

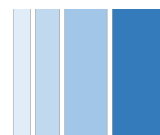
大健康与新医疗

BIG DATA Health
and New Medical

2023 年

第 22 期

上海科学技术情报研究所
上海市前沿技术发展研究中心
技术与创新支持中心 (TISC)



生成式人工智能在医疗领域的研发近况

编者按

近年随着 ChatGPT 和 DALL·E2 的火爆出圈，生成式人工智能正在走进我们的生活。从图像和音乐生成到新的机器学习算法的开发，生成式人工智能正在被添加到无数新的应用程序中。在生物技术领域它也被很多人看好，目前数十家公司在这个领域耕耘，未来将会开启生物医学一个全新的时代。本期简报将从学术前沿、临床应用、市场动态、政策指导四个方面来追踪该领域的研发近况。

目 录

学术前沿 3

➤ 第 25 期前沿技术论坛成功举办

3

➤ 生物界「CHATGPT」首次实现从零合成全新蛋白

4

临床应用 5

➤ 首款由生成式人工智能设计的原创新药进入二期临床试验

5

➤ 新一代 AI 设计抗癌疗法挺进临床开发

7

➤ 英矽智能通过生成式 AI 发现强效选择性 MYT1 抑制剂

9

市场动态 11

- VANTAI 与百时美施贵宝达成合作，总价值 6.74 亿美元

11

政策指导

12

- 医疗行业“未来已来” 打通“政产学研医资”全链条成关键

12

- 欧盟严厉的人工智能法对研究和 CHATGPT 意味着什么

15

第 25 期前沿技术论坛成功举办

第 25 期前沿技术论坛于 2023 年 11 月 14 日在中国中医科学院医学实验中心成功举办，会议主题为“大模型与医学数据”。

该论坛邀请了清华大学计算机系许斌研究员作主题演讲。许斌研究员系统讲述了预训练大模型（生成式 AI 时代的基座）ChatGLM 的研究进展，带领听众解读了 Stanford 报告的世界主流大模型的评测。ChatGLM 是一系列由清华大学开发的国产大模型之一，其中千亿 GLM-130B 是可与 GPT-3 对标的开源数据模型，是中国唯一入选模型，其准确性、恶意性与 GPT-3 持平，鲁棒性和校准误差在所有模型中表现最佳，同时可以同步处理图片和文字等多模态信息的大模型开源基座也已经诞生。许教授在介绍 ChatGLM 系列模型在不同领域进展的基础上，介绍了基于 GLM 大模型的数字中医的应用与成果，结合标准的中医古籍、教材、诊疗记录等数据，实现中医智能问答、辅助诊疗等，为医药领域的研究和应用提供参考和启示。这些应用展示了大模型在提高医疗效率等方面的潜力，引起了与会者的高度关注。



图 1 许斌研究员主题演讲 PPT 信息页

资料来源：第 25 期前沿技术论坛-大模型与医学数据. <http://www.merc.ac.cn/news/766.html> (访问日期：2024-2-26)

生物界「ChatGPT」首次实现从零合成全新蛋白

最近，加州伯克利的一家初出茅庐的初创公司再次取得了惊人的进步。

科学家们采用类似 ChatGPT 的蛋白质工程深度学习语言模型——Progen，

首次实现了 AI 预测蛋白质的合成。

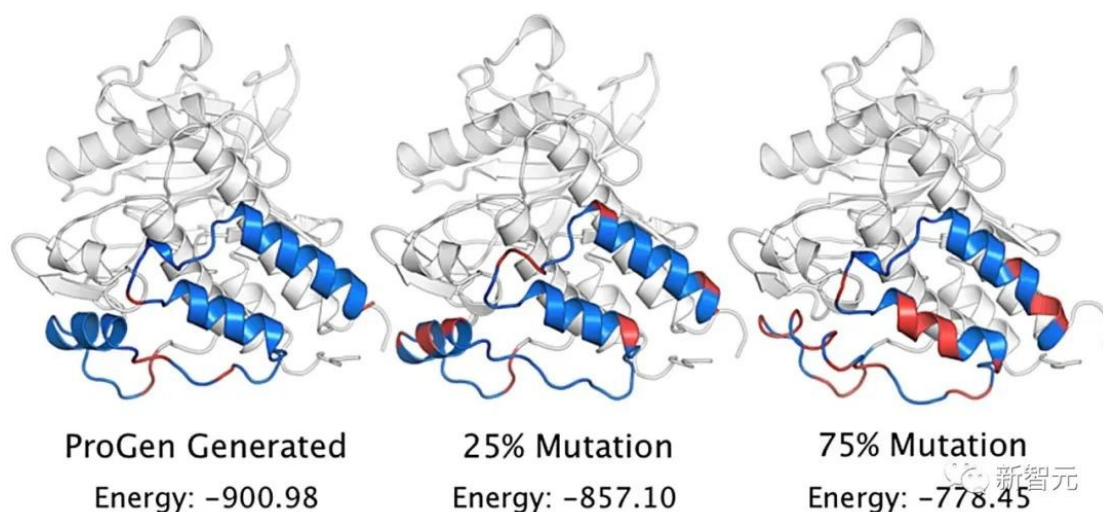


图 2 采用 Progen 合成的蛋白质

这些蛋白质不仅与已知的完全不同，相似度最低的甚至只有 31.4%，但和天然蛋白一样有效。这项工作已经正式发表于 Nature 子刊。论文地址：<https://www.nature.com/articles/s41587-022-01618-2>

该研究用实验表明，自然语言处理虽然是为读写语言文本而开发的，但它也可以学习生物学的一些基本原理。

研究人员表示，这项新技术可能会变得比定向进化（获得诺贝尔奖的蛋白质设计技术）更加强大——“它将通过加快开发可用于从治疗剂到降解塑料等几乎所有用途的新蛋白质，为有 50 年历史的蛋白质工程领域注入活力”。

这家公司名叫 Profluent，由前 Salesforce AI 研究负责人 Ali Madani 创立，已获得 900 万美元的启动资金，用于建立一个集成的湿实验室，并招募机器学习

科学家和生物学家。Madani 在采访中表示，Profulent 已经设计出了多个家族的蛋白质。这些蛋白质的功能与样本蛋白（exemplar proteins）一样，因此是具有高活性的酶。

以往，在自然界中挖掘蛋白质，或者调整蛋白质到所需功能，都十分费力。

Profulent 的目标是，让这个过程变得毫不费力。

资料来源：超越诺奖？生物界「ChatGPT」首次实现从零合成全新蛋白，登 Nature 子刊！. <https://www.51cto.com/article/745739.html>（网上更新日期：2023-02-06，访问日期：2024-2-26）

临床应用

首款由生成式人工智能设计的原创新药进入二期临床试验

据悉，英矽智能作为在沪外资企业，其研发的治疗特发性肺纤维化（IPF）候选药物“INS018_055”已完成 II 期临床试验中首例患者给药，标志着首款由生成式人工智能（AIGC）完成新靶点发现和分子设计的候选药物已推进至 II 临床试验阶段。

据中国医学科学院北京协和医院主任医师、INS018_055 中国 II 期临床试验

牵头研究者徐作军教授介绍，特发性肺纤维化是一种慢性瘢痕性肺部疾病，其特点是肺功能进行性和不可逆的下降，影响全球约 500 万人的生命健康。由于发病、进展较为隐秘，多数患者确诊时病情已发展到中晚期，确诊后中位生存期为 3 年。鉴于现有疗法有限、预后整体较差等情况，IPF 领域仍存在较大的未被满足临床需求。

作为一家人工智能驱动的生物医药企业，英矽智能开发了一个“端到端”人工智能制药系统，包括靶点发现平台 PandaOmics、小分子化合物设计和生成平台 Chemistry42、临床试验结果预测平台 InClinico。

治疗特发性肺纤维化的候选新药“INS018_055”就诞生于这个系统。它的作用靶点是全新的，由 PandaOmics 平台发现。这个平台通过对组学数据和文本数据池的分析，找到病人与健康人群的基因差异，再结合信号通路、科学文献、意见领袖等多方面信息的综合分析，找到靶点与疾病的关联，进而发现某种适应证的新靶点。

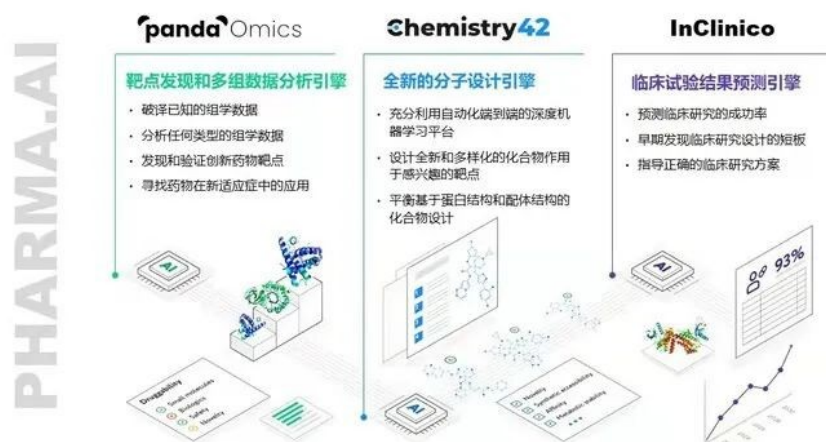


图 3 英矽智能“端到端”药物研发平台 Pharma.AI

针对新靶点，Chemistry42 平台通过“生成式对抗网络”等深度学习算法，设计并合成了一批小分子化合物。研发人员对这批化合物进行实验测试，并将测试数据输入 Chemistry42。这个平台对数据进行深度学习后，再合成一批更优的化合物。就这样，经过数轮“设计—合成—测试—优化”后，英矽智能确定了临床前候选化合物。

“INS018_055 具有抗纤维化和抗炎症潜力，也是第一款由 AI 设计的进入临床阶段的药物。”英矽智能联合首席执行官兼首席科学官任峰博士说，“完成 II 期临床试验的首例给药，是 AI 制药领域的重要进展。我们期待它为全球患者带来新的选择。”

据了解，II 期临床试验采用随机、双盲、安慰剂对照方法，旨在评价“INS018_055”口服给药 12 周用于治疗特发性肺纤维化受试者的安全性、耐受性、药代动力学和有效性，包括 4 个平行队列。为了在更大人群中检验候选药物，英矽智能计划在中国和美国近 40 个研究中心招募 60 例特发性肺纤维化患者，同步开展 II 期临床试验。

英矽智能创始人兼首席执行官亚历克斯·扎沃洛科夫博士表示：“我们将在临床试验中探索人工智能发现和设计的药物对患者的作用，这将对生成式人工智能平台的真正验证。”

资料来源：【药闻】人工智能系统生成的原创新药进入二期临床试验. <https://kcb.sh.gov.cn/html/1/168/151/270/3714.html>（网上更新日期：2023-06-27，访问日期：2024-2-26）

新一代 AI 设计抗癌疗法挺进临床开发

人工智能（AI）模型 AlphaFold 的问世变革了蛋白质三维结构预测和蛋白设计领域。利用 AI 模型赋能创新药物的发现和开发也已经成为生物医药领域的热点。近日，Iambic Therapeutics 公司宣布，利用该公司独有的生成式 AI 平台

NeuralPLexer 预测蛋白质/配体复合体三维结构，以及加入药物分子对这些复合体构象影响的研究在 Nature Machine Intelligence 上发表。新闻稿指出，NeuralPLexer 在多项指标的表现优于其它前沿 AI 模型，包括 AlphaFold2。

[nature](#) > [nature machine intelligence](#) > [articles](#) > [article](#)

Article | [Published: 12 February 2024](#)

State-specific protein–ligand complex structure prediction with a multiscale deep generative model

[Zhuoran Qiao](#) , [Weili Nie](#), [Arash Vahdat](#), [Thomas F. Miller III](#)  & [Animashree Anandkumar](#) 

[Nature Machine Intelligence](#) (2024) | [Cite this article](#)

[Metrics](#)

图 4 Iambic Therapeutics 公司研究论文信息页

蛋白质与小分子配体结合生成的复合体对生命来说至关重要。NeuralPLexer 平台的目标是只基于蛋白质的氨基酸序列和配体分子结构，直接预测蛋白质/配体复合体的结构。这一计算平台利用深度生成式 AI 模型，在原子分辨率的层面对复合体的三维结构以及构象变化进行采样和分析，并且在算法中加入了重要的生物物理学限制。

在这项研究中，NeuralPLexer 与其它已有模型相比，在预测蛋白质/配体盲对接（protein-ligand blind docking）和预测灵活结合位点结构方面均表现优异。

在预测包含大规模构象变化的蛋白质/配体复合体或最新发现的配体结合蛋白的整体结构方面，NeuralPLexer 与 AlphaFold2 相比表现更优。

该公司在 2023 年 10 月训练的 NeuralPLexer2 模型对已有的方法论进行了大幅度改良。除了改进对创新靶点的预测精确度，研究团队还显著扩展了这一模型的应用范围，包括几乎所有生物结构类别，添加了蛋白/蛋白复合体，辅因子（cofactors），翻译后修饰和蛋白质/核酸复合体，并涵盖了蛋白质数据库（PDB）中的几乎所有结构。

在预测表现上，NeuralPLexer2 预测蛋白质/配体复合体结构的成功率达到 75%，而当给模型提供小分子配体附近的氨基酸信息后，成功率可提高到 93%。

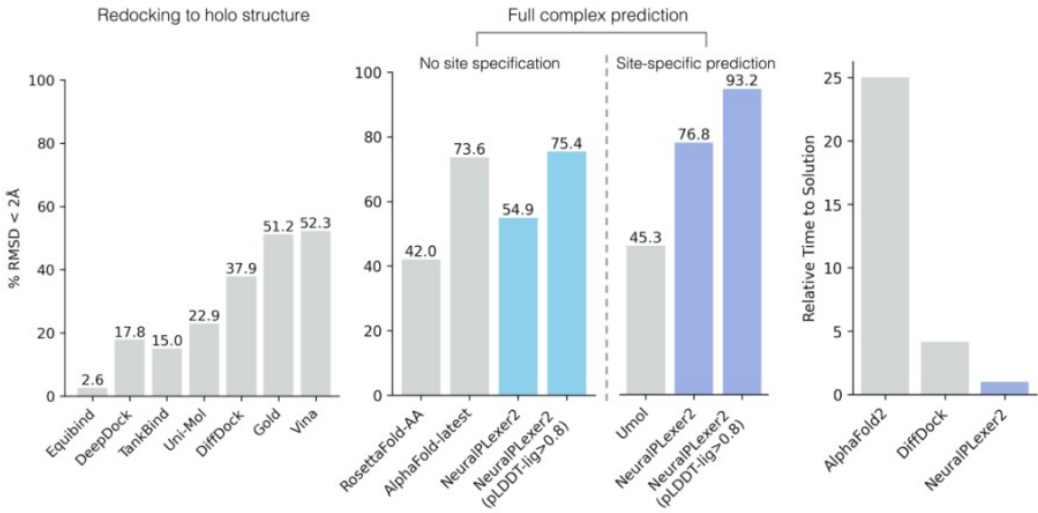


图 5 NeuralPLexer2 在预测蛋白质/配体复合体结构方面的表现

值得一提的是，NeuralPLexer2 的结构预测速度也大幅度提高。Iambic 发表的白皮书指出，NeuralPLexer2 与 AlphaFold2 相比预测速度提高 50 倍！提供了进行超大规模的虚拟筛选和在蛋白组范围内研究配体结合诱发构象变化的机会。

Iambic 公司已经将这一系统成功用于创新药物的发现和开发。该公司的主打在研药物 IAM1363 就是由这一技术平台发现。它是一款选择性，大脑渗透性 HER2 小分子抑制剂，可以在抑制野生型和致癌 HER2 突变蛋白的同时，避免对表皮生长因子受体（EGFR）的脱靶抑制效应，从而提高药物的安全性、扩展治疗窗口。在临床前研究中，IAM1363 对 HER2 的选择性是 EGFR 的 1000 倍以上。IAM1363 的 IND 申请已经获得美国 FDA 的许可，预计在 2024 年初进入临床开发阶段。

资料来源：超越 AlphaFold2? 新一代 AI 设计抗癌疗法挺进临床开发. <https://new.qq.com/rain/a/20240215 A00ZF700>（网上更新日期：2024-02-15，访问日期：2024-2-26）

英矽智能通过生成式 AI 发现强效选择性 MYT1 抑制剂

近日，生成式人工智能（AI）驱动的临床阶段药物研发公司英矽智能

(Insilico Medicine) 在药物化学领域 Top 期刊 Journal of Medicinal Chemistry 期刊发表了一篇题为：Discovery of Tetrahydropyrazolopyrazine Derivatives as Potent and Selective MYT1 Inhibitors for the Treatment of Cancer 的研究论文。

在英矽智能人工智能驱动的生成生物学和生成化学平台支持下，该研究发现了 MYT1 作为乳腺癌和妇科癌症的治疗靶点的潜力，并设计了一系列靶向 MYT1 的新型、强效、高选择性抑制剂。



图 6 英矽智能公司研究论文信息页

在全球范围内，乳腺癌和妇科癌症严重威胁着女性的健康和整体生活质量。

为了确定新疗法的潜在靶点，生成式人工智能（AI）驱动的临床阶段药物研发公司英矽智能（Insilico Medicine）的研究团队利用专有的由人工智能驱动的靶

点识别平台 PandaOmics，分析了五种妇科癌症的数据，包括卵巢癌、子宫内膜癌、宫颈癌、乳腺癌，以及三阴性乳腺癌。值得注意的是，在所有疾病中，MYT1 靶点和这些疾病的相关性始终位居前列。

MYT1 是 Wee1-激酶家族的成员，该蛋白在大多数正常组织中很少表达，但在多种癌症中高度表达。研究发现，MYT1 抑制和 CCNE1 扩增构成合成致死效应，在细胞周期调控中发挥关键作用。这表明抑制 MYT1 是一种具有前景的、治疗 CCNE1 扩增肿瘤的药物开发策略。

然而，MYT1 与 Wee1 高度同源，这使得设计高选择性的 MYT1 抑制剂充满挑战性。在本研究中，英矽智能在人工智能驱动的小分子生成平台 Chemsitry42 的支持下克服了这个难点。英矽智能利用 Chemistry42 基于结构的药物设计（SBDD）策略，并设置严格的相似性和选择性筛选指标，从头开始设计了一系列靶向 MYT1 的化合物。在这些结构新颖的化合物中，有一个系列脱颖而出被选为苗头化合物。

英矽智能随后对该复合物进行了 X 射线晶体结构分析，发现细微的化学结构改变对化合物活性有重大影响。这些发现为进一步分子优化提供了指导，并

帮助研究者发现了先导化合物。该分子具有良好的 MYT1 活性，对 Wee1 和其他激酶具有出色的选择性，从而降低了脱靶效应的潜在风险，使得化合物更为安全。在临床前研究中，该化合物还显示出理想的体内抗肿瘤功效，和具有开发潜力的 ADME 和 PK/PD 性质。

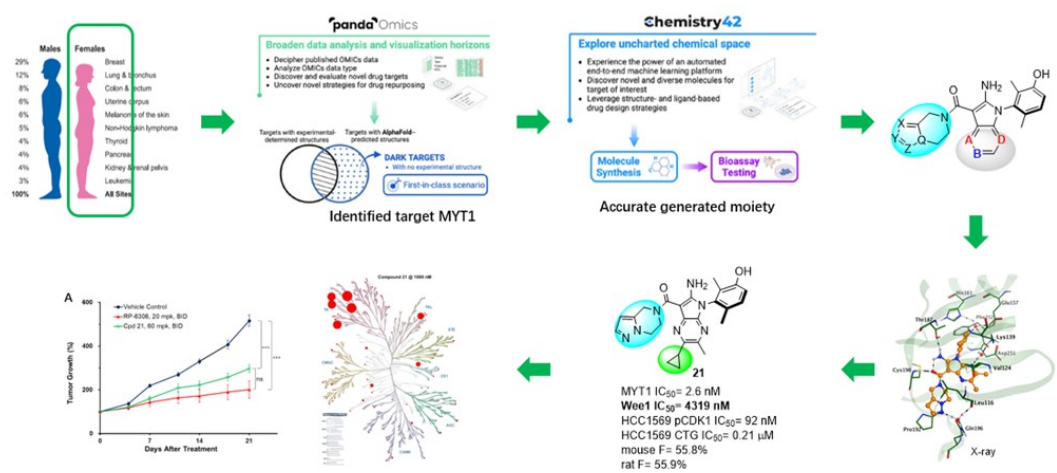


图 7 MYT1 项目的研究流程

论文第一作者、英矽智能 MYT1 项目药物化学负责人王亚洲博士表示，这个项目的创新方法不仅提供了一种有效的靶点识别策略，还发现了一种具有开发潜力的高选择性 MYT1 抑制剂。这个发现既拓展了公司的合成致死管线，也为寻找一种惠及患者的更安全、更有效的治疗策略提供了新思路。

资料来源：英矽智能通过生成式 AI 发现治疗乳腺癌等癌症的强效选择性 MYT1 抑制剂。
<https://news.bioon.com/article/92568106988a.html>（网上更新日期：2024-01-24，访问日期：

2024-2-26)

市场动态

VantAI 与百时美施贵宝达成合作，总价值 6.74 亿美元

2024 年 2 月 13 日，专注于生成式人工智能药物发现的公司 VantAI 与领先的制药公司百时美施贵宝（Bristol Myers Squibb）达成战略合作，为感兴趣的治疗靶点发现新的分子粘合剂。该合作利用 VantAI 的几何深度学习能力和百时美施贵宝在靶向蛋白质降解方面的专业知识，发现和开发新的小分子疗法。双方的合作将侧重于利用 VantAI 的生成式人工智能平台设计作为小分子疗法的分子粘合剂。这些努力推动了 VantAI 的使命，即发掘近程调控的潜力，将其作为针对治疗靶点的有力工具。VantAI 有资格从百时美施贵宝获得高达 6.74 亿美元的发现、开发、临床、监管和销售里程碑付款以及分阶段特许权使用费，并可选择进一步扩展到其他治疗项目。

资料来源：AI+分子粘合剂发现 | VantAI 与百时美施贵宝达成合作，总价值 6.74 亿美元.

<https://bydrug.pharmcube.com/news/detail/001a839f51c71b717d5bfbf3b142c32d>（网上更新

日期：2024-02-19，访问日期：2024-2-26)

政策指导

医疗行业“未来已来”打通“政产学研医资”全链条成关键

今年以来以 GPT-4 为代表的大模型点燃人工智能发展浪潮，国内多家企业先后发布大模型产品。AI 大模型的强力出圈，引爆生成式人工智能（AIGC）时代的到来。

上海当前正在建设人工智能“上海高地”，发挥人工智能的“头雁效应”，打造一批具有影响力的人工智能产业集群。同时，上海先后发布新一代人工智能算法创新行动计划，全国首个人工智能地方标准体系建设指导意见，全国首部 AI 省级法规《上海市促进人工智能产业发展条例》等一系列相关制度和配套创新政策，对人工智能产业发展起到了重要的引领作用。

数字经济作为全球经济发展的重要引擎之一，正迅速推动各行各业的深度融合。而生物医药作为社会经济发展的重点领域之一，越来越重视数据在推动创新和发展中的不可替代作用。

- 数字化医疗高速发展

近年来，以大数据、云计算和人工智能等为代表的新一代数字技术的日新月异，数字化医疗领域正经历着深刻变革。传统医疗模式面临重构，预示着中国数字医疗行业进入新的发展期。数字化医疗作为将计算机技术和信息技术应用于整个医疗过程的一种现代化医疗方式，不仅能够优化患者就诊流程、提高诊断准确率和完善病人病例信息，还可以实现医疗设备与医疗专家的资源共享，建立起完善的健康信息数据库和权威的健康信息系统，从而打造更高质量、更精确的医疗服务体系。

数字医疗领域的发展将更加迅速、更加深入。数字技术将在改善医疗服务、提高治疗效果，甚至预防疾病发生等方面发挥更加重要的作用。数字医疗将更深地融入人们的日常生活，成为健康管理不可或缺的一部分。随着精准数据的不断积累，数字医疗的发展将不断向精准治疗的方向延伸。

全球数字医疗规模庞大，中国被视为其中最有可能性的市场之一。随着“健康中国 2030”的目标实现渐近，医药产业结构改革和重塑，以及数字化医疗技术日新月异的发展，中国的生物制药行业迎来了变革的时机。另一方面，医疗行业人员和患者也在呼吁和期盼更多以数字化科技驱动的创新诊断方式和疗法，

为行业发展提出了多元化的挑战和需求。数字化转型已经成为生物医药技术创新和长期增长的关键因素。

而从后续发展来看，上海也将进一步推动有效促进生物医药科研成果转化和数据行业专业人才培养等的措施。而在此期间，数字化手段将发挥重要作用。

● 数字化赋能产业布局

据 IDC 数据预测，到 2025 年人工智能技术应用市场总值将达到 1270 亿美元，其中医疗行业将占市场规模的五分之一。在中国，医疗与数字化技术的结合发展势头迅猛，日益显现出巨大潜能。人工智能赋能健康解决方案也成为跨国药企进行本土化战略布局的重点方向之一。

有券商分析师对 21 世纪经济报道记者表示，目前，人工智能数字化技术在药物研发方面已进行了一系列布局。首先，在早期研发阶段，数据科学+AI 通过机器学习、疾病数据库与生物医学知识图谱相结合，能够帮助更早发现新的潜在药物靶点；其次，在新药开发阶段，可以将人工智能技术与化工自动化相结合，使得原本需要数月才能研制完成的先导分子现在只需数周即可完成，而且无需人工干预；再者，在临床研发阶段，可以利用人工智能技术辅助患者分

类。人工智能还提高了临床试验用药需求的预测效率，从而降低了试验中心供应药物的成本，避免造成浪费；此外，在研发后期，可以通过“真实世界证据数据计划”（real-world evidence data initiatives）拓宽对患者的洞察，依靠机器学习方法，更加准确、高效地获取患者治疗效果。

科睿唯安生命科学与医疗健康事业部资深产品负责人刘梦佳也指出，药物研发过程是缓慢而昂贵的，《Nature》最新数据也指出，一款小分子创新药物研发需要花费 20 亿美金历经 15 年才能研发出来，这么高的投入背后还面临进入临床 I 期后 90% 的失败率。

“制药行业相对于 IT 行业来说，还是一个对新技术比较保守的行业。很多时候与企业自建能力相比，也更希望与 AI 平台公司的合作。事实上，从数据上也可以看出，2018 年开始，这种合作交易有了大幅增加。特别是随着新靶点、新生物标志物的发现，数字化技术对于新药物靶点机制的筛选、提升研发效率等层面均具有一定的优势，而这也将成为药企规避研发失败风险的一大方向。”刘梦佳说。

武田制药全球高级副总裁单国洪近日也在行业论坛上对 21 世纪经济报道表

示，数字化发展正在重塑医疗行业，企业也需要通过建立开放、创新的生态体系，与产业集群共同发力，推动创新产品的商业化进程和落地。

“数字化科技是赋予企业韧性、推动企业可持续发展的关键要素，并能够减少患者的就医障碍。在武田最近的临床试验中，公司正在利用数字化能力，整合患者和运营数据，提供虚拟和混合试验方案，从而增加试验参与者的多样性。”单国洪说，最近，武田中国与复旦大学智能医学研究院签署合作备忘录，共同打造武田数智创新研究院。与此同时，武田中国还凭借 TakedaSpark 创新孵化平台，吸引了 200 多家本土创业公司参与，并先后有 12 个创新方案进入最终的验证性试点合作阶段。

从目前市场发展趋势来看，全新作用靶点和机制越来越少，制药企业需要投入更多人力及动力去打造 First-In-Class 产品，如何才能提高研发效率，如何用已有的技术手段赋能药物研发，数字化技术的作用已是越来越大。

如何打通最后“一公里”？

美国药品研究和制造商协会（PhRMA）的一项研究表明，进入临床试验的药物中只有约 12% 能够获得监管部门的批准。而 TechEmergence 一份报告称 AI

技术能够将临床新药研发的成功率从 12%提高到约 14%，这几个百分点的提高将为生物制药业节省数十亿美元的研发经费。此外，有相关数据显示，AI 技术在化合物合成和筛选方面比传统手段可节约 40%-50%的时间，每年为药企节约 260 亿美元的化合物筛选成本；在临床研究阶段，可节约 50%-60%的时间，每年可节约 280 亿美元的临床试验费用。

不过，在推动数字化技术与医疗产业发展的结合过程中，也可以发现，数字化技术目前也存在一定的局限性：一方面，生物学本身太复杂，产业界对生物学的知识概念理解太局限，这也导致临床研究结果带有一定的不确定性；另一方面，人工智能现有大数据的结构、质量都存在一定缺陷。这需要业界多方共同去解决这些问题。

如何解决？为此，我国从顶层设计层面在人工智能和药物研发方面均出台了一系列政策，鼓励和支持人工智能应用于药物研发，推动产业高质量发展。国家“十四五”规划纲要提出，要着眼于抢占未来产业发展先机，培育先导性和支柱性产业。2023 年 4 月 28 日，中央政治局会议提出，要“重视通用人工智能发展，营造创新生态”。7 月，上海发布了《上海市生物医药产业数字化转型实

施方案（2023-2025 年）》，全面、深入地支持生物医药产业的数字化转型。

除了政策支持，加强产学研医多方融合也成为一大趋势。复旦大学生物医学研究院党委书记储以微教授指出，现在高校项目开展的临床试验分为两部分，一部分叫干实验，一部分叫湿实验。干实验就是一种数据分析，可以获得相对正确的靶点。有了大数据，可以帮助发现一些新的免疫细胞亚群和相关功能。

这也意味着，大数据在基础研究和临床应用都是必不可少的，尤其对今后临床诊断而言具有较为重要的作用。此外，预测性的学术合作也非常重要，但是需要学院、医院、药厂加强合作，需要构建产学研医合作桥梁，适应多方发展需求。

“想要引领新医科的发展，提升医科学科水平，需要加强开展与产业界的深度合作，整合行业资源，探索创新课题研究和建立概念模型。通过与企业的合作，多方合作，可以为医疗行业带来变革性的影响，在学术界和产业界之间架起一座桥梁，为医学研究和患者治疗带来有价值、有意义的成果。”储以微教授说，中国是数字化创新的策源地和领先者，正持续为数字化技术的发展注入新的活力。

资料来源：医疗行业“未来已来” 打通“政产学研医资”全链条成关键. <https://www.haocew.com/news/detail/jpfkyk9vbd> (网上更新日期：2023-10-18，访问日期：2024-2-26)

欧盟严厉的人工智能法对研究和 ChatGPT 意味着什么

欧盟国家准备通过世界上第一套全面的人工智能（AI）监管法律。《欧盟人工智能法》对风险最大的人工智能模型做出了最严厉的规定，旨在确保人工智能系统的安全，并尊重基本权利和欧盟价值观。

加利福尼亚州斯坦福大学研究人工智能社会影响的里希-博马萨尼（Rishi Bommasani）说：“就我们如何看待人工智能监管和开创先例而言，该法案影响巨大。”

这项立法是在人工智能飞速发展的背景下出台的。预计今年将推出新版本的生成式人工智能模型，如加州旧金山 OpenAI 公司开发的 GPT，它为 ChatGPT 提供了动力，而现有的系统正被用于诈骗和传播错误信息。中国已经使用了一系列法律来指导人工智能的商业应用，美国的监管也在进行之中。去年 10 月，乔-拜登总统签署了美国首个人工智能行政命令，要求联邦机构采取

行动管理人工智能风险。

欧盟各国政府于 2 月 2 日批准了这一立法，该法律现在需要欧盟三大立法机构之一的欧洲议会的最终签署；预计这将在 4 月份实现。如果文本保持不变，正如政策观察者所预期的那样，该法将于 2026 年生效。

一些研究人员对该法鼓励开放科学的潜力表示欢迎，而另一些人则担心它会扼杀创新。《自然》杂志将探讨该法将如何影响科研工作。

欧盟的做法是什么？

欧盟选择根据人工智能模型的潜在风险对其进行监管，对风险较高的应用适用更严格的规则，并对具有广泛和不可预测用途的通用人工智能模型（如 GPT）制定单独的法规。

该法禁止使用具有“不可接受风险”的人工智能系统，例如使用生物识别数据推断敏感特征（如性取向）的系统。高风险应用，如在招聘和执法中使用人工智能，必须履行一定的义务；例如，开发者必须证明他们的模型是安全的、透明的、可向用户解释的，并且遵守隐私法规，不存在歧视。对于风险较低的人工智能工具，开发者仍需告知用户何时与人工智能生成的内容进行交互。该法

律适用于在欧盟运营的模式，任何违反规定的公司都有可能被处以高达其全球年利润 7% 的罚款。

意大利米兰博科尼大学（Bocconi University）计算机科学家德克-霍维（Dirk Hovy）说：“我认为这是一个好办法”。他说，“人工智能已迅速变得强大且无处不在。建立一个框架来指导其使用和发展是绝对有意义的”。

总部位于柏林的非营利组织“算法观察”（AlgorithmWatch）是一家研究自动化对社会影响的机构，该组织的政治学家基利安-维特-迪特曼（Kilian Vieth-Ditlmann）说，有些人认为这些法律走得不够远，为军事和国家安全目的的豁免留下了“缺口”，也为人工智能在执法和移民中的使用留下了漏洞。

它对研究人员的影响有多大？

理论上，影响很小。去年，欧洲议会在法案草案中增加了一项条款，规定纯粹为研究、开发或原型设计而开发的人工智能模型可以免责。柏林赫蒂学院（Hertie School）研究人工智能及其监管的乔安娜-布赖森（Joanna Bryson）说，“欧盟一直在努力确保该法案不会对研究产生负面影响。他们真的不想切断创新，所以如果这会成为一个问题，我会感到非常震惊。”

霍维说，但该法案仍有可能产生影响，它将促使研究人员思考透明度、如何报告他们的模型以及潜在的偏见。他说：“我认为这将渗透下去，促进良好做法的形成。”

德国慕尼黑工业大学的医生、旨在实现机器学习民主化的非营利组织 LAION（大规模人工智能开放网络）的联合创始人罗伯特-卡茨马尔奇克（Robert Kaczmarczyk）担心，这项法律可能会阻碍推动研究的小公司，它们可能需要建立内部结构来遵守法律。他说：“作为一家小公司，适应法律确实很难。”

这对 GPT 等功能强大的模型意味着什么？

经过激烈的讨论，政策制定者选择将功能强大的通用模型--如创建图像、代码和视频的生成模型--纳入自己的两级类别进行监管。

第一层涵盖所有通用模型，但仅用于研究或根据开源许可发布的模型除外。这些模型将受到透明度要求的约束，包括详细说明其训练方法和能源消耗，并且必须表明它们尊重版权法。

第二层更为严格，将涵盖被认为具有“高影响能力”的通用模型，它们会带来

更高的“系统性风险”。博马萨尼说，这些模型将承担“一些相当重要的义务”，包括严格的安全测试和网络安全检查。开发者将被要求公布其架构和数据源的详细信息。

对欧盟来说，“大”实际上等同于危险：任何在训练中使用超过 1025 FLOPs（计算机运算次数）的模型都属于高影响模型。Bommasani 说，用如此大的计算能力来训练一个模型需要花费 5000 万到 1 亿美元，因此这是一个很高的标准。它应该包括 OpenAI 当前的模型 GPT-4 等模型，也可能包括 Meta 的开源竞争对手 LLaMA 的未来迭代。这一级别的开源模型都要接受监管，不过纯研究模型可以免于监管。

一些科学家反对对人工智能模型进行监管，他们更愿意关注如何使用这些模型。德国尤里希超级计算中心（Jülich Supercomputing Centre）的人工智能研究员、LAION 的另一位联合创始人杰尼亚-吉采夫（Jenia Jitsev）说：“更聪明、能力更强并不意味着危害更大”。Jitsev 补充说，“以任何能力衡量标准作为监管依据都没有科学依据。他们打了个比方，把使用一定人时的化学都定义为危险化学品。这样做毫无益处”。

该法案会支持开源人工智能吗？

欧盟政策制定者和开源倡导者希望如此。霍维说，该法案鼓励人工智能信息的可用性、可复制性和透明性，这几乎就像是“宣读开源运动的宣言”。博马萨尼说，有些模式比其他模式更开放，目前还不清楚该法案的语言将如何解释。但他认为，立法者有意将通用模型（如 LLaMA-2 和巴黎初创公司 Mistral AI 的模型）排除在外。

Bommasani 说，“欧盟鼓励开源人工智能的做法与美国的策略明显不同。欧盟的思路是，开源对于欧盟与美国和中国竞争至关重要”。

如何执行该法案？

欧盟委员会将成立一个人工智能办公室，在独立专家的建议下监督通用模型。该办公室将制定评估这些模型能力和监控相关风险的方法。但是，即使像 OpenAI 这样的公司遵守了相关规定并提交了庞大的数据集等资料，Jitsev 也质疑公共机构是否有足够的资源对提交的资料进行审查。“透明化的要求非常重要”，他们说“但对于如何执行这些程序却很少考虑”。

资料来源：《自然》：欧盟严厉的人工智能法对研究和 ChatGPT 意味着什么。

<https://zhu anlan.zhihu.com/p/682489808> (网上更新日期：2024-02-17 ,
访问日期：2024-2-26)



地址：上海市永福路 265 号

邮编：200031

编辑：赵晓勤

责编：姚恒美

编审：林鹤

电话：021-64455555