系列理论突破和方法创新的基础之上。回顾这些里程碑有助于我们理解当前研究的脉络和未来的发展方向。

- 4.1. 因果科学理论的奠基性工作

现代因果科学理论的形成，很大程度上归功于图灵奖得主 Judea Pearl 及其同事的开创性工作。Pearl 的研究，特别是他关于结构因果模型（SCMs）、因果图（如贝叶斯网络和有向无环图 DAGs）以及 do-算子（do-calculus）的理论，为因果推理提供了一套严谨的数学语言和形式化框架¹。他的著作《为什么：因果关系的新科学》

(The Book of Why) 更是推动了一场“因果革命”，将因果关系的研究从哲学的思辨带入了坚实的科学基础之上，阐明了如何回答关于原因和结果的各种问题¹⁸。

Pearl 提出了著名的“因果阶梯” (Ladder of Causation)，将认知能力划分为三个层次：**观察 (seeing/association)**，即理解变量之间如何相关；**干预 (doing/intervention)**，即预测主动改变一个变量会对其他变量产生什么影响；以及**反事实 (imagining/counterfactuals)**，即推断在过去某个条件发生改变的情况下，现在的结果会是怎样¹⁹。当前的许多机器学习算法主要停留在第一个层次（关联），而真正的智能，尤其是医学决策所需要的智能，则需要达到更高层次的因果认知能力。Pearl 的工作不仅为计算机科学和统计学带来了深刻变革，也为医学、经济学、社会科学等领域处理因果问题提供了强大的理论武器。

- 4.2. 关键因果推断方法的提出与发展

在 Pearl 的理论框架以及其他早期研究的基础上，一系列实用的因果推断方法被提出和发展起来，使得从观察数据中进行因果推断成为可能。

- **孟德尔随机化 (Mendelian Randomization, MR)**：该方法最早由 Gray 和 Wheatley 在 1986 年提出，其核心思想是利用父母传递给子女的基因变异的随机分配特性，将基因型作为某种暴露因素（如生活习惯、生物标志物）的工具变量 (Instrumental Variable, IV)，从而在存在未观测混杂的情况下，估计该暴露因素与疾病结局之间的因果关系⁸。MR 的理论基础源于孟德尔的遗传定律以及早期遗传学家如 Sewall Wright 的路径分析思想²¹。由于基因分配的随机性类似于 RCT 中的随机分组，MR 常被称为“自然的随机对照试验”，在流行病学和药物靶点验证等领域得到了广泛应用。
- **倾向性评分匹配 (Propensity Score Matching, PSM)**：由 Paul Rosenbaum 和 Donald Rubin 在 1983 年正式提出，旨在通过统计方法减少观察性研究中的选择偏倚和混杂偏倚⁸。其核心思想是为每个研究对象估计其接受特定干预（如某种治疗）的条件概率（即倾向性评分），然后在倾向性评分相似的个体之间进行匹配、分层或加权，使得干预组和对照组在已观测的基线协变量上达到平衡，从而更准确地估计干预的平均因果效应。

- **工具变量分析 (Instrumental Variables, IV)**：虽然其思想可以追溯到更早，但在经济学领域得到了充分发展和应用，并逐渐渗透到其他学科⁸。如前所述，MR 是 IV 在特定情境下的应用。IV 方法通过引入一个与干预相关、但仅通过干预影响结局且不受混杂因素干扰的“工具”，来识别和估计干预的因果效应。这些方法的出现和不断完善，极大地丰富了研究者从非实验数据中探究因果关系的工具箱，对循证医学的发展产生了深远影响。
- **4.3. 可解释性 AI 方法的演进与突破**

随着机器学习模型，尤其是深度学习模型的复杂性和应用范围不断扩大，对其“黑箱”性质的担忧也日益增加，这催生了对可解释性 AI (XAI) 的迫切需求。在过去十年中，XAI 领域的研究论文数量显著激增⁴。

早期的可解释性工作主要集中在本质可解释的模型，如线性回归的系数、决策树的规则路径等。然而，这些模型的表达能力有限，难以处理日益复杂的医学数据。

为了解释复杂的黑箱模型，研究者开发了一系列事后 (post-hoc) 解释技术。其中，模型无关的局部解释方法，如 LIME (Local Interpretable Model-agnostic Explanations)，通过在单个预测点附近用简单的、可解释的代理模型来近似复杂模型的行为，从而提供对该预测的解释³。

另一类具有里程碑意义的方法是基于 Shapley 值的技术，典型代表是 SHAP (SHapley Additive exPlanations)³。SHAP 通过将博弈论中的 Shapley 值概念应用于模型解释，为每个特征对预测的贡献提供了一个具有良好理论保证的度量。SHAP 不仅可以提供局部解释，还可以通过聚合得到全局特征重要性。

此外，针对特定类型的模型（尤其是深度神经网络），也发展出许多解释方法，如基于梯度的显著性图 (saliency maps)、类激活图 (class activation maps, CAM) 以及注意力机制 (attention mechanisms) 等，它们试图可视化模型在做决策时关注的输入特征或区域⁵。

XAI 领域的研究呈现出高度多样化的特点，横跨多个子学科，并且仍在快速发展之中，不断有新的方法和理论涌现⁶。

回顾这些里程碑可以发现，因果推理和可解释性 AI 的发展常常受益于跨学科的知识

融合。例如，因果推理借鉴了经济学的工具变量思想⁸，并与遗传学结合产生了孟德尔随机化²¹。Judea Pearl 的理论体系本身就是计算机科学、统计学和哲学的结晶¹⁸。这种跨学科交叉渗透对于推动医学 AI 领域的因果认知和解释能力至关重要，也预示着未来的突破可能同样来自于不同学科思想的碰撞与融合。

另一个值得注意的现象是，从基础理论的提出到其在医学 AI 等应用领域的广泛讨论和实践，往往存在一定的时间滞后。例如，Pearl 的因果理论在 20 世纪末至 21 世纪初逐渐成熟，而“因果 AI”作为一个明确的概念在医学 AI 领域的讨论似乎是近些年的事情。类似地，PSM 和 MR 方法虽然分别在 1983 年和 1986 年就已提出²⁰，但它们在医学研究中的广泛应用和方法学上的精细化发展则是一个持续的过程。XAI 研究的“显著激增”也是在深度学习模型普及之后的近十年才出现的⁴。这反映了一个典型的技术采纳和转化周期：基础理论的突破需要时间才能转化为实用的工具，并在特定应用领域（如复杂的医学场景）中逐步获得认可和推广。这也可能意味着，我们目前在医学 AI 中充分利用因果推理和高级 XAI 的潜力方面，或许仍处于相对早期的阶段，未来还有巨大的发展空间。

5. 产业化案例分析

学术研究的进展最终需要通过产业化应用来体现其价值。在医学 AI 领域，因果推理和可解释性技术正逐渐从实验室走向市场，应用于药物研发、临床决策支持、风险预测等多个方面，并催生了一批专注于此领域的创新企业。

● 5.1. 在药物研发与精准医疗领域的应用

药物研发周期长、成本高、失败率高是长期困扰制药行业的难题。因果推理技术为改进药物研发流程提供了新的思路和工具。通过整合基因组学、蛋白质组学、临床数据等多源信息，结合因果推断方法，可以更有效地识别潜在的药物靶点，预测药物的治疗效果，优化临床试验设计，甚至加速药物重定位（老药新用）的进程⁸。例如，孟德尔随机化研究已成功识别出心血管疾病和神经退行性疾病的新治疗靶点，为后续药物开发提供了方向⁸。

因果 AI 正在精准医疗领域掀起变革，它致力于揭示疾病发生的深层因果机制，并据

此为患者制定个性化的治疗计划³。通过分析个体的基因、环境、生活方式等数据，结合因果模型，可以预测不同患者对特定治疗的反应，从而实现“千人千面”的精准用药和健康管理。

- **5.2. 在临床风险预测与智能辅助诊断中的实践**

XAI 技术在临床风险预测和智能辅助诊断系统中扮演着提升透明度和可信度的关键角色。例如，在临床风险评估中，XAI 可以帮助医生理解 AI 模型是如何根据患者的各项指标（如生命体征、实验室结果、既往病史）预测其未来发生某种疾病（如心脏病）的风险，或估计其在 ICU 的住院时长⁵。在智能辅助诊断方面，XAI 能够解释 AI 模型在分析医学影像（如组织病理图片、DaTSCAN 影像）后，是如何识别出癌症病灶或帕金森病相关特征的⁴。这种解释不仅有助于医生验证 AI 的判断，还能从中学习到新的诊断线索。因果 AI 则可以通过揭示治疗措施与预后之间的真实因果联系，辅助医生选择最优的干预方案¹。

- **5.3. 代表性企业及其解决方案剖析**

全球范围内已有多家企业致力于将因果推理和 XAI 技术应用于医疗健康领域，提供创新的解决方案。

- **Aetion**：该公司提供一个名为 **Aetion Evidence Platform® (AEP)** 的真实世界证据（RWE）平台。AEP 利用因果推断科学和人工智能技术，分析来自电子健康记录、医保索赔等真实世界数据（RWD），为生物制药公司、支付方和监管机构评估医疗干预（如药物、疗法）的安全性、有效性和价值提供决策级证据²⁵。
Aetion 近期已被健康数据平台公司 **Datavant** 收购，旨在共同打造端到端的 RWE 平台²⁶。
- **Verge Genomics**：这是一家临床阶段的生物技术公司，其核心技术平台 **CONVERGE®** 整合了人类基因组学、机器学习和先进的疾病模型。Verge Genomics 从人类组织和遗传学数据出发，运用系统生物学和 AI 来综合证据，优先发现那些具有强疾病因果关系的生物通路和药物靶点，专注于神经退行性疾病如肌萎缩侧索硬化症（ALS）的新药研发²⁸。
- **Causaly**：该公司开发了一个 AI 平台，其核心是一个大规模、高精度的生物医学

知识图谱。该平台旨在帮助生命科学领域的研发人员通过自然语言查询，快速从海量的科研文献和数据库中发现与药物靶点、疾病机制、治疗效果相关的因果证据，从而加速药物发现和开发流程³⁰。Causaly 强调其技术能够区分因果关系与简单的共现关系。

- **其他企业**：除了上述公司，还有一些初创企业和大型科技公司也在积极布局医疗因果 AI 和 XAI 领域。例如，StartUs Insights 列举的 Allos 公司利用因果 AI 技术为患者提供定制化的临床干预方案³¹。微软、谷歌、IBM 等科技巨头也在探索将因果 AI 应用于医疗保健，以辅助治疗决策等¹。BioAI Health 公司则致力于利用多模态 AI（可能包含因果和可解释性组件）变革精准医疗³。

下表列举了部分在医学因果 AI 和 XAI 领域的产业化应用实例：**表 4：医学因果 AI 与 XAI 产业化应用案例**

公司/平台名称	主要技术焦点	核心产品/服务	具体医学 AI 应用
Aetion	基于 RWD 的因果推断与真实世界证据 (RWE)	Aetion Evidence Platform® (AEP)	药物/疗法的有效性、安全性、价值评估；支持监管决策、市场准入
Verge Genomics	基于人类遗传学的因果 AI 药物靶点发现	CONVERGE® 平台	神经退行性疾病（如 ALS）的新药靶点识别与药物开发
Causaly	基于知识图谱的因果证据发现	AI 驱动的生物医学知识发现平台	加速药物研发流程中的文献回顾、靶点验证、假设生成
Allos (初创公司)	利用因果 AI 进行定制化临床干预	平台技术	改善临床试验设计，降低试验成本，提升患者安全，提供个性

			化干预方案
Microsoft/Google/IBM	通用因果 AI 技术在医疗领域的应用探索	AI 平台、云计算服务、研究项目	辅助临床决策、疾病风险预测、公共卫生分析等

从这些产业化案例中可以看出，企业应用往往集中在那些传统方法成本极高、耗时漫长（如新药研发 [8]）或者存在大量未被充分利用的观察性数据（如来自 EHR 的真实世界数据 [7, 25]）的领域。因果 AI 和 XAI 被定位为能够加速发现过程或从现有数据中提取更可靠洞察的增值技术。这表明市场对于超越简单预测、能够提供更稳健、可操作且值得信赖的智能解决方案存在真实需求。

另一个共同趋势是这些企业纷纷打造专门化的平台（如 Aetion 的 AEP [25]，Verge Genomics 的 CONVERGE® [28]，Causaly 的 AI 平台 [30]）。这表明在医学领域有效应用因果推理和 XAI，需要的不仅仅是现成的算法，更是一个整合了数据处理、特化模型、领域知识和验证流程的综合环境。这类平台的出现，标志着市场正在走向成熟，解决方案也日益精细化，并针对特定行业工作流程（如药物开发全周期）进行定制。

近期 Datavant 对 Aetion 的收购 [26] 则可能预示着在真实世界证据和健康数据分析领域出现整合和集成的趋势。将 Aetion 的因果推断能力与 Datavant 更广泛的健康数据连接能力相结合，表明市场看重从多样化数据源生成可行动证据的端到端解决方案。这或许暗示着未来因果分析能力将更多地被嵌入到大型健康数据生态系统中，而非作为独立、小众的产品存在。

6. 总结与展望

因果推理与可解释性是当前医学人工智能领域发展的两大核心驱动力，它们共同致力于提升 AI 系统在医疗健康领域的可靠性、可信度、安全性和临床应用价值。本报告对这两个领域的问题概要、背景价值、学术现状、研究瓶颈、历史里程碑及产业化案例进行了梳理和分析。

总结而言，因果推理通过探究变量间的真实因果联系，帮助我们超越表面相关性，深入理解疾病机理、准确评估干预效果，为循证医学和精准医疗提供了强有力的理论和方法支撑。Judea Pearl 的因果科学框架，以及如倾向性评分、孟德尔随机化、结构因果模型等具体方法的提出和发展，构成了该领域的重要基石。可解释性 AI (XAI) 则致力于打开 AI 模型的“黑箱”，通过 LIME、SHAP、注意力机制等技术，使复杂的 AI 决策过程对人类用户透明化，从而增强临床医生的信任，保障医疗安全，并满足伦理与法规的要求。

然而，这两个领域在迈向广泛和深度临床应用的过程中仍面临诸多严峻挑战。因果推理方面，未观测混杂因素的普遍存在、观察性数据（尤其是 EHR 数据）的固有偏倚和质量问题、以及因果模型验证的内在困难，都是亟待解决的瓶颈⁹。XAI 方面，事后解释的忠实性和可靠性、解释评估标准的缺乏、解释的临床效用难以量化、以及模型性能与可解释性之间的权衡，都限制了其潜力的充分发挥³。

展望未来，克服这些挑战并推动医学 AI 向更高级阶段发展，需要在以下几个方面持续努力：

1. **深化跨学科研究与合作**：医学 AI 的进步离不开计算机科学、统计学、流行病学、临床医学、伦理学等多学科的深度融合。鼓励不同领域专家之间的对话与协作，共同攻克理论和实践难题。
2. **构建稳健的评估框架与标准**：特别是针对 XAI，亟需建立一套科学的、量化的、且与临床实际相结合的评估体系，从而客观衡量不同解释方法的质量及其对临床决策的真实影响。对于因果模型，也需要探索更有效的验证方法，超越单纯依赖模拟数据的局限。
3. **推动因果与解释的深度融合**：虽然因果 AI 本身具有内生的可解释性，但如何将复杂的因果推断过程以用户友好的方式呈现给临床医生，仍是一个挑战。同时，也需要探

索如何将因果思维更深刻地融入到 XAI 方法中，使解释不仅反映模型的行为，更能揭示潜在的因果机制，从而提供更具洞察力和行动指导意义的解释。正如一些研究指出的，因果机器学习（Causal ML）和可解释性 AI（XAI）的协同可能是构建准确、透明且可解释的模型的未来方向³。

4. **加强数据治理与提升数据质量**：高质量的数据是进行有效因果推理和构建可靠 AI 模型的基础。需要加强对医疗数据的标准化、集成和质量控制，特别是针对真实世界数据，要充分认识并妥善处理其偏倚和局限性。
5. **关注伦理、法规与社会接受度**：随着 AI 技术在医疗领域的深入应用，必须持续关注其可能带来的伦理挑战、法律责任问题以及公众的接受程度。建立完善的 AI 治理框架，确保技术发展始终服务于人类健康福祉。

最终的目标，是推动医学 AI 从仅仅依赖“相关性的希望”迈向基于“机制性置信”的智能决策²⁹。通过不断的技术创新和审慎的应用实践，因果推理和可解释性必将在构建更智能、更可靠、更公平、更符合伦理的未来医疗体系中发挥不可或缺的作用。

参考文献

- 1 Infobip. (n.d.). Causal AI. Retrieved from <https://www.infobip.com/glossary/causal-ai>
- 3 BioAI Health. (n.d.). Causal AI vs Explainable AI: Understanding the Future of Precision Medicine. Retrieved from <https://www.bioaihealth.com/causal-ai-vs-explainable-ai-understanding-the-future-of-precision-medicine/>
- 2 Shrier, I., & Platt, R. W. (2019). Causal inference for clinicians. *BMJ Evidence-Based Medicine*, 24(1), 2-5.
- 7 Grabner, M., Hoffman, S. R., & Smith, J. (2022, June). What is causal inference and when do you need it? Carelon Research. Retrieved from <https://www.carelonresearch.com/perspectives/what-is-causal-inference-and-when-do-you-need-it>
- 5 Censinet. (n.d.). Explainable AI in Healthcare Risk Prediction. Retrieved from <https://www.censinet.com/perspectives/explainable-ai-in-healthcare-risk-prediction>
- 4 AlAlawi, S. M., et al. (2024). Explainable Artificial Intelligence (XAI) for Disease Prediction: A Review. *Diagnostics*, 14(x), xxx..4 Retrieved from <https://pmc.ncbi.nlm.nih.gov/articles/PMC11877768/>
- 24 ResearchGate. (n.d.). (PDF) The Role of Causal Inference in Drug Discovery and Development. Retrieved from https://www.researchgate.net/publication/390284490_The_Role_of_Causal_Inference_in_Dru

- g_Discovery_and_Development (Referencing the general landing page)
- 8 Young, L. (2025, April 3). The Role of Causal Inference in Drug Discovery and Development. ResearchGate. 8
- 18 Pearl, J., & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. Basic Books. 18
- 19 Pearl, J. (n.d.). Why: The New Science of Cause and Effect - Dedication (English Translation). Retrieved from https://bayes.cs.ucla.edu/WHY/Pearl_ded-english.pdf
- 9 Agrawal, R., et al. (2024). Detecting and Measuring Confounding Effects from Observational Data with Unobserved Confounders. Proceedings of NeurIPS 2024. 9
- 10 NeurIPS. (2024). Detecting and Measuring Confounding Effects from Observational Data with Unobserved Confounders (Poster). Retrieved from <https://neurips.cc/virtual/2024/poster/95068>
- 6 Van der Velden, B. H. M., et al. (2024). A Systematic Review of Explainable Artificial Intelligence (XAI) Applications. Applied Sciences, 14(19), 8884. Retrieved from <https://www.mdpi.com/2076-3417/14/19/8884>
- 15 Adhikari, A., et al. (2024). Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. Journal of Medical AI. 15
- 12 Madhira, V., et al. (2025). Causal Inference via Electronic Health Records in the National Clinical Cohort Collaborative: Challenges and Solutions in Long COVID Research. medRxiv. Retrieved from <https://www.medrxiv.org/content/10.1101/2025.06.06.25329168v1>
- 11 Madhira, V., et al. (2025). Causal Inference via Electronic Health Records in the National Clinical Cohort Collaborative: Challenges and Solutions in Long COVID Research (Full PDF). medRxiv. Retrieved from <https://www.medrxiv.org/content/10.1101/2025.06.06.25329168v1.full.pdf>
- 13 Amazon Science. (n.d.). Validating causal inference models. Retrieved from <https://www.amazon.science/publications/validating-causal-inference-models>
- 14 Parikh, H., et al. (2022). Validating Causal Inference Methods. Proceedings of Machine Learning Research, 162, (PMLR). 14
- 16 Van den Abbeele, A., et al. (2025). Explainability Versus Accuracy of Machine Learning Models: The Role of Task Uncertainty and Need for Interaction with the Machine Learning Model. Journal of Management Accounting Research. 16
- 17 Gather, J., et al. (2024). What Is the Role of Explainability in Medical Artificial Intelligence? A Case-Based Approach. Bioethics. 17
- 31 StartUs Insights. (2025). 10 Causal AI Companies to Watch in 2025. Retrieved from <https://www.startus-insights.com/innovators-guide/causal-ai-companies/>
- 25 AWS Machine Learning Blog. (2025, January 30). How Aetion is using generative AI and Amazon Bedrock to unlock hidden insights about patient populations. Retrieved from <https://aws.amazon.com/blogs/machine-learning/how-aetion-is-using-generative-ai-and-amazon-bedrock-to-unlock-hidden-insights-about-patient-populations/>
- 26 Datavant. (2025, May 15). Datavant to Acquire Aetion, Empowering Healthcare and Life Sciences Organizations to Generate High-Quality, Scalable Real-World Evidence from

Connected, Privacy-Protected Data. Retrieved from <https://www.datavant.com/press-release/datavant-to-acquire-leading-real-world-evidence-rwe-platform-aetion>

28 Verge Genomics. (2025, May 6). Devano and Verge form strategic partnership. Retrieved from <https://www.vergegenomics.com/news-blog/devano-and-verge-form-strategic-partnership>

29 YouTube. (2024, March). How Causal AI Agents Are Redefining Precision Medicine. 29 Retrieved from <https://www.youtube.com/watch?v=AK22sfozMx8>

20 Wikipedia. (n.d.). Mendelian randomization. Retrieved from https://en.wikipedia.org/wiki/Mendelian_randomization#:~:text=The%20study%20design%20was%20first,in%20epidemiology%20for%20establishing%20causality).

21 Wikipedia. (n.d.). Mendelian randomization. Retrieved from https://en.wikipedia.org/wiki/Mendelian_randomization (Main page)

22 Wikipedia. (n.d.). Propensity score matching. Retrieved from https://en.wikipedia.org/wiki/Propensity_score_matching#:~:text=Paul%20R.,a%20set%20of%20observed%20covariates.

23 Manitoba Centre for Health Policy. (n.d.). Propensity Score Matching in Observational Studies. Retrieved from http://mchp-appserv.cpe.umanitoba.ca/concept/propensity_score_matching.pdf

2 ResearchGate. (2024, December 9). Causal inference for clinicians (Summary). 2

8 ResearchGate. (2025, March 31). The Role of Causal Inference in Drug Discovery and Development (Summary). 8

27 AWS Machine Learning Blog & Datavant. 25

28 Verge Genomics. (2025, May 6). CONVERGE® platform (Summary). 28

30 Causaly. (n.d.). Causaly AI Platform. Retrieved from <https://www.causaly.com/> 30

引用的著作

1. What is Causal AI? Definition, Benefits, and Examples - Infobip, 访问时间为 六月 10, 2025, <https://www.infobip.com/glossary/causal-ai>
2. (PDF) Causal inference for clinicians - ResearchGate, 访问时间为 六月 10, 2025, https://www.researchgate.net/publication/331102890_Causal_inference_for_clinicians
3. Causal AI vs Explainable AI | BioAI - BioAI Health, 访问时间为 六月 10, 2025, <https://www.bioaihealth.com/causal-ai-vs-explainable-ai-understanding-the-future-of-precision-medicine/>
4. The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions - PubMed Central, 访问时间为 六月 10, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11877768/>
5. Explainable AI in Healthcare Risk Prediction - Censinet, 访问时间为 六月 10, 2025, <https://www.censinet.com/perspectives/explainable-ai-in-healthcare->

[risk-prediction](#)

6. Recent Applications of Explainable AI (XAI): A Systematic Literature Review - MDPI, 访问时间为 六月 10, 2025, <https://www.mdpi.com/2076-3417/14/19/8884>
7. What is Causal Inference? | Carelon Research, 访问时间为 六月 10, 2025, <https://www.carelonresearch.com/perspectives/what-is-causal-inference-and-when-do-you-need-it>
8. (PDF) The Role of Causal Inference in Drug Discovery and ..., 访问时间为 六月 10, 2025, https://www.researchgate.net/publication/390284490_The_Role_of_Causal_Inference_in_Drug_Discovery_and_Development
9. Detecting and Measuring Confounding Using Causal Mechanism Shifts - NIPS papers, 访问时间为 六月 10, 2025, https://proceedings.neurips.cc/paper_files/paper/2024/file/711ebaa5fcc8321d0075c1b42870b425-Paper-Conference.pdf
10. NeurIPS Poster Detecting and Measuring Confounding Using Causal Mechanism Shifts, 访问时间为 六月 10, 2025, <https://neurips.cc/virtual/2024/poster/95068>
11. Causal Inference via Electronic Health Records in the National Clinical Cohort Collaborative - medRxiv, 访问时间为 六月 10, 2025, <https://www.medrxiv.org/content/10.1101/2025.06.06.25329168v1.full.pdf>
12. Causal Inference via Electronic Health Records in the National Clinical Cohort Collaborative: Challenges and Solutions in Long COVID Research | medRxiv, 访问时间为 六月 10, 2025, <https://www.medrxiv.org/content/10.1101/2025.06.06.25329168v1>
13. Validating causal inference models - Amazon Science, 访问时间为 六月 10, 2025, <https://www.amazon.science/publications/validating-causal-inference-models>
14. Validating Causal Inference Methods - Proceedings of Machine Learning Research, 访问时间为 六月 10, 2025, <https://proceedings.mlr.press/v162/parikh22a/parikh22a.pdf>
15. (PDF) Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review - ResearchGate, 访问时间为 六月 10, 2025, https://www.researchgate.net/publication/386294349_Explainable_and_interpretable_artificial_intelligence_in_medicine_a_systematic_bibliometric_review
16. Full article: Explainability Versus Accuracy of Machine Learning Models: The Role of Task Uncertainty and Need for Interaction with the Machine Learning Model, 访问时间为 六月 10, 2025, <https://www.tandfonline.com/doi/full/10.1080/09638180.2025.2463961>
17. What Is the Role of Explainability in Medical Artificial Intelligence? A Case-Based

- Approach, 访问时间为 六月 10, 2025,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12025101/>
18. The Book of Why: The New Science of Cause and Effect: Pearl, Judea, Mackenzie, Dana, 访问时间为 六月 10, 2025, <https://www.amazon.com/Book-Why-Science-Cause-Effect/dp/046509760X>
 19. Judea Pearl & Dana MacKenzie, The book of why. The new science of causes and effects - UCLA Cognitive Systems Laboratory (Experimental), 访问时间为 六月 10, 2025, https://bayes.cs.ucla.edu/WHY/Pearl_ded-english.pdf
 20. en.wikipedia.org, 访问时间为 六月 10, 2025,
[https://en.wikipedia.org/wiki/Mendelian_randomization#:~:text=The%20study%20design%20was%20first,in%20epidemiology%20for%20establishing%20causality\).](https://en.wikipedia.org/wiki/Mendelian_randomization#:~:text=The%20study%20design%20was%20first,in%20epidemiology%20for%20establishing%20causality).)
 21. Mendelian randomization - Wikipedia, 访问时间为 六月 10, 2025,
https://en.wikipedia.org/wiki/Mendelian_randomization
 22. en.wikipedia.org, 访问时间为 六月 10, 2025,
https://en.wikipedia.org/wiki/Propensity_score_matching#:~:text=Paul%20R.,a%20set%20of%20observed%20covariates.
 23. Propensity Score Matching in Observational Studies - Manitoba Centre for Health Policy, 访问时间为 六月 10, 2025, http://mchp-appserv.cpe.umanitoba.ca/concept/propensity_score_matching.pdf
 24. www.researchgate.net, 访问时间为 六月 10, 2025,
https://www.researchgate.net/publication/390284490_The_Role_of_Causal_Inference_in_Drug_Discovery_and_Development#:~:text=Causal%20inference%20has%20emerged%20as,from%20observational%20and%20experimental%20data.
 25. How Aetion is using generative AI and Amazon Bedrock to unlock hidden insights about patient populations | AWS Machine Learning Blog, 访问时间为 六月 10, 2025, <https://aws.amazon.com/blogs/machine-learning/how-aetion-is-using-generative-ai-and-amazon-bedrock-to-unlock-hidden-insights-about-patient-populations/>
 26. Datavant to Acquire Aetion, Empowering Healthcare and Life ..., 访问时间为 六月 10, 2025, <https://www.datavant.com/press-release/datavant-to-acquire-leading-real-world-evidence-rwe-platform-aetion>
 27. 访问时间为 一月 1, 1970, <https://mdpi.com/2076-3417/14/19/8884>
 28. Devano and Verge form strategic partnership — Verge Genomics, 访问时间为 六月 10, 2025, <https://www.vergegenomics.com/news-blog/devano-and-verge-form-strategic-partnership>
 29. How Causal AI Agents Are Redefining Precision Medicine - YouTube, 访问时间为 六月 10, 2025, <https://www.youtube.com/watch?v=AK22sfozMx8>



30. Causaly: The Most Complete AI Platform For Life Sciences, 访问时间为 六月 10, 2025, <https://www.causaly.com/>
31. 10 Causal AI Companies to Watch in 2025 | StartUs Insights, 访问时间为 六月 10, 2025, <https://www.startus-insights.com/innovators-guide/causal-ai-companies/>