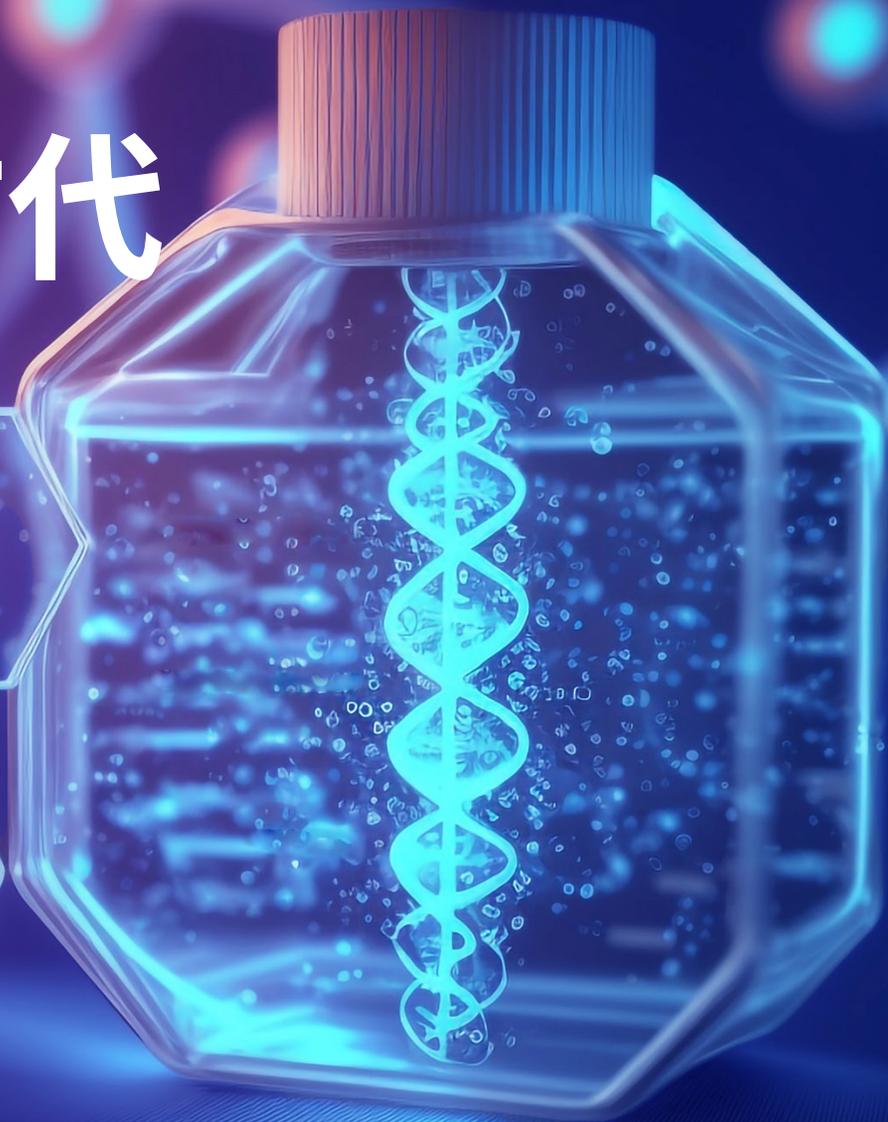


精准医疗新时代

生成式人工智能将产生颠覆性影响



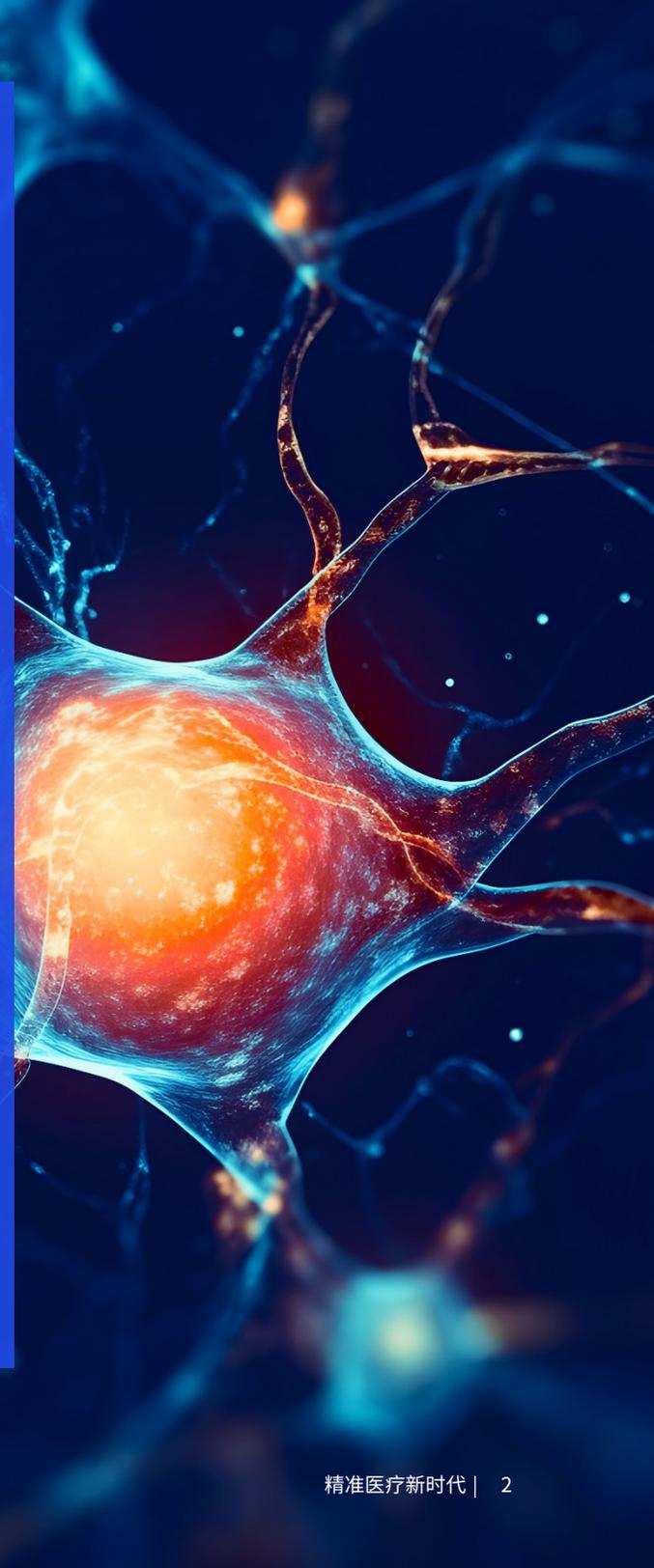
前言

精准医疗代表着医疗健康和生命科学领域业务模式的范式转变，旨在为个体患者提供量身定制的防治策略。在这一重要临床学科不断发展的过程中，人工智能，尤其是生成式人工智能，很可能将成为其创新基石和推进动力。

当前，人工智能在患者风险评估、筛查和诊断中的应用正取得前所未有的进展。然而，我们认为，人工智能驱动的治疗决策潜藏着更多机遇。医疗健康服务提供商和辅助技术专家（如数据科学家、机器学习工程师等）可以使用人工智能分析复杂的数据特征，确定最佳的患者治疗方式，预测治疗反应，并提供个性化的医疗健康服务体验。

在运营方面，人工智能对于增强基于研究和临床的医疗健康服务运营潜力巨大。但也面临着一些有待解决的挑战，包括保护数据隐私、管理道德影响、获得监管批准以及保障基础设施投资等。此外，还需通过扎实的验证研究，建立医疗健康专家和患者对人工智能的可靠性和透明度的信任。

本文深入探讨了目前人工智能对精准医疗领域的作用。针对有碍在更大范围内应用人工智能的问题，通过例证案例探究了潜在的以新兴学习模型为基础，以人工智能为驱动的解决方案。我们认为，在构建强大、可扩展和能有效保护隐私的精准医疗生态系统中，联邦学习将发挥尤其关键的作用。



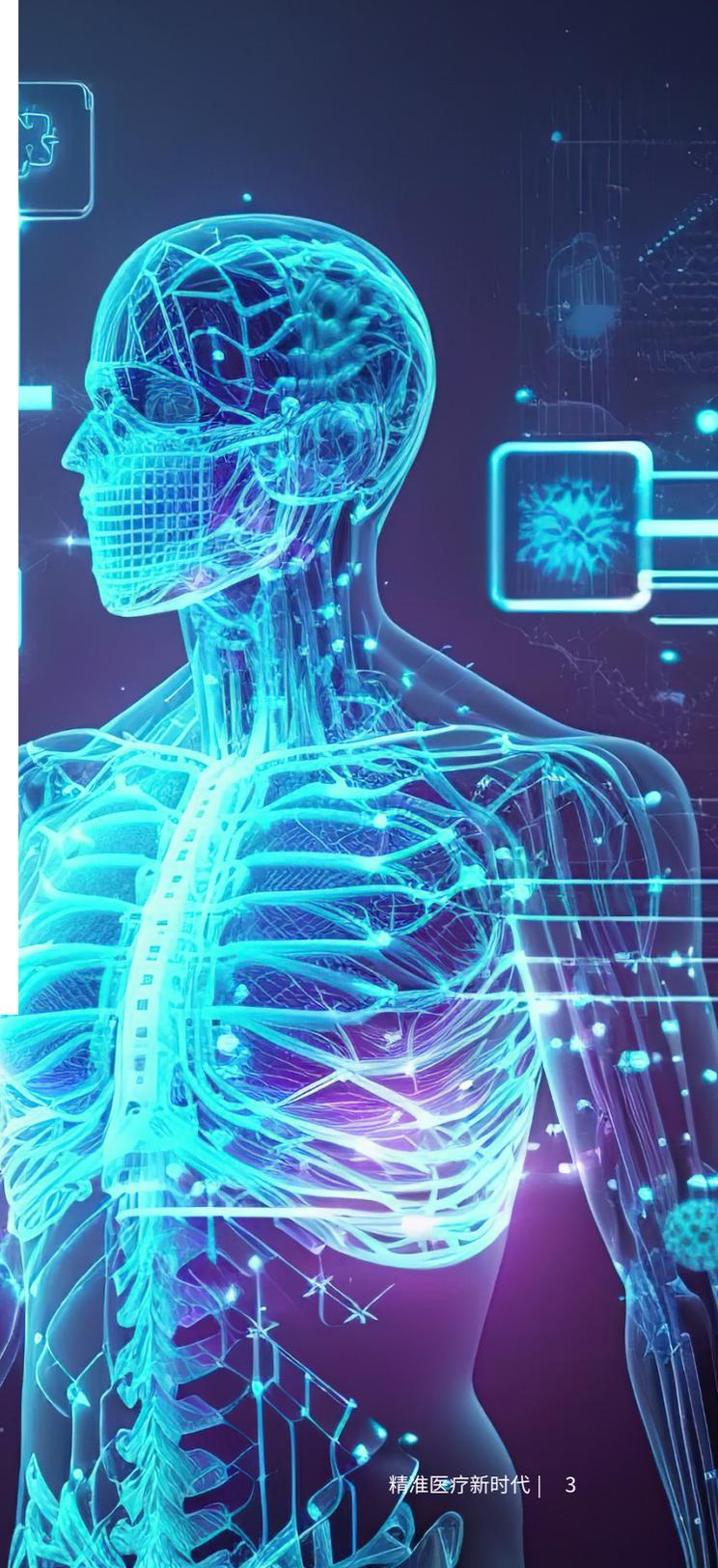
当前人工智能在精准医疗发展过程中的应用情况

生物标志物的识别和理解对精准医疗至关重要，因为它们是生物过程、疾病状态和治疗干预反应的可测量指标。其重要性在精准医疗的方方面面都得以体现，能够左右与疾病诊断、预后、患者分层、治疗选择、药物开发、治疗监测和疾病预防相关的决策。

人工智能（包括机器学习和深度学习/神经网络）¹ 显著增强了生物制药行业处理和分析大量复杂的多组学数据

（例如基因组学、蛋白质组学和代谢组学）的能力。² 这些技术有助于阐明疾病发展变化背后的分子途径、遗传变异和生物过程。

这种洞察能为精准医疗的各个方面提供支持，从识别新的候选生物标志物³，到基于个体独特分子特征制定个性化治疗计划等（参见第10页对ArteraAI的介绍）。



¹资料来源：Stefano A. Bini医学博士，“人工智能、机器学习、深度学习和认知计算：这些意味着什么？它们将如何影响医疗健康领域？”（Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care?）《关节成形外科杂志》（The Journal of Arthroplasty）第8期第33卷，AAHKS研讨会，ScienceDirect, 2018年7月19日

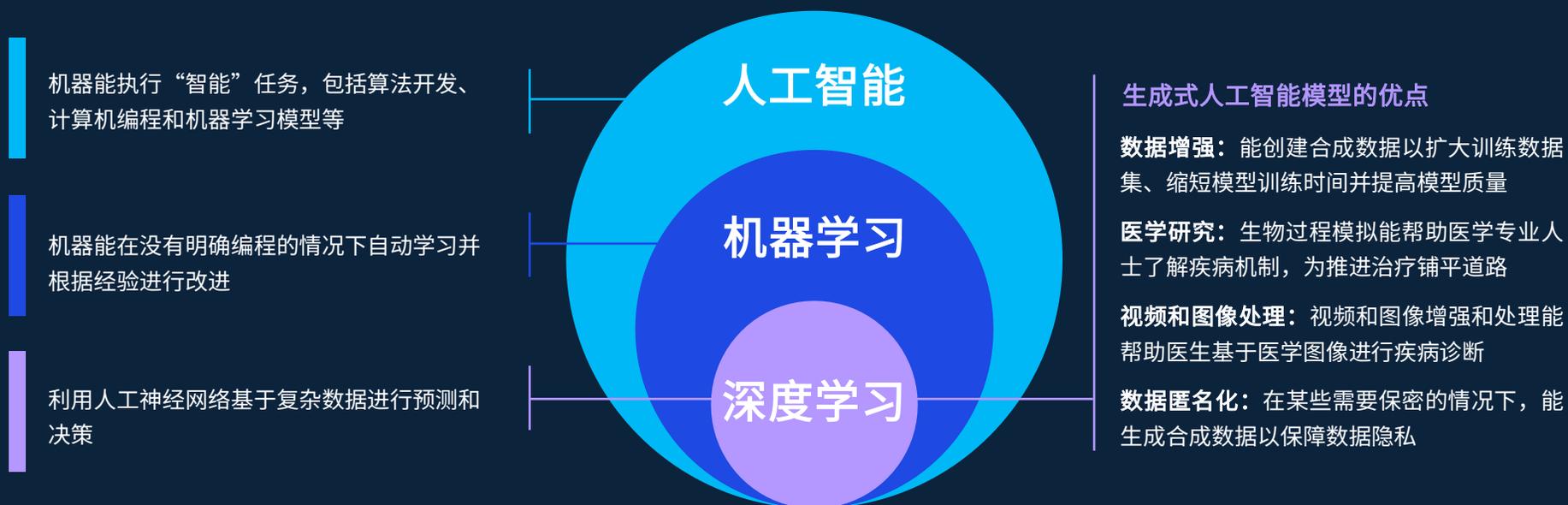
²资料来源：Matthias Mann、Chanchal Kumar、Wen-Feng Zeng和Maximilian T. Strauss，“用于识别蛋白质组学和生物标志物的人工智能”（Artificial intelligence for proteomics and biomarker discovery），《细胞系统》第8期第12卷观点文章，ScienceDirect, 2021年8月18日

³同上

深入探究不同类型的人工智能模型

目前，将人工智能成果和模型用于解决精准医疗领域挑战已经取得了稳步的进展，包括较常见的基于监督学习的机器学习模型（其中各数据点都有相关标签），⁴ 以及最近的生成式模型，如生成式对抗性网络和变分自动编码器⁵（图1）。这些技术具有独特的能力，能够在数据缺失的情况下使用，并解析复杂数据，以助力生物标志物识别、患者分层和药物再利用等领域。

图1：生成式模型的优点



⁴ 资料来源：Stefano A. Bini医学博士，“人工智能、机器学习、深度学习和认知计算：这些意味着什么？它们将如何影响医疗健康领域？”（Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care?）《关节成形外科杂志》（The Journal of Arthroplasty）第8期第33卷，AAHKS研讨会，ScienceDirect, 2018年7月19日

⁵ 资料来源：Bilal Ahmad、Jun Sun、Qi You、Vasile Palade和Zhongjie Mao，“综合利用变分自动编码器和生成式对抗性网络进行脑肿瘤分类”，《生物医学》第8期第12卷，MDPI, 2022年1月21日

人工智能在患者精准医疗中的应用

人工智能技术的发展提高了精准医疗领域的效率，使医疗专业人士得以进一步利用先进的预测性建模和决策辅助工具协助制定个性化治疗策略。精准医疗过程可分解为一系列步骤，这些步骤对应于患者旅程中的关键节点，各节点均可能通过人工智能加以优化（图2）。

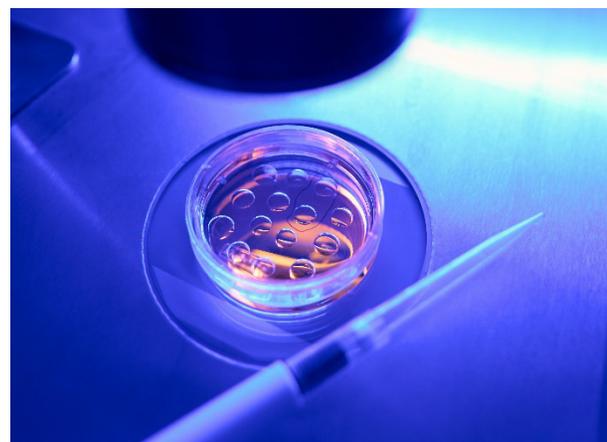


图2：精准医疗中患者旅程的各个阶段都可能可以利用人工智能进行优化⁶

	风险评估	筛查	诊断	分期与预后	治疗选择	监测
阶段描述	根据个体遗传和其他生物标志物数据、临床发现和环境影响因素评估患者风险	根据预定节奏进行测试，以助早期疾病识别	通过个体生物标志物和其他独特数据提高疾病诊断的准确性	通过个体预后生物标志物评估疾病进展、严重程度、态势和复发风险	使用多组学数据以及病史、社会因素和环境动态定制治疗方案	通过个体生物标志物数据监测治疗的安全性、副作用和反应
人工智能影响举例	通过基因型数据预测表型表达，并通过图像分析评估疾病风险	通过基于图像分析和临床数据的神经网络建模，提供量身定制的筛查方案和频率建议	通过快速全基因组测序和自然语言处理表型，加快新生儿基因变异分析和致病变异鉴定	通过血液检查、成像和电子健康记录数据分析，更快、更准确地预测新冠的预后和严重程度	使用活检图像、生物标志物测试和临床数据等多模式分析更好地预测治疗反应	使用基于多变量、单核苷酸多态性的模型预测化疗毒性风险

此处为部分影响举例。

⁶ 资料来源：毕马威内部分析

阶段1：风险评估

传统上，精准医疗的风险评估侧重于利用易感性生物标志物测试、基因组数据、内部扫描和其他数据来辅助传统临床和危险因素评估。由于生物学数据和基因型-表型关联的复杂性，制定强有力的风险评估方案一向颇具挑战。

人工智能可通过有效解释大量遗传信息和预测基因表达来帮助应对这一挑战。对基因组变异及其与疾病表现、成功治疗和预后之间的联系的了解，有助于更好地评估患者风险。请参考右边的第一个用例。

人工智能还被用于加强风险评估，而非仅用于解释基因组学或其他生物标志物。请参考第二个用例，以了解如何使用图像数据来训练风险评估模型。

风险评估用例：

利用统计基因组学和机器学习预测乳腺癌和卵巢癌风险

生物医学研究所和基因组调控中心之间通过合作确定了42种遗传基因，这些基因会使个体容易产生更多突变。这些突变会导致较高的癌症发生率，特别是乳腺癌和卵巢癌。研究人员使用统计基因组学和被称为“自动编码器”神经网络的机器学习模型，在复杂数据（特别是来自欧洲癌症患者的11,000个基因组序列）中挖掘特征，将某些基因与表明癌症风险增加的特定体细胞突变联系起来。

资料来源：Fran Supek博士和Nahia Barberia，“新研究中详述的增加癌症突变可能性的遗传因素”（Hereditary factors that increase the likelihood of cancer mutations detailed in new study），科学新闻，生物医学研究所（IRB），巴塞罗那，2022年7月5日

利用人工智能和影像学预测肺癌风险

哈佛医学院的研究人员和麻省理工学院的研究人员在马萨诸塞州总医院开展了一项合作，研究如何建立一个深度学习模型，以便在无需额外提供人口统计学或临床数据的情况下，评估肺部扫描影像并预测个人风险。哈佛大学/麻省理工学院的团队使用三组低剂量计算机断层扫描（LDCT扫描）数据、来自“美国国家肺癌筛查试验”参与者的6,282个LDCT扫描数据、来自马萨诸塞州总医院的8,821个LDCT扫描数据以及来自长庚纪念医院的12,280个LDCT扫描数据（包括一系列吸烟者和非吸烟者），训练了一个3D卷积神经网络架构。在几家大型医疗健康企业和投资者的资助下，他们的模型Sybil已被证明可以通过单次低剂量计算机断层扫描准确预测吸烟者和非吸烟者未来的肺癌风险。

资料来源：Peter G. Mikhael、Jeremy Wohlwend、Adam Yala、Ludvig Karstens、Justin Xiang、Angelo K. Takigami、Patrick P. Bourgouin、PuiYee Chan、Sofiane Mrah、Wael Amayri、Yu-Hsiang Juan、Cheng-Ta Yang、Yung-Liang Wan、Gigin Lin、Lecia V. Sequist、Florian J. Fintelmann和Regina Barzilay，“Sybil：经验证可通过单次低剂量胸部计算机断层扫描预测未来肺癌风险的深度学习模型”（Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography），《临床肿瘤杂志》第12期第41卷，期刊列表，2023年1月12日

癌症诊断焦点

监督机器学习和深度学习算法可通过以下方式帮助评估遗传性癌症风险：

- 分析大量基因数据并识别高危基因
- 根据基因图谱对患者进行分层
- 通过人工智能决策支持工具辅助临床决策

阶段2：筛查

传统精准医疗筛查包括基于预定的时间间隔对高危患者进行检测，以进行早期疾病识别。

当前，人工智能正通过各种方式辅助筛查，特别是通过提高医学成像的准确性和效率。人工智能算法，尤其是卷积神经网络等深度学习技术，在分析医学图像（如乳房X光检查、计算机断层扫描（CT）和磁共振成像（MRI）扫描）方面显示出光明的前景，可有助于检测癌症的早期迹象（如肿瘤、异常组织生长等）。

例如，麻省理工学院和马萨诸塞州总医院联合开发了一个名为“Mirai”的深度学习模型，该模型可以使用乳房X光检查数据提前五年预测潜在的乳腺癌患者。Mirai接受了马萨诸塞州总医院超过200,000次的测试训练，并基于马萨诸塞州总医院、卡罗林斯卡研究所和长庚纪念医院的测试集进行了验证。该模型在预测癌症风险⁸、识别高危人群和对患者进行分层以进一步筛查等方面优于传统方法⁷。与Tyrer-Cuzick模型⁹相比，Mirai在高危人群中识别出的潜在癌症患者几乎达前者的两倍。¹⁰值得注意的是，该模型在不同年龄

组、乳腺密度类别、癌症亚型和种族等方面表现出了一致的准确性。鉴于有色人种女性的乳腺癌发病率较高（见下文方框），Mirai有效证明了人工智能能够克服传统筛查模型中固有的一些偏见。



解决精准医疗中的偏见问题

与白人女性相比，黑人女性乳腺癌死亡率高出40%。¹¹ 用例表明，Mirai等深度学习模型大致解决了基于人工智能的精准医疗和医药中存在的健康公平和模型偏见等重要问题。这表明它实现了良好的适用性，并对白人和黑人女性都表现出了同样的准确性。

⁷ 资料来源：传统癌症筛查方法包括盖尔模型、乳腺癌风险预测工具（BCRAT）、Tyrer-Cuzick风险评估计算器等

⁸ 资料来源：Fran Supek博士和Nahia Barberia，“新研究中详述的增加癌症突变可能性的遗传因素”，科学新闻，生物医学研究所（IRB），巴塞罗那，2022年7月5日

⁹ 资料来源：Tyrer-Cuzick是一种风险评估模型，通过提出个人和家族病史问题来确定患乳腺癌的可能性。其结果包含10年风险评分和终身风险评分。Tyrer-Cuzick风险评估计算器，magview.com，MagView

¹⁰ 资料来源：Rachel Gordon，“未来癌症预测的强大人工智能工具”（Robust artificial intelligence tools to predict future cancer），麻省理工学院计算机科学与人工智能实验室（MIT CSAIL），麻省理工学院新闻，2021年1月28日

¹¹ 资料来源：“黑人女性乳腺癌死亡率仍属最高”（Breast Cancer Death Rates Are Highest for Black Women—Again），美国癌症协会，2022年10月3日

阶段3：诊断

诊断测试用于疾病的最终诊断，并为精准医疗中患者旅程的下一步（即疾病管理）提供支持信息。

当前，人工智能正通过各种方式辅助诊断，特别是通过提高诊断测试的准确性、效率和客观性。例如，人工智能能够从电子健康档案数据、临床笔记、实验室结果和成像数据中提取相关信息并识别特征，从而帮助临床医生做出更准确和及时的诊断。¹² 此外，整合各种数据源（如筛查章节所述）可有助于更全面地了解患者的病情。¹³ 除了帮助更准确进行诊断外，这

种综合方法还可为制定个性化治疗策略提供辅助信息。¹⁴

值得注意的是，临床上对个体表型遗传变异的理解正在成为影响罕见遗传病基因组诊断成本和时间的最大因素。¹⁵ 人工智能可整合预测方法并更好地识别遗传病及其原因，从而显著加快和简化基因组解释。

此外，人工智能驱动的诊断在新生儿护理中也具有特别的价值。全球每年有超过800万婴儿出生时患有能危及生命的遗传病¹⁶，早期诊断对他们的存活至关重要。通过使用自然语言处理进

行自动表型分型，并采用全基因组测序，人工智能可以在新生儿紧急护理时快速提供关键诊断信息。请参考下面的用例。

最后，自然语言处理程序能自动提取患者的电子健康档案数据，通过病历文件将其表型与潜在病因进行匹配。自然语言处理集成已被证明确实有效，即使是在特定程序缺乏足够训练数据以有效减少错误的情况下（即自然语言处理集成可通过多数规则确定输出）。

诊断用例：

对婴儿遗传性疾病进行更快速的诊断和干预

为加快基因组解释，犹他大学医院、Fabric Genomics和雷迪儿童医院共同开发了名为“Fabric GEM”的基于人工智能的算法，用于新生儿遗传病诊断。Fabric GEM将诊断准确性提高到一个新的水平，诊断时90%以上会将致病变异列为首要或次要因素，与现有工具相比是一大进步。该工具可减轻基因变异分析工作，从而提高婴儿诊断速度和准确性。

资料来源：快速结构变异识别 | 利用人工智能加速疾病诊断 (fabricgenomics.com)；人工智能可以快速识别新生儿疾病的遗传诱因 —— 犹他大学医院

¹² 资料来源：“人工智能在眼科电子健康档案数据中的应用” (Applications of Artificial Intelligence to Electronic Health Record Data in Ophthalmology)，《转化视觉科学与技术》，2020年2月27日。

¹³ 资料来源：“减少人工智能工具在个性化医疗中的偏见” (Limiting Bias in Artificial Intelligence Tools, Personalized Medicine)，HealthITAnalytics，2021年12月9日

¹⁴ 资料来源：“精准医疗、人工智能以及个性化医疗未来发展” (Precision Medicine, AI, and the Future of Personalized Health Care)，《临床与转化科学》，2021年1月

¹⁵ 资料来源：Francisco M De La Vega、Shimul Chowdhury、Barry Moore、Erwin Frise、Jeanette McCarthy、Edgar Javier Hernandez、Terence Wong、Kiely James、Lucia Guidugli、Pankaj B Agrawal、Casie A Genetti、Catherine A Brownstein、Alan H Beggs、Britt-Sabina Löscher、Andre Franke、Braden Boone、Shawn E Levy、Katrin Öunap、Sander Pajusalu、Matt Huentelman、Keri Ramsey、Marcus Naymik、Vinodh Narayanan、Narayanan Veeraraghavan、Paul Billings、Martin G Reese、Mark Yandell和Stephen F Kingsmore，“人工智能可对罕见遗传病进行全面基因组解释并提出候选诊断方案” (Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases)，PMCID: PMC8515723，PubMed Central生物医学和生命科学数据库，美国国立卫生研究院 (NIH)，美国国家医学图书馆 (NLM)，2021年10月14日

¹⁶ 资料来源：“2023年世界出生缺陷日：通过全球努力提高公众意识并提供家庭支持” (World Birth Defects Day 2023: Global Efforts to Raise Awareness and Support Families)，cdc.gov，2023年2月27日

阶段4：分期与预后

在精准医疗中，分期和预后包括使用个体预后生物标志物来更好地评估疾病进展、严重程度、态势和复发风险。

目前，机器学习和深度学习可通过分析预后生物标志物、疾病影像和其他疾病数据来增强此类流程。例如，机器学习/深度学习算法擅长分析基因表达、蛋白质水平和其他分子数据，以识别致病特征。

针对病理学研究，切片影像可被转换为数字数据，并被卷积神经网络“汇集”到最相关的卷积层。然后，此类卷积层将形成一个“扁平”的数据集，通过传统的人工神经网络流程来进行个性化疾病特征评估（图3）。

分期和预后用例：

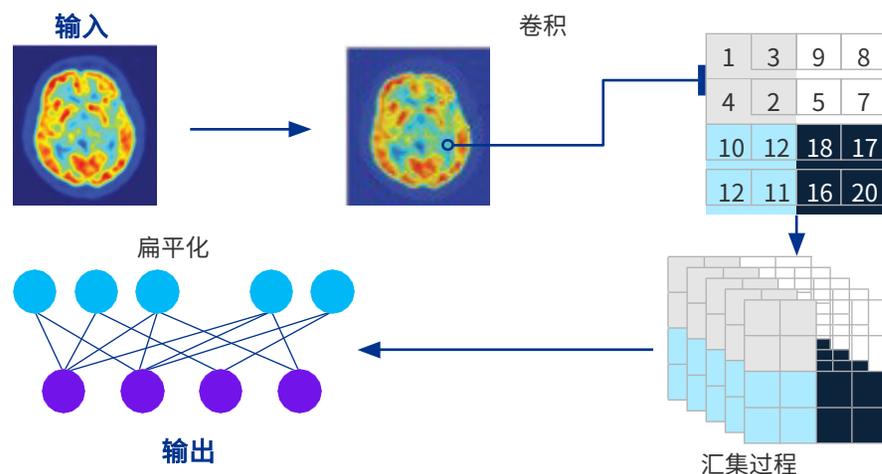
首个人工智能驱动的糖尿病肾病预后测试平台

Renalytix是肾脏健康生物诊断™新领域全球领导者，其人工智能驱动的预后测试平台KidneyIntelX™已获得美国食品药品监督管理局的新型医疗器械（De Novo）上市授权。该平台基于西奈山伊坎医学院开发并许可Renalytix使用的技术。KidneyIntelX能帮助医生详细了解慢性糖尿病肾病早期患者五年内肾功能持续衰竭的速度。它将患者分为三个风险等级：低、中、高，并使用人工智能算法，针对基于血液的生物标志物和临床变量的组合生成结论。该算法能提供可靠和可操作的信息，以指导大规模高危患者群体的护理。

自推出以来，该平台已在美国约10,000名患者身上使用。鉴于超过3,000万美国人患有2型糖尿病¹⁷，且肾脏疾病对全球超过8.5亿人构成影响，因此其应用前景潜力巨大。¹⁸

资料来源：“FDA给予KidneyIntelX.dkd新型医疗器械上市授权，以帮助评估糖尿病和早期肾病成年患者的进行性肾功能衰竭风险”（FDA Grants De Novo Marketing Authorization for KidneyIntelX.dkd to Assess Risk of Progressive Kidney Function Decline in Adults with Diabetes and Early-Stage Kidney Disease），新闻稿，Renalytix，2023年6月29日

图3：卷积神经网络将数据进行分层以评估疾病特征



资料来源：Milecia McGregor，“什么是卷积神经网络？机器学习和深度学习入门教程”（What Is a Convolutional Neural Network? A Beginner’s Tutorial for Machine Learning and Deep Learning），#Machine Learning，freeCodeCamp.org，2021年2月4日

¹⁷ “2型糖尿病”（Type 2 Diabetes），美国疾病控制与预防中心，www.cdc.gov

¹⁸ “潜在的流行病：全球超8.5亿人患有肾脏疾病”（The hidden epidemic: Worldwide, over 850 million people suffer from kidney diseases），美国肾脏病学会、欧洲先进肾脏病学会（Leading European Nephrology）和国际肾脏病学会，2018年6月27日

阶段5：治疗选择

传统个性化治疗选择基于多组学数据，结合病史、社会因素和环境变化。

人工智能包括机器学习、深度学习和神经网络技术，有助于通过各种途径优化治疗选择，包括预测患者对特定治疗的反应、确定潜在的药物靶点和优化治疗方案。而应用时主要的考虑因素包括数据的质量和可用性，以及隐私和安全方面的道德和监管因素。

例如，ArteraAI前列腺测试是一种具有突破性的人工智能驱动的检测，旨在识别可能可以进行强化治疗的局部前列腺癌患者。该测试由包括Coatue、Johnson & Johnson Innovation、Koch Disruptive Technologies、Walden Catalyst Ventures、TIME Ventures和Breyer Capital在内的知名药企和医疗健康投资机构组成的联盟研发，采用了多模态人工智能架构，结合了临床和组织病理学影像数据。¹⁹

这种创新方法在多次大型III期临床试验中得到了验证，与美国国家综合癌症网络（NCCN）模型等传统风险模型相比，在预测生化复发、远处转移、前列腺癌特异性生存率和总体生存率等方面表现出了更优越的性能。²⁰

阶段6：监测

在精准医疗中，个体生物标志物数据被用于监测治疗安全性、副作用变化和疾病进展。

人工智能可以帮助医生监测治疗效果和安全性，对疾病进展进行有根据的判断，并预测副作用变化。关于后者，部分疾病如急性淋巴细胞白血病，其治疗并发症的发病率增加了发病后出现康复缓慢甚至早逝的可能性。因此，能够预测哪些患者会出现负面治疗反应和持续的副作用非常重要，护理团队可据此确保对这些患者进行密切监测。²¹

例如，佛罗里达大学的研究人员最近开发了一种基于人工智能的工具，能够预测急性淋巴细胞白血病患者产生化疗药物毒性的风险。²² 研究人员使用佛罗里达大学的患者数据对人工智能模型进行训练，使其能够预测哪些单核苷酸多态性性状位点和其他遗传变异的组合可能产生毒性，并最终得出个体患者的“毒性评分”。²³ 人工智能驱动的多变量分析用于高效得出大量单核苷酸多态性性状位点和基因变异的潜在组合，以帮助确定哪些组合可能增加化疗对患者的伤害，其结果在随后的治疗中得到了验证。²⁴

¹⁹ 资料来源：“Artera投资9,000万美元打造多模态人工智能以实现个性化癌症治疗”（Artera Launches with \$90 Million in Funding to Personalize Cancer Therapy with Multimodal AI），Artera新闻，《美国商业资讯》，2023年3月21日

²⁰ 资料来源：“人工智能驱动的生物标志物对男性高风险前列腺癌患者的预后预测结果优于NCCN风险组”（AI-Powered Biomarker Predicts Outcomes Better than NCCN Risk Groups For Men with High-Risk Prostate Cancer），美国临床肿瘤学会每日新闻，2023年2月16日

²¹ 资料来源：“儿童急性淋巴细胞白血病幸存者治疗的晚期反应”（Late Effects of Therapy in Childhood Acute Lymphoblastic Leukemia Survivors），《土耳其血液学杂志》——土耳其血液学学会官方杂志，2019年2月7日

²² 资料来源：Leah Buletti，“佛罗里达大学研究人员开创白血病药物并发症预测方法”（UF researchers create method to predict leukemia drug complications），佛罗里达大学药学院，2023年3月24日

²³ 资料来源：Trisha Larkin（医学博士）、Reema Kashif（医学博士）、Abdelrahman H. Elsayed（博士）、Beate Greer（学士）、Karna Mangrola（医学博士）、Roya Raffiee（博士）、Nam Nguyen（药学博士）、Vivek Shastri（博士）、Biljana Horn（医学博士）和Jatinder K. Lamba（博士），“多基因药物基因组标记物作为急性淋巴细胞白血病治疗中毒性表型的预测因子：单中心研究”（Polygenic Pharmacogenomic Markers as Predictors of Toxicity Phenotypes in the Treatment of Acute Lymphoblastic Leukemia: A Single-Center Study），《JCO® Precision Oncology》杂志第7卷，期刊列表，2023年3月23日

²⁴ 资料来源：Sophia C. Kamran和Kent W. Mouw，“在放射肿瘤学中应用精准肿瘤学原理”（Applying Precision Oncology Principles in Radiation Oncology），《JCO® Precision Oncology》杂志第2卷，期刊列表，2018年5月14日

在精准医疗中使用人工智能的关键考虑因素

随着人工智能在精准医疗中应用的不断深化，企业应对自学人工智能、生成式人工智能以及（可能最关键）联邦学习加以考虑。下文三个小节提供了相关的指引。

01 自学人工智能的考虑因素

协作开发工具：人工智能模型协同开发所需的基础设施（包括共享编码平台）是操作可能会与机器学习框架协同工作的内部人工智能模型的必要工具。²⁵

合作伙伴：精准医疗生态系统中的企业应准备好与包括医院、研究人员和生物制药企业在内的各种利益相关方建立合作伙伴关系，以获取必要的数据、扩大规模并辅助临床实施。²⁶

通过道德考量及尽最大努力减少偏见：必须对人工智能在精准医疗中的道德使用给予大力关注。这包括维护患者的隐私和知情同意权，以及努力确保培训数据集不会进一步导致偏见。另一要点在于必须建立透明的数据处理和分析方案，以尊重个人对数据自主性和机密性的要求，同时确保人工智能驱动的医疗干预措施公平可得。此外，数据必须用于预期目的，而不会落到可能出于自身利益使用该等数据的人之手。

法规变化：在遵守与患者数据共享和数据隐私相关的潜在合规、法律和监管要求方面，生物制药企业应提前做好准备。正如美国医保与医助服务中心要求付款人必须使用安全的、基于标准的应用程序编程接口，使患者得以访问其索赔和遭遇数据一样，医疗机构很快也必须遵循同样的指引要求。²⁷ 此外，随着精准医疗的日益普及且相关数据变得越发复杂，对违规供应商进行公告正成为常态。²⁸

构建生态系统所需的强大数据共享协议：精准医疗中的人工智能必须确保以安全、符合隐私规定的方式处理各类数据，包括遗传、表型和生活方式数据。此外，助力机构间实现数据共享安全和合规的政策的重要性也与日俱增。

²⁵ 资料来源：例如，TensorFlow正利用Keras应用程序编程接口构建神经网络，此类工作通常由机器学习工程师和数据科学家推动。而该接口不涉及更细粒度的神经网络细节，因此不具备高级数据科学知识的工作团队成员也能够理解相关的流程和模型。

²⁶ 资料来源：Onconova Therapeutics与聚焦肿瘤领域的机器学习公司Pangea Therapeutics建立了合作研究关系。Onconova将使用Pangea的专有ENLIGHT AI平台来识别会对力格赛狄（rigosertib）产生反应的生物标志物，rigosertib是Onconova用于治疗各种实体瘤的小分子药物之一。这一合作关系将有助于加快试验，开发适当的配套诊断措施，并最终在大规模患者群体应用中取得更大的商业成功。

²⁷ 资料来源：最新的FHIR R5标准提升了数据交换和互操作性”（Latest FHIR Standard R5 Elevates Data Exchange, Interoperability），ehrintelligence.com，2023年4月18日

²⁸ 资料来源：公告对临床结果的影响：系统综述和元分析”（The impact of Public Reporting on clinical outcomes: a systematic review and meta-analysis），《BMC健康服务研究》，2016年7月22日

02 生成式人工智能的考虑因素

基础设施和算力：人工智能，尤其是生成式人工智能，需要极为强大的算力。生物制药企业可能需要在高性能计算以及数据存储和开发工具等领域进行能力投资。高性能计算专为大规模数据处理和建模而设，需要根据模型开发时间表进行大量硬件投资。

大幅增强数据存储能力：网络连接存储（NAS）和存储区域网络（SAN）等存储系统通常用于满足中小型存储需求。相比之下，基于云的基础设施能提供生成式人工智能解决方案所需的大规模存储能力。

模型验证：在临床环境下验证生成式人工智能模型至关重要，因为要使这些技术民主化，研究人员必须使用合成数据。例如，如使用生成式人工智能来扩大临床试验对照组并进行虚拟试验，则需加强所用数据的质量保证。

人力因素：设计和部署用于精准医疗的独特生成式人工智能系统，需要一支由数据科学家、机器学习工程师、软件开发人员、用户体验设计师和专业项目经理组成的团队。面对当前科技行业人才缺口，生物制药企业应积极主动，尽早启动招聘活动，还必须适应当代就业模式，在适用的情况下考虑虚拟组织结构，以吸引该领域的顶尖人才。

03 聚焦联邦学习：数据隐私保障的关键考虑因素 29、30、31、32、33

为成功建立精准医疗生态系统，必须具备一系列关键条件，如数据隐私、互操作性和对不同数据集的有效利用，联邦学习可能是达成此类条件的适当方式。联邦学习是一种机器学习方法，即在多个去中心化节点上训练全局模型，每个节点都包含自己的本地数据（图4）。

在生成式人工智能领域，联邦学习有助于在本地生成和优化新版模型，从而确保数据隐私并尽量减少数据传输需求。此举能增强全局模型，提升其在大规模数据源基础上产生创新和多样化输出的能力。尽管联邦学习在精准医疗领域颇具潜力，但重要的是要解决实施过程中面临的挑战，包括系统架构可变性以及协议获取和打标方法标准化的需求。

²⁹ 资料来源：Mohammed Aledhari、Rehman Razzak、Reza M. Parizi和Fahad Saeed，“联邦学习：关于底层技术、协议和应用的调查”（Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications），PMCID: PMC7523633，HHS作者手稿，期刊列表，PubMed Central(生物医学和生命科学数据库，美国国家医学图书馆（NLM），2020年9月29日

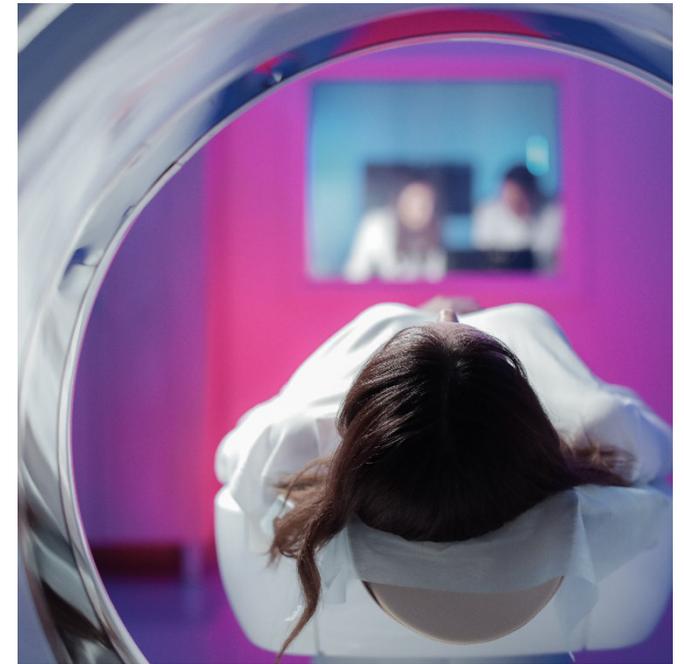
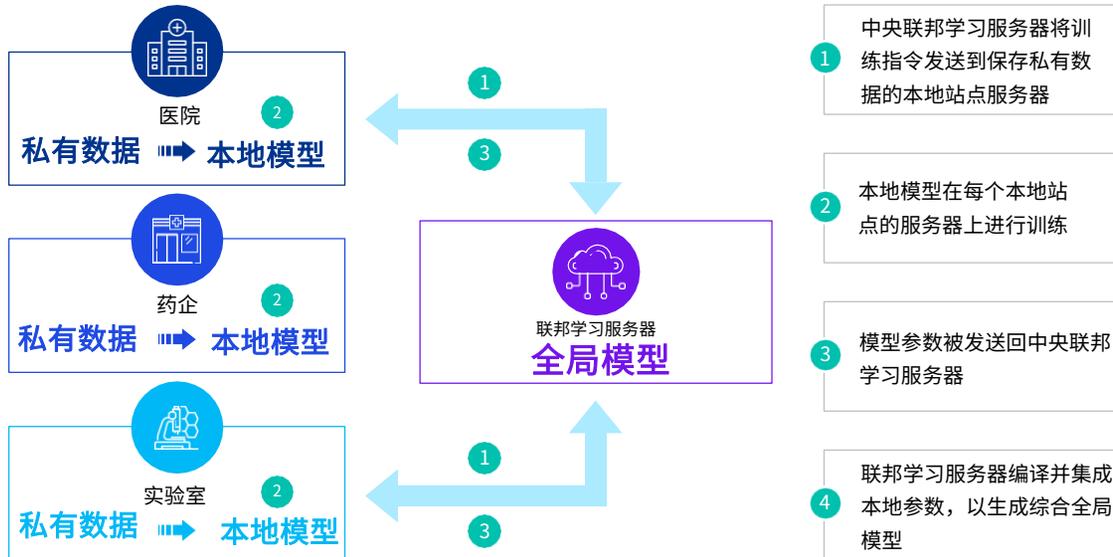
³⁰ 资料来源：“挖掘用于机器学习的分布式健康数据”（Unlocking Distributed Health Data for Machine Learning），《白皮书》，integrate.ai

³¹ 资料来源：Srinivasa Rao Chalamala、Naveen Kumar Kummari、Ajeet Kumar Singh、Aditya Saibewar和Krishna Mohan Chalavadi，“通过联邦学习遵守数据保护法规”（Federated learning to comply with data protection regulations），《CSI Transactions on ICT》文章，2020年3月15日

³² 资料来源：Jie Ding、Eric Tramel、Anit Kumar Sahu、Shuang Wu、Salman Avestimehr和Tao Zhang，“联邦学习的挑战与机遇：前景展望”（Federated Learning Challenges And Opportunities: An Outlook），arXiv: 2202.00807v1[cs.LG]，arXiv，康奈尔大学，2020年2月1日

³³ 资料来源：Tian Li、Anit Kumar Sahu、Ameet Talwalkar和Virginia Smith，“联邦学习：挑战、方法和未来方向”，arXiv: 1908.07873v1[cs.LG]，arXiv，康奈尔大学，2019年8月21日

图4：联邦学习模型应用



资料来源：“挖掘用于机器学习的分布式健康数据”（Unlocking Distributed Health Data for Machine Learning），《白皮书》，integrate.ai

联邦学习用例

HealthChain项目：HealthChain希望开发和部署涵盖法国四家医院的联邦学习框架，以预测乳腺癌和黑色素瘤患者的治疗反应。这项工作将有助于肿瘤学家根据组织学影像或皮肤镜检查影像确定每位患者的最有效治疗方法。³⁴

ATHENA项目（以新型分析增强治疗效果）：ATHENA是一个汇集了学术界、医院和行业领袖的多领域合作网络，他们使用机器学习进行肿瘤预测分析。³⁵

MELLODDY项目（机器学习分类帐编排的药物发现）：MELLODDY项目涉及10家大型药企，它们与英伟达、Owkin等公司签署协议携手建立共享平台，以便通过联邦学习在数据集上共同训练人工智能，而无需共享专有数据。³⁶

³⁴ 资料来源：Nicola Rieke、Jonny Hancox、Wenqi Li、Fausto Milletari、Holger R. Roth、Shadi Albarqouni、Spyridon Bakas、Mathieu N. Galtier、Bennett A. Landman、Klaus Maier-Hein、Sébastien Ourselin、Micah Sheller、Ronald M. Summers、Andrew Trask、Daguang Xu、Maximilian Baust和M. Jorge Cardoso，“融合联邦学习的数字医疗的未来展望”（The future of digital health with federated learning），PMCID: PMC7490367，v.3，2020，NPJ Digit Med，期刊列表，PubMed Central生物医学和生命科学数据库，美国国家医学图书馆（NLM），2020年9月14日

³⁵ 资料来源：“以新型分析增强治疗效果”（Augmenting Therapeutic Effectiveness through Novel Analytics），ATHENA项目，ATHENA联盟，portal.athenafederation.org

³⁶ 资料来源：“机器学习分类帐编排的药物发现”（MachinE Learning Ledger Orchestration for Drug Discovery），MELLODDY契约ID: 831472，地平线2020计划，欧洲社区研究与发展信息服务组织

总结

随着个性化医疗健康新时代的发展，人工智能和精准医疗之间相互促进的重要性与日俱增。这种模式有助于采取可能更为有效的个性化治疗和疾病预防策略。本文中，我们探讨了人工智能在风险评估、筛查、诊断、预后、治疗选择和监测等方面的应用潜力，揭示了其在整个医疗健康领域可能起到的颠覆性作用。尽管如此，业界仍需扫除几个障碍，以促进人工智能的大规模应用，包括解决数据隐私、道德影响、监管批准和基础设施投资等问题。通过应对上述挑战，建立对人工智能的信任，培养该领域的人才，我们可以为人工智能和精准医疗之间的融合发展铺平道路，以实现更快速、以患者为中心的个性化医疗干预。

毕马威能够提供的服务

凭借自身的独特优势，毕马威能够通过战略合作关系帮助生物制药企业深入了解生成式人工智能将给精准医疗带来的深度变革。通过下文概述的各项服务，我们协助客户驾驭精准医疗和其他药物开发的复杂局面，识别趋势，评估潜在影响，并制定策略，从而应对生成式人工智能带来的机遇和威胁。

- **战略咨询：** 聚焦业务战略对人工智能战略的需求，帮助客户梳理业务场景和业务转型模式和战略，开展案例研究和业务对标服务。
- **业务机会挖掘和评估：** 在精准医疗和生成式人工智能交叉领域寻找潜在收购或合作机会。考虑的因素包括市场定位、投资组合协同效应、与先进精准医疗计划的一致性以及预期投资回报。
- **商业尽职调查：** 包括评估目标公司的市场地位、商业模式、客户关系和增长前景。
- **市场和竞争情报：** 对生成式人工智能发展进行持续监测，包括其对精准医疗领域的影响。帮助客户深入了解市场动态趋势、精准医疗领域竞争对手对生成式人工智能的使用情况、生成式人工智能发展导致的监管变化以及其他影响商业环境的关键变化。
- **整合规划和并购后整合：** 交易完成后，帮助客户整合收购的企业或资产，着重开展技术的无缝整合，包括生成式人工智能技术。可能包括确定潜在的协同机制、制定整合计划或协助管理整合进程。
- **数字化赋能：** 通过人工智能等数字化技术的导入，实施以及数字化平台运营，全面优化公司的经营管理、客户服务，促进公司收入增长，运营成本降低和工作效率提升。



本文作者



George Stavropoulos

医疗健康和生命科学交易咨询与战略总监
毕马威美国

617-637-5114

gstavropoulos@kpmg.com



Kristin Pothier

医疗健康和生命科学全球主管和美国
交易咨询和战略主管
毕马威美国

617-549-2779

kpothier@kpmg.com



Jeff Stoll PhD

生命科学美国战略主管
毕马威美国

857-334-8768

jeffreystoll@kpmg.com

鸣谢:

Yuma Schuster、Jack Verity、Harsh Kumar、David Goldenthal、Elizabeth Gotfried、Catherine Mcdermott

联系我们



张庆杰

人工智能主管合伙人
数字化赋能主管合伙人
毕马威中国

电话: +86 (10) 8508 5000

邮箱: qingjie.zhang@kpmg.com



于子龙

生命科学行业主管合伙人
毕马威中国

电话: +86 (10) 8553 3588

邮箱: cz.yu@kpmg.com



马卓然

生命科学行业战略与运营合伙人
毕马威中国

电话: +86 (10) 8553 3070

邮箱: jz.ma@kpmg.com



季刚

生命科学行业数字化赋能业务合伙人
毕马威中国

电话: +86 (10) 8508 4430

邮箱: andrew.ji@kpmg.com



戴阳阳

交易战略咨询合伙人
毕马威中国

电话: +86 (21) 2212 3256

邮箱: effie.dai@kpmg.com



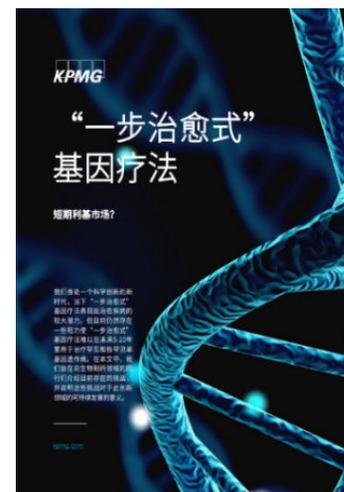
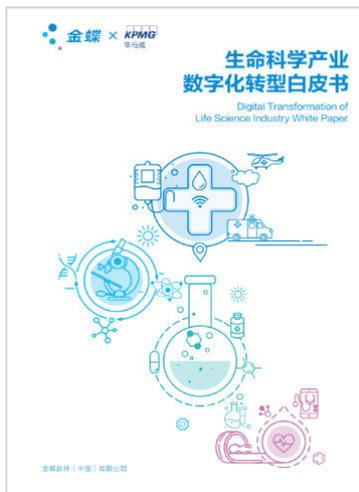
杨阔

交易战略咨询总监
毕马威中国

电话: +86 (21) 2212 3157

邮箱: sk.yang@kpmg.com

毕马威相关洞察报告：



本刊物经毕马威国际授权翻译，已获得原作者（及成员所）授权。

本刊物为毕马威国际发布的英文原文 “A new era of precision medicine” 的中文译本。如本中文译本的字词含义与其原文刊物不一致，应以原文刊物为准。

所载资料仅供一般参考用，并非针对任何个人或团体的个别情况而提供。虽然本所已致力提供准确和及时的资料，但本所不能保证这些资料在阁下收取时或日后仍然准确。任何人士不应在没有详细考虑相关的情况及获取适当的专业意见下依据所载资料行事。

© 2024 毕马威华振会计师事务所(特殊普通合伙) — 中国合伙制会计师事务所，毕马威企业咨询(中国)有限公司 — 中国有限责任公司，毕马威会计师事务所 — 澳门特别行政区合伙制事务所，及毕马威会计师事务所 — 香港特别行政区合伙制事务所，均是与毕马威国际有限公司(英国私营担保有限公司)相关联的独立成员所全球组织中的成员。版权所有，不得转载。

毕马威的名称和标识均为毕马威全球组织中的独立成员所经许可后使用的商标。

DASD-2023-13146.

2024年6月

kpmg.com/cn/socialmedia

