arXiv:2601.03482v1 [cs.AI] 7 Jan 2026

面向健康干预的大基础模型个性化

Stefan Konigorski[1,2], Johannes E. Vedder[1], Babajide Alamu Owoyele[1], İbrahim Özkan[1]

[1]Hasso Plattner数字工程研究所，波茨坦大学，德国波茨坦[2]Windreich人工智能与人类健康系，西奈山伊坎医学院，美国纽约 stefan.konigorski@hpi.de

摘要

大型基础模型（LFMs）正在改变人工智能在医疗保健领域的应用，涵盖预防、诊断和治疗。然而，LFMs 能否提供真正个性化的治疗建议，仍是一个悬而未决的问题。近期研究揭示了个性化面临的多重挑战，其中包括根本性的可泛化性悖论：模型在一项临床研究中达到很高的准确率，在其他研究中的表现却仅处于随机猜测水平，这表明个性化与外部效度之间存在张力。这体现了人工智能驱动的医疗保健中更广泛的矛盾：隐私—性能悖论、规模—特异性悖论以及自动化—共情悖论。另一个挑战在于，个性化建议所需的因果理解程度，与 LFMs 仅具备预测能力之间的区别，仍有待厘清。单一个体试验——交叉自我实验，也是个性化医学中个体因果推断的金标准——通过提供个体内因果证据，同时借助本地实验保留隐私，从而化解这些张力。尽管 LFMs 能力出众，本文认为它们无法取代单一个体试验。我们认为，LFMs 与单一个体试验相辅相成：LFMs 擅长利用多模态数据，从群体模式中快速生成假设；而单一个体试验则擅长对特定个体进行因果验证。我们提出一种结合二者优势的混合框架，以实现个性化并应对已发现的悖论：LFMs 生成带有不确定性估计的干预候选方案排序，并以此触发后续的单一个体试验。明确预测与因果关系之间的界限，并正视这些悖论所体现的张力，对于在个性化医学中负责任地整合人工智能至关重要。

引言

大型基础模型（LFMs）是利用自监督目标，在海量异构数据上训练的大规模神经网络，旨在学习通用表征（Vaswani等，2017）。它们具有很强的迁移能力，无需从头开始重新训练，即可通过微调或提示适配各种下游任务。它们能够捕捉不同模态中的广泛统计结构和语义结构，因而成为现代人工智能系统的统一基础（Bommasani 等人，2022）。

版权所有 © 2026年，已被 AAAI 2026年大型基础模型时代个性化研讨会（PerFM）接收。

在医疗保健领域，LFMs 推动了一场范式转变，使人工智能系统更加以数据为驱动且具有泛化能力。LFMs 已应用于多个领域（Moor 等人，2023），人工智能、机器学习领域以及知名医学期刊上的相关论文数量也日益增加。例如，基于大量电子健康记录（EHR；Du et al. 2026；Steinberg et al. 2024）预训练的 LFMs 、用于医学影像和病理学的 LFMs （ Xu et al. 2024 ；Kondepudi et al. 2025；Kim et al. 2024）、应用于基因组学及其他分子领域的 LFMs（Ali et al. 2025；Dalla-Torre et al. 2025；Fu et al. 2025），以及影像—文本模型（Kim et al. 2024）。这类 LFMs 可用于疾病风险预测（Belyaeva 等人 2023）、疾病诊断（Xiang et al. 2025）、治疗建议（ Alkaeed et al. 2025）以及治疗反应建模（Chen et al. 2025）。医疗保健领域还有许多其他 LFMs，例如基于面部图像训练的 LFMs（Haugg et al. 2025）、基于可穿戴设备数据训练的 LFMs（Erturk et al. 2025），以及用于生物医学研究的 LFMs，例如蛋白质折叠（Jumper et al. 2021）。LFMs 在医疗保健领域的应用也带来了特定挑战，包括数据隐私、偏差、幻觉、可解释性、较高的计算成本，以及与伦理和监管标准的一致性。

本文聚焦于利用大型基础模型为健康干预措施提供个性化服务，包括医疗治疗、生活方式干预，以及任何旨在维持或改善健康的行为。例如，个人应采取什么措施来缓解头痛，或采用哪种药物、何种剂量才能最佳地控制高血压。这带来了大型基础模型面临的两个关键挑战：因果推理和个性化。尽管大型基础模型擅长识别总体人群中的统计模式，却可能缺乏必要的反事实证据，无法确定个体的因果治疗效应。这引出了一个问题：如何才能真正实现对基于总体数据训练的大型基础模型的个性化，使其能够在个体层面给出有因果依据的建议？

面向健康干预的大型基础模型个性化

个性化假设

可以定义一些一般条件，在这些条件下，群体层面的大型基础模型能够为特定患者开出最优治疗建-

议。如果方框 1 中的三个条件有任意一个得到满足，就可以获得最优的个性化治疗建议。否则，大型基础模型给出的建议无法保障最优治疗，且可能存在严重偏差，甚至导致不良健康结局。只有在少数情境下，条件 1 才能得到满足。条件 2 和 3 适用于更广泛的医疗和行为情境，但需要详细了解具体背景，并建立定义明确的模型；该模型可能需要、也可能不需要准确掌握治疗与健康结局的因果结构，同时还需要充分了解患者的个人特征和具体背景。如果治疗效果会随时间变化，情况可能会更加复杂。罕见疾病也带来了另一项挑战：顾名思义，可用于学习群体层面大型基础模型的数据很少，甚至完全没有。

方框 1：个性化假设

满足个性化治疗建议的充分条件包括以下任一项：

1. 该治疗始终对所有人有效。2. 从上下文中可以获得充分信息，并且药物在这种情况下始终对所有人有效。3. 从上下文和患者处可以获得充分信息，并且所学习的模型足够复杂，能够针对每组患者特征和每种上下文，学习出最优的干预措施。

贡献

这些条件带来了额外的挑战。近期研究揭示了一个根本性的可泛化性悖论：在一项临床研究中达到高准确率的模型，在其他研究中的表现却仅处于随机水平，这表明个性化与外部效度之间存在冲突。这体现了人工智能驱动的医疗保健领域中更广泛的矛盾，我们将其概括为：隐私—性能悖论（个性化需要全面数据，但隐私要求将数据最小化）、规模—特异性悖论（基础模型需要庞大的人群，但必须服务于个体），以及自动化—共情悖论（人工智能的效率可能使其旨在改善的以患者为中心的护理失去人性化）。

在医学和生物统计学文献中，针对治疗建议获取个性化证据的金标准是单一个体试验（Nikles 和 Mitchell，2015）。在单一个体试验中，个体会在交叉期内采用不同的治疗方案，并通过自我报告或借助可穿戴设备被动收集关注的健康结局。随后，可对由此产生的单人时间序列进行统计分析和因果推断（Piccininni 等，2024年；Daza，2018年；Konigorski、Ried-Larsen 和 Schmid，2024年）。这种开展新型个性化自我实验的方法，其优势在于能够应对上述许多挑战。不过，如何扩大这种方法的应用规模，使其广泛惠及人群，仍有待探讨。

并广泛应用于人群。

我们认为，结合大型基础模型与单一个体试验，可以为个性化健康干预开辟新途径。下文首先介绍当前医疗保健领域基础模型个性化的实例，然后详细阐述将其用于个性化健康干预所面临的挑战与矛盾，最后讨论结合大型基础模型与单一个体试验的混合框架。该框架应对了已识别的悖论：利用群体先验解决冷启动问题，同时逐步汇聚个体证据；通过选择性验证，在规模与特异性之间取得平衡；并通过透明且以实验为依据的个性化，维持以人为本的护理。

LFM 中现有个性化实践示例

为实现健康干预的个性化，已有若干大型基础模型被提出，它们利用大规模多模态数据，为个体患者量身定制医疗决策。这些模型整合了多种数据来源，包括电子健康记录、医学影像、基因组学和可穿戴设备数据，以提高医疗服务提供的精准度。表1总结了具有代表性的方案：基于现有群体层面数据或新采集的个体数据，在模型构建过程中直接纳入个性化的方案；添加微调层以实现个性化的方案；以及间接方案和假设无需个性化的方案。

大型基础模型个性化的挑战

如表1所述，已有若干模型被提出用于生成个性化健康建议。不过，除了必须满足的一般假设之外，当前模型还面临一些挑战和局限（见方框 1）。

数据需求与冷启动。个性化需要足够的个体内数据来捕捉个体异质性。大多数部署在信号不足的情况下启动，导致估计不稳定、校准效果不佳。当新用户没有历史记录时，冷启动会限制初期效用并延迟收益。纵向可穿戴设备数据和自我报告数据可能包含缺失数据点、伪影和未佩戴偏差；若不加以明确处理，这些问题可能降低建模效果（Dewulf 和 Stock，2019年）。

隐私与安全。个性化系统会汇总敏感的多模态数据轨迹，包括电子健康记录、文本、医学影像、基因组学和可穿戴设备数据。即使经过去标识化，再识别风险仍然存在，且整个开发流程中的数据保管责任仍不明确。持续学习的模型给现有合规框架带来挑战，例如《健康保险流通与责任法案》、《通用数据保护条例》和美国食品药品监督管理局的上市后监管，尤其是在记录音视频数据时（Price 和 Cohen，2019年）。

偏差与公平性。基础模型可能会继承并放大现有的不平等。将医疗成本等代理指标用作需求指标，可能会重现系统性差异。特定人口群体代表性不足，会导致错误率不均，并造成获取机会不平等。医学领域的视觉语言模型所呈现的公平性差距大于人类专家，而且能够从图像中推断受保护属性。许多

No.	模型/方法	训练数据	领域	个性化方法
1	CausalMed（Li 等人 2024年）	EHR	药物推荐 建议	因果发现与整合 模型中的纵向患者数据
2	HeLM（Belyaeva 等人 2023年）	临床特征，疾病 标签，肺量测定 try	疾病 risk 预测，治疗 recommenda- 干预	个性化 建议 基于群体特征的 tics
3	PH-LLM （Khasentino 等，2025年）	Gemini 大语言模型 微调 用于文本理- 解 and 推理	专家领域知识，健康 建议，预测	微调大语言模型，预测患者报告的结局（基于测量的.....） 测得的可穿戴设备数据
4	PhysioLLM（Fang 等人 2024年）	无（提供 Fitbit 数据以供 GPT-4-turbo 在提示词中)	对已实现的个人- 化的评分	与拥有获取机会的大语言模型聊天 个人的 Fitbit 汇总数据
5	Time2Lang（Pillai 等人 2025年)	时间合成数据 周期性的时间序列合成数据	男性中的分类任务- 心理健康分类任务	大语言模型可应用于个人的 可穿戴设备数据
6	联邦微调（Li 等人 2025年）	无（综述论文）	不同的生物医学应用 阳离子	提出联邦大语言模型能够实 现个性化模型微调
7	MedAgentSim (Al- mansoori, Kumar, and Cholakkal 2026年)	无（用户提示 评估中）	患者-医生 对话 内容	通过创建基于智能体的方式间接实现 考虑个人 因素的模拟
8	语言辅助药物 推荐 (Zhao 等人，2025年)	微调 现有的 基于电子健康记录数据的大 语言模型	处方推荐 ，个性化并非 首要目标	忽略个人特征 大语言模型提示，但包含微调后的模型 整体药物-疾病关系
9	UniCure（Chen 等人 2025年）	整合组学 & 化学领域大基础模型	癌症治疗预测	获取个性化药物排序 向量，基于预测的转录组 扰动

表1：医疗保健领域大型基础模型中现有个性化方法总结。模型利用多种数据源，并采用包括因果推断、多模态整合、联邦学习和实时传感器分析在内的各种技术，为患者量身定制干预措施。

已发布的高性能基准模型和工具仍缺乏多样化的多中心验证（Bommasaniet al. 2022年）。

幻觉与可靠性。大型语言模型可能生成流畅但错误的输出，缺乏对最新医学证据的了解，并错误判断自身的不确定性。如果没有强有力的验证、检索和保障机制，这种行为在临床情境中并不安全（Moor 等人，2023）。

大型基础模型个性化面临的四种结构性张力，需要采用混合方法

我们现在指出，人工智能驱动的个性化存在四种内在的悖论性张力，这些张力源于基于人群训练的模型与个体因果推断之间的根本性错配。这些张力促使我们构建混合式个性化系统，并以设计原则和要求为指导，将基础模型与医疗保健领域的实验验证相结合。

张力一：个性化与外部效度

张力所在。针对特定情境优化模型的个性化预测能力，必然会削弱其向新情境泛化的能力，这反映的是一种结构性局限，而非校准问题。

实证证据。（Chekroud 等人，2024）利用试验数据训练机器学习模型，以预测精神分裂症治疗结局。在试验内部的预测表现出色（曲线下面积（AUC）>0.70），但当应用于采用相同治疗方案且人群相似的独立试验时，表现则降至随机水平（AUC ≈ 0.50）。

成因。模型估计平均效应，但如果没有个体层面数据，就无法判断个体属于哪个亚组。此外，模型还会对无法迁移的情境特有偶然因素过度拟合（Subbaswamy、Schulam 和 Saria，2019；Hern´an 和 Robins，2020年）。

设计要求：系统必须量化预测不确定性，并在不确定性阈值（针对治疗效果估计，或治疗效果差异的估计）超出时触发实验验证。

张力 2：数据需求与隐私保护

这种张力在于：有效的个性化需要全面的个体数据，而隐私保护则要求数据最小化（Price 和 Cohen，2019年）。这造成了一种循环依赖：用户不信任系统就不会共享数据，但系统没有数据又无法证明其值得信赖。

为什么技术解决方案还不够。差分隐私会降低准确性（Abadi 等人，2016年），联邦学习会通过梯度泄露信息（Nasr、Shokri 和 Houmansadr，2019年）；多模态健康数据难以去标识化，基因组数据本质上可识别（Gymrek 等人，2013年），而行为模式会形成独特指纹（de Montjoye 等人，2013年）。采用掩蔽方法对实验记录进行去标识化，是在治疗生态系统内共享这些记录的可行方案（Owoyele 等人，2024年）。

设计要求：必须在用户设备上本地进行个体验证，并尽量减少数据传输。只有在获得同意后，匿名化结果才能用于集体学习。

张力 3：大规模人群训练与个体应用

张力所在：基础模型需要庞大人群的数据进行训练（Moor 等人，2023），但临床决策针对的是个体。当异质性显著时，群体估计无法很好地预测个体的反应（Kravitz、Duan 和 Braslow，2004）。

从认识论角度看，“平均患者”是一种数学抽象。若缺乏个体层面证据，在数百万数据上训练的模型就无法识别某个个体属于哪个亚组。从经济角度看，随着治疗方案日益精准化，若将开发成本分摊到规模更小的人群中，成本将变得难以承受（Kimmelman 和 Tannock，2018年）。

设计要求：利用多模态数据进行选择性验证（Schneider、G¨artner 和 Konigorski，2023年；Fu 等人，2023年），仅在高风险或高不确定性、群体知识不足的情境下开展实验。

张力 4：算法效率与以人为本的护理

这种张力。人工智能承诺通过自动化提高效率，但医学需要倾听、理解价值观、建立信任，并以同情心回应痛苦（Pot、Kieusseyan 和 Prainsack，2021年）。算法决策可能将个人视为数据点，从而使医疗服务失去人性化（Morrow 等人，2023年）。

具体表现。基础模型难以把握疾病的叙事维度和存在维度。黑箱的不透明性妨碍了有意义的解释，削弱了共同决策（Kocaballi 等人，2019年；Mahesh 等人，2024年）。如果人工智能提供诊断，临床诊疗互动的治疗价值可能会降低。

设计要求：实验验证提供可解释证据，让患者和临床医生能够共同理解。患者成为知识生成的积极参与者，从而维护其自主性和临床问责制（Konigorski 等人，2022年）。

总结：混合方法为何不可或缺这些张力是将基于人群训练的模型用于个体因果关系时固有的结构性特征。表2总结了混合框架如何应对每种张力。混合系统将基于人群的假设生成（基础模型的优势）与个体层面的实验验证相结合，后者能够提供观察性学习无法实现的因果保证。

通过单一个体试验个性化大型基础模型 单一个体试验

单一个体试验是确定哪些干预对特定个体有效的金标准。不同于估计人群平均效应的传统研究，单一个体试验聚焦于单一个体。如图 1 所示，

	张力	为何会发生	混合解决方案
1	个性化与外部相比有效性	群体模式无法预测个体个体 re-反应	LFM 生成/规提出假设；N-of-1 试验验证当不确定性 high
2	Data 要求-条件 vs. 隐私保护 tion	全面需要数据，但会带来风险	本地实验需要最小化个人数据共享
3	群体训练与个体应用应用	平均值 效应不会 确定个体 re-反应	选择性化 for 效度 高风险／不确定性案例
4	效率与以人为本 care	算法叙事度 lack 维度-	实验性的证据是；患者是积极参与者 ev- 可解读的

表2：四种结构性张力与混合解决方案。

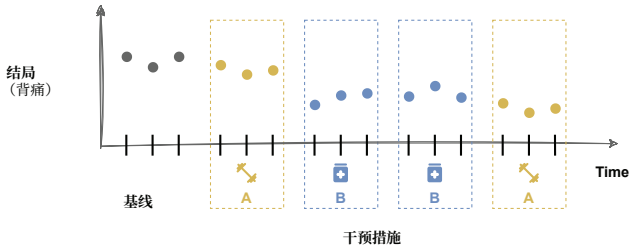

图1：示意图展示了单人试验设计：在基线期之后交替实施干预措施（A 和 B），以评估个体治疗反应随时间的变化。

该个体参与交替进行的重复受控干预期，在不同治疗或条件之间进行切换，并随着时间推移系统地记录结局。这会生成关于什么对该特定个体最有效的因果证据。例如，一名患有慢性疼痛的患者可能在数周内每周交替使用两种药物，并每天评估自己的疼痛程度。对其个人数据进行统计分析，可以揭示哪种药物对该患者更有效，而不论哪种药物对大多数患者“平均而言”更有效。

基于单一个体试验个性化语言模型基础模型的混合框架

我们提出将单一个体试验方法应用于基础模型的个性化。其核心思路是将基础模型视为基线“数字孪生”，用于构建初始假设并提出治疗建议；随后，在需要时通过个体化实验进行动态调整。图2展示了所提出的混合框架。下面我们将逐步说明这一方法在实践中的运作方式。

步骤 1：作为基线的基础模型。一个在大规模人群数据上训练的基础模型作为起点——捕捉从数百万个示例中学到的广泛模式和能力。这代表了“平均的”

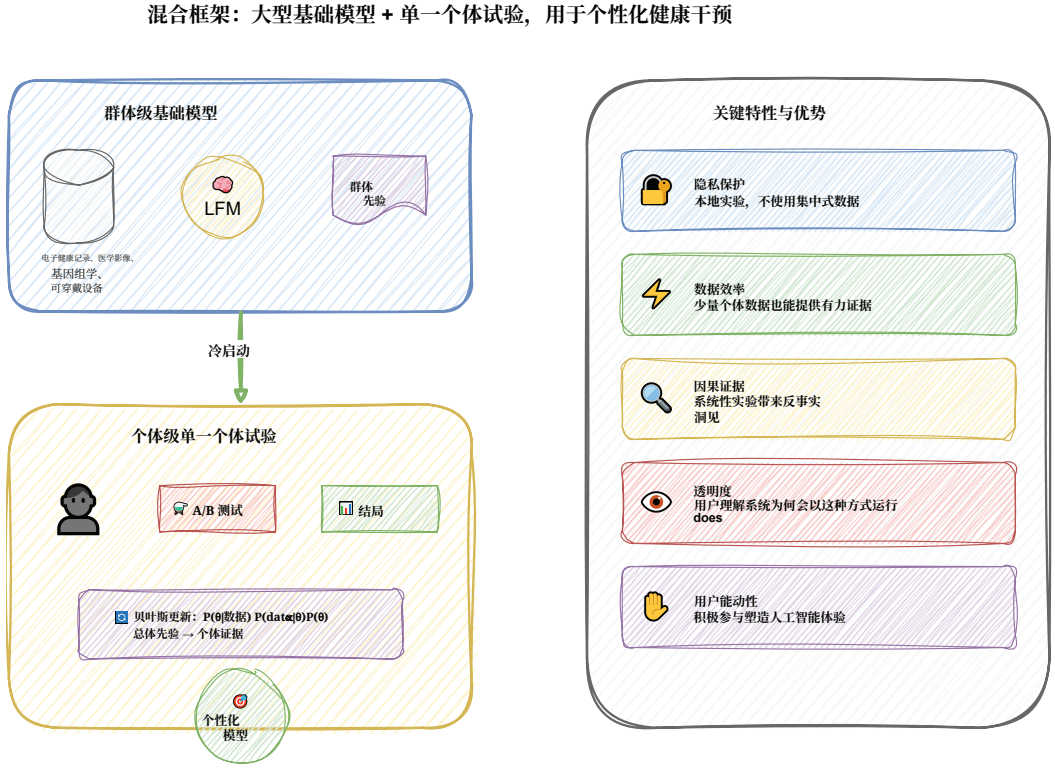

图2：混合框架的可视化示意图，该框架将群体层面的大型基础模型与个体级单一个体试验相结合，以证据为依据实现健康干预的个性化。

知识，类似于医学中的群体层面的治疗效果。

步骤2：单一个体试验。系统会在需要时针对每位用户开展对照实验，以确定最佳个性化策略。系统不会假设对大多数用户有效的方法也适用于这位用户，而是直接针对该个体测试不同方案。具体而言，系统会针对同一用户开展对照实验，比较不同的个性化方法：

- 干预方案变化：比较提示微调版本A与检索增强生成版本B；启用记忆与无记忆响应；或正式与随意的沟通风格。
- 时间交替：在不同会话或不同日期之间切换条件，就像临床试验会在不同治疗方案之间交替一样。
- 个体分析：衡量结局——用户满意度、任务成功率、参与度——并针对特定个人分析结果，而不是对所有用户取平均。

所开展的试验可以是图1所示的固定单一个体试验，也可以是自适应单一个体试验（Shrestha 和 Jain 2021年；Senarathne、Overstall 和 McGree 2020年；Meier、Ensari 和 Konigorski 2023年），以更高效地比较两个以上的干预候选方案：

- 贝叶斯整合：从群体层面的先验（对大多数用户有效的方法）开始，然后结合个体证据进行更新。随着个人数据不断积累，个体模式的影响会越来越大。
- 情境优化：使用情境赌博机等方法，在探索（尝试新方法）与利用（采用已知有效的方法）之间取得平衡，并根据每位用户的情境进行调整。
- 不确定性感知自适应：当对个性化的信心较低时，开展更多试验。当模式变得清晰时，持续一致地应用已确立的偏好。

作为第二步中个性化定制大型基础模型的另一种方式，单一个体试验的相关概念或许能为如何利用用户行为提供持续的实验性证据提供启示。可以考虑纳入明确的用户反馈，例如评分、纠正和明确表达的偏好；也可以纳入隐式信号，例如参与时长、任务完成率、澄清请求的频率、对建议的接受程度；还可以纳入更复杂的行为模式，例如时段偏好、查询复杂度趋势和交互风格。这些指标都有助于提供个性化的性能指标，揭示哪些模型配置最适合该用户。

步骤 3：动态更新。根据实验结果，模型会针对每位用户调整自身行为，创建一个将群体知识与个体证据相结合的个性化数字孪生。在后续的迭代步骤中，模型应保持对不确定性的认知，并知道何时进行试验、何时应用已确立的模式。

这种混合方法结合了群体规模学习（大型基础模型）的优势与个体层面的因果推断（单一个体试验），从而实现既有效又值得信赖的个性化。它还有助于应对前文所述的矛盾，解决人工智能个性化中的关键问题：在用户设备上本地开展实验，保障隐私。个人数据无需集中存储或共享。只有在用户明确同意后，匿名化结果才会用于集体学习。这种方法数据高效，因为系统化实验只需少量个体数据就能获得有力证据，无需全面收集个人数据。用户可以了解系统为何以特定方式运行，因为系统依据的是实验结果，而非不透明的算法决策，从而提升透明度。最后，该框架支持用户能动性：个人可以选择参与结构化实验，掌控自己的数据，并主动参与塑造自己的人工智能体验，而不是被动接受算法建议。

先前研究提出的用于健康干预的数字孪生框架（Qian 等，2021年；Holt、Liu 和 van der Schaar，2024年；Sad´ee 等，2025年）可能包含因果推断，并利用个体层面数据更新模型，这与我们的提议类似。与我们的提议不同，现有框架依赖于更新或微调这些模型，却不生成有关试验有效性的个体层面实验性证据。这些方法在疗效和差异方面的表现，可能源于方框 1 中概述的假设。

示例案例研究：慢性偏头痛管理

为展示我们的混合框架在实践中的应用，我们以慢性偏头痛管理为例进行说明——这种疾病的特点是患者间的治疗反应存在高度差异（Lipton 等，2007）。本案例研究展示了混合框架如何应对已识别的矛盾：LFM 高效地生成假设，而单一个体试验则提供个体层面的因果证据，同时兼顾隐私并维护患者自主权。

临床背景

以患者 Alice 为例，尽管尝试了多种预防性药物，她每月仍有 12 天偏头痛发作。人群层面的证据表明，有多种潜在干预方案（如 β受体阻滞剂、降钙素基因相关肽抑制剂和生活方式调整），但临床试验报告的应答率仅为 40–60%，这表明个体异质性显著。

步骤 1：大型基础模型假设生成

一个基于电子健康记录、可穿戴设备数据和临床试验结果训练的 LFM 会处理 Alice 的个人资料：

- 输入特征：人口统计学特征、合并症、既往用药史、来自可穿戴设备的睡眠模式、自我报告的诱因、遗传标记（如有）
- 输出：按优先级排列的干预候选方案及其不确定性估计。在此情境中，这些估计值表示该治疗方案在所有候选治疗方案中为最佳方案的估计概率，该概率基于估计疗效的不确定性以及与其他治疗方案相比的估计疗效。

LFM 可能会生成以下建议：

干预	预测值疗效	概率是最优（σ）	触发因素单一个体试验?
镁补充剂	0.72	0.30	Yes
睡眠规律性	0.68	0.32	Yes
普萘洛尔 40mg	0.65	0.15	No
减少咖啡因摄入	0.61	0.23	No

表3：LFM 为示例患者 Alice 生成的干预候选方案及示例数值。σ 表示在所有候选治疗方案中成为最佳治疗方案的概率；当不确定性阈值 $\tau = 0.25$ 触发单一个体试验中的验证时。

基于群体证据，无法直接推荐普萘洛尔和减少咖啡因摄入，因为其达到最佳效果的概率为 σ，低于预先设定的 τ =0.25 阈值。然而，镁和睡眠规律性超过了不确定性阈值，因此触发单例验证。

步骤 2：单人试验设计

Alice 参加一项序贯单一个体试验，比较镁补充、睡眠规律性和安慰剂：

- 设计：6 个周期，每个周期 $\times$ 2 周，采用区组随机化，例如 (ABC)(BCA) 或 (BAC)(ABC) 序列
- 主要结局：每个周期的偏头痛发作日数（通过移动应用自我报告）
- 次要结局：疼痛强度（0–10 分量表）、功能障碍、药物使用
- 数据收集：每日头痛日记、可穿戴设备追踪的睡眠质量

步骤 3：贝叶斯更新与结果

12周后，对 Alice 的个体数据进行分析：

$$P(\theta_{\text{Alice}}|D_{\text{Alice}}) \propto P(D_{\text{Alice}}|\theta_{\text{Alice}}) \cdot P(\theta_{\text{Alice}}|\theta_{\text{pop}}) \quad (1)$$

其中，θ_{pop} 表示 LFM 的总体先验，D_{Alice}表示 Alice 的试验数据。

结果：例如，对于 Alice，每月偏头痛发作日减少至少 2 天的后验概率为 90%，而在

实施睡眠规律性方案时，相同概率为 70%。她的个性化数字孪生会相应更新；如有需要，系统可能会在后续试验中继续测试其他候选干预措施。

隐私保护示例

在整个过程中，所有试验数据都保留在 Alice 的设备上；只有经过汇总和匿名化的效应估计会在征得同意后共享，以改进群体先验；原始健康记录不会离开本地环境。

隐私保护型实现

混合框架的一项关键优势是，个体层面的试验可以在本地进行，从而最大限度地减少数据暴露。表 4 总结了所有框架组件中的隐私保护型架构。

安全、监管与公平考量

在医疗保健领域部署混合型 LFM-N-of-1 系统，会带来重要的安全、监管和公平性问题。表5和表6总结了我们负责任的实施方法。

结论

基于总体数据训练的大型基础模型在提供个性化健康干预时面临四种结构性张力：个性化与外部效度（在特定情境中优化会降低模型在不同情境间的泛化能力）、数据隐私（个性化需要全面数据，但隐私要求尽量减少数据收集）、群体与个体规模（模型需要庞大的人群数据，却必须服务于异质性各异的个体），以及算法效率与人文关怀（自动化可能使倾听、共情和建立信任等医学必不可少的要素失去人性化）。我们提出通过混合框架化解这些张力：将群体级基础模型与个体层面的 N-of-1 试验相结合。在这一框架中，自适应数字孪生体继承总体先验，针对个体干预策略开展系统性实验，并依据个体结果通过贝叶斯推断进行更新。这种方法能够量化不确定性，并在必要时触发验证；通过本地实验保护隐私；仅在群体知识不足的高风险情境下开展成本高昂的实验；并通过生成可解释证据维护患者自主权。除了医疗保健领域，这一框架也为任何需要个性化人工智能系统的领域提供了一条有原则的路径：平衡大规模学习的力量与个体层面因果推断的严谨性，打造不仅能大规模展现智能，而且值得信赖、透明，并真正针对每个个体实现个性化的人工智能。

参考文献

Abadi, M.；Chu, A.；Goodfellow, I.；McMahan, H. B.；Mironov, I.；Talwar, K.；以及 Zhang, L. 2016年。《差分隐私深度学习》。载于《2016年 ACM SIGSAC 计算机与通信安全会议论文集》，CCS'16，第308–318页。ACM。

Ali, S.；Qadri, Y. A.；Ahmad, K.；Lin, Z.；Leung, M. F.；Kim, S. W.；Vasilakos, A. V.；以及 Zhou, T. 2025年。《基因组学中的大型语言模型——关于个性化医学的观点》。《生物工程》（巴塞尔，瑞士），12(5)：440。Alkaeed, M.；Abioye, S.；Qayyum, A.；Mekki, Y. M.；Berrou, I.；Abdallah, M.；Al-Fuqaha, A.；Bilal, M.；以及 Qadir, J. 2025年。《医疗领域的开放基础模型：生成式人工智能驱动的个性化处方所面临的挑战、悖论与机遇》。arXiv:2502.04356。

Almansoori, M.; Kumar, K.; and Cholakkal, H. 2026年. MedAgentSim：面向真实临床交互的自我进化的多智能体模拟. 载于 Gee, J. C.; Alexander, D. C.; Hong, J.; Iglesias, J. E.; Sudre, C. H.; Venkataraman, A.; Golland, P.; Kim, J. H.; and Park, J. 编，《医学影像计算与计算机辅助介入 – MICCAI 2025》，362–372. 尚：施普林格·自然瑞士. ISBN 978-3-032-05114-1.

Belyaeva, A.; Cosentino, J.; Hormozdiari, F.; Eswaran, K.; Shetty, S.; Corrado, G.; Carroll, A.; McLean, C. Y.; and Fur- lotte, N. A. 2023年. 基于个体特异性数据的医疗保健多模态大语言模型. 载于《多模态医疗保健数据机器学习研讨会论文集》，86– 102. 施普林格.

Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M. S.; Bohg, J.; Bosse- lut, A.; Brunskill, E.; Brynjolfsson, E.; Buch, S.; Card, D.; Castellon, R.; Chatterji, N.; Chen, A.; Creel, K.; Davis, J. Q.; Demszky, D.; Donahue, C.; Doumbouya, M.; Durmus, E.; Ermon, S.; Etchemendy, J.; Ethayarajh, K.; Fei-Fei, L.; Finn, C.; Gale, T.; Gillespie, L.; Goel, K.; Goodman, N.; Gross- man, S.; Guha, N.; Hashimoto, T.; Henderson, P.; Hewitt, J.; Ho, D. E.; Hong, J.; Hsu, K.; Huang, J.; Icard, T.; Jain, S.; Jurafsky, D.; Kalluri, P.; Karamcheti, S.; Keeling, G.; Khani, F.; Khattab, O.; Koh, P. W.; Krass, M.; Krishna, R.; Kuditipudi, R.; Kumar, A.; Ladhak, F.; Lee, M.; Lee, T.; Leskovec, J.; Levent, I.; Li, X. L.; Li, X.; Ma, T.; Malik, A.; Manning, C. D.; Mirchandani, S.; Mitchell, E.; Munyikwa, Z.; Nair, S.; Narayan, A.; Narayanan, D.; Newman, B.; Nie, A.; Niebles, J. C.; Nilforoshan, H.; Nyarko, J.; Ogut, G.; Orr, L.; Papadimitriou, I.; Park, J. S.; Piech, C.; Portelance, E.; Potts, C.; Raghunathan, A.; Reich, R.; Ren, H.; Rong, F.; Roohani, Y.; Ruiz, C.; Ryan, J.; R´e, C.; Sadigh, D.; Sagawa, S.; Santhanam, K.; Shih, A.; Srinivasan, K.; Tamkin, A.; Taori, R.; Thomas, A. W.; Tram`er, F.; Wang, R. E.; Wang, W.; Wu, B.; Wu, J.; Wu, Y.; Xie, S. M.; Yasunaga, M.; You, J.; Zaharia, M.; Zhang, M.; Zhang, T.; Zhang, X.; Zhang, Y.; Zheng, L.; Zhou, K.; 以及 Lia ng, P。2022年。《基础模型的机遇与风险》。arXiv:2108.07258。

Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H. B.; Patel, S.; Ramage, D.; Segal, A.; and Seth, K. 2017年. 面向隐私保护机器学习的实用安全聚合。载于《2017年 ACM SIGSAC 计算机与通信安全会议论文集》，CCS '17，1175–1191。美国纽约州纽约：计算机协会。国际标准书号 978-1-4503- 4946-8。

Chekroud，A. M.；Hawrilenko，M.；Loho，H.；Bondar，J.；Gueorguieva，R.；Hasan，A.；Kambeitz，J.；Corlett，P. R.；

组件	位置	处理的数据	隐私 机制 **nism**	保障
设备端（本地）组件				
原始数据存储	用户设备	可穿戴设备、自我报告，电子健康记录摘录	本地 加密（AES-256）	数据绝不会离开设备
试验执行	用户设备	随机化，out-摄入追踪	完全本地计算	完全隐私
后验更新	用户设备	个体效应估计值	贝叶斯 更新设备端	无需传输
数字孪生	用户设备	个性化 模型权重	本地 微调 only	用户控制
服务器端组件				
LFM 推理	服务器	特征 嵌入（非原始数据）	嵌入投影-tion	抗重构
总体先验	服务器	聚合 统计数据 only	安全聚合 (Bonawitz et al. 2017年)	个人贡献隐藏
先前贡献	服务器（选择加入）	裁剪后的 梯度 + 噪声	差分隐私 (Abadi 等人，2016)	(ϵ, δ)-DP
试验模板	服务器	通用协议	公开，non-敏感	N/A

表 4：混合框架的隐私保护架构。设备端组件在本地处理所有敏感个人数据，而服务器端组件仅处理经过隐私保护的汇总数据或非敏感模板。

类别	关注事项	缓解策略	实施
干预风险分层			
第1级：低风险（示例：生活方式、补充剂）		如果干预可以进行测试，则完全符合单一个体试验的资格— —验	知情同意 +教育材料— 试验
第2级：中等风险（例如：常用药物）		在临床监督下进行单一个体试验	医疗服务提供者整合；机构审查委员会 对系统性部署进行审查
第2级：高（例如：手术、化疗）		仅限 LFM假设	RCT 需要证据；禁止自行 实验
不良事件监测			
安全性信号	意外症状	自动停止规则	预先定义的严重程度阈值触发 试验终止
升级	严重事件	医疗服务提供者警 报	向医疗保健提供实时通知 team
残留 ef-效应	期与期之间的污染 tion	洗脱期执行或统计 建模	干预阶段之间的强制 洗脱期或试验后建模 eling
LFM 特定防护措施			
幻觉	错误建议	检索增强	以经验证的临床 知识库为依据
禁忌证 不安全的建议		自动检查	根据患者特定的 con- 进行验证 禁忌事项列表
过度自信	校准失当的确定性	置信度阈值	标记低可靠性建议 供临床医生审核
法规合规			
SaMD 分类	美国食品药品监督管理局/欧盟医疗器械法规适用范围	预期用途规范	决策支持与自主建议 建议界定
持续学习	上市后更新	监测方案	监测性能漂移；记录 模型更新
透明度	算法不透明性	文档要求	架构，训练 数据，不确定-性方法已披露

表 5：混合型 LFM-N-of-1 系统的安全与监管框架。干预措施按照风险等级进行分层，并规定相应的监管要求和技术保障措施。

类别	挑战	缓解策略	实施
算法偏见			
代表性不足	不均衡 错误 率 群体间	分层不确定性监测器-ing	代表性不足群体的不确定性升高-受影响群体的诱因 单一个体试验验证更容易
有偏代理变量	历史 差异 编码	公平性感知先验	通过正则化，减少对偏倚结果代理变量的依赖（Obermeyer 等人（2019年））
验证差距	多样化测试有限	包容性试验设计；摘要试验数据反馈至群体 LFM	多语言界面；文化适宜的结局测量指标
数字鸿沟			
设备获取	No 智能手机/穿戴式设备 able	低技术替代方案	纸质结局追踪，配合定期数字化
数字素养	有限 tech 熟练	社区卫生工作者	代录数据；现场支持以参与试验
网络连接	网络有限或无网络	离线功能	本地试验执行；机会性在联网时同步
成本障碍	设备/数据费用	设备借用计划	医疗保健 系统 合作伙伴关系；补贴网络连接
包容性设计			
视力障碍	基于屏幕的界面	屏幕阅读器兼容性	符合 WCAG 2.1 AA 标准；基于语音的交互
运动障碍	Fine 运动 要求-项	简化输入方式	大型触控目标；语音输入；开关访问
认知负荷	复杂试验流程	自适应简化	分步指导；提醒系统；照护者模式
语言障碍	仅支持英语的系统	本地化	多语言用户界面；经文化适配内容
治理			
利益分配	仅为系-统带来效率提升	以患者为中心的结果	透明报告；患者倡导者参与治-治理倡导者
问责	责任不明确	治理结构	明确的责任框架；开放源组件以供验证

表6：公平与可及性框架，确保混合系统服务于多元人群。该框架关注算法公平性、数字接入障碍、包容性设计原则和治理结构。

Koutsouleris, N.; Krumholz, H. M.; Krystal, J. H.; 和 Paulus, M. 2024年。《临床预测模型泛化能力的假象》。《科学》，383(6679)。Chen, Z.; Tian, S.; Pei, J.; Gu, R.; Li, Y.; Ding, S.; Xu, Y.; Zheng, X.; Liu, M.; Du, X.; Zhu, J.; Zou, J.; Xu, J.; Jiang, W.; Ye, C.; Dong, B.; Zhang, Q.; Ren, S.; Wang, S.; Wen, H.; Zhang, W.; 和 Chen, L. 2025年。UniCure：用于预测个体化癌症治疗反应的基础模型。bioRxiv，2025(2025.06.14.658531)。Dalla-Torre, H.; Gonzalez, L.; Mendoza-Revilla, J.; 等。2025年。核苷酸Transformer：构建并评估用于人类基因组学的稳健基础模型。《自然·方法》，22：287–297。Daza, E. J. 2018年。利用反事实框架对单一个体试验中的自我追踪时间序列数据进行因果分析。《医学信息学方法》，57(1)：e10–e21。de Montjoye, Y.-A.; Hidalgo, C. A.; Verleysen, M.; 和 Blondel, V. D. 2013年。人群中的唯一性：人类移动性的隐私界限。《科学报告》，3：1376。Dewulf, B.; 和 Stock, M. 2019年。数据驱动预测中的冷启动问题。载于《推荐系统手册》。施普林格。Du, X.; Zhou, Z.; Wang, Y.; Chuang, Y. W.; Li, Y.; Yang, R.; Hong, P.; Bates, D. W.; 和 Zhou, L. 2026年。电子健康记录应用中生成式大语言模型的测试与评估：系统综述。《国际医学信息学杂志》，205：106091。Erturk, E.; Kamran, F.; Abbaspourazad, S.; Jewell, S.; Sharma, H.; Li, Y.; Williamson, S.; Foti, N. J.; 和 Futoma, J. 2025年。超越传感器数据：可穿戴设备行为数据基础模型提升健康预测。arXiv:2507.00191。Fang, C. M.; Danry, V.; Whitmore, N.; Bao, A.; Hutchison, A.; Pierce, C.; 和 Maes, P. 2024年。PhysioLLM：利用可穿戴设备和大语言模型支持个体化健康洞察。载于2024年IEEE生物医学与健康信息学国际会议（BHI），1–8。Fu, J.; Liu, S.; Du, S.; Ruan, S.; Guo, X.; Pan, W.; Sharma, A.; 和 Konigorski, S. 2023年。多模态单人试验：一种新型个体化医疗设计。arXiv预印本 arXiv:2302.07547。版本v1，2023年2月15日。Fu, X.; Mo, S.; Buendia, A.; Laurent, A. P.; Shao, A.; del Mar Alvarez-Torres, M.; Yu, T.; Tan, J.; Su, J.; Sagatelian, R.; Ferrando, A. A.; Ciccia, A.; Lan, Y.; Owens, D. M.; Palomero, T.; Xing, E. P.; 和 Rabadan, R. 2025年。跨人类细胞类型的转录基础模型。《自然》，637：965–973。Gymrek, M.; McGuire, A. L.; Golan, D.; Halperin, E.; 和 Erlich, Y. 2013年。通过姓氏推断识别个人基因组。《科学》，339(6117)：321–324。Haugg, F.; Lee, G.; He, J.; N¨urnberg, L.; Bontempi, D.; Bitterman, D. S.; Catalano, P.; Prudente, V.; Glubokov, D.; Warrington, A.; Pai, S.; Ruysscher, D. D.; Guthier, C.; Kann, B. H.; Gladyshev, V. N.; Aerts, H. J.; 和 Mak, R. H. 2025年。利用面部照片进行健康识别的基础人工智能模型（FAHR-Face）。arXiv预印本 arXiv:2506.14909。

Hern´an, M. A.; 和 Robins, J. M. 2020年。《因果推断：假如》。博卡拉顿：Chapman & Hall/CRC。Holt, S.; Liu, T.; 和 van der Schaar, M. 2024年。自动学习动力系统的混合数字孪生。载于第38届神经信息处理系统大会论文集（NeurIPS 2024）。加拿大温哥华。获 NeurIPS 2024 Spotlight 论文。Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; ˇ Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Z´ıdek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; 和 Hassabis, D. 2021年。AlphaFold 实现高精度蛋白质结构预测。《自然》，596：583–589。Khasentino, J.; Belyaeva, A.; Liu, X.; Yang, Z.; Furlotte, N. A.; Lee, C.; Schenck, E.; Patel, Y.; Cui, J.; Schneider, L. D.; Bryant, R.; Gomes, R. G.; Jiang, A.; Lee, R.; Liu, Y.; Perez, J.; Rogers, J. K.; Speed, C.; Tailor, S.; Walker, M.; Yu, J.; Althoff, T.; Heneghan, C.; Hernandez, J.; Malhotra, M.; Stern, L.; Matias, Y.; Corrado, G. S.; Patel, S.; Shetty, S.; Zhan, J.; Prabhakara, S.; McDuff, D.; 和 McLean, C. Y. 2025年。用于睡眠与健身指导的个人健康大语言模型。《自然·医学》，31：3394–3403。

Kim, C.; Gadgil, S. U.; DeGrave, A. J.; Omiye, J. A.; Cai, Z. R.; Daneshjou, R.; 和 Lee, S.-I. 2024年。基于医学文献的图像—文本基础模型实现透明的医学图像人工智能。《自然·医学》，30：1154–1165。

Kimmelman, J.; 和 Tannock, I. 2018年。精准医学的悖论。Nature Reviews. Clinical Oncology，15(6)：341–342。

Kocaballi, A. B.; Berkovsky, S.; Quiroz, J. C.; Laranjo, L.; Tong, H. L.; Rezazadegan, D.; Briatore, A.; 和 Coiera, E. 2019年。医疗保健领域对话代理的个性化：系统综述。《医学互联网研究杂志》，21(11)：e15360。

Kondepudi, A.; Pekmezci, M.; Hou, X.; Scotford, K.; Jiang, C.; Rao, A.; Harake, E. S.; Chowdury, A.; Al-Holou, W.; Wang, L.; Pandey, A.; Lowenstein, P. R.; Castro, M. G.; Ko- erner, L. I.; Roetzer-Pejrimovsky, T.; Widhalm, G.; Camelo- Piragua, S.; Movahed-Ezazi, M.; Orringer, D. A.; Lee, H.; Freudiger, C.; Berger, M.; Hervey-Jumper, S.; and Hollon, T. 2025年。用于快速无标记检测胶质瘤浸润的基础模型。《自然》，637：439–445。

Konigorski, S.; Ried-Larsen, M.; and Schmid, C. H. 2024年。数字化单人试验及其在实验生理学中的应用。arXiv预印本 arXiv:2412.15076。

Konigorski, S.; Wernicke, S.; Slosarek, T.; Zenner, A. M.; Strelow, N.; Ruether, D. F.; Henschel, F.; Manaswini, M.; Pottb¨acker, F.; Edelman, J. A.; Owoyele, B.; Danieletto, M.; Golden, E.; Zweig, M.; Nadkarni, G. N.; and B¨ottinger, E. 2022年。StudyU：一个用于设计和开展创新型数字化单人试验的平台。Journal of Medical Internet Research，24(7)：e35884。

Kravitz, R. L.; Duan, N.; and Braslow, J. 2004。循证医学、治疗效果异质性以及平均值带来的问题。《米尔班克季刊》，82(4)：661–687。Li, X.; Liang, S.; Lei, Y.; Li, C.; Hou, Y.; Zheng, D.; and Ma, T. 2024年。CausalMed：以患者健康状态为核心的基于因果关系的个性化用药建议。载于第33届ACM信息与知识管理国际会议论文集，CIKM '24，1276–1285。美国纽约：计算机协会。ISBN 9798400704369。Li, X.; Peng, L.; Wang, Y.-P.; and Zhang, W. 2025年。迈向生物医学医疗保健的联邦基础模型：开放挑战与机遇。BioData Mining，18(1)：2。Lipton, R. B.; Bigal, M. E.; Diamond, M.; Freitag, F.; Reed, M. L.; Stewart, W. F.; and AMPP顾问组。2007。偏头痛患病率、疾病负担以及预防性治疗的必要性。《神经病学》，68(5)：343–349。Mahesh, N.; Devishamani, C. S.; Raghu, K.; Mahalingam, M.; Bysani, P.; Chakravarthy, A. V.; and Raman, R. 2024年。推进医疗保健：人工智能和基础模型的作用与影响。《美国转化研究杂志》，16(6)：2166–2179。Meier, D.; Ensari, I.; and Konigorski, S. 2023年。为单一个体试验中的体育锻炼建议设计并评估在线强化学习智能体。arXiv预印本。Moor, M.; Banerjee, O.; Abad, Z. S. H.; Krumholz, H. M.; Leskovec, J.; Topol, E. J.; and Rajpurkar, P. 2023年。通用型医疗人工智能的基础模型。《自然》，616(7956)：259–265。Morrow, E.; Zidaru, T.; Ross, F.; Mason, C.; Patel, K. D.; Ream, M.; and Stockley, R. 2023年。医疗保健中的人工智能技术与同情心：系统性范围综述。《心理学前沿》，13。Nasr, M.; Shokri, R.; and Houmansadr, A. 2019年。深度学习的全面隐私分析：针对集中式和联邦学习的被动与主动白盒推理攻击。载于2019年IEEE安全与隐私研讨会（SP），739–753。IEEE。Nikles, J.; and Mitchell, G., 编。2015。《健康领域单病例试验实用指南》。施普林格。ISBN 978-94-017-7199-3。Obermeyer, Z.; Powers, B.; Vogeli, C.; and Mullainathan, S. 2019年。剖析用于管理人群健康的算法中的种族偏见。《科学》（纽约），366(6464)：447–453。Owoyele, B. A.; Schilling, M.; Sawahn, R.; Kaemer, N.; Zherebenkov, P.; Verma, B.; Pouw, W.; and Melo, G. d. 2024年。MaskAnyone工具包：提供在视听数据归档中最大限度降低隐私风险、最大限度提高效用的策略。arXiv:2408.03185 [计算机科学]。Piccininni, M.; Stensrud, M. J.; Shahn, Z.; and Konigorski, S. 2024年。单一个体试验的因果推断。arXiv预印本 arXiv:2406.10360。Pillai, A.; Spathis, D.; Nepal, S.; Collins, A. C.; Mackin, D. M.; Heinz, M. V.; Griffin, T. Z.; Jacobson, N. C.; and Campbell, A. 2025年。《超越提示：Time2Lang——连接时间序列基础模型与大语言模型以实现健康感知》。载于第六届健康、推断与学习会议论文集，机器学习研究论文集第287卷，268–288页。PMLR。Pot, M.; Kieusseyan, N.; and Prainsack, B. 2021年。并非所有偏见都有害：机器学习与放射学中的公平与不公平偏见。《影像洞察》，12(1)：13。Price, W. N. I.; and Cohen, I. G. 2019年。医疗大数据时代的隐私。《自然·医学》，25(1)：37–43。Qian, Z.; Zhang, Y.; Bica, I.; Wood, A.; and van der Schaar, M. 2021年。SyncTwin：利用纵向结局估计治疗效应。载于《神经信息处理系统进展》第34卷，3178–3190页。加拿大温哥华。Sad´ee, C.; Testa, S.; Barba, T.; Hartmann, K.; Schuessler, M.; Thieme, A.; Church, G. M.; Okoye, I.; Hernandez-Boussard, T.; Hood, L.; Shmulevich, I.; Kuhl, E.; and Gevaert, O. 2025年。医疗数字孪生：助力精准医学与医疗人工智能。《柳叶刀·数字健康》，7(7)：e100864。Schneider, J.; G¨artner, T.; and Konigorski, S. 2023年。结合无监督学习与统计推断开展多模态单人试验。arXiv预印本，arXiv:2309.06455。Senarathne, J.; Overstall, A. M.; and McGree, J. 2020年。用于估计群体和个体治疗效应的贝叶斯自适应单病例试验。《医学统计》，39(29)：4420–4440。Shrestha, S.; and Jain, S. 2021年。用于单病例临床试验的贝叶斯多臂老虎机自适应设计。《医学统计》，40(7)：1825–1844。Steinberg, E.; Fries, J. A.; Xu, Y.; and Shah, N. H. 2024年。MOTOR：结构化医疗记录的事件时间基础模型。载于国际学习表征会议（ICLR）论文集。OpenReview。Subbaswamy, A.; Schulam, P.; and Saria, S. 2019年。防止数据集偏移导致的失效：学习可迁移的预测模型。arXiv:1812.04597。Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Go mez, A. N.; Łukasz Kaiser; and Polosukhin, I. 2017年。《注意力就是一切》。CoRR，abs/1706.03762。Xiang, J.; Wang, X.; Zhang, X.; Xi, Y.; Eweje, F.; Chen, Y.; Li, Y.; Bergstrom, C.; Gopaulchan, M.; Kim, T.; Yu, K.-H.; Willens, S.; Olguin, F. M.; Nirschl, J. J.; Neal, J.; Diehn, M.; Yang, S.; and Li, R. 2025年。用于精准肿瘤学的视觉语言基础模型。《自然》，638：769–778。Xu, H.; Usuyama, N.; Bagga, J.; et al. 2024年。基于真实世界数据构建的数字病理全切片基础模型。《自然》，630：181–188。Zhao, Z.; Fan, C.; Gao, C.; Feng, F.; and He, X. 2025年。无需过度处方的通用药物推荐：大语言模型的细粒度对齐。ArXiv:2503.03687。