

2023

中国AI制药企业白皮书

White Paper on Chinese AI Pharmaceutical Enterprises

成都高投生物医药园区管理有限公司

药融云数字科技（成都）有限公司

总编辑：王中健

2023年11月30日



免责声明

DISCLAIMER

本报告基于本公司（药融云数字科技（成都）有限公司）及其研究人员认为可信的公开资料和文献，反映了研究人员的研究观点，结论不受任何第三方的授意或影响。但本公司对这些信息的真实性、准确性、可靠性和完整性不作任何保证，且本报告中的资料、意见、预测均反映报告初次公开发布之前的判断，可能会随时调整。本公司可在不发出通知的情形下对本报告所含信息做出修改，读者应当自行关注相应的更新或修改。

本报告所有的信息或观点仅作为读者的参考，不构成任何投资、法律、会计或税务的建议，也不构成任何广告或要约。在任何情况下，本公司不对任何人因使用本报告中的内容所致的损失承担任何责任。

本报告相关的知识产权归本公司所有。未经本公司事先授权，任何人不得对本报告进行任何形式的翻译、发布或复制。如需引用，应注明出处为“药融云数字科技（成都）有限公司”以及“中国AI制药企业白皮书”名称，且不得对本报告进行有悖原意的删节或修改。如需联系本公司，请联系：yangjianyun@bcpmdata.com。

编制说明

作为近年来技术变革的重要力量，AI在生物医药领域展示了极大的应用潜力。国家发改委2022年发布的《“十四五”生物经济发展规划》提出，要利用云计算、大数据、人工智能等信息技术，对治疗适应症与新靶点验证、临床前与临床试验、产品设计优化与产业化等新药研制过程进行全程监管，实现药物产业的精准化研制与规模化发展。随着AI浪潮席卷全球，生物制药也被视为大模型有望率先落地的场景之一。

基于此，2023年10月，成都高投生物医药园区管理有限公司联合药融云数字科技（成都）有限公司，共同编写了《中国AI制药企业白皮书》，希望为业界厂商、政府机构等相关方提供有益思考，共同推动产业高质量发展。

由于时间仓促，水平所限，我们的工作还有不足之处。下一步，我们还将广泛采纳各方面的建议，进一步深入相关研究，持续完善白皮书内容，在已有版本的基础上，适时修订发布新版。

出品方：药融云数字科技（成都）有限公司
成都高投生物医药园区管理有限公司

总编辑：王中健

副总编辑：肖海波 戴祁 江丹

责任编辑：张祯杰

执行主编：王婷

副主编：赖庆刚 李亚玲 杨建云

编写专家：王婷 衡星 黄镭 梁崇晟

版面设计：陈悦

2023年11月



成都高投生物医药园区管理有限公司

成都高投生物医药园区管理有限公司（以下简称“高投生物园公司”）成立于2016年8月，系成都高新投资集团有限公司下属子集团成都高投产城建设集团有限公司的全资子公司。作为国内知名生物医药产业园区天府生命科技园和成都前沿医学中心的运营管理公司，高投生物园公司定位为国内顶尖生物产业高品质科创空间整合运营商。

高投生物园公司现管理运营天府生命科技园、成都前沿医学中心等专业载体面积约83万 m^2 。



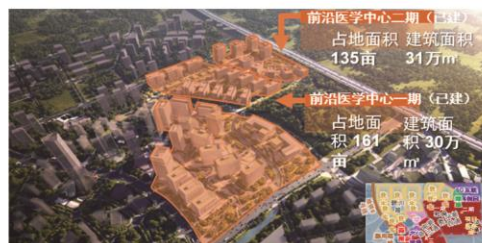
天府生命科技园

其中，天府生命科技园位于成都高新区科园南路88号，占地面积111亩，总建筑面积22万 m^2 ，于2010年正式启用。定位生物医药研发创新中心和产业孵化中心，主要聚焦创新生物制剂、化学药研发、中药现代化、细胞治疗、生物材料、医疗器械等细分领域。



成都前沿医学中心

成都前沿医学中心目前由一期、二期两个园区组成，三期正在建设中。一期前沿医学转化校地合作示范园区布局高性能医疗器械、创新药研发、干细胞治疗等，二期“BT+IT”融合产业示范园区建设国内首个全周期新药创制大科学装置平台，布局医疗大数据智算中心、生物样本库等，三期打造TECH+TOD“双T”产业结构，天府锦城实验室核心承载区，布局精准医学、前沿生物技术，创新医疗器械、疾病动物模型库等方向。



药融咨询介绍

药融咨询是享融智云旗下生物医药专业咨询服务品牌，由深耕医药领域多年的专业人士组成，核心成员均来自国际顶级咨询机构和行业标杆企业，涵盖立项、市场、战略、投资等从业背景，依托药融圈丰富的外部专家资源及药融云全面的医药产业链数据库，为客户提供专业咨询服务与定制化解决方案。



40000+

医药行业精英人脉资源



1300+

业内高端专家资源



300+

生态合作伙伴



200+

项目案例



115+

疾病领域



10+

行业白皮书

服务业务 · 企业端

聚焦行业痛点，引领医药创新，为企业提供一站式咨询服务

咨询业务总览



产品管线规划



产品立项评估



企业竞争调研



市场潜力评估



企业战略咨询



外部环境观察

服务业务 · 政府端

依托生物医药产业数据优势
为政府规划与监管深入落地提供依据



产业现状

战略定位

总体目标

生物医药
产业规划

生物医药
产业监督

产业概览

审批现状

优势分析

目录

第一章 人工智能（AI）制药行业概览

人工智能的基本概念	09
AI制药是AI产业化中重要的应用场景	10
AI制药行业发展历史	11
AI制药行业配套政策分析	13
AI制药行业投融资现状分析	14

第二章 AI+药物发现

AI助力药物发现阶段概览	16
AI助力药物靶点发现与验证	16
AI驱动药物分子设计、优化	23
AI辅助药物分子筛选	34

第三章 AI+临床前研发

AI助力药物临床前研发概览	38
AI辅助药物理化特性预测	38
AI辅助药物剂型设计	39
AI助力药物ADMET性质预测	40
AI+临床前研发国内企业布局	42

第四章 AI+临床研究

AI助力新药临床试验概览	46
AI助力受试者招募管理	50
AI助力预测临床试验效果	53
AI助力临床试验数据管理	57
AI赋能药物注册申报	59

第五章 AI+药物生产及商业化

AI赋能药物生产支持	61
AI赋能药物市场开拓与商业化	65
AI逐步应用于药物警戒领域	67

第六章 AI制药行业总结及未来展望

现状分析及展望	70
上市企业商业化模式	73
全球主要AI制药上市企业	74

报告缩略词中英文对照表

英文缩写	英文全称	中文译文
AI	Artificial Intelligence	人工智能
AIDD	Artificial Intelligence Drug Discovery & Design	AI制药
CADD	Computer-aided Drug Design	计算机辅助药物设计
SBDD	Structure-based Drug Design	基于结构的药物设计
HTS	High-throughput Sequencing	高通量测序
ML	Machine Learning	机器学习
DL	Deep Learning	深度学习
RL	Reinforcement Learning	强化学习
TL	Transfer Learning	迁移学习
NLP	Natural Language Processing	自然语言处理
GAN	Generative Adversarial Networks	生成对抗结构
RNN	Recurrent Neural Network	循环神经网络
VAE	Variational Autoencoder	变分自编码器
SMILES	Simplified Molecular Input Line Entry System	分子反应的线性符号
DTI	Drug Targets Interactions	药物-靶标相互作用
GPT	Generative Pre-Trained Transformer	生成式预训练Transformer模型
CSP	Crystal Structure Prediction	晶型预测
ADMET	Absorption、Distribution、Metabolism、Excretion、Toxicity	药物吸收，分配，代谢，排泄和毒性
QSAR	Quantitative structure-Activity Relationship	定量结构活性关系
PK	Pharmacokinetics	药代动力学
PLM	Protein Language Model	蛋白质语言模型

01

第一章

人工智能（AI）制药行业概览

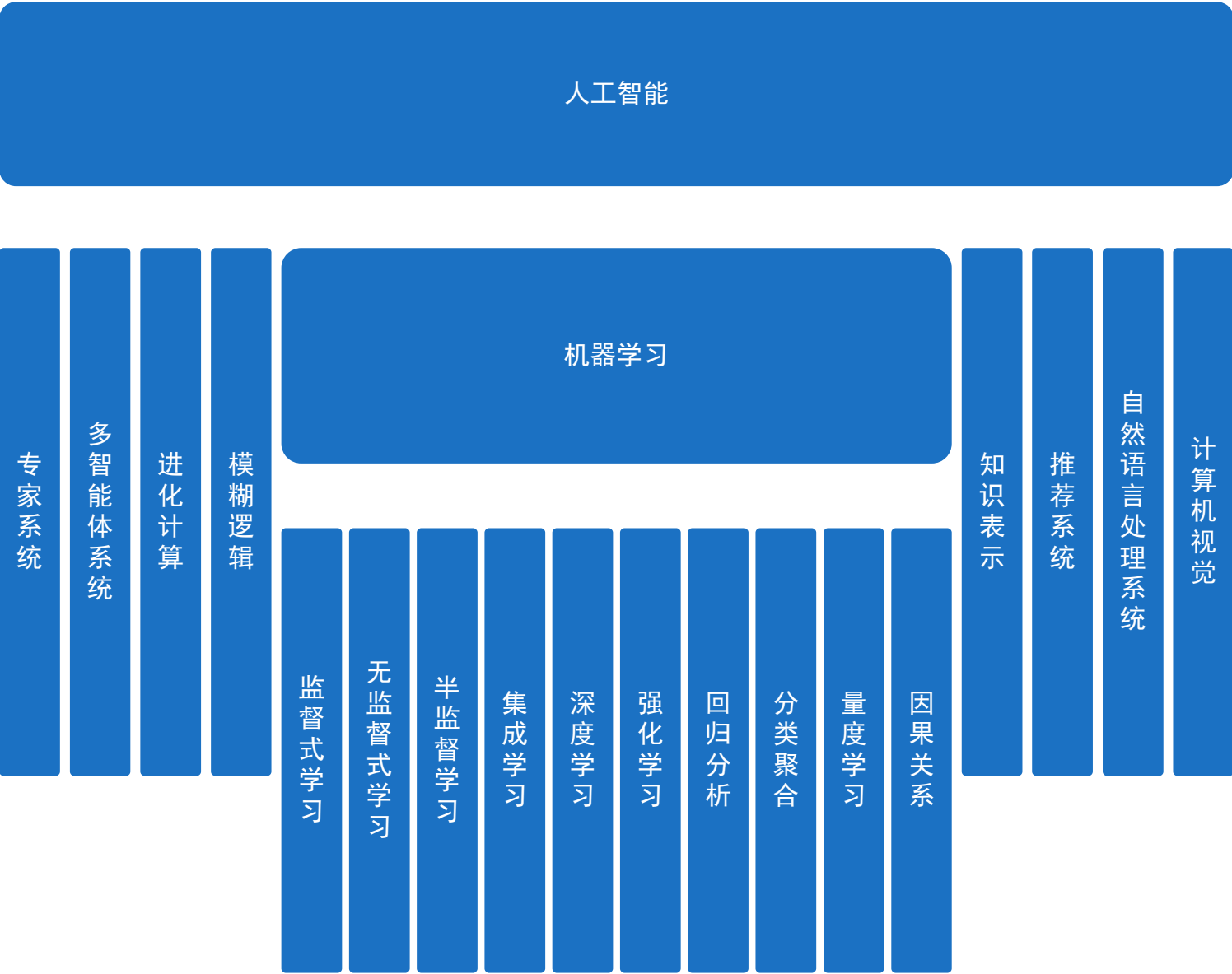
■ 人工智能的基本概念

人工智能（Artificial Intelligence，AI），是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器。

人工智能是一个广泛的领域，包括专家系统、多智能体系统、进化计算、模糊逻辑、机器学习、知识表示、推荐系统、自然语言处理系统和计算机视觉等多个方面。其中，机器学习技术是应用最广泛的一种技术，具体包括以下几种方法：监督式学习、无监督式学习、半监督学习、集成学习、深度学习、强化学习、回归分析、分类聚合、量度学习、因果关系等。各类机器学习技术在各个领域都有广泛的应用，推动了人工智能的发展和应用。

人工智能作为新一轮产业变革的核心驱动力，将进一步释放历次科技革命和产业变革积蓄的巨大能量，并创造新的强大引擎，重构生产、分配、交换、消费等经济活动各环节，形成从宏观到微观各领域的智能化新需求，催生新技术、新产品、新产业、新业态、新模式，引发经济结构重大变革，深刻改变人类生产生活方式和思维模式，实现社会生产力的整体跃升。

图表1 实现人工智能的主要技术汇总^[1]

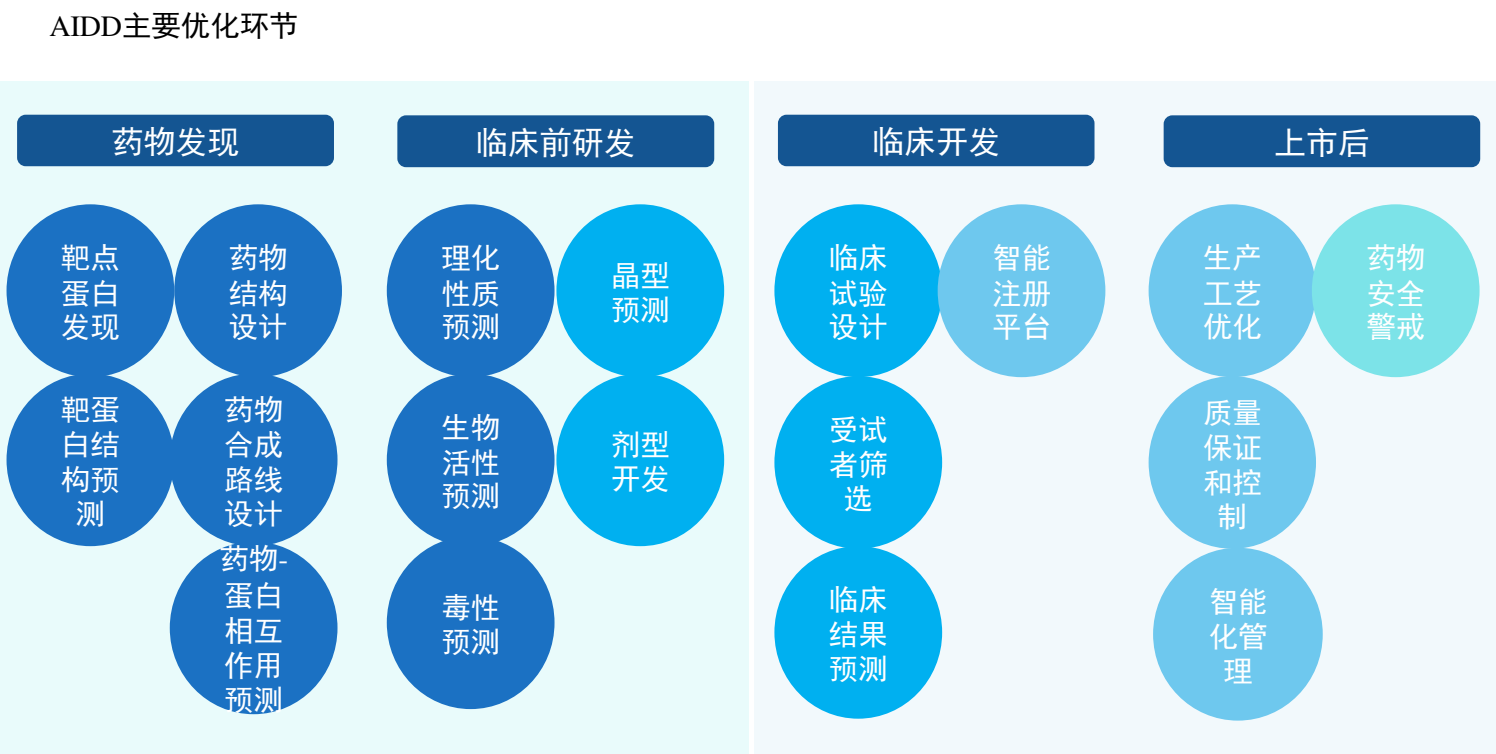


[1]整理自药融咨询

■ AI制药是AI产业化中重要的应用场景

AI制药（AIDD）是指利用AI技术在药物研发、药物设计、药物筛选、临床试验和药物生产等各个环节中应用的制药领域。AI在药物研发中可以通过数据挖掘、机器学习和深度学习等技术，加速药物发现和设计过程，提高研发效率和成功率。AI还可以在药物筛选中帮助挑选出具有潜在疗效的候选药物，降低研发成本和时间。在临床试验中，AI可以帮助优化试验设计、招募适合的患者群体，并提供数据分析和预测，加快药物上市进程。此外，AI还可以应用于药物生产中的质量控制、流程优化和智能化管理等方面，提高药物的生产效率和质量。

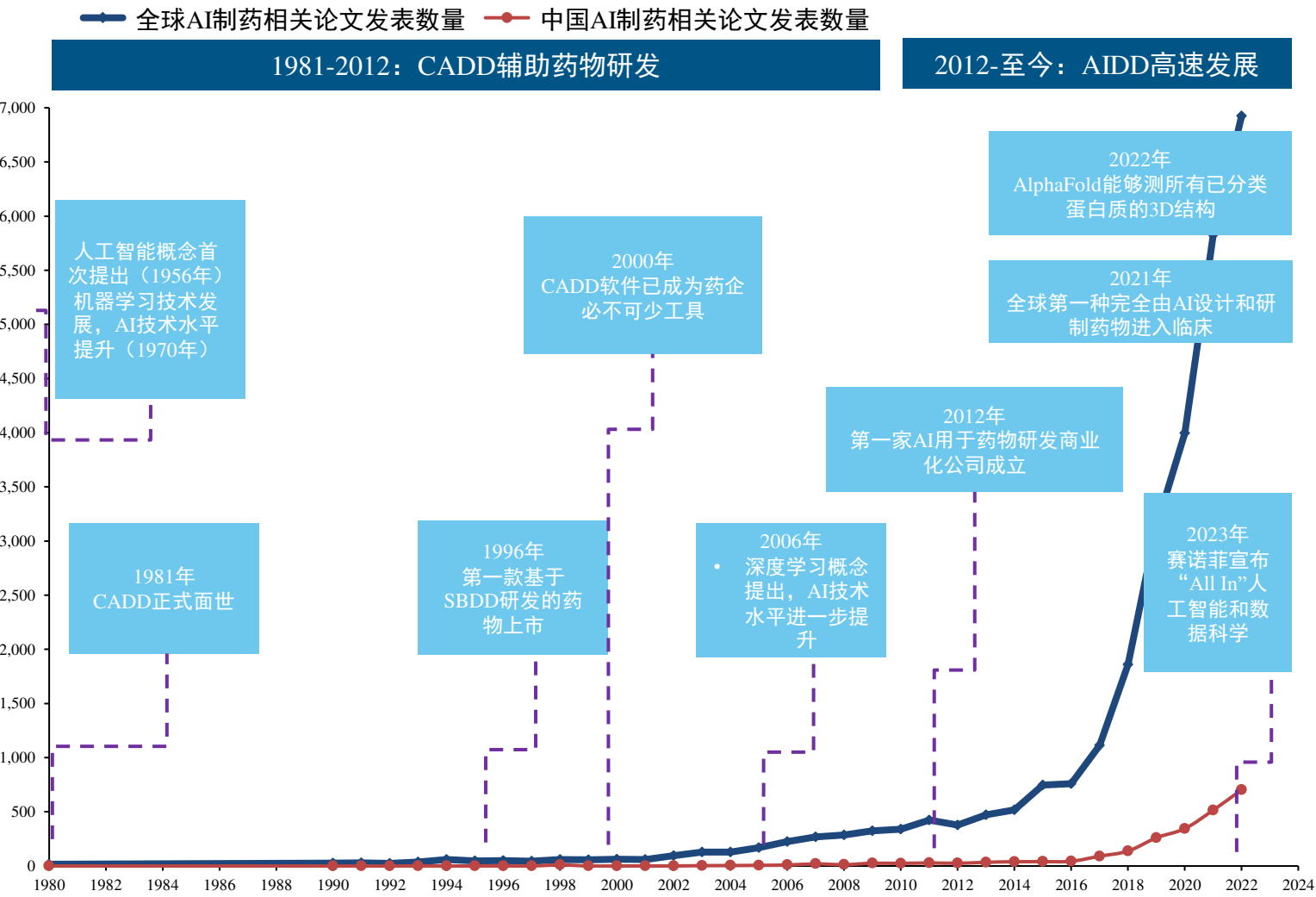
图表1 AI技术在药品全周期中的主要应用



随着AI技术的发展，AI技术基本实现了药物研发与市场化全流程的覆盖，特别是在药物发现和临床前阶段取得了显著进展。药物发现和临床前研发阶段是AIDD的主要优化环节，也是AI制药企业主要的研发及商业化方向。利用AI技术，企业能够在药物发现和临床前研发阶段，节约大量湿试验成本与时间成本，并实现高通量筛选，助力更多新靶点的发现与“老药新用”的适应症拓展。未来AI技术将持续发展，相信能够在药品临床开发、上市后生产与商业化等方面更进一步。

AI制药行业发展历史

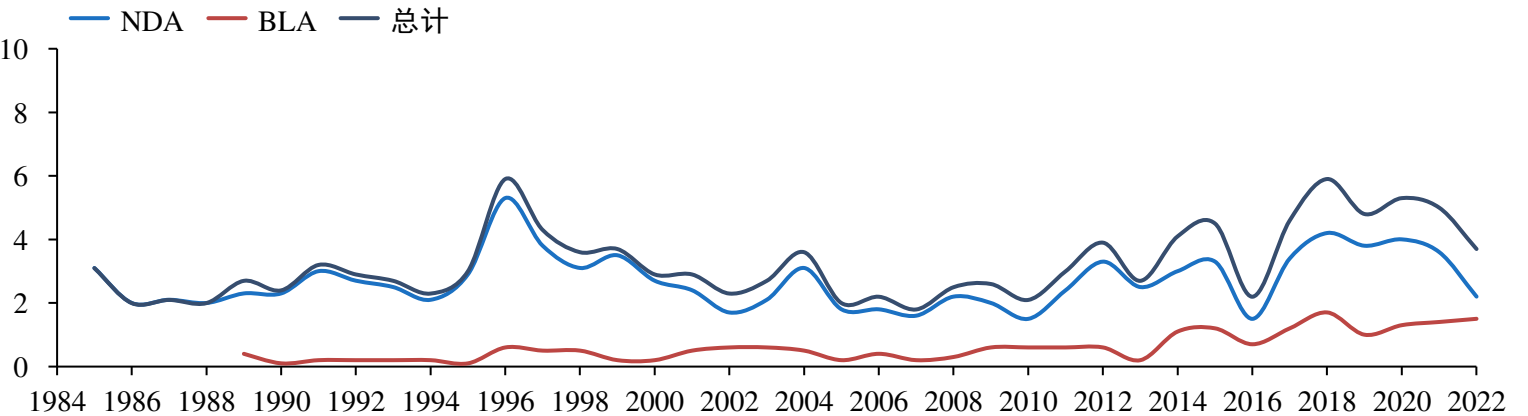
图表1 AIDD行业发展历程^[1]



纵观AIDD行业的发展，主要分为3个时期：

- 1956年~1981年的技术积累期，该时期是相关概念的早期构想阶段，但由于技术限制未能成功实施；
- 1981年~2012年计算机辅助药物设计（CADD）研发时期，CADD的虚拟筛选，可用于寻找有希望成药的苗头化合物进行药物开发，但CADD是药物发现过程的支持部门，而不是核心部门。CADD仅缩短药物发现时间，受限于工艺优化、临床、注册等环节，新药数量并未出现激增。

图表2FDA CDER批准新药数量统计

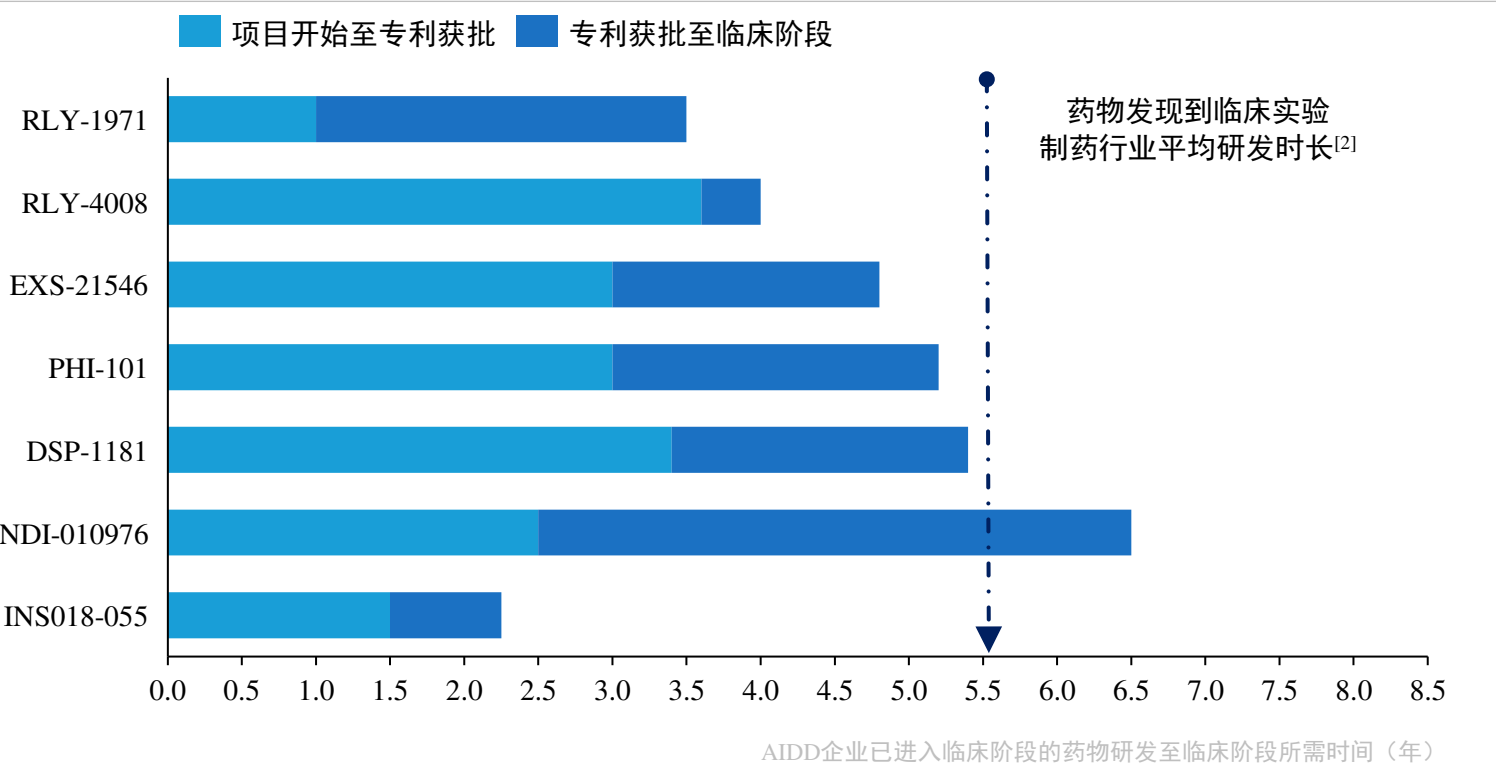


[1] 数据来源于Web of Science Science Citation Index Expanded 数据库

2012年~至今的AIDD高速发展期，该时期迎来的AI技术的“井喷式”发展，AIDD相关论文数量迎来“指数级”增长，相关技术栈逐步丰富完善。

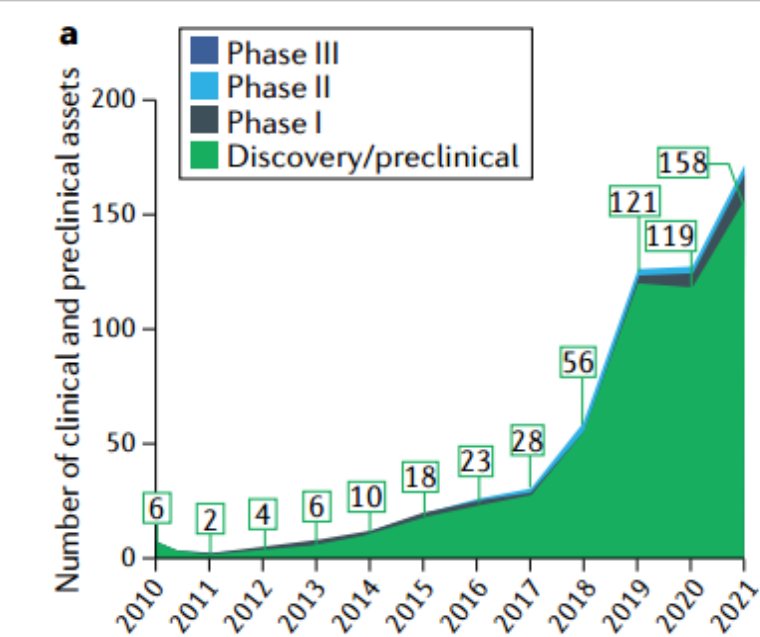
在药物发现及临床前研究阶段，基于从前的CADD技术，结合人工智能、机器学习和深度学习等技术，可以处理大量的生物数据，从而在药物开发全过程中减少时间和成本。

图表1 AIDD有效减少药物发现及临床前研究时间^{[1][2][3]}

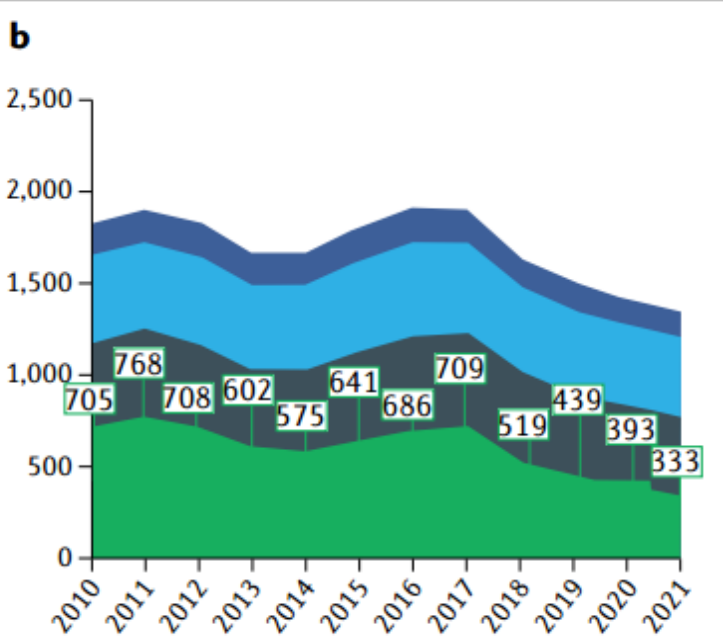


同时，区别于CADD研发时期主要聚焦于药物发现阶段，AI技术逐步应用于药物开发全流程。如：临床前研究中的剂型开发、工艺优化等；临床研究阶段的受试者筛选及结果预测；药物生产过程中的智能化生产管理、生产工艺优化。AIDD行业相对起步较晚，但管线数量逐年增加，头部AI药企临床前研究管线中药物数量已接近传统头部药企的50%，并已经有多款药物进入临床阶段。

图表2 全球主要24家AIDD企业研发管线数量汇总^[1]



图表3 全球TOP20 “BIG PHARMA” 企业研发管线汇总^[1]



[1] Jayatunga MKP, Xie W, Ruder L, Schulze U, Meier C. AI in small-molecule drug discovery: a coming wave?. Nat Rev Drug Discov. 2022;21(3):175-176.
[2] Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9(3):203-214
[3] 根据公开资料整理

■ AI制药行业配套政策

AIDD行业，是在近年才开始崭露头角的新兴领域，相比传统制药行业，它起步较晚。这主要是因为AI技术的发展和应用在医药领域的探索相对较新。然而，随着人工智能技术的迅猛发展和在医疗领域的广泛应用，AIDD行业正逐渐崭露出巨大的潜力。

尽管AIDD行业发展迅速，但相关的配套政策相对较少。由于AI技术的复杂性和涉及的伦理、法律等问题，政府和监管机构需要时间来理解和评估其潜在影响，以制定相应的政策和法规。美国、欧洲等国家及地区，行业起步较早，经过多年的发展及规划初版的监管政策于近年才推出。

图表1 国外AIDD行业相关监管政策（含草案）汇总

时间	部门	政策名称	AIDD相关政策内容
2023年5月	美国FDA	《Using Artificial Intelligence Machine Learning in the Development of Drug Biological Products》-讨论文件	涵盖 AI 和 ML 的当前和潜在用途，包括药物发现、临床和非临床研究、上市后安全性监测和高级药品上市。
2023年7月	欧洲EMA	《Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle》-草案	概述在整个药品生命周期和基于风险的背景下使用AI/ML开发与生产药物的考量，并针对AI/ML何时可以安全有效地开发、监管药物以及药物上市后使用等方面提供了建议。

目前，国内AIDD行业相关政策起始于“十四五”时期，主要以政府宏观政策为主，相应的执行发展政策及监管政策，还需要随着国内行业逐步深入发展而进行细化、完善。

图表2 国内AIDD行业相关宏观政策（含草案）汇总

时间	政策名称	AIDD相关政策内容
2022年1月	《“十四五”医药工业发展规划》	以新一代信息技术赋能医药研发。探索人工智能、云计算、大数据等技术在研发领域的应用，通过对生物学数据挖掘分析、模拟计算，提升新靶点和新药物的发现效率。在实验动物模型构建、药物设计、药理药效研究、临床试验、数据分析等环节加强信息技术应用，缩短研发周期、降低研发成本。推进健康医疗大数据的开发应用和整合共享，探索建立统一的临床大数据平台，为创新药研发及临床研究提供有力支撑。
2022年5月	《“十四五”生物经济发展规划》	信息技术支撑新药研制。利用云计算、大数据、人工智能等信息技术，对治疗适应症与新靶点验证、临床前与临床试验、产品设计优化与产业化等新药研制过程进行全程监管，实现药物产业的精准化研制与规模化发展。提升制药装备的自动化、数字化和智能化水平，发展基于人工智能的药物结晶、超临界萃取和色谱分离、固体制剂生产在线检测、连续培养生物反应器、蛋白质大规模纯化、冷链储存运输等信息化制药装备。
2022年7月	《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》	推动人工智能技术成为解决数学、化学、地学、材料、生物和空间科学等领域的重大科学问题的新范式，充分发挥人工智能技术在文献数据获取、实验预测、结果分析等方面作用，重点围绕新药创制、基因研究、生物育种研发、新材料研发、深空深海等领域，以需求为牵引谋划人工智能技术应用场景，融合人工智能模型算法和领域数据知识，实现重大科学问题和发现的研究突破。

各省都紧跟国家政策的方向，因地制宜出台了各地的特色政策。以AIDD行业产业布局最密集上海为例：

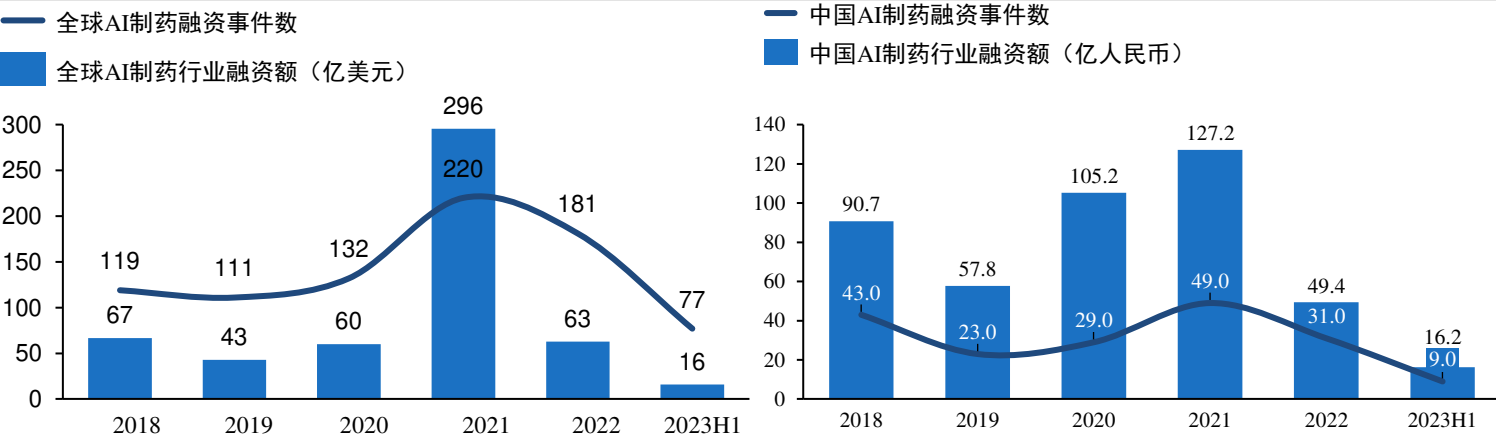
图表1 上海市部分AIDD行业产业政策

时间	政策名称	AIDD相关政策内容
2023年5月	《上海市计算生物学创新发展行动计划（2023—2025 年）》	1.发展人工智能助力的分子动力学和增强采样算法，开发我国独立自主的大分子结构预测与设计的算法和软件，构建大分子结构解析及其互作的分析计算流程和预测模型，构建大分子工程AI设计平台，指导大分子改造实验和结构优化。 2.发展AI药物研发的原创性算法、模型与计算平台，开发小分子药物大模型和生成算法，建立人工智能增强的抗体药物和肽类药物大模型与设计平台，发展中药复方药理药效预测模型，研发具有协同药效的现代复方。
2023年7月	《上海市生物医药产业数字化转型实施方案（2023—2025 年）》	瞄准研发、生产、流通、管理等产业重点环节，深化数字赋能。到2025 年，推动5个以上人工智能辅助研发药物进入临床，实现30 个以上高端智能医疗器械产品获批上市。 深化人工智能技术赋能医药创新研发；提升生物医药研发生产自动化智能化水平

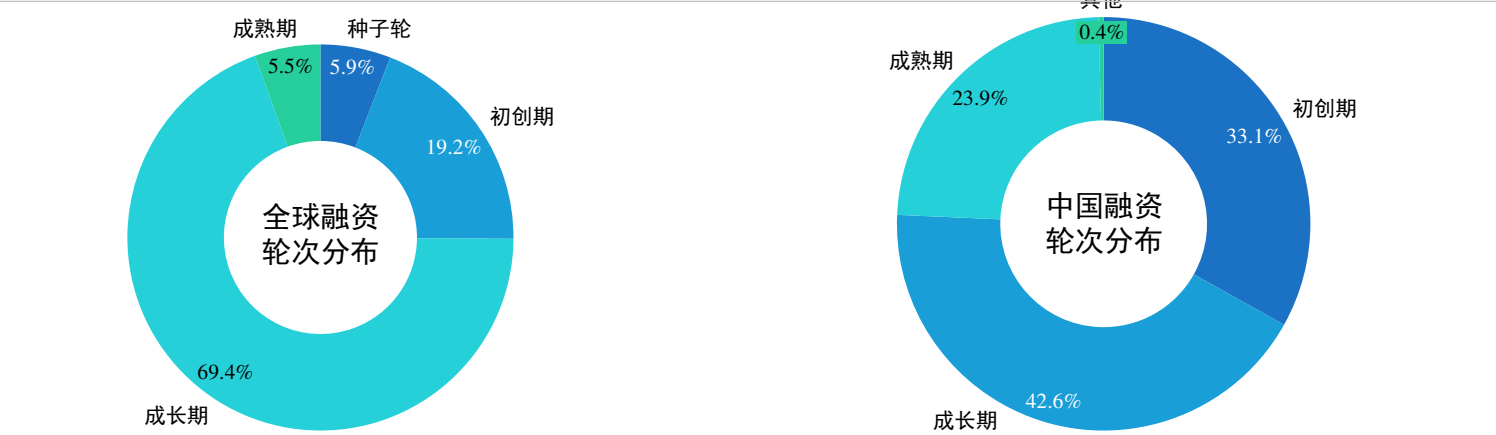
■ AI制药行业投融资现状

近年来，AIDD行业曾短暂成为投资的热点，其中2021年行业融资金额达历史高点，超过290亿美元。2022年以来，受资本环境影响，行业投资“热潮”已经有所退却，回落至相对稳定水平，预计融资金额将处于下行水平。AIDD资金回报周期长，全球投资者偏好选择成长期（69.4%）和初创期（19.2%）企业，中国投资者选择初创期、成长期、成熟期比例分别为33.1%，42.6%，23.9%^[1]。

图表2 AIDD行业投资金额统计分析



图表3 AIDD行业投资策略分布



[1] 整理自药融咨询，数据来源自药融云投融资数据库

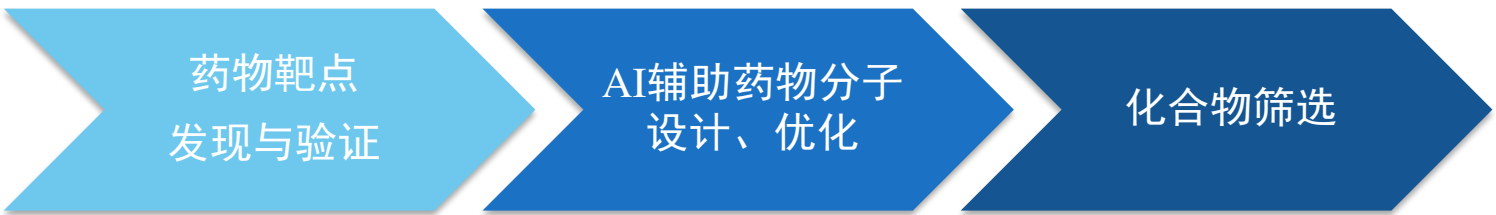
02

第二章

AI+药物发现

AI助力药物发现阶段概览

图表1 AI技术通过三方面助力药物发现阶段研发

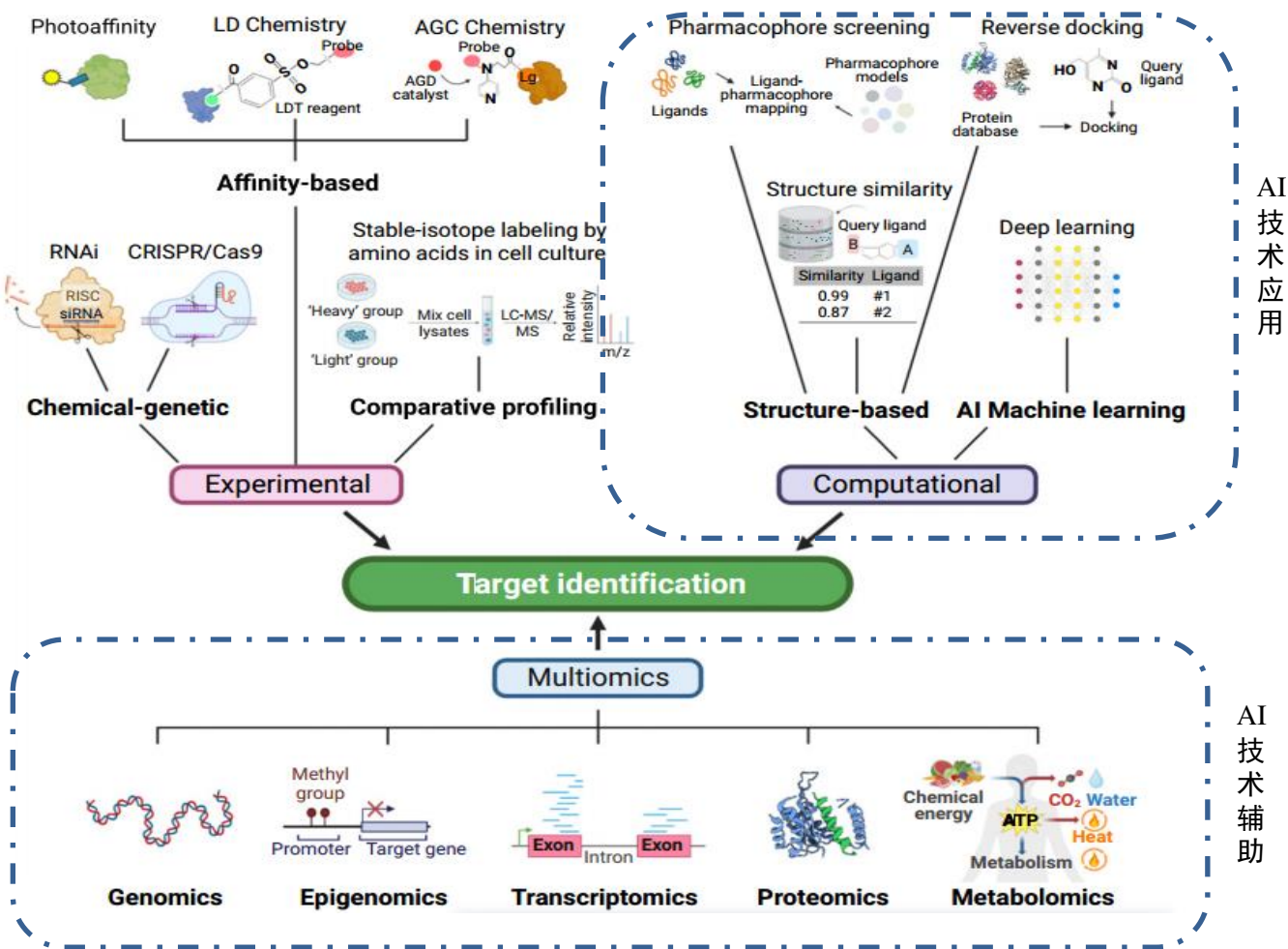


AI助力药物靶点发现与验证

随着微阵列、RNA-seq和高通量测序（HTS）技术的发展，高校、研究机构及企业实验室产生大量的生物医学数据，当代药物发现也因此进入了生物大数据时代。随着生物大数据分析，药物发现研究的重点逐渐转向靶向药物发现，即以“基因-药物-疾病”为中心，通过AI技术分析并疾病高度相关的靶点，发现酶、蛋白质或其他基因产物，以及针对该靶点的小分子。

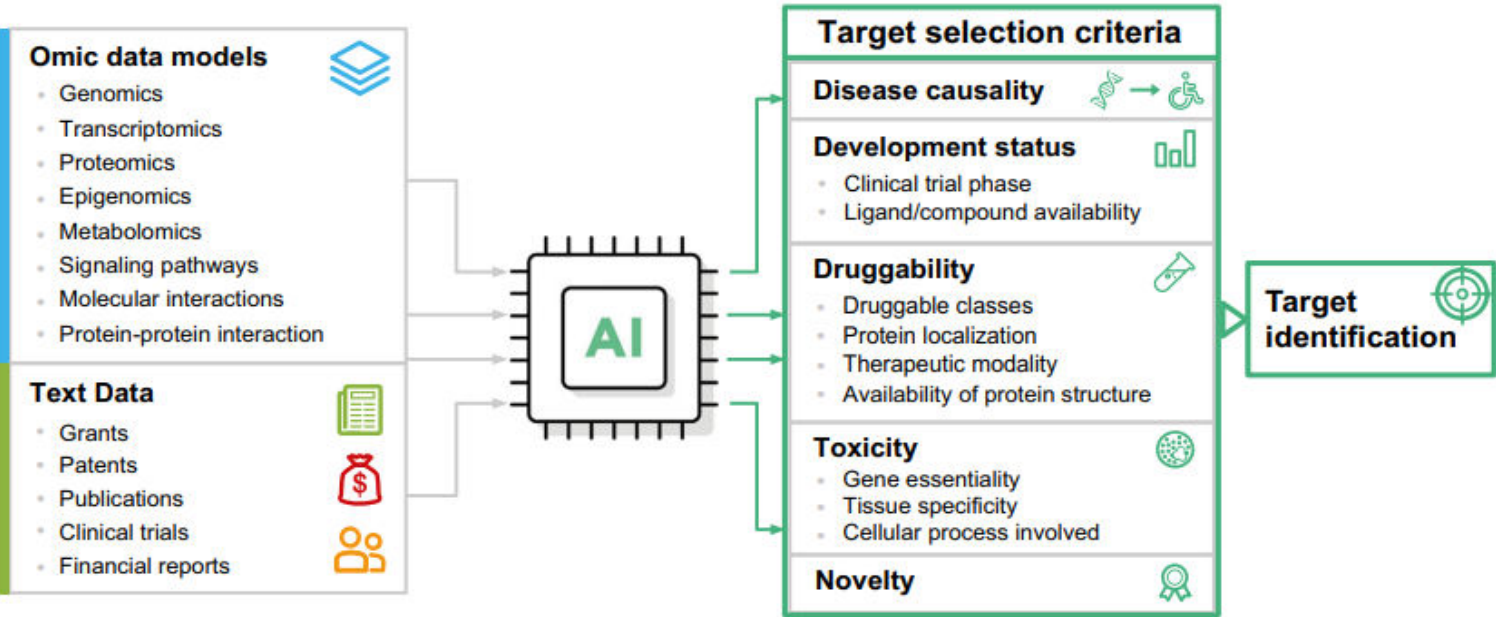
药物靶点发现策略包括实验方法、多组学分析方法和AI计算方法。实验方法受限于通量和成本，而基于AI技术的多组学分析方法和AI计算方法能够高效、有效地发掘具有潜力成为药物靶点的分子。多组学分析方法整合生物学数据，揭示疾病关键靶点，而AI计算方法利用机器学习和数据挖掘技术筛选化合物库，加速药物靶点发现。这些方法在药物研究和开发中具有重要作用，并在未来发展中将变得更加关键。

图表2 AI技术通过三方面助力药物发现阶段研发^[1]



[1] Pun, Frank W., Ivan V. Ozerov, and Alex Zhavoronkov. "AI-powered therapeutic target discovery." Trends in Pharmacological Sciences (2023).

图表1 AI技术驱动药物靶点发现与验证工作流程^[1]



■ 最新技术举例

■ 利用多组学数据训练机器学习/深度学习模型进行靶点发现：

近年来，基于机器学习的算法，特别是深度学习方法，在制药领域取得了显著成果并引起了广泛关注。深度学习，也被称为深度神经网络，由多个隐藏层节点组成，通过级联方式逐步进行数据处理和特征提取。与传统的机器学习方法相比，近期更加先进的基于深度学习的架构，如生成对抗网络（GANs）、循环神经网络和迁移学习技术，在制药方面广泛应用，例如新型小分子设计、衰老研究以及基于转录数据的药物药理预测。此外，利用公开可用的多组学数据和文本挖掘，深度学习还在未满足临床需求的疾病方面发挥着重要作用。

时间	研究机构/公司	相关论文	相关工作及发现
2018年	AgeX Therapeutics	Use of deep neural network ensembles to identify embryonic-fetal transition markers: repression of COX7A1 in embryonic and cancer cells	利用>12,000个胚胎和成年细胞的转录组数据训练了一个深度学习集成模型；发现控制胚胎转变的新靶点COX7A1。
2020年	利物浦大学/肯特大学	Using deep learning to associate human genes with age-related diseases	建立了一种基于深度学习的方法，该方法采用新颖的模块化架构，通过学习从基因或蛋白质特征中检索的模式，识别与多种年龄相关疾病相关的靶点。
2020年	云深智药	GCN-MF: Disease-Gene Association Identification By Graph Convolutional Networks and Matrix Factorization	提出了一种新的疾病-基因关联任务的框架，将图卷积网络（GCN）和矩阵分解相结合，命名为GCN-MF。借助其帮助，研究者可以捕捉非线性交互并利用测量的相似性。

[1] Pun, Frank W., Ivan V. Ozerov, and Alex Zhavoronkov. "AI-powered therapeutic target discovery." Trends in Pharmacological Sciences (2023).

■ 利用机器学习/深度学习模型进行分子建模

时间	研究机构/公司	相关论文	相关工作及发现
2021年	冰洲石生物	Molecular modeling with machine-learned universal potential functions	展示了神经网络可用于训练能量势函数的通用近似值。通过采用全自动训练过程，能够在大规模晶体结构上训练出平滑、可微分和可预测的势函数。
2021年	华盛顿大学 David Baker 团队	Accurate prediction of protein structures and interactions using a three-track neural network	报道了新蛋白结构预测工具 RoseTTAFold，其采用了注意力机制使整个DL能够同时学习到蛋白一级/二级/三级结构不同维度的信息，预测准确率与AlphaFold2不相上下。
2022年	DeepMind	AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models	真正解决了蛋白质三维结构预测的算法困境，大幅度降低了药物新靶点发现的难度，对于生命科学研究具有革命性的潜在意义。
2022年	百图生科	Helixfold-Single: Msa-Free Protein Structure Prediction By Using Protein Language Model As An Alternative	提出将大规模蛋白质语言模型（PLM）与 AlphaFold2 优越的几何学习能力相结合，并开发了新的蛋白结构预测模型 HelixFold-Single

■ 大语言模型通过文献挖掘发现治疗发现

基于大型语言模型的聊天功能，如微软的BioGPT和Insilico Medicineiv的ChatPandaGPT，通过对从数百万论文中提取的大量文本数据进行预训练，可以将疾病、基因和生物过程联系起来，从而快速识别疾病发生和进展的生物学机制，并发现潜在的药物靶点和生物标志物。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2023年	微软	BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining	BioGPT：针对特定领域的生成性 Transformer语言模型，在大规模生物医学文献上进行了预训练。文章在六个生物医学自然语言处理任务上评估了 BioGPT，并证明BioGPT在生物医学文献上的优势
2023年	英矽智能	ChatPandaGPT	Insilico Medicine 将 ChatGPT 引入其靶点发现平台 PandaOmic
2023年	麻省理工学院	Contrastive learning in protein language space predicts interactions between drugs and protein targets	团队设计了一种基于大语言模型的AI算法ConPLex，可以利用预先训练的蛋白质语言模型（PLex），并采用蛋白质锚定的对比共嵌入，超越目前最先进的算法，并尝试将药物靶点相互作用（DTI）推进到未知药物和靶点类型，同时又能区分具有不同结合特征的结构高度相似的药物候选分子。

■ 使用AI生成的合成数据进行靶点识别

“合成数据”是指通过人工生成的数据，它可以模拟真实世界中的模式和特征。借助人工智能算法，能够创建合成数据，以模拟各种生物数据，从而帮助研究人员探索和分析新药物的可能。这种方法在实验数据稀缺或难以获得的治疗领域尤其有价值。举例来说，在罕见疾病或患者数据有限的情况下，人工智能可以基于现有的知识和模式生成合成数据。然后，我们可以利用这些合成数据来训练人工智能模型，并识别可能被忽视的潜在治疗靶点。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2020年	英矽智能	Molecular Generation for Desired Transcriptome Changes With Adversarial Autoencoders	提出的模型可以产生新的分子结构，可以诱导给定的基因表达变化或在给定分子结构孵育后预测基因表达差异
2022年	剑桥大学	Adversarial generation of gene expression data	开发一种基于条件生成对抗网络的方法，为大肠杆菌和人类生成真实的转录组学数据。我们评估我们的方法在几种组织和癌症类型的表现

■ 计算生物学大模型用于靶点发现

时间	研究机构/公司	相关论文/成果	相关工作及发现
2023年	格拉德斯通研究所	Transfer learning enables predictions in network biology	基于迁移学习的Transformer模型开发了一个理解基因相互作用的基础模型

■ 国内企业布局

AI靶点发现需要大量的生物数据来训练模型，并需要专业人员对数据进行清洗、预处理，去除冗余和错误数据，并对数据进行标准化和格式化。对于AIDD初创企业而言，与大型制药企业相比，他们面临着生物实验场景数据严重不足的挑战。在短时间内难以积累足够的数据量。因此，AI靶点发现技术在初创企业中具有一定的门槛。

国内能够进行靶点发现的AIDD企业约32家^[1]，主要分布在北京、上海、深圳等地区，具体企业名称及布局技术领域如下表所示。值得注意的是，AI靶点发现技术具有一定的局限性，靶点发现的企业并非完全利用AI技术：例如，水木未来能够利用冷冻电镜技术解析蛋白质结构进而完成靶点的发现；药物牧场利用IDInvivo+转座子靶点发现平台，结合Medchem5人工智能药物化学平台在药物靶点发现上进行开拓。

能够进行靶点发现的AIDD企业主要分为3类：AIDD Biotech企业、具备AI技术的CRO企业、AIDD技术平台企业。

图表1 国内主要布局AI药物靶点发现与验证初创企业汇总^[1]

公司名称	公司位置	主要技术领域	企业类型	企业技术/投资背景
水木未来	北京	冷冻电镜	CRO	高校/研究所背景
英飞智药	北京	小分子药物	Biotech/CRO	高校/研究所背景
百奥智汇	北京	单细胞测序	CRO	高校/研究所背景
哲源科技	北京	生命科学	Biotech/CRO	高校/研究所背景
星亢原生物	北京	大分子药物	Biotech/CRO	名校博士创业项目
百图生科	北京	算法开发	技术平台	互联网巨头项目
普瑞基准	北京	多组学研发	CRO	创投基金孵化
角井生物	北京	大分子药物	Biotech	资深医药专家创业
智睿医药	广州	小分子药物	CRO	资深医药专家创业
西湖欧米	杭州	蛋白质组	CRO	高校/研究所背景
西湖云谷智药	杭州	基因编辑	Biotech	高校/研究所背景
高维医药	杭州	药物发现	Biotech	高校/研究所背景
呈元科技	杭州	合成肽新药研发	Biotech	创投基金孵化
红云生物	南京	小分子药物	Biotech	高校/研究所背景
燧坤智能	南京	小分子药物	Biotech/CRO	高校/研究所背景
劲风生物	青岛	TILs疗法	Biotech	高校/研究所背景
冰洲石生物	上海	小分子药物	Biotech	名校博士创业项目
药物牧场	上海	小分子药物	Biotech/CRO	资深医药专家创业
爱信智耀	上海	药物研发	Biotech	资深医药专家创业
湃隆生物	上海	小分子药物	Biotech/CRO	创投基金孵化

[1] 公开资料整理

公司名称	公司位置	主要技术领域	企业类型	企业类型
科镁联	上海	寡核苷酸药物	Biotech	创投基金孵化
焕一生物	上海	免疫学	CRO	跨界创业项目
新合生物	深圳	RNA药物	Biotech/CRO	名校博士创业项目
未知君	深圳	肠道微生物药物	Biotech	名校博士创业项目
溪砾科技	深圳	靶向RNA小分子	Biotech	创投基金孵化
华为云EIHealth	深圳	算法开发	技术平台	互联网巨头项目
云深智药	深圳	算法开发	技术平台	互联网巨头项目
科辉智药	深圳	小分子药物	Biotech	创投基金孵化
予路乾行	苏州	分子模拟	CRO	资深医药专家创业
科因生物	苏州	小分子药物	Biotech	创投基金孵化
睿健医药	武汉	诱导多能干细胞	Biotech	资深医药专家创业

■ 国内代表企业



■ 英矽智能

英矽智能是一家由生成式人工智能驱动的临床阶段药物研发公司，通过下一代人工智能系统连接生物学、化学和临床试验分析，利用深度生成模型、强化学习、转换模型等现代机器学习技术，构建强大且高效的人工智能药物研发平台，识别全新靶点并生成具有特定属性分子结构的候选药物。英矽智能聚焦癌症、纤维化、免疫、中枢神经系统疾病、衰老相关疾病等未被满足医疗需求领域，推进并加速创新药物研发。

英矽智能AI靶点发现技术全球领先，具有大量应用实例。自主研发人工智能生物靶点发现平台 PandaOmics™，公司利用该平台确定了许多此前从未报道过的肌萎缩侧索硬化症（ALS）潜在治疗靶点^[1]。



■ 药物牧场

药物牧场成立于2015年，主要利用遗传学和人工智能技术开发针对乙肝、肿瘤和自身免疫疾病的创新疗法。利用IDInvivo+转座子靶点发现平台，结合Medchem5人工智能药物化学平台在药物靶点开拓的“卡脖子”领域攻关突破。前者基于转座子进行基因筛选，在具有完整免疫系统的活体动物中直接发现新药物靶点；后者利用深度学习技术进行药物分子的设计。

药物牧场已经发现20余个创新靶点；结合专有的人工智能（AI）药物发现平台，公司原创新药DF-006已经进入I期国际多中心临床。



■ 百图生科

百图生科是中国首家生物计算技术驱动的生命科学平台公司，致力于用高性能生物计算和多组学数据技术加速创新药物和早筛早诊等精准生命科学产品的研发。

业务将分为两大阶段，第一阶段利用前沿AI技术构建完整的生物计算平台，并与提供新的数据轴和新的数据分析、药物设计工具的初创企业与研究机构携手，构建生物计算生态，为生命科学企业和科研用户提供丰富的工具能力和完整的解决方案。第二阶段，将深度参与或主导发起新型精准药物和精准诊断产品的研发，携手合作伙伴，研发生命科学产品。百图生物成果入选美国癌症研究协会（AACR）2023 年会，研究内容覆盖全新抗体设计的蛋白结构预测模型、靶点发现^[2]，以及高通量自动化蛋白表达/表征平台的构建。



■ 予路乾行

予路乾行是一家以人工智能、量子力学及分子模拟算法为核心技术，通过模拟运算赋能新药研发的交叉学科技术驱动型生物医药公司。公司结合自主创新的AI技术，结合高精度分子动力学模拟核心算法，建立了国际领先的新一代人工智能药物研发平台。该AI研发平台主要针对未解析靶点与高柔性靶点进行高精度结构解析及生理动态过程还原模拟，从而实现难成药靶点的新药研发。

[1] Identification of Therapeutic Targets for Amyotrophic Lateral Sclerosis Using PandaOmics – An AI-Enabled Biological Target Discovery Platform

[2] Identification of Molecular Targets 2 / New Nonclinical Models for Targets

■ AI驱动药物分子设计、优化

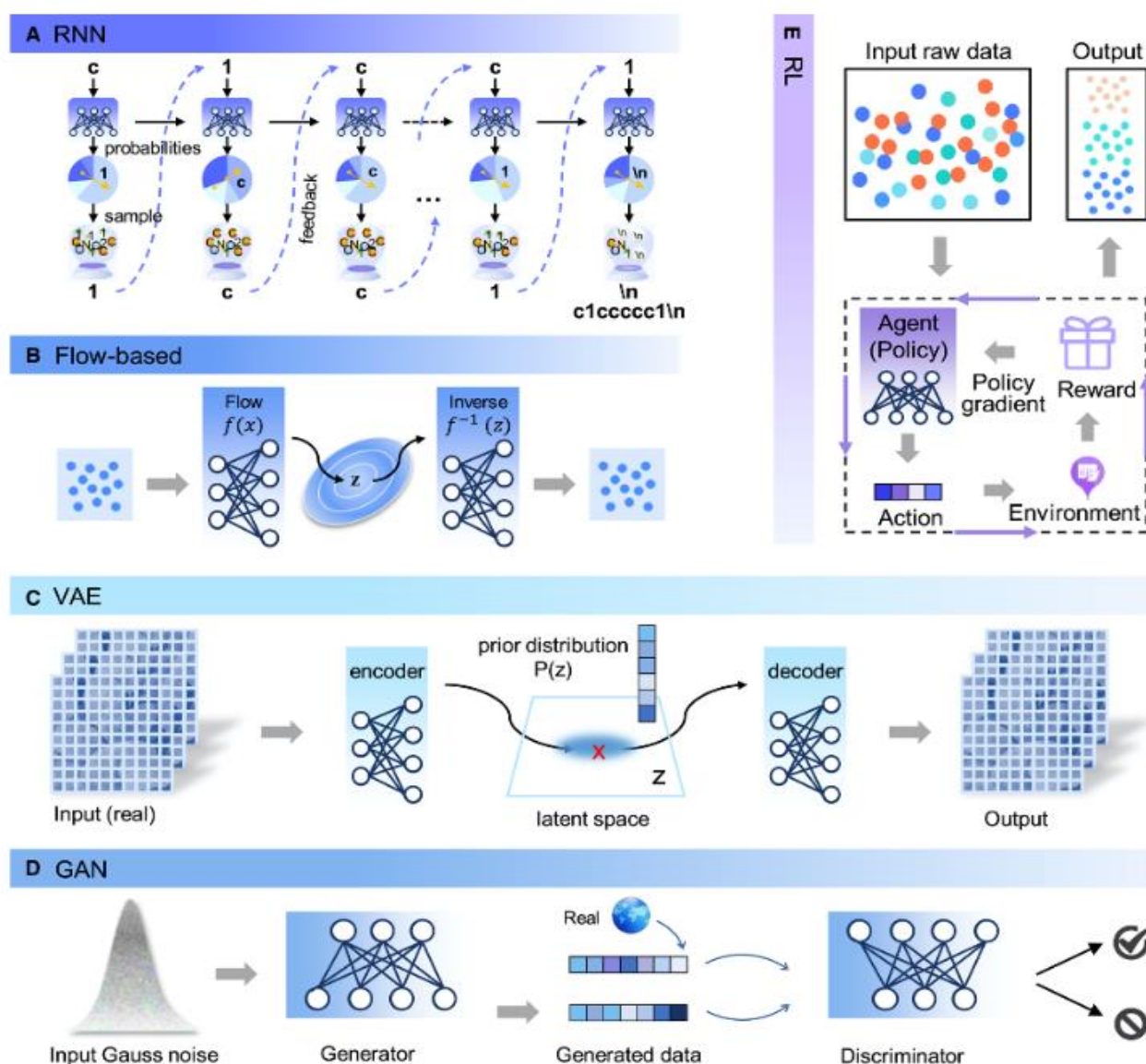
药物设计的目标是开发出具有特定性质、符合特定标准的药物，包括疗效、安全性、合理的化学和生物特性以及结构的新颖性。近年来，利用深度生成模型和强化学习算法进行新药设计被认为是一种有效的药物发现方法。这种方法能够绕过传统基于经验的药物设计模式的限制，让计算机通过自主学习药物靶点和分子特征，以更快的速度和更低的成本生成符合特定要求的化合物。这种新方法为药物设计领域带来了更多的创新和可能性。

■ 最新技术举例

■ AI生成模型在药物设计领域应用

AI生成模型在全新药物设计领域具有巨大的优势，因为它们在生成分子时不需要先验的化学知识。这些模型可以在更广阔的未知化学空间中进行搜索，超越现有分子结构的限制，自动设计新的分子结构。在新药设计中，广泛使用的深度生成模型包括基于循环神经网络（RNN）的生成模型、变分自编码器（VAE）和生成对抗网络（GAN）等。利用生成模型设计分子的过程具有高度的随机性，生成的分子在结构上呈现千变万化，质量也参差不齐。强化学习可以通过微调模型参数进行有针对性的优化，使生成的分子具备特定的药物分子特性。这种结合生成模型和强化学习的方法为全新药物设计提供了新的途径和可能性。

图表1 主要AI生成模型框架示意^[1]



[1] Zeng, Xiangxiang, et al. "Deep generative molecular design reshapes drug discovery." Cell Reports Medicine (2022).

■ 简单AI生成模型：基于循环神经网络（RNN）的生成模型设计药物分子

基于RNN的生成模型可以生成具有全新结构的化合物，同时保持与样本化合物具有相似的生化特性。训练过程首先利用大型化学数据库对RNN模型进行训练，使其学习如何生成正确的化学结构。然后使用强化学习算法对 RNN 参数进行微调，使模型能够将生成的化学结构映射到指定的化学空间。强化学习使基于 RNN 的生成模型能够生成具有良好药理特性的新分子，并确保生成分子的结构多样性。

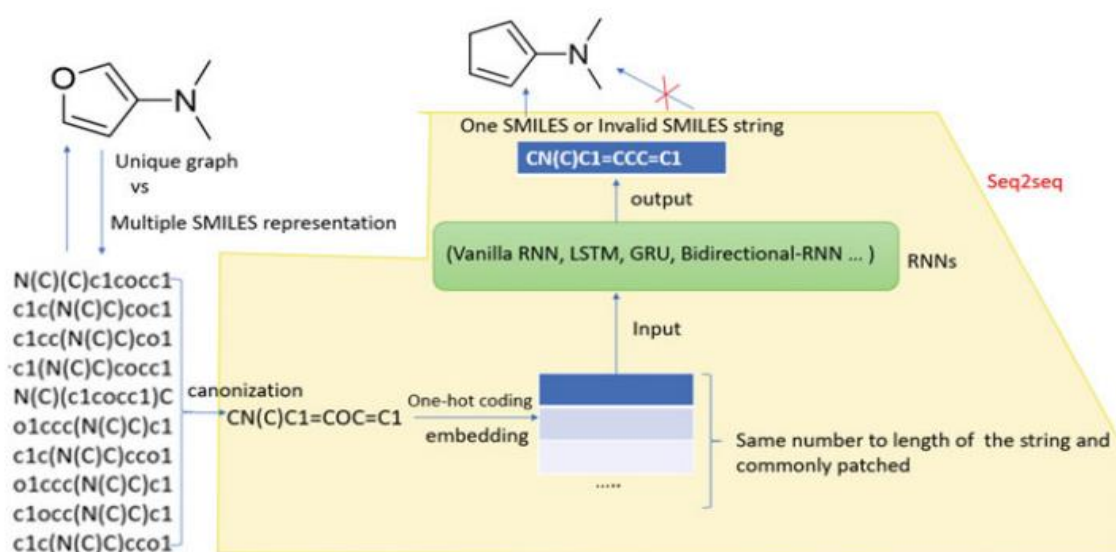
时间	研究机构/公司	相关论文/成果	相关工作及发现
2017年	阿斯利康	Molecular de-novo design through deep reinforcement learning	开发了一种基于序列的新药设计方法 REINVENT。研究者从 ChEMBL 数据库中收集了 150 万个符合特定要求的分子，并使用这些分子的 SMILES 来训练 RNN 模型，以学习活性分子的特征并生成新分子。

■ 简单生成模型：基于循环神经网络（RNN）的生成模型的先导化合物优化

基于 RNN 的生成模型药物在先导化合物的优化方面发挥着作用。研究者提出了一种名为“Scaffold-constrained molecular generation (SAMOA)”的新分子生成算法，用于解决先导化合物优化问题中的支架约束问题：使用 RNN 生成模型生成新分子的 SMILES 序列，然后使用精细采样程序实现支架约束并生成分子。此外，还应用了一种基于策略的强化学习算法来探索相关的化学空间，并生成与预期相匹配的新分子。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2019年	阿斯利康	Deep Reinforcement Learning for Multiparameter Optimization in de novo Drug Design	研究者开发了DeepFMPO 框架，能够从初始先导化合物开始，通过替换部分先导分子片段来修改这些先导分子的结构。这项研究证实了基于 RNN 的生成模型在分子优化领域的广泛应用。

图表1 RNN分子生成模型主要工作流程^[1]



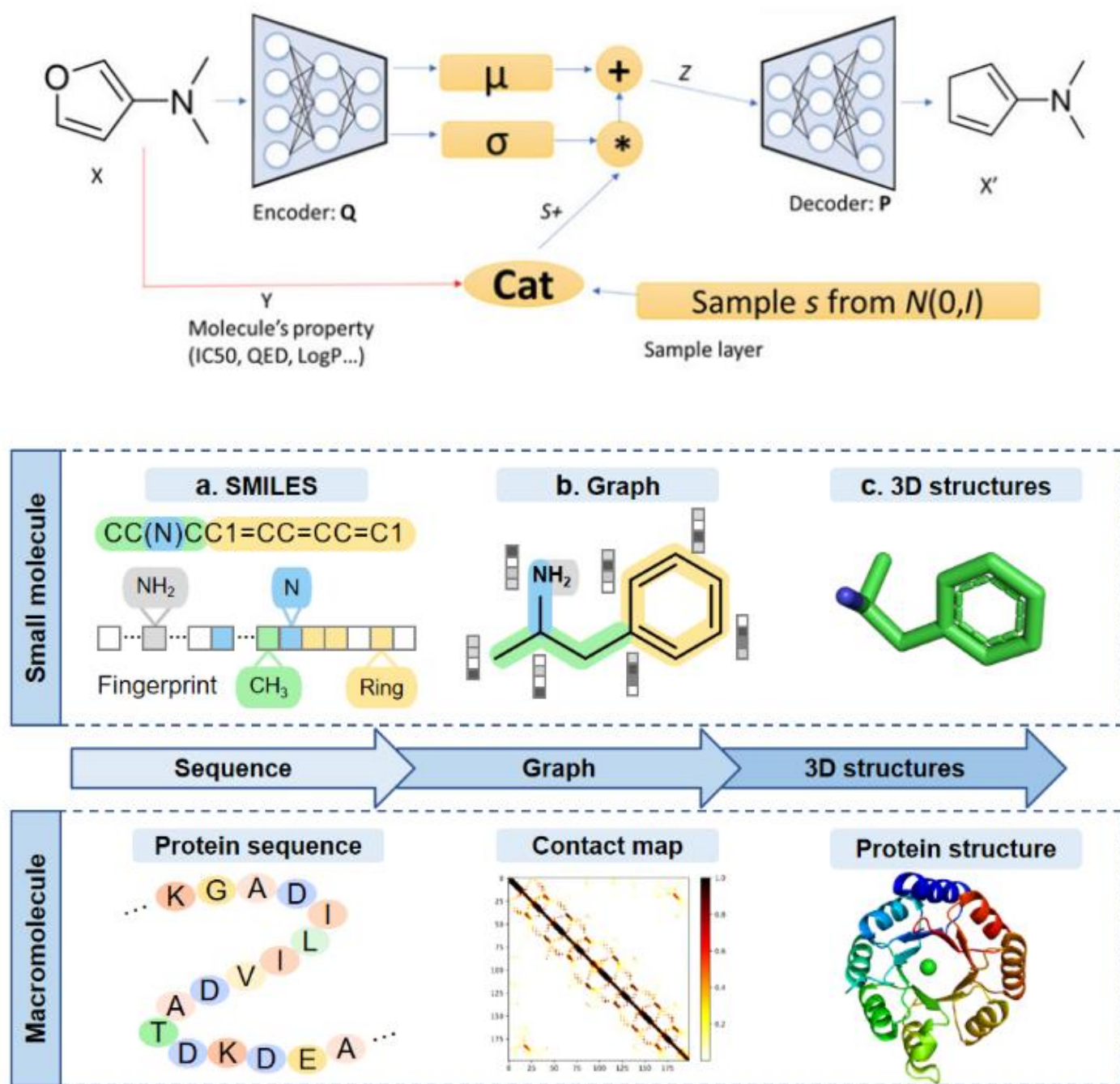
[1] Tang, Bowen, John Ewalt, and Ho-Leung Ng. "Generative AI models for drug discovery." Biophysical and Computational Tools in Drug Discovery. Cham: Springer International Publishing, 2021. 221-243.

■ 进阶AI生成模型：基于变分自动编码器（VAE）的生成模型

VAE已成为复杂数据分布的无监督学习最流行的方法之一。VAE模型能够完成小分子的全新设计和肽序列的生成。VAE有许多不同的变体，但其核心架构始终不变（如下图所示）。

VAE的变体主要可以分为SMILES-VAE、Graph-VAE和3Dgrid-VAE三种类型。SMILES-VAE是应用最为广泛，因为其容易实现，例如英矽智能构建的GENTRL算法。尽管GENTRL生成的分子顺利通过实验验证，但仍难以保证生成100%验证的SMILES。Graph-VAE和3Dgrid-VAE能够解决这些问题，由于3D结构训练数据有限，目前商业化应用较少。3D技术将对未来的AIDD项目具有巨大价值。

图表1 VAE模型的核心架构^[1-2]



[1] Zeng, Xiangxiang, et al. "Deep generative molecular design reshapes drug discovery." Cell Reports Medicine (2022).

[2] Tang, Bowen, John Ewalt, and Ho-Leung Ng. "Generative AI models for drug discovery." Biophysical and Computational Tools in Drug Discovery. Cham: Springer International Publishing, 2021. 221-243.

时间	研究机构/公司	相关论文/成果	相关工作及发现
2019年	英矽智能 药明康德	Deep Reinforcement Learning for Multiparameter Optimization in de novo Drug Design	开发了一种深度生成模型--生成式张量强化学习（GENTRL），用于全新小分子设计。GENTRL 可优化合成可行性、新颖性和生物活性。利用该模型在 21 天内发现了4种DDR1抑制剂。抑制剂在体内/体外实验均得到验证。
2020年	麻省理工学院	Hierarchical Generation of Molecular Graphs using Structural Motifs	报道了Graph-VAE相关工作，提出了 JT-VAE，其将在化学上有效的分子子结构视为图结构中的一个节点，因此可在不显式加入化学规则的情况下实现分子有效性达到 100% 的效果。解决了部分SMILES-VAE算法的局限性。
2020年	麻省理工学院	Multi-Objective Molecule Generation using Interpretable Substructures	RationaleRL 训练了一个基于图的强化学习模型，将预先选择的分子子图完成为一个整体分子，该分子具有多种属性，如针对多个靶点的生物活性（如 GSK3 β 和 JNK3）、药物相似性的定量估计和合成可及性。
2021年	云深智药	A Novel Scalarized Scaffold Hopping Algorithm with Graph-Based Variational Autoencoder for Discovery of JAK1 Inhibitors	开发了一种基于图的变异自动编码器与高斯混杂隐藏空间（Graph-GMVAE），可以选择性地从给定的参考化合物中准确生成分子。利用Graph-GMVAE研究者生成了更强效的JAK1抑制剂，并在生化实验中得到验证。
2023年	华为 复旦大学	Q-Drug: a Framework to bring Drug Design into Quantum Space using Deep Learning	提出了一种名为 Q-Drug（药物的量子启发优化算法）的分子优化框架，该框架利用量子启发算法来优化离散二元域变量上的分子

■ 进阶AI生成模型：基于通用对抗性网络（GAN）的生成模型

GAN主要由生成器和鉴别器组成，它们相互博弈，互相对抗，通过对抗生成高质量的数据。GAN能够生成与真实数据分布相似的新样本，并能够使用网络来辨别网络，删除不需要的内容。在AIDD领域，GAN 通常与特征学习和强化学习等技术相结合，在蛋白质功能预测、小分子生成等方面发挥了重要作用。基于 GAN已经构建多种分子生成模型，如Mol-CycleGAN、ORGANIC、RANC。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2017年	哈佛大学	Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC)	ORGANIC 是著名的分子生成模型，已成为其他模型的比较基准模型。它结合了 GAN 模型和强化学习算法，可以生成新颖有效的分子。
2018年	英矽智能 莫斯科物理技术学院	Reinforced Adversarial Neural Computer for de Novo Molecular Design	基于生成对抗网络（GAN）范式和强化学习（RL），研究者开发了RANC的深度神经网络（DNN）架构，用于新型小分子有机结构的从头设计。
2020年	Ardigen	Mol-CycleGAN: A generative model for molecular optimization	研究者开发了Mol-CycleGAN模型，它能生成与原始化合物结构相似度高的优化化合物。在给定一个分子的情况下，Mol-CycleGAN模型会能够生成结构相似的分子，并对所考虑的属性值进行优化。

■ 进阶AI生成模型：基于对抗性自动编码器（AAE）的生成模型设计药物分子

AAE在VAE基础上增加了判别模型，对采样分子和真实样本进行区分，基于对抗的思想训练生成模型和判别模型，扩展了分子的生成空间，一定程度上弥补了VAE在生成分子时结构新颖性方面的缺陷。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2017年	英矽智能 Mail.Ru	The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology	首次将AAE应用于根据确定的参数生成新的分子指纹。AAE 使用二进制指纹向量和分子浓度作为输入和输出，使用了 NCI-60 细胞系的检测数据训练模型。AAE 用于筛选 PubChem 中的 7200 万种化合物，并成功筛选出具有潜在抗癌特性的候选分子。
2020年	英矽智能	Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery	构建了一种新的AAE模型，即条件AAE，其能够基于指定条件（如药物分子的靶标特异性、溶解性、可合成性等）生成相应的分子。其中，基于长短时记忆网络（LSTM）分别构建编码器和解码器，使用多层的FNN作为判别模型，判断采样分子是否符合真实数据分布以及是否具备所需的理化性质，并基于半监督学习方法优化模型。利用该模型成功发现了一种新型JAK3抑制剂。

■ 基于强化学习（RL）的生成模型设计药物分子

强化学习（RL）通过智能体在与环境的交互过程中，不断优化人工智能模型，以最大限度地提高预期的累积回报。与有监督学习和无监督学习不同，RL可以在不受人类先验知识和训练数据束缚的情况下达到惊人的表现,例如最著名的应用是AlphaGo。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2017年	阿斯利康	Molecular De-Novo Design through Deep Reinforcement Learning	首次将RL应用于药物发现领域，使用强化学习算法对生成的分子进行评分，以微调 RNN 参数，从而生成对特定靶点具有活性的新化合物。这种方法已经应用于几种不同的分子生成任务。
2019年	斯坦福大学 Google	Optimization of Molecules via Deep Reinforcement Learning	利用化学知识与强化学习技术，提出用于分子优化的框架）—MoIDQN。MoIDQN是第一次在分子生成中使用标准的强化学习框架。通过直接定义添加/缺失原子或化学键的分子修饰，MoIDQN确保100%的化学有效性。
2020年	北卡罗来纳大学	Deep reinforcement learning for de novo drug design.	构建了结构演化的增强学习（ReLeaSE）可用于产生针对单个期望性质或多个性质而优化的新型化合物的靶向化学文库。两种模型均采用RL方法进行训练，以使新化学结构的生成偏向具有所需物理和/或生物学特性的化学结构。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2022年	星药科技 中山大学	Accelerated rational PROTAC design via deep learning and molecular simulations	提出一种新颖的深度生成模型 PROTAC-RL，能够在低资源的化学环境下设计出合理的PROTAC，并且结合深度强化学习使得该模型最终能够设计出具有良好药物代谢动力学性质的 PROTAC。在短短49天内发现了具有高降解活性及高药代性能的新型先导化合物并完成湿实验验证。
2023年	上海交通大学	De novo drug design by iterative multiobjective deep reinforcement learning with graph-based molecular quality assessment	提出一种新颖的基于质量评估的全新多目标药物设计方法（QADD），该方法将迭代改进框架与新颖的基于图的药物潜力分子质量评估模型相结合。QADD 设计了一个多目标深度强化学习管道，用于迭代生成具有多种所需特性的分子，其中引入了一个基于图神经网络的模型，用于对药物潜力进行精确的分子质量评估，以指导分子的生成。应用 QADD 生成与生物靶蛋白 DRD2 结合的新型分子也证明了该算法的有效性。

■ 综合多种技术的AI药物分子小分子生成平台技术成熟并成功在相关工作中应用

时间	研究机构/公司	相关论文/成果	相关工作及发现
2017年	以色列海法理工学院	Prototype-Based Compound Discovery Using Deep Generative Models	提出了一种分子模板驱动的神经网络，该网络结合了 VAE、CNN 和 RNN，可生成与模板分子性质相似的化学结构，同时具有结构多样性。研究者发现，生成的分子中有效分子的比例显著提高。
2020年	英矽智能	Chemistry42: An AI-Driven Platform for Molecular Design and Optimization	推出用于从头设计和优化小分子的平台 Chemistry42，将多种最先进的生成 AI 算法与药物和计算化学方法连接起来，来生成具有优化特性的新型分子结构。
2023年	英矽智能	Discovery of Potent, Selective, and Orally Bioavailable Small-Molecule Inhibitors of CDK8 for the Treatment of Cancer	通过多模态生成式强化学习平台 Chemistry42 基于结构生成化学分子的药物设计方法，成功发现一款有效的、具有选择性的、口服CDK8抑制剂，有潜力用于癌症的治疗。

■ AI技术应用于大分子药物分子设计

目前，国内外的AI药企主要专注于技术相对成熟的小分子药物研发，并且已经有多款由AI设计或辅助优化的小分子药物进入临床阶段。然而，近年来，随着新药物形式的兴起，布局抗体药物、核酸药物、多肽药物、基因细胞治疗药物、合成生物学等领域的企业比例也在不断增加。这些新药物形式带来了新的治疗理念和手段，为疾病的治疗提供了更广阔的前景。AI技术在这些领域的应用也在逐渐发展，为新药物形式的研发和优化提供支持。随着技术的进一步发展和应用的推广，未来可以期待AI在大分子药物领域的应用为医药行业带来更多的创新和突破。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2021年	华盛顿大学 David Baker团队	De novo protein design by deep network hallucination	通过深度神经网络进行全新蛋白质设计。
2021年	晶泰科技 IDEA 研究院	基于AI的高通量高精度抗体结构预测模型idealFold(IDEa-xtALpi-Fold)	依托晶泰科技自主搭建的 XupremAb™ 下一代抗体发现平台和 IDEA 研究院的人工智能技术，将先进的 AI 算法与大分子药物研发专家经验有机结合，并运用于抗体结构预测中。
2022年	华盛顿大学 David Baker团队	Robust deep learning-based protein sequence design using ProteinMPNN	开发了一种新的深度学习工具：ProteinMPNN。ProteinMPNN借鉴了图像识别使用的神经网络，能够明确与特定结构相对应的序列，以更加精准、快速地设计蛋白质分子，将设计蛋白质的时间长度从“月”缩短至“秒”。
2022年	华盛顿大学 David Baker团队	Design of protein-binding proteins from the target structure alone	实现了在没有靶蛋白三维结构信息的情况下，设计能与其特定位点结合的蛋白。
2023年	华盛顿大学 David Baker团队	Top-down design of protein architectures with reinforcement learning	开发了一种基于强化学习的蛋白质设计软件，并证明了它有能力创造有功能的蛋白质。
2023年	华盛顿大学 David Baker团队	De novo design of protein structure and function with RFdiffusion	开发了一种能从头设计全新蛋白质的深度学习方法——RoseTTAFold Diffusion。该方法能生成各种功能性蛋白质，包括在天然蛋白质中从未见过的拓扑结构。
2023年	百图生科	基于跨模态大模型“xTrim”开发了AI Generated Protein平台	设置了3类功能模块，分别是Function to Protein Design、Protein to Protein Design，以及Cell to Protein Design

■ 国内企业布局

国内能够进行药物分子发现、设计、优化的AIDD企业60余家^[1]，主要分布在北京、上海、深圳、苏州、杭州等地区，具体企业名称及布局技术领域如下表所示。国内的药物分子设计、优化AI发展水平参差，具备药物分子从头设计的AI技术的企业数量较少，而大多数企业根据现有数据库进行/提供药物发现或“老药新用”服务。

图表1 国内主要布局AI药物分子发现、设计、优化初创企业汇总^[1]

公司名称	公司位置	主要技术领域	企业类型	企业技术/投资背景
中大唯信	北京	小分子药物	CRO/技术平台	高校/研究所背景
亿药科技	北京	小分子药物	Biotech	高校/研究所背景
英飞智药	北京	小分子药物	Biotech/CRO	高校/研究所背景
华深智药	北京	大分子药物	CRO	高校/研究所背景
百奥几何	北京	大分子药物	Biotech	高校/研究所背景
康迈迪森	北京	小分子药物	CRO	高校/研究所背景
分子之心	北京	蛋白质设计	技术平台	高校/研究所背景
星亢原生物	北京	大分子药物	CRO/Biotech	名校博士创业项目
深势科技	北京	新药研发	CRO/软件	名校博士创业项目
域新说	北京	蛋白质设计	技术平台	名校博士创业项目
深度智耀	北京	小分子药物	技术平台	资深医药专家创业
角井生物	北京	大分子药物	Biotech	资深医药专家创业
百图生科	北京	多组学/抗体	技术平台	互联网巨头项目
望石智慧	北京	小分子药物	CRO	跨界创业项目
佩德生物	成都	多肽药物	技术平台	高校/研究所背景
成都先导	成都	小分子药物/核酸药物	CRO	资深医药专家创业
智睿医药	广州	小分子药物	Biotech/CRO	资深医药专家创业
信华生物	广州	大分子药物	Biotech/技术平台	资深医药专家创业
费米子	广州	小分子药物	Biotech/CRO	跨界创业项目
西湖欧米	杭州	蛋白质组	CRO	高校/研究所背景
西湖云谷智药	杭州	基因编辑	Biotech	高校/研究所背景
赛得康生物	杭州	蛋白药物	Biotech	高校/研究所背景
高维医药	杭州	药物发现	Biotech	高校/研究所背景
剂泰医药	杭州	药物递送	Biotech	名校博士创业项目
德睿智药	杭州	小分子药物	Biotech/CRO	资深医药专家创业
呈元科技	杭州	合成肽新药研发	Biotech	创投基金孵化
碳硅智慧	杭州	药物研发	CRO/技术平台	跨界创业项目

[1] 公开资料整理

公司名称	公司位置	主要技术领域	企业类型	企业技术/投资背景
索智生物	嘉兴	小分子药物	Biotech	资深医药专家创业
红云生物	南京	小分子药物	Biotech	高校/研究所背景
燧坤智能	南京	小分子药物	CRO/Biotech	高校/研究所背景
宇道生物	上海	变构药物	Biotech/CRO	高校/研究所背景
寻百会生物	上海	抗体药物	Biotech	高校/研究所背景
数因信科	上海	新药研发	CRO	高校/研究所背景
圆壹智慧	上海	小分子药物	Biotech/CRO	高校/研究所背景
冰洲石生物	上海	小分子药物	Biotech	名校博士创业项目
宇耀生物	上海	分子胶药物	Biotech	名校博士创业项目
智峪生科	上海	蛋白质预测	技术平台	名校博士创业项目
元星智药	上海	抗衰老产品	Biotech	名校博士创业项目
药物牧场	上海	小分子药物	Biotech/CRO	资深医药专家创业
胜普泽泰	上海	多肽药物	CRO/CDMO	资深医药专家创业
锐格医药	上海	癌症/免疫药物	Biotech/CRO	资深医药专家创业
湃隆生物	上海	小分子药物	Biotech/CRO	创投基金孵化
科镁联	上海	寡核苷酸	Biotech	创投基金孵化
焕一生物	上海	免疫	CRO	跨界创业项目
莱芒生物	深圳	CAR-T疗法	Biotech	高校/研究所背景
阿尔法分子	深圳	新药研发	Biotech	高校/研究所背景
新槌生物	深圳	药物研发	Biotech/CRO	高校/研究所背景
晶泰科技	深圳	小分子药物	CRO	名校博士创业项目
新合生物	深圳	RNA药物	Biotech/CRO	名校博士创业项目
未知君	深圳	肠道微生物态	Biotech	名校博士创业项目
智药科技	深圳	小分子药物为主	CRO/技术平台	名校博士创业项目
星药科技	深圳	小分子药物	CRO	名校博士创业项目
华为云 EIHealth	深圳	药物研发	技术平台	互联网巨头项目
云深智药	深圳	药物研发	技术平台	互联网巨头项目
科辉智药	深圳	小分子药物	Biotech	创投基金孵化
双运生物	苏州	小分子药物	CRO	高校/研究所背景
医图生科	苏州	药物研发	CRO	高校/研究所背景
Aureka	苏州	大分子药物	Biotech	高校/研究所背景
阿尔脉生物	苏州	小分子药物	CRO/Biotech	高校/研究所背景

公司名称	公司位置	主要技术领域	企业类型	企业技术/投资背景
朗睿生物	苏州	小分子药物	Biotech	资深医药专家创业
予路乾行	苏州	分子模拟	CRO	资深医药专家创业
立妙达	苏州	小分子药物	Biotech/CRO	资深医药专家创业
科因生物	苏州	小分子药物	Biotech	创投基金孵化
腾迈医药	苏州	新药研发	CRO	创投基金孵化
睿健医药	武汉	多能干细胞	Biotech	资深医药专家创业
英矽智能	香港	小分子药物	Biotech/技术平台	资深医药专家创业

■ 国内代表企业

■ 云深智药



云深智药是由腾讯自主研发的AI驱动的药物发现平台。该公司开发的iDrug平台能够实现小分子的从头合成。现有的模型通过对ZINC drug-like中超过680万个小分子数据进行训练学习，来达到学习分子空间的目的。目前支持对319个Kinase和52个GPCR靶点进行分子生成。在分子生成这一过程中，腾讯的iDrug平台的算法能够针对不同靶点在分子空间中的映射进行针对性采样，从而生成新的可能对靶点有活性的分子。

■ 晶泰科技



晶泰科技是一家以智能化、自动化驱动的药物研发科技公司，为全球生物医药企业提供药物发现一体化解决方案。晶泰科技首创智能计算、智能实验和专家经验相结合的药物研发新模式，致力于打造三位一体的研发平台，助力客户缩短药物研发的周期，提高药物研发成功率。

晶泰科技具备小分子药物发现技术，基于智能计算、自动化实验和专家经验的“三位一体”研究模式，以XtalCryo靶点确证、XcelaHit分子设计、XceptionOp成药性优化和XcelDev可开发性优化四大技术平台覆盖新药研发从靶点到获得PCC的完整流程，提供一体化药物研发服务。此外，晶泰科技具备大分子药物发现技术，将AI算法与湿实验进行有机结合拥有从靶点识别到先导抗体药发现的全流程一体化能力。

■ 星药科技



星药科技是一家从临床需求出发、以AI为核心技术驱动药物研发的生物科技公司。

星药致力于使用人工智能驱动药物研发，布局未成药、难成药靶点的药物研发管线，解决未满足临床需求。以“AI驱动的研发管线”和“AI赋能的计算平台”为双循环，赋能药物研发产业。星药人工智能药物发现平台Pyxir®使用AI前沿算法，结合计算化学、药物化学和生物学的工具及经验，全方位攻克小分子药物早期研发流程中的疑难问题，快速发现活性高、成药性好且结构新颖的候选分子。星药的智能计算平台M1结合了AI与经典物理学原理，可快速准确地描述分子和蛋白间的相互作用、精确地计算目标药物分子与特定靶标的结合自由能，突破经典计算模拟方法，达到世界顶尖水平。

■ AI辅助药物分子筛选

除了传统的实验方法，计算机辅助药物设计（Computer-Aided Drug Design, CADD）自问世以来在药物研发阶段发挥了巨大的作用。药物-靶标相互作用（DTI）是药物发现的重要基础，准确有效的DTI预测能极大地助力药物研发，加速先导或苗头化合物发现。计算机模拟DTI并非全新技术，世面已有许多成熟的开源程序或商业化软件。将AI技术引入CADD中，能够一定程度提高分子筛选的效率与覆盖度。

近年来基于机器学习/深度学习的预测DTI的方法陆续被报道。以已知化合物的活性和结构信息作为训练数据，并对数据进行预处理和特征提取，通过选择合适的机器学习算法进行模型构建和训练。后续利用已构建的机器学习模型对新化合物进行预测和筛选，并结合化学规则和药物性质等要求对化合物进行优化和筛选。

■ 最新技术举例

■ 基于机器学习/深度学习的药物-靶标相互作用预测

时间	研究机构/公司	相关论文/成果	相关工作及发现
2019年	弗洛里达大学	DeepAtom: a framework for protein-ligand binding affinity prediction	构建了DeepAtom模型，用于预测药物-靶标亲和力。除了具备三维CNN模型的各种优势，DeepAtom模型同时通过深度可分离卷积解决了三维CNN模型参数过多的问题，并利用多个小的卷积核代替单个大卷积核，达到减少参数的同时增加网络复杂度的目的。该模型表现出较强的预测能力。
2020年	中国科技大学	The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology	首次提出利用GAN模型预测蛋白质残基-残基接触图，并在基准测试集上表现出不错的预测效果。该模型被命名为GANcon，GANcon通过对抗性学习策略训练生成模型和判别模型，最终能够生成接近真实数据分布的接触图。
2021年	四川大学	Layer-wise relevance propagation of InteractionNet explains protein-ligand interactions at the atom level	提出了一种多注意力模块MATT_DTI，首先通过相对自注意模块提取药物的化合物原子间联系，用CNN模块分别学习药物和靶标的隐含信息，最后通过多头注意力模块和全连接层提取相互作用信息并给出预测结果。
2022年	西湖大学 厦门大学 德睿智药	Pre-Training of Equivariant Graph Matching Networks with Conformation Flexibility for Drug Binding	首创研发了能够刻画蛋白质构象变化与亲和力预测的AI模型—ProtMD

■ 生物医药大模型助力药物分子筛选

时间	研究机构/公司	相关论文/成果	相关工作及发现
2022年	华为 上海药物所	PanGu Drug Model: Learn a Molecule Like a Human	开发了新的深度学习架构，称为盘古药物模型 PanGu Drug Model。它使用图到序列的不对称条件变异自动编码器，可以从两种表征中适当地描述分子，提高下游药物发现任务的性能。
2023年	百图生科 清华大学	xTrimoPGLM: Unified 100B-Scale Pre-trained Transformer for Deciphering the Language of Protein	百图生科与清华大学合作开发了一款名为xTrimoPGLM的蛋白质语言模型，该模型的参数量高达1000亿，是目前蛋白质领域首个达到这一规模的“通用大模型”。

国内企业布局

国内能够进行药物分子设计、优化的AIDD企业约55家，主要分布在北京、上海、深圳、苏州、杭州等地区，具体企业名称及布局技术领域如下表所示。

中国AIDD行业起步较欧美等国家略有滞后，虽然并不缺乏新药管线，但大多数管线仅处于早期研发阶段，具有临床管线的AIDD企业数量较少。由于交叉学科的特性，早期AIDD公司成立后一般需要经历漫长的团队建设、平台完善和技术验证的时期，自主设计分子并推进到后期临床的公司凤毛麟角。除“老药新用”外，目前国内研发进度进展临床试验最快的AI药物是英矽智能研发的ISM001-055，该药物的研发进度同样位于全球前列。

现阶段AI药企普遍数据生产能力和干湿实验结合能力不足，这限制了新平台的应用。AI药企在新治疗方式领域同样面临数据稀缺性和孤岛效应，而且生物药的数据生产成本更高，兼具不同药物形式和AI的人才储备更少，所以未来AI药企需要持续加强数据资产的挖掘和累积。

图表1 国内主要布局AI药物分子筛选初创企业汇总^[1]

公司名称	公司位置	主要技术领域	企业类型	企业技术/投资背景
中大唯信	北京	小分子药物	CRO/技术平台	高校/研究所背景
水木未来	北京	技术平台	CRO	高校/研究所背景
亿药科技	北京	小分子药物	Biotech	高校/研究所背景
英飞智药	北京	小分子药物	Biotech/CRO	高校/研究所背景
华深智药	北京	大分子药物	CRO	高校/研究所背景
百奥几何	北京	大分子药物	技术平台	高校/研究所背景
超维知药	北京	小分子药物	CRO/Biotech	高校/研究所背景
康迈迪森	北京	小分子药物	CRO	高校/研究所背景
分子之心	北京	蛋白质设计	技术平台	高校/研究所背景
星亢原生物	北京	大分子药物	CRO/Biotech	名校博士创业项目
深势科技	北京	新药研发	CRO/技术平台	名校博士创业项目
域新说	北京	蛋白质药物	技术平台	名校博士创业项目
深度智耀	北京	小分子药物	CRO/技术平台	资深医药专家创业
角井生物	北京	大分子药物	Biotech	资深医药专家创业
望石智慧	北京	小分子药物	CRO	跨界创业项目
佩德生物	成都	多肽药物	技术平台	高校/研究所背景
智睿医药	广州	小分子药物	Biotech/CRO	资深医药专家创业
信华生物	广州	大分子药物	软件/Biotech	资深医药专家创业
费米子	广州	小分子药物	Biotech/CRO	跨界创业项目
西湖欧米	杭州	蛋白质组	技术平台	高校/研究所背景
高维医药	杭州	药物发现	Biotech	高校/研究所背景
剂泰医药	杭州	药物递送	Biotech	名校博士创业项目
德睿智药	杭州	小分子药物	Biotech/CRO	资深医药专家创业
呈元科技	杭州	合成肽	Biotech	创投基金孵化

[1] 公开资料整理

公司名称	公司位置	主要技术领域	企业类型	企业技术/投资背景
碳硅智慧	杭州	技术平台	CRO/技术平台	跨界创业项目
索智生物	嘉兴	小分子药物	Biotech	资深医药专家创业
红云生物	南京	小分子药物	Biotech	高校/研究所背景
燧坤智能	南京	小分子药物	CRO/Biotech	高校/研究所背景
宇道生物	上海	变构药物	Biotech/CRO	高校/研究所背景
寻百会生物	上海	抗体药物	Biotech	高校/研究所背景
数因信科	上海	新药研发	Biotech	高校/研究所背景
圆壹智慧	上海	小分子药物	Biotech/CRO	高校/研究所背景
冰洲石生物	上海	小分子药物	Biotech	名校博士创业项目
宇耀生物	上海	分子胶药物	Biotech	名校博士创业项目
智峪生科	上海	蛋白质预测	CRO	名校博士创业项目
元星智药	上海	抗衰老产品	Biotech	名校博士创业项目
药物牧场	上海	小分子药物	Biotech/CRO	资深医药专家创业
锐格医药	上海	免疫药物	Biotech/CRO	资深医药专家创业
湃隆生物	上海	小分子药物	Biotech/CRO	创投基金孵化
阿尔法分子	深圳	新药研发	Biotech	高校/研究所背景
新樾生物	深圳	DEL赋能药物研发	Biotech/CRO	高校/研究所背景
晶泰科技	深圳	小分子药物	CRO	名校博士创业项目
新合生物	深圳	RNA药物	Biotech/CRO	名校博士创业项目
未知君	深圳	肠道微生物药物	Biotech	名校博士创业项目
智药科技	深圳	小分子药物为主	CRO/软件	名校博士创业项目
云深智药	深圳	新药研发	技术平台	互联网巨头项目
科辉智药	深圳	小分子药物	Biotech	创投基金孵化
中以海德	深圳	乙肝药物	Biotech	资深医药专家创业
双运生物	苏州	小分子药物	CRO	高校/研究所背景
医图生科	苏州	药物研发	CRO	高校/研究所背景
Aureka	苏州	大分子药物	Biotech	高校/研究所背景
阿尔脉生物	苏州	小分子药物	CRO/Biotech	高校/研究所背景
朗睿生物	苏州	小分子药物	Biotech	资深医药专家创业
予路乾行	苏州	技术平台	CRO	资深医药专家创业
立妙达	苏州	小分子药物	Biotech/CRO	资深医药专家创业
科因生物	苏州	小分子药物	Biotech	创投基金孵化
腾迈医药	苏州	药物研发	CRO	创投基金孵化
睿健医药	武汉	多能干细胞	Biotech	资深医药专家创业
英矽智能	香港	小分子药物	Biotech	资深医药专家创业

03

第三章

AI+临床前研发

■ AI助力药物临床前研发概览

AI在药物临床前研发侧重于非临床药理学、药动学和毒理学研究。药物的物理化学特性及其 ADMET 特性对于药代动力学和毒理学研究至关重要。候选药物的不良特性可能导致昂贵的药物开发阶段失败。利用AI技术通过对候选药物的相关特性进行早期评估，可以降低临床研究的失败率和损失。

■ AI辅助药物理化特性预测

候选药物的ADMET特性直接受其物理化学特性的影响，对于药物成功上市至关重要。例如，药物分子的溶解度会严重影响药物的制剂方法，而药物分子在不同pH条件下的ADMET特性则深受其电荷状态的影响。尽管并非所有具有潜在药物价值的苗头化合物最终都能成功上市，但通过对苗头化合物理化特性的研究，能够提供大量实验数据，完善AI预测模型，助力药物晶型的设计与改良，制剂的设计与优化，并为其他药物的设计提供经验与灵感。

- 最新技术举例
- AI技术辅助药物晶型预测

多晶型现象是一种物质可以存在多种不同晶体结构中的现象，对于化学药物分子，几乎所有固体药物都具有多态性。由于晶型的变化可以改变固体化学药物的许多物理性质和化学性质。药物发展史上，存在一些药物由于晶型问题而导致上市失败的情况。

晶型预测（CSP）是指给定分子的二维结构式通过计算模拟获得它的所有可能的稳定晶型。晶型预测在制药工业中具有重要意义。使用AI有效地动态配置药物晶型可以完全预测小分子药物的所有可能的晶型，与传统的药物晶型研发相比，制药公司不必担心缺少重要的晶型。此外，晶型预测技术大大缩短了晶体的发展周期，更有效地选择了合适的药物晶型，缩短了开发周期并且降低成本。现常见的方法或工具有Crystalline Sponge Method、Dimorphite、ChemML等。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2021年	阿斯利康 晶泰科技	Current State-of-the-art In-house and Cloud-Based Applications of Virtual Polymorph Screening of Pharmaceutical Compounds: A Challenging Case of AZD1305	AZD1305 实验中，发现常温下物理稳定性非常接近的两种晶型，并借助晶泰科技的晶型预测和稳定性评估技术对该多晶型现象进行系统评估，印证了实验结果。 对于晶型预测技术在辅助实验解决复杂多晶型筛选问题，帮助研究者快速判断是否筛选出了在多种条件下最稳定的固体形态等方面提供了积极的建议。
2023年	默克 晶泰科技	Effect of Polymer Additives on the Crystal Habit of Metformin HCl	晶泰科技通过分子动力学模拟考察了溶剂和添加剂与各个晶面间的相互作用，基于校正吸附能模型定量分析了不同添加剂品种对晶面生长的特异性影响，并成功预测出不同结晶环境中盐酸二甲双胍的晶体形态，预测结果与实验结果高度匹配。

■ AI技术辅助药物热力学溶解度预测技术

利用AI技术能够识别并预测新分子溶解度等理化性质，在确定研发成本或临床前终止研发至关重要。实验中测量溶解度存在较大的不确定性，导致测得的溶解度可能被高估。基于物理模型的第一性原理计算预测溶解度需要高额的计算成本，可行性较低。利用人工智能技术建立基于物理模型的热力学溶解度预测流程，实现了轻量级晶型预测技术，这对于预测分子的热力学溶解度具有重要意义。这些方法可以帮助提高溶解度预测的准确性，加速药物开发过程，并为决策提供更可靠的依据。

时间	研究机构/公司	相关论文/成果	相关工作及发现
2023年	AbbVie 晶泰科技	Novel Physics-Based Ensemble Modeling Approach That Utilizes 3D Molecular Conformation and Packing to Access Aqueous Thermodynamic Solubility: A Case Study of Orally Available Bromodomain and Extraterminal Domain Inhibitor Lead Optimization Series	晶泰科技通过分子动力学模拟考察了溶剂和添加剂与各个晶面间的相互作用，基于校正吸附能模型定量分析了不同添加剂品种对晶面生长的特异性影响，并成功预测出不同结晶环境中盐酸二甲双胍的晶体形态，预测结果与实验结果高度匹配。

■ AI技术辅助指导实验制备新晶型

时间	研究机构/公司	相关论文/成果	相关工作及发现
2020年	南安普敦大学	Minimizing Polymorphic Risk through Cooperative Computational and Experimental Exploration	将最先进的计算晶体结构预测（CSP）技术与广泛的实验结晶方法相结合，以了解和探索药物的晶体结构，并最大限度地降低后期出现意外多晶体的风险。

■ AI辅助药物剂型设计

■ 最新技术举例

时间	应用	相关论文/成果	相关工作及算法
2022年	预测药物释放	Impact Assessment of the Variables Affecting the Drug Release and Extraction of Polyethylene Oxide Based Tablets.	ANN, SVM, 回归树集合, 决策树
2021年	开发3D打印片剂	A machine learning method for defect detection and visualization in selective laser sintering based on convolutional neural networks	ANN、自组织映射、RF、SVM和CNN
2022年	检测片剂缺陷	Real-time coating thickness measurement and defect recognition of film coated tablets with machine vision and deep learning	CNN, YOLOv5
2022年	估算崩解速率	Disintegration testing augmented by computer Vision technology	RF, XGBoost, ANN, CNN
2022年	药物粒度检查	UV/VIS imaging-based PAT tool for drug particle size inspection in intact tablets supported by pattern recognition neural networks.	模式识别神经网络

时间	应用	相关论文/成果	相关工作及算法
2021年	设计吸入制剂	Machine learning approach for carrier surface design in carrier-based dry powder inhalation	RF, XGBoost, LightGBM, SVM, KNN, ANN, CNN
2020年	改善喷雾剂相容性	The application of machine learning algorithms in understanding the effect of core/shell technique on improving powder compactability	SVM , ANN
2022年	预测聚集程度	A Novel Computational Approach Coupled with Machine Learning to Predict the Extent of Agglomeration in Particulate Processes	SVM、RF和偏最小二乘回归
2022年	估算崩解速率	Disintegration testing augmented by computer Vision technology	RF, XGBoost, ANN, CNN
2022年	检测胶囊内颗粒缺陷	A micro-XRT image analysis and machine learning methodology for the characterisation of multiparticulate capsule formulations.	SVM
2017年	控制制粒过程	Artificial Intelligence Tools for Scaling Up of High Shear Wet Granulation Process	神经模糊逻辑与遗传编程
2016年	预测粒度分布	A Computational intelligence modeling of granule size distribution for oscillating milling.	多元线性回归和遗传编程
2021年	预测溶解速率和曲线	A novel AI-based computational platform for solid dispersion formulation design.	RF, SVM, LightGBM, XGBoost

■ AI助力药物ADMET 性质预测

■ 最新技术举例

ADMET（药物吸收，分配，代谢，排泄和毒性研究）是当代药物设计和药物筛选中十分重要的方法。临床试验的失败通常是由于药物的 ADMET 问题，而不是缺乏疗效。传统的ADMET研究非常耗时且昂贵，但同时又是能否成药一个很重要的条件。AI可提取化合物相关结构特征，评估多个ADMET参数间的关系和趋势，有效提升ADMET性质预测的准确性。

AI在药物ADMET预测的原理主要基于机器学习算法。这些算法使用了大量已知药物的ADMET数据来训练预测模型，并通过预测药物分子的化学和物理属性来推测其ADMET性质。数据的质量和数量是ADMET预测的巨大障碍。大多数预测模型由数百到数千个小型化学数据集组成，无法覆盖足够的化学空间。生物活性测定的数据受实验平台的影响，容易产生实验误差，导致后续的数据管理产生困难。当前数据驱动的ADMET预测研究依赖于文献衍生数据及公开数据库，数据量稀缺且数据质量参差不齐。这既考验公司对数据管理能力，也考验模型的泛化能力。

■ 定量结构活性关系（quantitative structure–activity relationship, QSAR）模型预测ADMET性质

从人工智能中受益的药物发现的关键领域之一是ADMET性质预测，通过定量结构活性关系（quantitative structure–activity relationship, QSAR）模型来预测多种性质，从简单的物理化学性质到复杂的药代动力学（PK）、药效学（PD）和毒理学特性。重要的PK端点包括清除率、通透性和稳定性；重要的药效学端点包括药物靶标特异性和选择性；重要的毒理学端点包括细胞色素P450诱导和hERG抑制等。

时间	相关论文/成果	预测性质
2012年	hERG K+ channels: structure, function, and clinical significance	人乙醚相关基因（hERG）介导的心脏毒性
2017年	New experimental models of the blood-brain barrier for CNS drug discovery	血脑屏障
1998年	The drug efflux protein, P-glycoprotein, additionally protects drug-resistant tumor cells from multiple forms of caspase-dependent apoptosis	渗透性糖蛋白
2020年	Large-scale evaluation of cytochrome P450 2C9 mediated drug interaction potential with machine learning-based consensus modeling	细胞色素 P450
2018年	Multi-Descriptor Read Across (MuDRA): A Simple and Transparent Approach for Developing Accurate Quantitative Structure-Activity Relationship Models	诱变性
2017年	ADMET evaluation in drug discovery. Part 17: development of quantitative and qualitative prediction models for chemical-induced respiratory toxicity	呼吸毒性
2018年	In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts	口服毒性
2016年	Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature	致癌性
2018年	Multi-descriptor read across (MuDRA): a simple and transparent approach for developing accurate quantitative structure-activity relationship models	致突变性
2017年	ADMET evaluation in drug discovery. Part 17: development of quantitative and qualitative prediction models for chemical-induced respiratory toxicity	刺激性
2020年	Towards deep neural network models for the prediction of the blood-brain barrier permeability for diverse organic compounds	经典的前馈反向传播神经网络（BPNN）结构和重复双重交叉验证（rdCV）方法被结合起来，用于估算血脑屏障渗透率。
2019年	Modeling physico-chemical ADMET endpoints with multitask graph convolutional networks	神经网络方法预测了七种不同的 ADMET 特性：logD、溶解度、熔点、膜亲和力和人血清白蛋白结合力。

时间	相关论文/成果	预测性质
2020年	Deep learning based drug metabolites prediction	开发了一种 DL 模型来预测药物代谢物，其准确性优于流行的基于规则的潜在代谢物系统生成方法（SyGMa）
2018年	Chemi-Net: A molecular graph convolutional network for accurate drug property prediction	冰洲石生物与安进合作，开发 Chemi-Net，完全由数据驱动、无需领域知识的深度学习 ADME 性质预测方法。

■ AI+临床前研发国内企业布局

AI技术在药物临床前研发阶段有着广泛的应用。主要包括晶型预测、剂型设计和药物ADMET性质预测等，其中能够完成药物ADMET性质预测企业数量最多。

由于技术发展时间长，相对技术成熟，市面具有多款药物ADMET性质预测CADD软件免费开源/商业化软件，随着AI技术的发展，相关软件逐渐添加了AI功能辅助数据的提取与处理，承接相关业务门槛低，因此相关企业数量较多。但多数企业仅停留在对相关软件的应用，并未有创新的核心算法与技术的拓展。

CADD的能力局限于对已有的知识和数据的筛选与优化，相较之下，AIDD能够充分探索未知的分子结构空间，对更多未知的分子性质进行预测，给药物研发带来了颠覆性变革，但AIDD的发展需要高数量、高质量的生物数据作为支撑。目前国内企业的技术重点在于药物ADMET性质预测，布局晶型预测、剂型设计等方面国内企业数量较少。

公司名称	晶型预测	剂型设计	药物ADMET性质预测
燧坤智能			√
亿药科技			√
西湖欧米			√
医图生科			√
阿尔脉生物			√
晶泰科技	√	√	√
冰洲石生物			√
剂泰医药		√	
智峪生科			√
深度智耀			√
智睿医药			√
索智生物			√
云深智药			√
百图生科			√
科辉智药			√
费米子			√
碳硅智慧			√

[1] 公开资料整理

■ 药物发现阶段/临床前研发阶段国内企业竞争格局

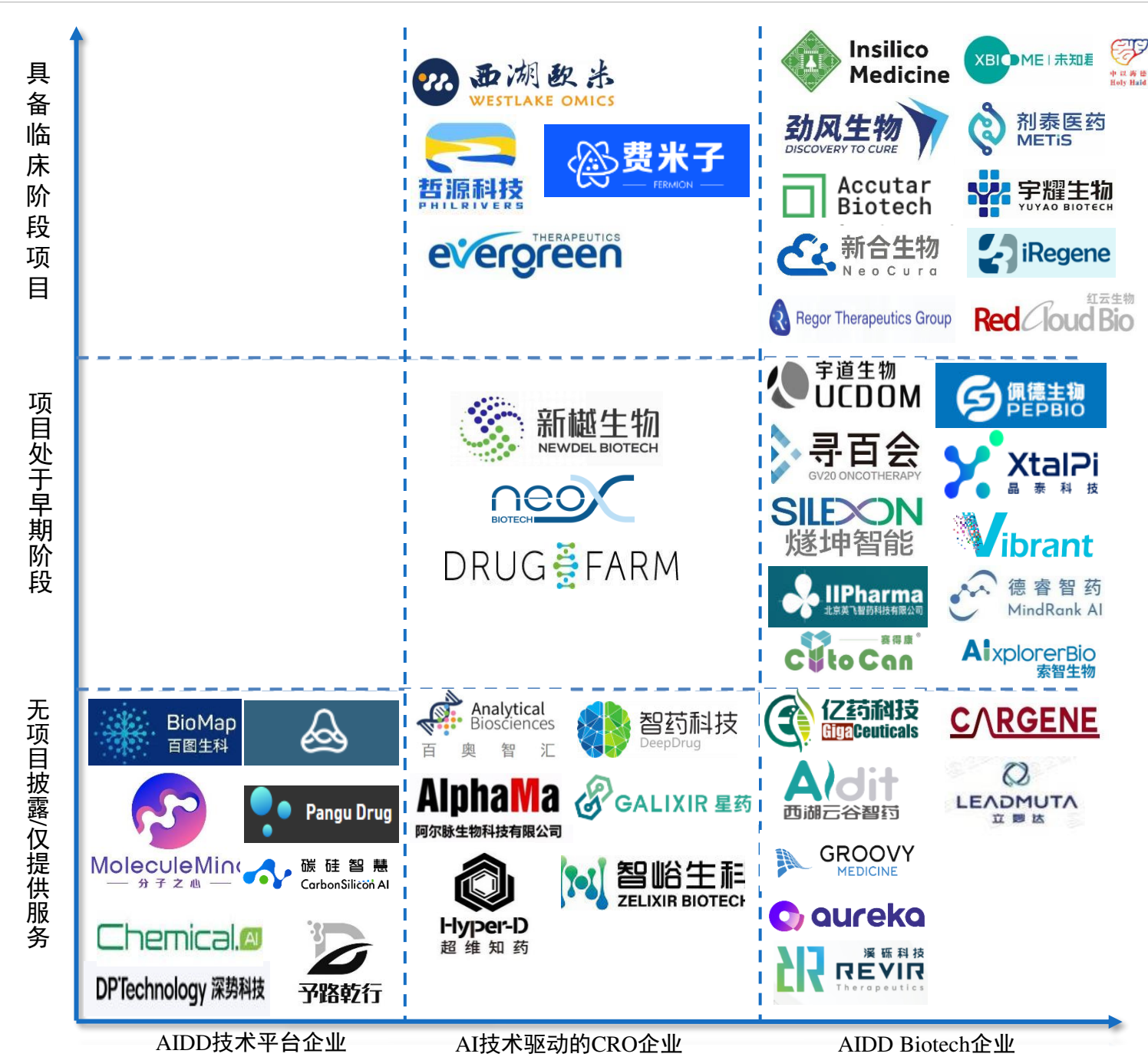
利用AI技术主导或辅助药物发现及临床前研发阶段是中国AIDD行业企业重点关注的技术，超过90%的行业企业在这方面有所投入。基于行业理解，我们将中国AIDD行业的企业主要分为AIDD Biotech企业、具备AI技术的药企/CRO企业、具备AI技术的药企/CRO企业、AIDD技术平台企业（分类不包括引进AI技术的大型制药企业）。

AIDD Biotech企业：即AI技术贯穿于新药发现全过程的企业，该类企业以新药管线研发为核心商业模式，此外部分Biotech企业与大型制药企业开展合作，承接药物筛选等服务工作。

AI技术驱动的CRO企业：该类企业以CRO服务为核心商业模式，该类型公司无研发管线/管线数量较少，主要通过对外提供研发服务作为盈利模式。此外，该类型的公司也通过对外授权/售卖早期管线实现营收。

AIDD技术平台企业：该类企业以软件工具服务为核心商业模式。

图表1中国主要AIDD行业企业格局^[1]



[1] 公开资料整理

AI新药研发的三要素是核心生物数据、基础算力和核心算法，AIDD技术平台企业基于本身在算法方面的优势，有望通过提供软件产品实现商业模式；平台型研发企业通过与大型药企/Biotech企业合作，完成项目的积累和自主研发能力边际的不断拓宽，在数据积累上会有较大优势，也有望衍生出AI新药研发的CRO企业，专注于建立平台提供服务，为不具备AI新药前端开发能力的企业提供支持；AI新药研发最终走的也是新药研发的逻辑，这也会催生一批在AI药物管线开发上具有突出能力的Biotech企业。

图表1中国主要AIDD行业与大型制药企业/Biotech企业合作关系网络^[1]



[1] 公开资料整理

04

第四章

AI+临床研究

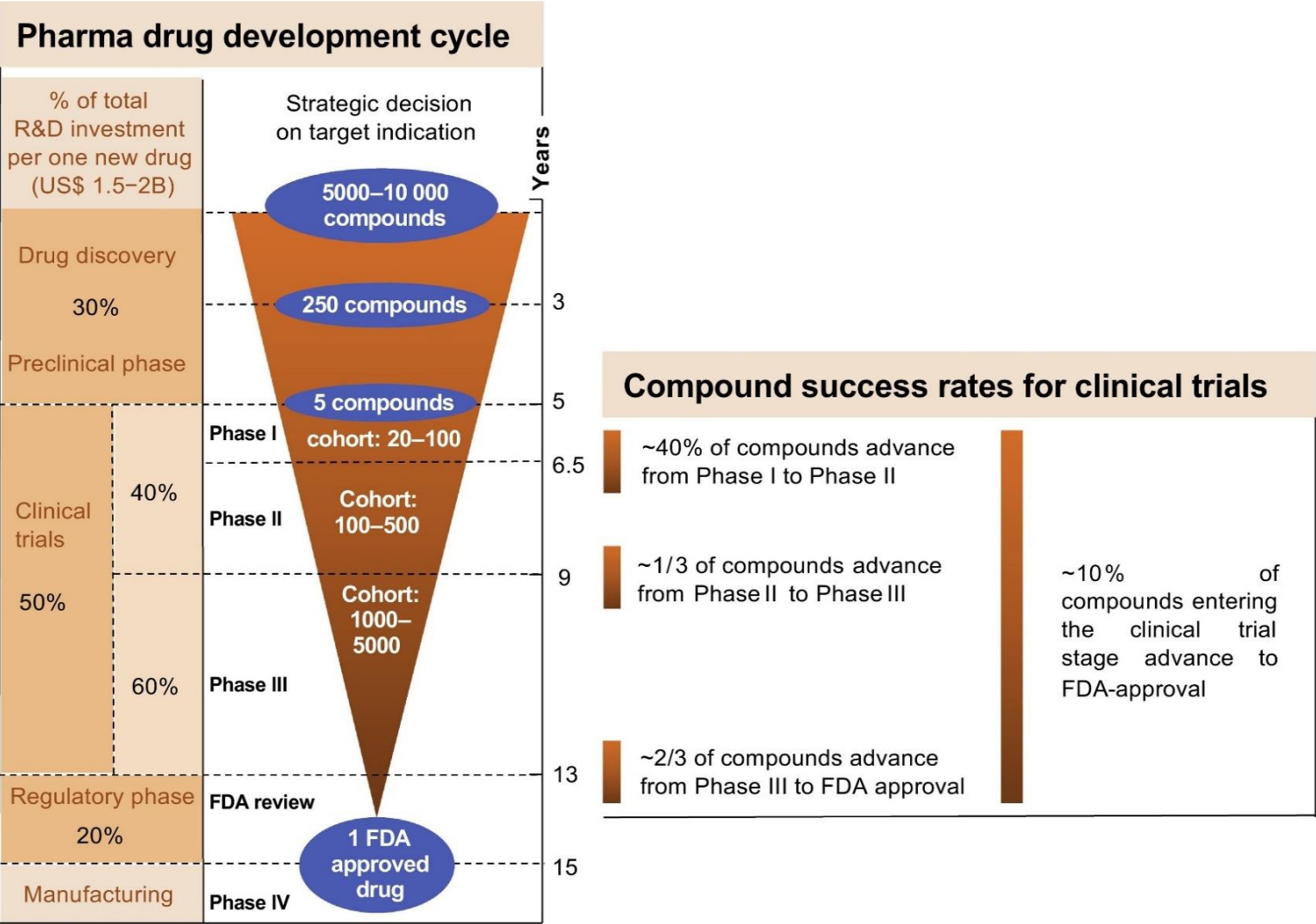
■ AI助力新药临床试验概览

新药开发是一个漫长、昂贵、高风险的过程，其中大约一半的时间和投资消耗在临床试验阶段，而临床试验的高失败率又使其成为新药研发中最困难的一步。临床试验是药物开发的基石，成功的药物开发依赖于准确且有效的临床试验。

据研究，约有90%的药物会在临床试验阶段被淘汰，进入临床II期的化合物中，不到三分之一能进入临床 III 期^[1]，在进入临床III期的化合物中，又有超过三分之一的化合物未能成功获批上市^[2]。由于II、III期临床试验处于整个药物开发阶段后期，且最复杂的 III 期试验成本约占整个试验的 60%，因此，每项失败的临床试验均会造成巨额损失。

一篇发表在Nature子刊《npj Digital Medicine》，关于数字化临床试验文章指出，目前传统临床试验存在参与率过低、临床试验进展缓慢等问题。在受试者招募、筛选、数据采集和质控等方面的效率低下，会给受试者以及研发企业带来经济和时间上的负担，且延长临床试验时间^[3]。

图表1 药物研发阶段及化合物在各临床试验阶段成功率



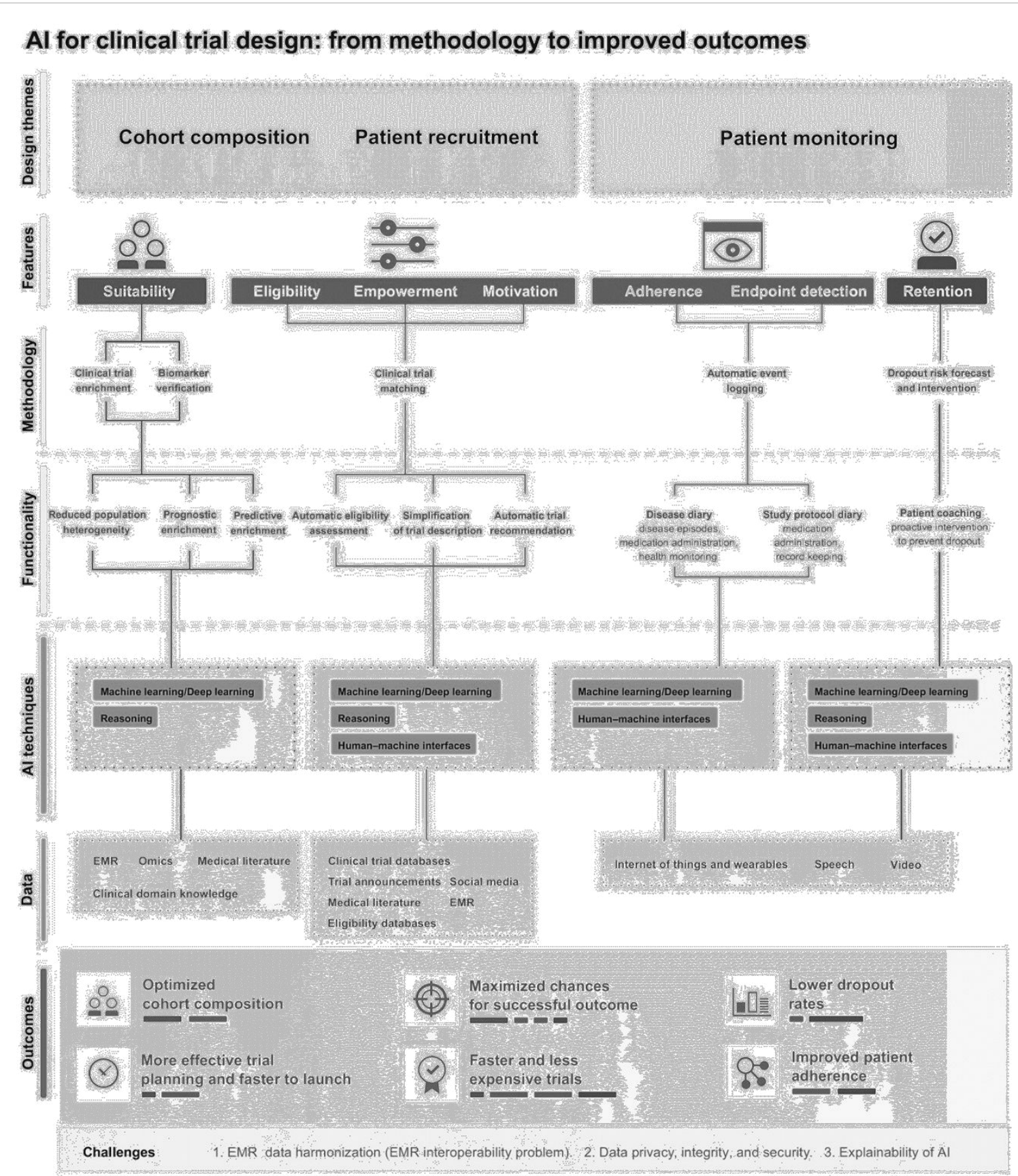
图源: <https://doi.org/10.1016/j.tips.2019.05.005>

Trends in Pharmacological Sciences

[1] doi.org/10.1038/nbt.2786
[2] [doi: 10.1093/biostatistics/kxx069](https://doi.org/10.1093/biostatistics/kxx069)
[3] [doi: 10.1038/s41746-020-0302-y](https://doi.org/10.1038/s41746-020-0302-y)

人工智能（artificial intelligence，AI）的发展，为新药研发带来了新的技术手段，尽管目前AI技术大多应用于药物发现及临床前阶段，但在临床试验环节，也可以利用ML、自然语言处理等技术辅助临床试验设计、患者招募和临床试验数据处理等。

图表1 用于临床试验设计的AI技术：从方法到结果改善



图源: <https://doi.org/10.1016/j.tips.2019.05.005>

■ AI在临床研究中的分析类型及目标

AI临床试验方案设计是指利用人工智能技术辅助和优化临床试验设计的过程。它利用ML（机器学习）、DL（深度学习）、自然语言处理（NLP）、决策支持系统(DSS)等技术，快速处理同类研究、临床数据和监管信息，以及读取临床试验等数据，以提供更有效的临床试验方案。

欧美在临床阶段对AI技术的应用起步较早，主要在应用自然语言处理（NLP）、改进病人分层、简化临床试验的操作流程三个大的层面。通过收到的超过300份AI临床研究申请，FDA归纳分析并对该技术进行了总结，将AI在临床阶段的应用分成了“六大分析类型”与“九大分析目标”。

图表1 AI在临床研究中的分析类型及目标



图注：1.来源于FDA官网2.药融咨询分析整理

近年来，AI在临床研究中不乏成功案例。如：成立于2016年的美国企业Owkin，使用联邦学习来训练和开发其机器学习模型，专门用于提高临床试验效率，并已经建立了一个高级模型目录，使他们能够从成像、基因组学和临床数据中识别新的生物标志物。Owkin致力于识别具有严重疾病进展特征的患者，这些患者可能对开发中的治疗反应最好，目前该公司总共筹集了2.54亿美元，其中包括赛诺菲、Bpifrance和Mubadala等投资机构。

另一家来自剑桥的公司GNS healthcare，成立于2000年，该公司的技术基于各种复杂的数据创建“虚拟患者”，这是一个准确的疾病计算机模型。GNS专注于肿瘤学、免疫学、中枢神经系统和心脏代谢疾病，他们的技术支持更好的病人分层，并能确定哪些病人应该接受一线或二线治疗。

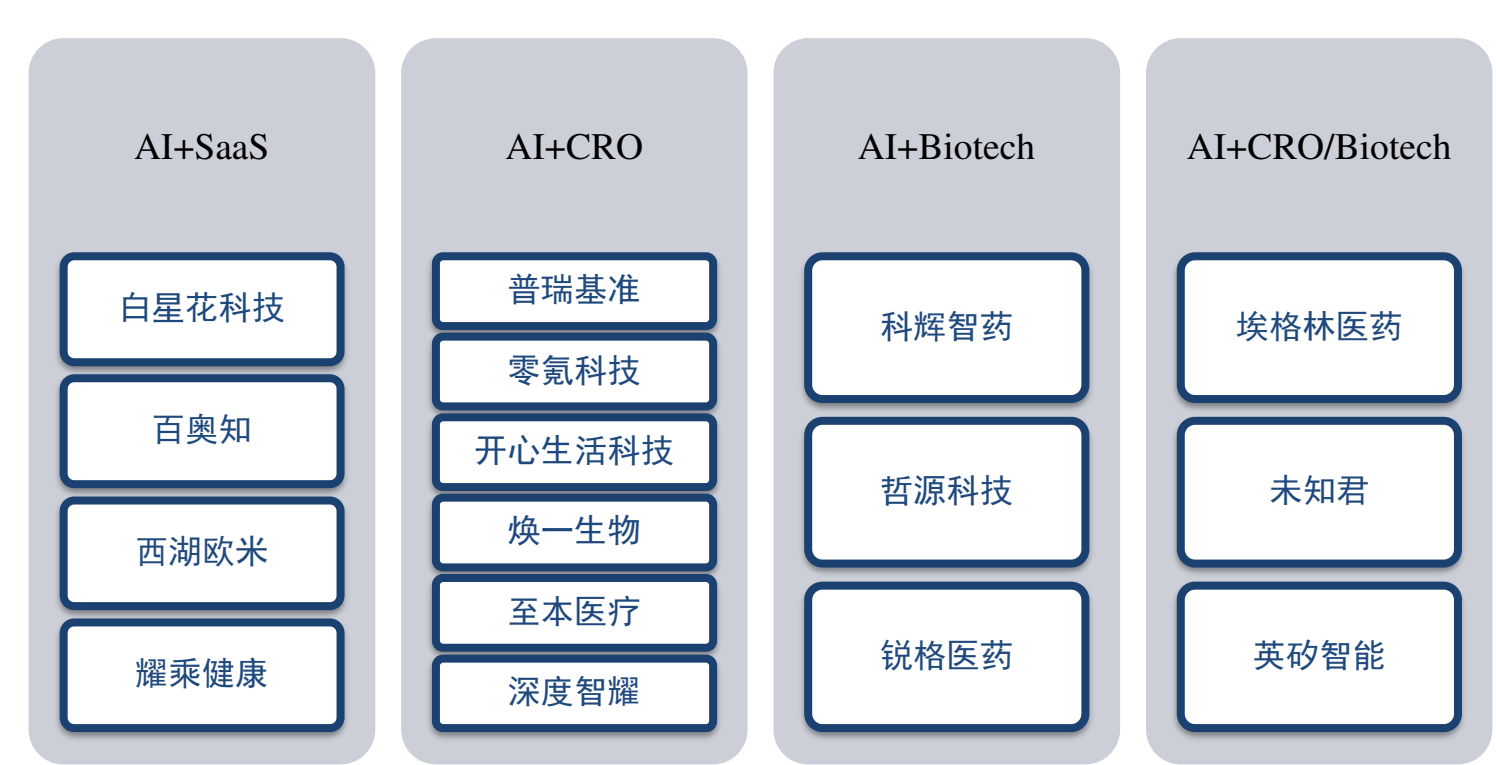
■ 临床研究AI国内企业布局

2022年8月，国家药品监督管理局药品审评中心（CDE）连发三项意见稿，支持远程智能临床试验方案的应用和发展，强调在合规且可行的范围内提高患者参与临床试验的便利度、减轻患者参与临床试验的负担，并对数字化、智能化等新技术在临床试验中的合规运用进行了细致的规定。而在此前，CDE已下发的2项建议中已强调ePRO在数据采集方面所体现出的高效性、实时性、灵活性、合规性、安全性和其在患者隐私保护等方面所具备的显著优势。

几项政策立足“以患者为中心”的理念，接轨国际诸多监管机构对临床试验智能化的政策，为中国临床试验进一步实现从“纸”到“云”的发展变革提供了方向。紧随技术发展趋势，基于国家政策，我国AI+临床试验企业迎来发展机遇。

据药融咨询统计，国内目前涉足AI+临床企业不多，截至2023.09.01，约16家，可大致分为AI SaaS 、AI Biotech、AI CRO以及AI Biotech+ CRO 四种商业模式。

图表1 国内涉足AI+临床研发主要企业



图注：1.各企业官网2.药融咨询分析整理

从十六家企业在临床阶段的具体业务类型来看，主要包括：临床试验方案设计，如患者筛选/招募等，以及临床结果预测、临床试验数据统计分析。从企业类型来看，可提供AI+临床试验企业多为向药企或CXO企业提供服务的企 业，此外，一部分企业实施“双业态”运营模式，仅在少数特定治疗领域内开发自有产品管线，而在大多数治疗领域对外提供AI赋能的临床设计服务。

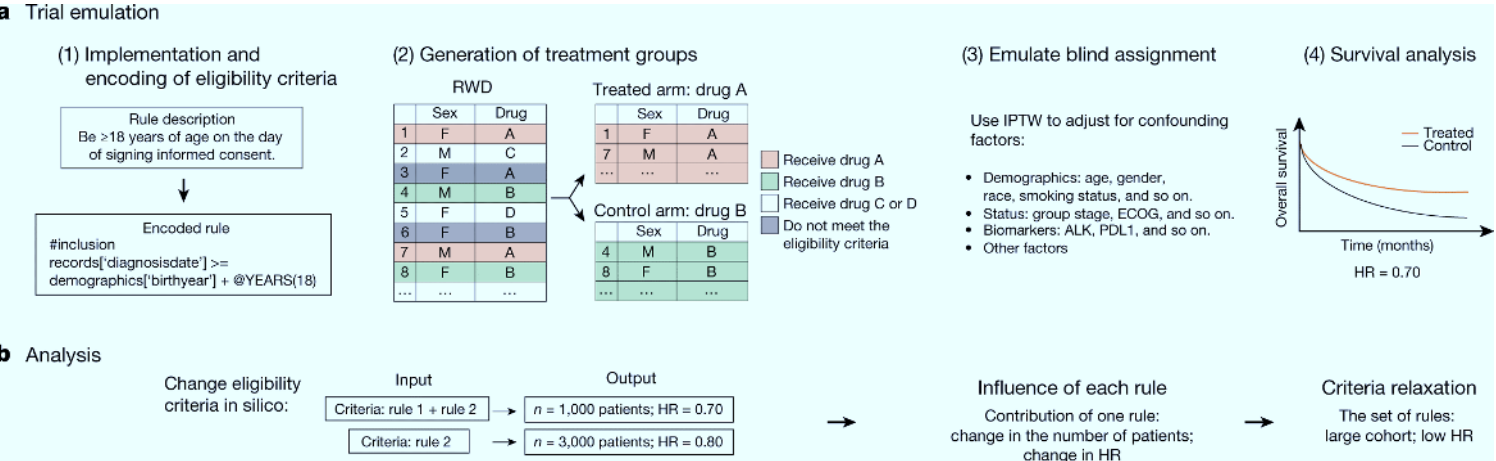
AI助力受试者招募管理

最新技术举例

患者招募是决定临床试验是否成功的关键因素之一，临床试验的入组标准通常有很严格的限制^[1]，研究显示，大约80%的进展期非小细胞肺癌（aNSCLC）患者不符合临床试验的资格标准，导致约86%的临床试验由于无法按时招募足够的受试者而终止^[2]。而大样本的随机对照临床试验（RCT）是临床证据的主要来源，但受试者的招募往往因为试验方案的入组标准严格而影响进度。

相关研究证明，结合真实世界数据，AI可自动筛选符合入组标准的患者，提高招募效率。AI辅助药物临床试验受试者招募，总体可以看作是利用自然语言处理、机器学习等技术，对不同来源的受试者信息和临床试验方案的入组/排除标准进行识别和匹配的过程。《Nature》发表的一项研究介绍了AI在简化临床试验入组标准方面的价值和应用，通过使用开源的人工智能工具Trial Pathfinder，从肿瘤患者的真实世界数据中学习，来优化入组标准的包容性，该工具可以使用EHR数据模拟临床试验，按照不同的入组标准自动整合EHR数据，并分析总体生存风险比率，同时，使用逆概率加权法来调整基线混杂因素和模拟随机化^[3]。分析结果显示，许多常见的入组标准，对临床试验的结果影响甚微。

图表1 用于临床试验设计的AI技术：从方法到改进结果



图源: <https://doi.org/10.1016/j.tips.2019.05.005>

Trial Pathfinder包括临床试验模拟和分析两部分。在临床实验模拟中，Trial Pathfinder将真实世界数据和目标试验方案（治疗和资格标准）作为输入，根据患者特征、诊断、实验室值、生物标志物和先前的治疗方法，以编程方式对不同的入组标准（从文本中提取）进行编码，并使用倾向评分的加权分析法进行临床试验模拟。然后对模拟的治疗组进行生存分析，并报告符合条件的患者数和由此产生的风险比。在分析中，Trial Pathfinder将标准的重要性分析与Shapley值相结合，评估了每个纳入/排除标准如何影响合格受试者的数量和试验结果。研究表明，几个常用的纳入/排除标准不会对试验的总生存期的风险比产生实质性影响，也不会潜在地降低试验的有效性。

此外，一项回顾性分析全面了解了AI在临床试验注册登记中的应用现状^[4]，并评估该技术筛选患者的准确性，这是第一项评估AI用于癌症临床试验登记的系统性回顾和meta分析。数据显示，除一个数据集外，所有数据集的准确性、灵敏度和特异性均超过80%，17个数据集中有5个的阳性预测值超过80%，所有数据集中的阴性预测值均超过80%，汇总敏感性为90.5%（95% CI:70.9%-97.4%）；总体特异性为99.3%（95% CI:81.8%—99.9%）。总的来说，AI在患者筛选方面的表现似乎与人工筛查持平，且AI拥有更高的效率，可花费更少的时间和资源。

[1] doi/10.1200/jco.2009.27.15_suppl.6538

[2] doi: 10.1016/j.cct.2018.01.003

[3] doi: 10.1038/s41586-021-03430-5

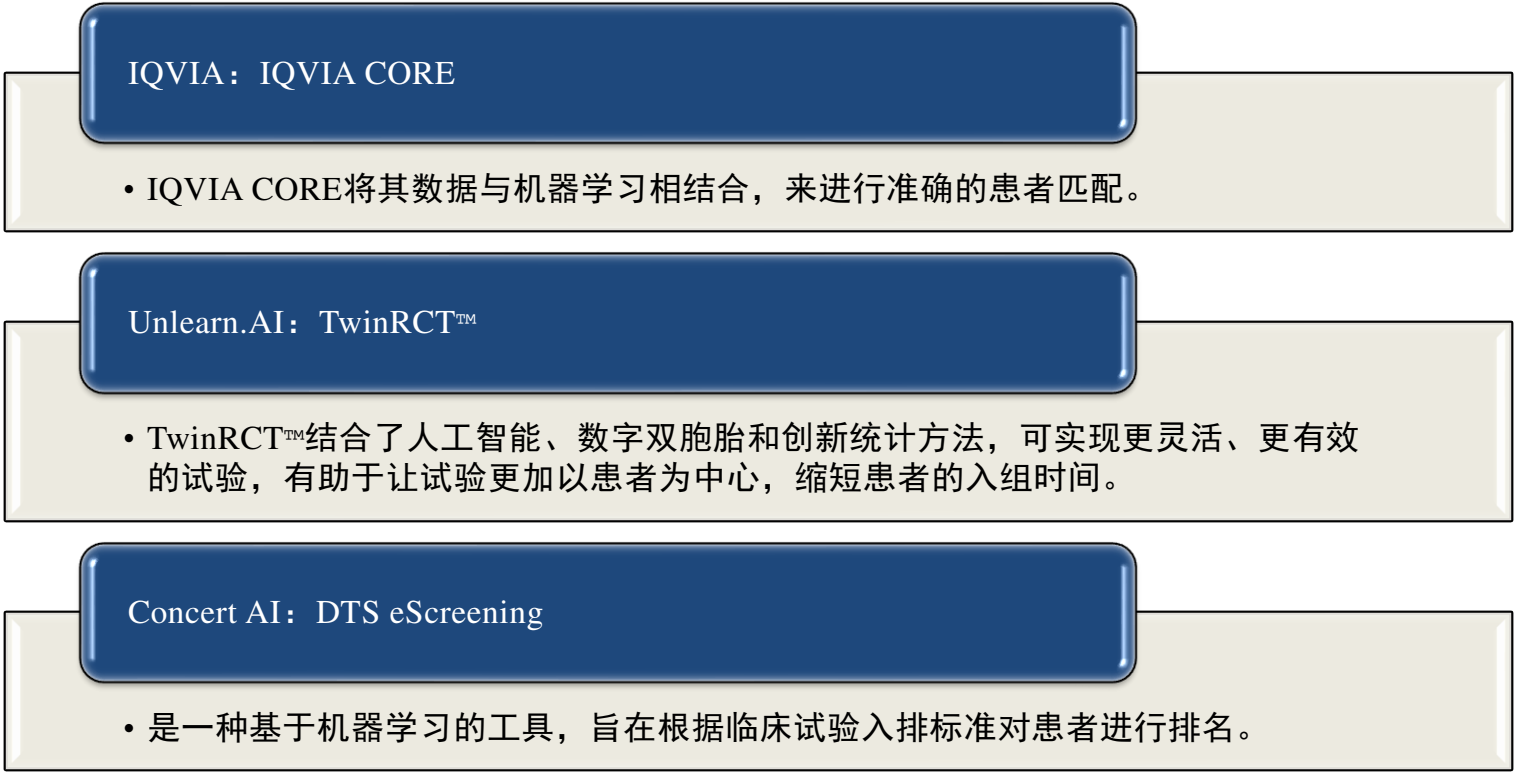
[4] doi: 10.1093/jnci/djad013

■ 国内企业布局

目前多家AI制药企业在助力患者入组的研究上已取得一定成就。如IQVIA、Unlearn、ConcertAI等，ConcertAI 的DTS eScreening采用了机器学习算法LightGBM来开发17个模型，这些模型根据各种标准预测患者参加临床试验的资格，包括ECOG评分（癌症研究中采用的标准衡量标准），使用的其他指标包括癌症分期、实验室结果和生命体征，平均预测准确率超过99%。

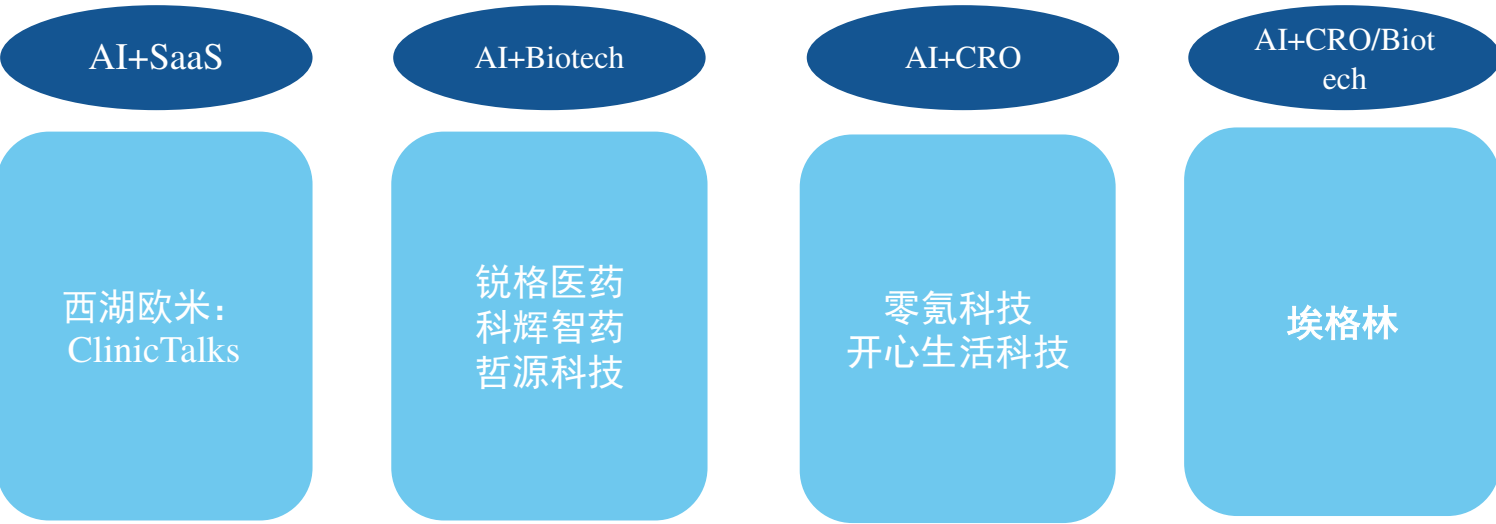
2020年，FDA同意Medidata Synthetic Control Arm® 合成对照组解决方案（“SCA®解决方案”）应用于美国临床阶段免疫治疗公司Medicenna Therapeutics 公司（“Medicenna”）一项MDNA55治疗复发性胶质母细胞瘤的III 期注册试验中。在此试验中，Medidata成功帮助Medicenna在此三期临床中减少招募了100个对照患者，这也是罕见病试验首次采用混合外部对照组，也开启了III期临床试验应用混合外部对照组的先河。

图表1 AI制药企业在临床试验患者入组方面的研究成果（举例）



图注：1.来源于公开信息2.药融咨询分析整理

国内企业运用AI助力临床试验患者招募/入组尚处于起步阶段，代表企业包括普瑞基准、埃格林、西湖欧米以及锐格医药等。这些企业并非专注于临床阶段或临床试验患者招募，而多覆盖药物开发多个流程，如科辉智药，其AI驱动的研发平台eLEAD是覆盖研发全流程，AI筛选患者人群是其中一部分。锐格医药主要通过自主研发的 rCARD™（计算机加速新药研发）平台来发现和开发临床差异化的创新药物。



哲源科技可进行虚拟临床试验，预测响应患者。据哲源科技官网披露，其核心技术包括TWIRIS（生物医学文献处理算法-大规模生物医学知识AI平台）、DAGG（基因组霰单枪损伤评估算法-细胞功能定量评估AI平台）、Dpverse（泛癌种功能数字化病人库MOA虚拟平台）。在研管线中，为全新机制、全新靶点药物管线共3条，ZY0001（胰腺癌、小分子）、ZY0002（泛癌种、大分子）、ZY0003（胰腺癌、小分子），均处于药物合成阶段。

零氪科技提供第三方服务，主要开展基于精准生命科学解决方案（LinkSolutions）的临床招募、真实世界研究、数据洞察等服务。其临床研究招募主要采用机器学习算法将已知的信号通路和药物靶点库进行匹配，以预测个性化治疗和患者耐药点的有效性，实现患者对特定临床试验的最佳适配性。

埃格林是具有从靶点到临床试验“端到端”能力的AI赋能创新药企，埃格林采取自研+对外服务的模式，一方面，从事自有药品管线的研发，另一方面，对有人工智能研发需求的制药企业，提供创新药开发服务与合作。

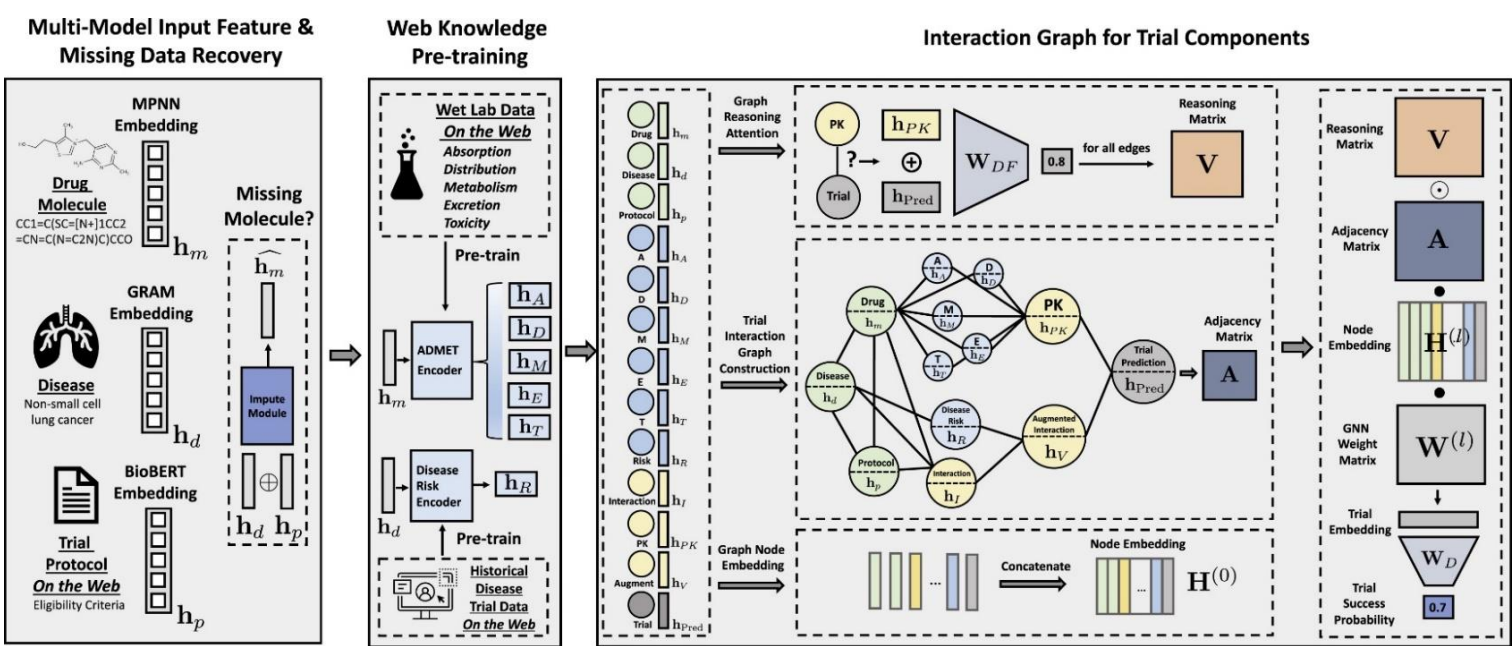
■ AI助力预测临床试验效果

■ 最新技术举例

鉴于临床试验的高失败率，以及药物开发的资金密集型性质，制药公司只能承担有限数量的项目，因此开发者通常不得不中止一些项目，以支持成功概率更高的药物进一步开发。对于继续研发药物的选择，目前常用的方法基本上是基于管线上药物的治疗类别，和所处开发阶段历史上其他药物的监管批准率，结合对项目特定风险因素的非结构化讨论和主观调整，来做出投资决策。

对于临床试验效果的预测，在药物开发中一直难以达到。AAPS于2022年一月发表的一篇关于在制药研发过程中应用机器学习/人工智能预测技术和监管成功概率的文章指出，关于临床试验效果预测尚未有经验证的报告^[1]。但AI技术发展迅速，此后，不少文章研究结果显示，尽管具有难度，但对于AI临床试验结果预测，从技术上看，正在朝着积极的方向前进。

图表1 HINT:端到端的神经网络模型



其中，《Cell Pattern》于2022年4月发布了首个临床试验结果预测公开数据集，提出了基于深度学习（图神经网络）的临床试验结果预测方法。作者设计了一种基于图神经网络的深度学习方法，可以显示的模拟不同的临床试验因素及其交互关系，从而预测试验结果，该方法被命名为Hierarchical interaction network (HINT)。HINT 在临床 I 期、II 期和 III 期的预测中分别获得了 0.665、0.620和0.847 的 F1 分数。

[1] doi: 10.1208/s12248-021-00644-3

[2] doi:10.1002/cpt.3008

[3] doi.org/10.1016/j.patter.2022.100445

图表1 部分AI预测临床试验效果相关文章

文章标题	文章结论	链接
Reusability report: Evaluating reproducibility and reusability of a fine-tuned model to predict drug response in cancer patient samples 可重复使用性报告：评估微调模型的再现性和可重复使用性，以预测癌症患者样本中的药物反应	可重复使用性结果表明，在大多数新的临床环境中，TCRP（细胞系反应预测转移）仍然是预测临床前和临床环境反应的优越方法。结果支持 TCRP 在临床前和临床环境中优于已建立的统计和机器学习方法。	Doi: 10.1038/s42256-023-00688-4
Prediction of Clinical Trials Outcomes Based on Target Choice and Clinical Trial Design with Multi-Modal Artificial Intelligence 基于多模态人工智能靶点选择和临床试验设计的临床试验结果预测	II 期临床试验的预测结果准确率达到 79%。	Doi: 10.1002/cpt.3008
Predicting clinical trial outcomes using drug bioactivities through graph database integration and machine learning 通过图形数据库集成和机器学习使用药物生物活性预测临床试验结果	作者提出了一种可以使用生物活性、化合物的物理化学性质、目标相关特征和基于 NLP 的化合物表示来准确地预测试验结果的机器学习，该模型在预测临床试验“是否成果”上表现良好，F1 达0.96。通过分析特征对预测能力，证明了生物活性在预测临床试验结果方面具有统计学意义。	Doi: 10.1111/cbdd.14092
用于临床试验结果预测的分层交互网络 HINT: Hierarchical interaction network for clinical-trial-outcome predictions	作者提出了分层交互网络（HINT）来预测临床试验结果。暗示在 1,160 项 I 期试验、4,449 项 II 期试验和 3,436 项 III 期试验中接受了培训和验证。它在 627 个 I 期、1,653 个 II 期和 1,140 个 III 期试验的单独测试集上分别获得了 0.665、0.620 和 0.847 F1 分数。HINT在大多数指标上明显优于最佳基线方法。	Doi: 10.1016/j.patter.2022.100445
In Silico Model Estimates the Clinical Trial Outcome of Cancer Vaccines 计算机模型估计癌症疫苗的临床试验结果	CRR（临床反应率）取决于广泛 T 细胞反应的诱导，并且当考虑与多个自体 HLA 结合的表位时，可以预测 IRR（免疫反应率）和 CRR。	Doi: 10.3390/cells10113048

国内企业布局

目前，国内可将AI技术用于临床试验结果预测企业较少，代表企业包括英矽智能、白星花科技、西湖欧米。2023年8月3日，国内领先的生成式人工智能（AI）驱动的生物医药科技公司英矽智能，利用其自主研发的基于Transformer的人工智能临床试验预测引擎inClinico，高度准确地预测了多项临床试验II期至III期的转化结果。这项研究成果已发表在《CLINICAL PHARMACOLOGY & THERAPEUTICS》上^[1]。

英矽智能

Pharma.AI

inClinico：预测临床试验II期至III期的转化成功率

白星花科技

Bx-SaaS平台：聚焦于临床数据、真实世界以及医药科研文献

西湖欧米

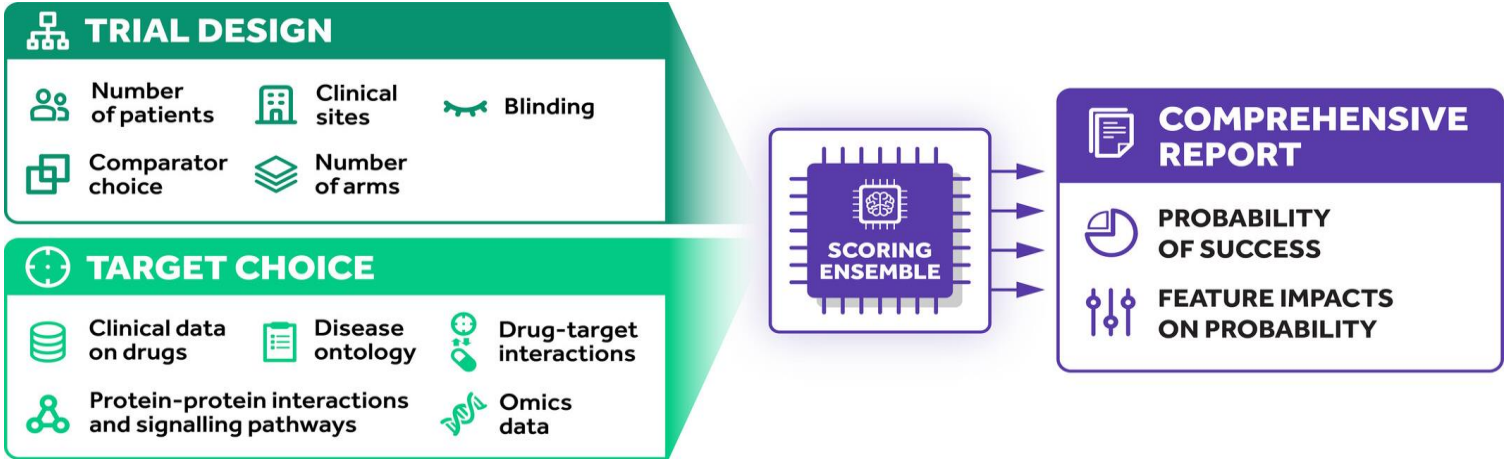
iTalks Series:西湖欧米AIDD平台

ClinicTalks：药物临床试验分析与结果预测

研究方向：蛋白质组学

2014年以来，英矽智能长期致力于开发用于II期到III期临床试验转化成功率预测的人工智能模型。英矽智能搭建的高效人工智能药物研发平台Pharma.AI，包括PandaOmics™、Chemistry42™、inClinico™三大组成部分，其中inClinico™ 是一个SaaS模式的在线软件系统，旨在预测临床试验II期至III期的转化成功率，它整合了多种评分方法、多模态数据源（包括文本、组学、临床试验设计和小分子特性）以及生物学背景，并与深度学习模型相结合。2022年11月以来，inClinico™以SaaS形式提供服务。

图表1 用于预测临床试验结果的人工智能框架和数据源概览



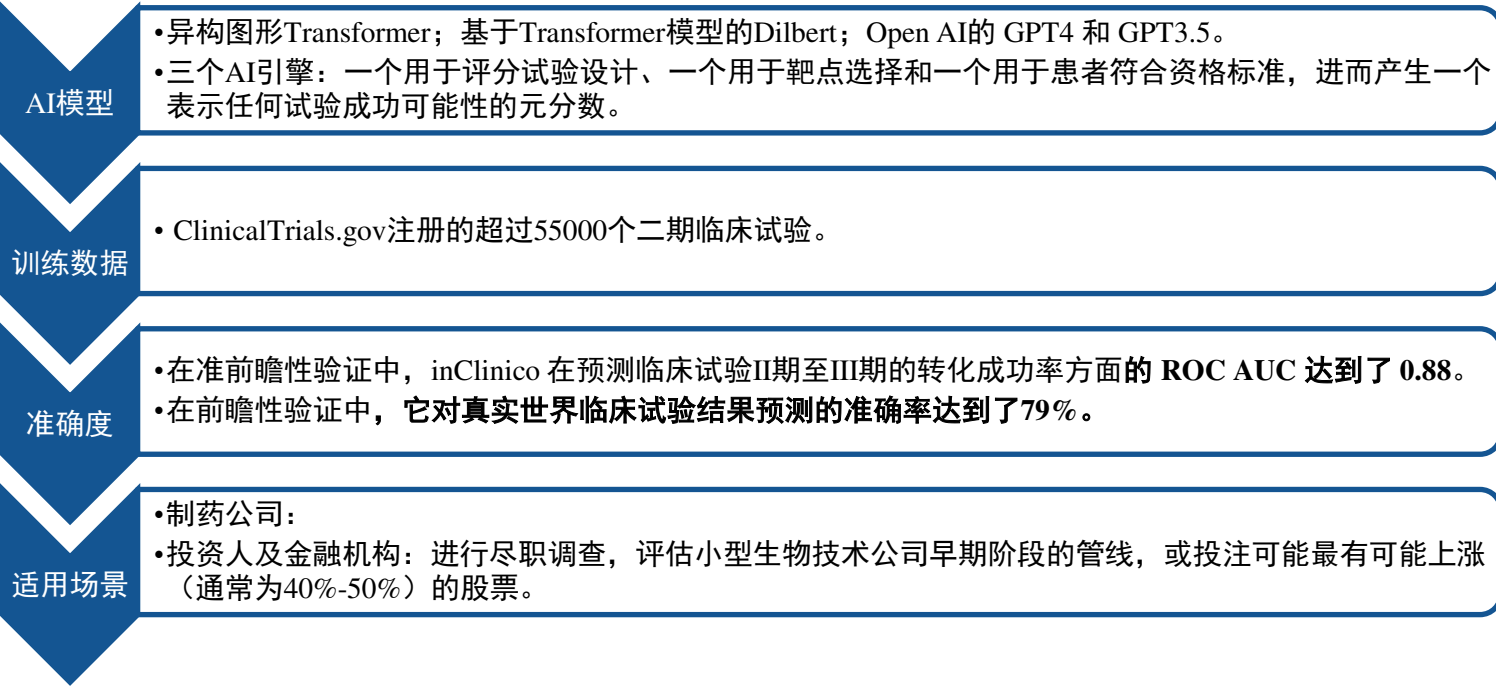
图源：Clin Pharma and Therapeutics, First published: 22 July 2023, doi (10.1002/cpt.3008)

[1] doi:10.1002/cpt.3008

根据英矽智能发表的论文，inClinico平台在55,000多项独特的II期临床试验中进行了训练，并在回顾性、准前瞻性和前瞻性验证研究中进行了验证。在准前瞻性验证中，inClinico 在预测临床试验II期至III期的转化成功率方面的ROC AUC 达到了0.88。在前瞻性验证中，它对真实世界临床试验结果预测的准确率达到了79%。

除了成功预测多项临床试验II期至III期的转化结果外，该研究结果还表明，靶点选择比临床试验设计更有可能影响临床试验结果预测。此外，论文还展示了inClinico的投资应用，使用了日期标记的虚拟交易组合，展示了35%的9个月投资回报率。以投资为目的的实践还表明，inClinico 可以为投资者提供技术上的尽职调查见解，并帮助制药公司确定药物开发项目的优先次序。

图表1 inClinico用于预测临床试验结果的人工智能框架和数据源概览



图表2 inClinico核心用途和功能^[2]

Core Features

inClinico is a platform for Data-driven multimodal **forecast of single clinical trial's probability of success (PoS)**.

- ✓ Predict clinical trials success rate
- ✓ Recognize the weak points in trial design
- ✓ Adopt the best practices in the industry

- ◆ **Scoring of custom trials:** compare designs for the clinical trial that are planned and optimize them for success
- ◆ **Database of historical and ongoing clinical trials:** find clinical catalysts that drive pharmaceutical growth, sort them by the PoS, and research the designs of successful trials
- ◆ **Clinical Trial Outcome Scores:** the ability to analyze performance for individual trials with simple data-driven metrics for the PoS
- ◆ **Comprehensive Report Generation**

[1] 来源于公开信息，药融咨询整理
[2] 来源于英矽智能官网

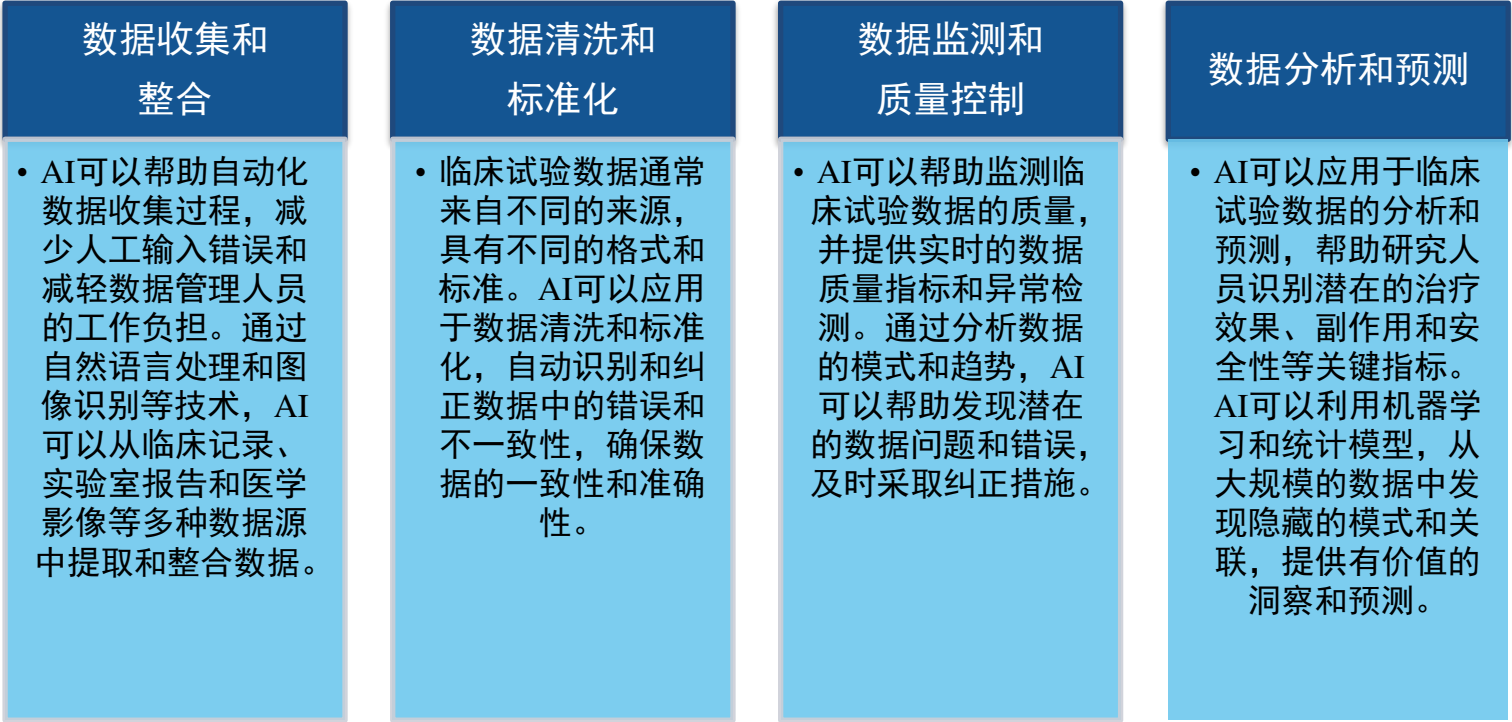
■ AI助力临床试验数据管理

■ 最新技术举例

新药研发是一个数据密集型的工作，虽然过去几年中AI取得了令人瞩目的进展，但研究者们仍然面临着海量数据的挑战。如临床研究数据管理需要包括从研究设计、研究注册、数据采集、患者随访、项目管理、安全预警、数据管理与初步分析的全链条管理。

单临床试验的纳入/排除标准可能多达六七十项，而这些标准可能包括实验室检查结果、过去的药物暴露史、药物反应等。此外，临床试验数据的固有复杂性和多样性面临着进一步的挑战。如在肿瘤治疗领域，临床试验通常包含多方面的设计，包括复杂的组织取样和分子处理要求。因此，医疗组织必须确保拥有具备经验丰富的医护人员，但传统的数据收集管理存在诸多问题，数据的可靠性、安全性和准确性得不到保证、数据易丢失、数据核查的工作量太大、数据管理的费用过高且过程较为繁琐等，整个过程会推迟临床研究进展。改革传统临床数据管理模式迫在眉睫。如何完整、及时、高效地采集数据变得尤为重要，人工智能和机器学习工具的使用也应运而生。

具体而言，AI在临床试验数据管理方面的应用包括数据收集和整合，数据清洗和标准化，数据监测和质量控制，以及数据分析和预测等。目前，在临床试验数据管理中，运用AI的主要技术包括：ML（机器学习）、DL以及NLP等。



近年来AI在临床试验数据管理方面已取得一定进展。如：在2023年美国临床肿瘤学会(ASCO)会议中，Concert AI分享了它如何成功地自动化了一项研究所需的55%的数据，该公司研发的DTS eScreening软件，是一种基于ML的工具，旨在根据临床试验入排标准对患者进行排名；2023年6月13日，领先的临床数据技术和自动化公司Beaconcure宣布获得1400万美元B轮融资，Beaconcure的人工智能平台Verify是与辉瑞联合开发的，它简化了临床试验数据统计分析过程并使之自动化；此外，生成临床研究报告 (clinical study reports, CSR) 是临床试验的重要组成部分，通常需要数周或数月，而 NLP处理先驱Yseop于2023年6月推出“Yseop Copilot”，据Yseop，使用该技术自动撰写CSR有助于将报告撰写时间平均减少40%。Yseop Copilot采用了预训练的大型语言模型，并结合了数据到文本（符号AI）和文本到文本（预训练的开源LLM）技术。

[1] 来源于公开信息，药融咨询整理

国内企业布局

国内，将AI技术应用到药物开发的临床及临床后阶段企业较少，在IND阶段后，对临床试验数据的管理、分析是应用相对较多的场景，主要有普瑞基准、耀乘健康等企业。

据公开信息披露：耀乘健康科技AI技术团队实施了 AuroraPrime 底层架构与 OpenAI 的对接，并着手在多条产品线中实际落地应用场景。包括： Prime Create、 Prime Coordinate、 Prime Catalog等等。Prime Catalog 是在 AuroraPrime 平台上构建的企业级文档管理系统，包含了 eTMF、项目公盘等文档产品，满足了企业级文档、项目级文档和项目 eTMF 多维度的文档管理诉求。在OpenAI助力下，可结合项目 eTMF 目录结构以及上传的文档名等关键信息，在用户操作过程中智能推荐文档的归档位置等，提升整体文档管理效率。

图表1 AuroraPrime 临床研究平台



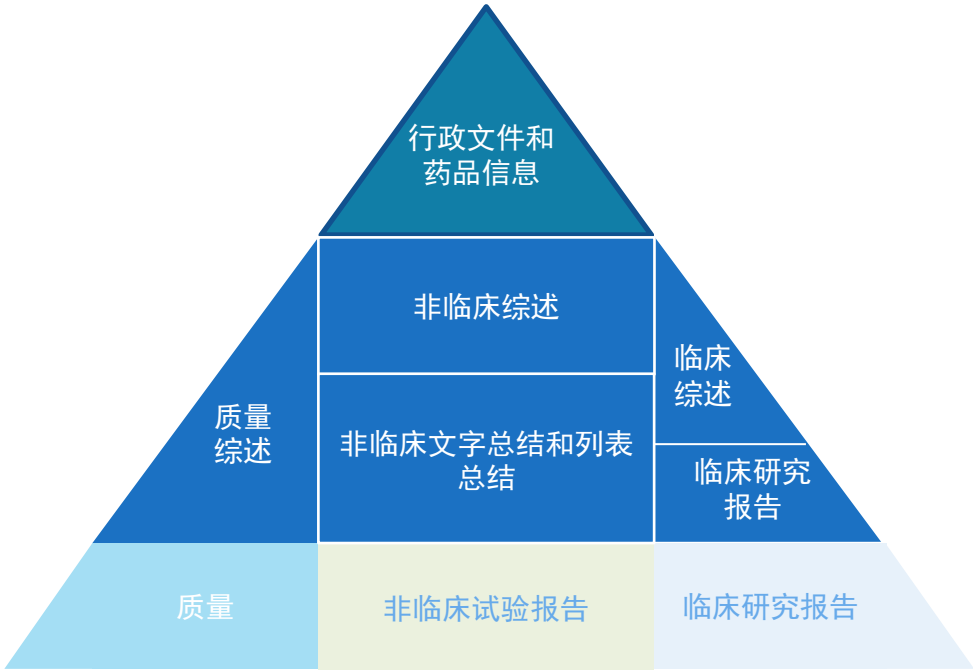
图表2 国内将AI应用于临床试验数据管理的代表企业及应用场景

企业	主要技术平台及应用场景
普瑞基准	AIBERT® AI 针对药物临床试验，提供临床试验统计分析服务，内容包括分析计划制订、数据治理、程序编写、图表制作、分析报告撰写等。 除了临床试验数据分析标准内容之外，亦可根据客户需求提供包括多组学数据挖掘在内的探索性分析：在全面剖析现有临床数据的同时，亦为下一步药物研发提供方向和依据。
耀乘健康	Prime Catalog 是耀乘健康科技 AuroraPrime 临床研究平台上用于临床试验全流程文档管理的软件系统。 基于 NLP 的归档目录推荐功能。
至本医疗	对话式AI医疗数据临床分析工具OMSmartData 通过结构化的整合进行深度挖掘和全面分析，以对话形式，充分挖掘用户诉求，智能输出文字和分析图表，满足不同用户、不同场景的数据需求。
深度智耀	电子数据采集系统（EDC） 临床数据管理系统 临床数据编程系统 提供完整的临床数据流解决方案，包括数据采集、转换、编码、核查等各业务环节。

■ AI赋能药物注册eCTD申报

CTD（The Common Technical Document，共同技术文件）是国际公认的文件编写格式，用来制作向药品注册机构递交的结构完善的注册申请文件。一共分为五个模块，模块一行政管理文件和药品信息；模块二通用技术文档总结；模块三质量；模块四非临床试验报告；模块五临床研究报告。其中，模块一为区域性要求，具体内容和格式由相应的监管机构规定，模块二、三、四和五为国际通用要求。电子通用技术文档（eCTD）是对CTD的一种电子化呈现和管理方式，通过可扩展标记语言（XML）技术将基于CTD文件结构和eCTD技术规范的文件加以组织，并用于药品注册申报和审评的一种注册文件技术格式。eCTD使申报资料编制、申报、受理、审评、全生命周期管理和档案存储更加便捷经济，既保障了资料的申报质量，又提高了审评效率，更重要的是使申报资料的格式、内容及标准要求保持全球通用性和一致性。

中国国家药品监督管理局已自2021年12月29日起全面开展和推进药品电子通用技术文档（eCTD）申报，正式打开了eCTD申报时代的大门。可以预见，随着 eCTD 在我国的全面实施，递交网络化及验证自动化、eCTD 4.0 试点以及云数据和 AI 自动化等应用也将逐步落地。



目前国内较为典型的是成立于2017年的深度智耀，其自主研发的AI智能系统和相关服务覆盖制药研发全链条，其中就包含全球注册申报以及文本智能（医学写作、医学翻译、eCTD递交、文档管理）等各模块。在临床研究阶段，包括 AI 驱动的药物警戒系统、注册事务系统、写作翻译系统、临床数据编程系统等产品。另外，公司的AI 自动翻译写作申报系统实现集医学写作、医学翻译和注册申报“三位一体”的人机交互自进化平台，改变以往上述三大环节的割裂状况，重构了传统的数据文件整合模式。



05

第五章

AI+药物生产及商业化

■ AI赋能药物生产支持

■ 最新技术举例

人工智能为药物生产也提供了许多可能性，包括但不限于优化工艺设计和工艺控制，智能监测和维护，以及趋势监测以推动持续改进等，使用人工智能来支持药品生产，可以与其他先进的生产技术一起部署，以实现预期的效益。人工智能是实施工业4.0模式的助推器，可以为生产商带来一个控制良好、超级连接、数字化的生态系统和医药价值链。

根据 FDA 药品审评和研究中心发布的《药品生产中的人工智能》讨论稿，AI用于药物生产可主要分为四个场景，包括工艺设计优化以及工艺放大、高级工艺控制、工艺过程的检测以及缺陷的检查、趋势分析和检测。

目前，AI用于药物生产尚不成熟，但已取得一些阶段性成果。如生物技术公司Pow.bio通过AI+连续发酵平台，朝着优化和自动化发酵的方向发展，使其既具有成本效益又精简。

Teledyne DALSA开发了一种特殊的盲文视觉检测系统，该系统使用带有面扫描摄像头的VICORE智能视觉系统的形状阴影技术，以及Sherlock形状阴影算法，从背景复杂的盲文中获取对比度较高的3D效果图像。然后，Sherlock应用预处理程序优化盲文点的残缺形状，从而使OCR算法能够读取盲文字符，以帮助盲人阅读药品说明书。

工艺设计优化以及工艺放大

可以利用ML（机器学习）等人工智能模型—使用工艺开发数据生成—更快速地确定最佳加工参数或放大工艺，减少开发时间和浪费。

高级工艺控制（APC）

APC允许对生产过程进行动态控制，以达到理想的产出。

人工智能方法也可用于开发过程控制，通过使用人工智能与实时传感器数据相结合，可以预测过程的进展情况。将对生产过程中发生的基本化学、物理和生物转化的理解与人工智能技术相结合的。

工艺过程的检测以及缺陷的检查

人工智能方法可用于监测设备，并检测出与正常性能不同的变化，从而触发维护活动，减少工艺停工时间。

也可用于监测产品质量，包括包装质量（例如，基于视觉的质量控制，使用包装、标签或玻璃瓶的图像，由基于人工智能的软件进行分析，以检测产品的特定质量属性要求的偏差）。

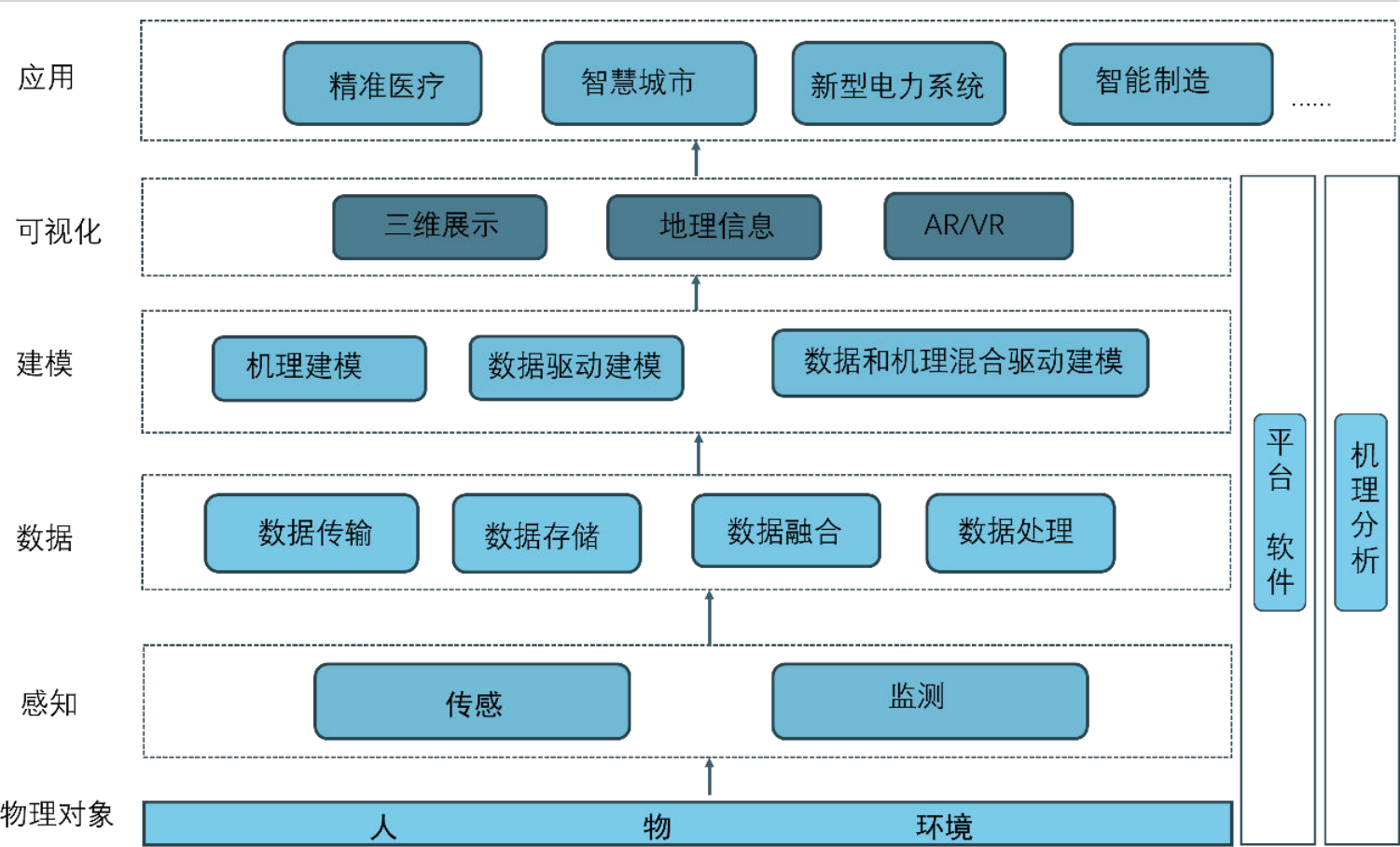
趋势分析和检测

人工智能可用于检查消费者投诉和包含大量文本的偏差报告，以确定集中问题领域并优先考虑持续改进的方向。这提供了识别生产相关偏差的趋势的优势，以支持更全面的根源识别。

与工艺性能和工艺能力指标相结合的人工智能方法可用于主动监测生产业务的趋势。这些方法还可以预测触发纠正和预防行动有效性评估的阈值。

在工艺优化设计中，可采用数字孪生技术。工艺数字孪生是物理过程的数字复制，用于更好地理解、分析、预测和优化工艺性能。对于分析以有限的开发数据为特征的制造工艺，数字孪生尤其有用。如：GSK与西门子公司合作^[1]，在中试规模验证数字孪生技术并获得成功；奥地利初创公司Novasign开发了基于混合模型的数字孪生系统，运用该系统对大肠杆菌表达超氧化物歧化酶的过程进行优化^[2]，加速发酵过程优化的进程；西门子公司2019年收购PSE公司过程数字化模型软件平台gPROMs，用于布局生物过程数字孪生系统。

图表1 数字孪生关键架构



图源：中国电力科学研究院人工智能应用研究所

在药品制造中，可以使用如神经网络等AI/ML技术，以实时过程数据为输入，来实现APC。这些方法也可以用于开发能够预测一个过程是否在一个控制状态下运行的过程控制，这需要结合实时传感器数据使用AI/ML工具，包括与生产线的智能监控相结合，以提高现有制造线的效率和产量。

对于药品质控，常常采用**机器视觉检测技术**。在制药生产过程中，对具有难点的产品进行检查是一项非常具有挑战性的工作。如：对于无法完全去除气泡的高粘度注射液，区别这些气泡与颗粒存在困难。在达到平衡的检测水平和误别率之前，这些情况通常需要长期进行**视觉算法**开发与优化。“AI图像视觉检测”，其本质是通过基于深度学习的AI算法，对检测物进行图像采集，并进行样本标注、训练，从而达到对检测物进行分类决策的目的，将良品与不良品进行分类处理。

对于趋势监控，AI/ML可以被用来辅助审查偏差报告，这些报告通常包含大量的数据或文本，用于分析制造相关的偏差趋势，聚类问题领域，并优先处理需要主动持续改进的领域。与过程性能（Ppk）和过程能力（Cpk）指标集成的AI/ML方法可以用来主动监控制造操作的趋势和失控事件，并预测触发CAPA（纠正和预防措施）有效性评估的阈值。

[1] Markarian J. Opportunities for digital twins. BioPharm International, 2021, 34(9): 30-33.

[2] doi: 10.1002/biot.202000121

■ 国内企业布局

具药融咨询统计，目前国内已将AI技术运用于制药环节企业不多。AI+制药企业主要包括沃时科技、大湾生物、胜普泽泰，三家企业研究领域有所不同，但在药物制造上，均主要用于工艺优化。

- 沃时科技聚焦于小分子化合物，在AI+逆合成与工艺场景优化上发力，服务客户主要为CRO、CDMO企业，帮助这些公司合成分子并在合成工艺上进行优化。
- 大湾生物将AI技术主要应用于生物药领域，公司已有3个智能化生物工艺开发平台进入商业化运作阶段，包括Klone4.0™、AlfaStaX®与AlfaMedX®。通过AI技术与生物技术，大湾生物进行抗体蛋白类药物CDMO工作，通过AI技术和生物工艺相结合，开发定点整合高表达细胞株。
- 胜普泽泰致力于多肽药物的研发和生产服务。该公司以大数据与AI驱动的Chemical Space新药发现技术在瑞士发展近20年，已合作药企众多，包括ROCHE、LONZA、BRACCO、人福药业、兴齐眼药等。

沃时科技

- 应用场景：小分子
- 制药工业场景AI决策平台
- 聚焦于智能制造和工艺优化等领域，致力于AI较少涉足的“分子合成”、“工艺设计”、“工艺生产放大”环节等。

大湾生物

- 应用场景：大分子
- 首创“AI+Bioprocessing”概念，致力于将AI赋能于生物工艺全过程。
 - ✓ AlfaCell®：高产 稳定 定点整合细胞株构建平台
 - ✓ AlfaMedX®：智能化培养基开发平台
 - ✓ AlfaOPA™：智能化无筛选培养基开发平台（细胞株关键成分代谢模型）

胜普泽泰

- 应用场景：多肽
- 专注于构建多肽、拟肽以及多肽偶联药物（PDC）从新药发现到工业化生产的一站式AI-CRDMO服务平台。
- 多肽CDMO：采用固相-液相组合工艺技术和拉曼-红外工艺过程控制技术，利用AI数据库找到最优的工艺参数，使得每步合成收率达到99.5%，总生产率提高15%，成本降低。

我国制药行业年产量高，且种类繁多，胶囊、片剂等在生产及铝塑封装过程中不可避免会产生缺粒、凹坑、破损、批号不清等缺陷。在制药过程中，除生产工艺需要优化外，对于药品的质检也具有迫切需求。一些通过人工智能技术提供先进生产力的企业，涉足领域较广，也包含生命科学领域。如：镁伽科技、科亿科技等，此类企业多为药企提供产品的缺陷检测。

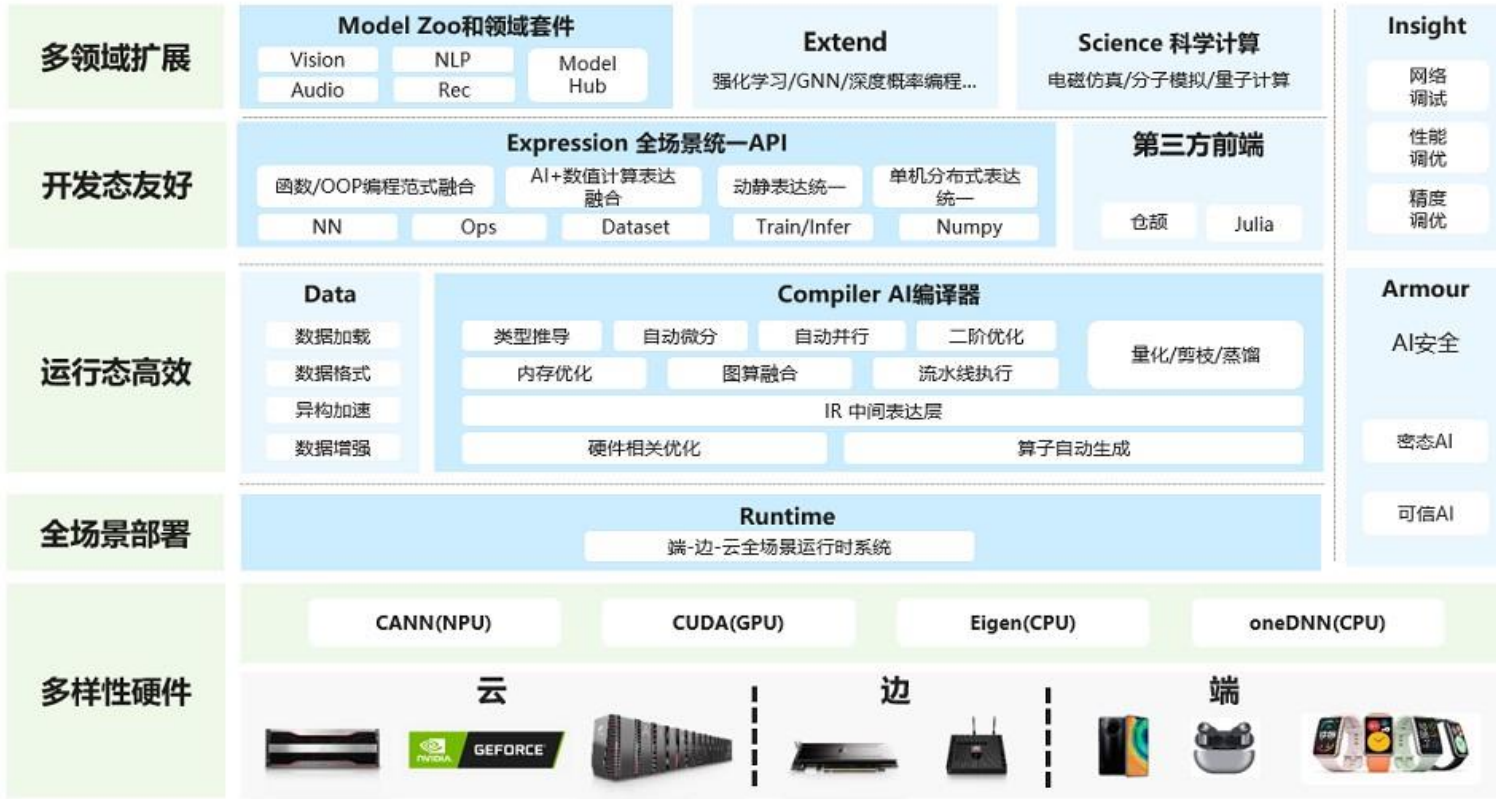
企业	主要技术及应用场景
镁伽科技	MEGAPLANT:镁伽针对药物发现、化合物合成和工业产品表面缺陷检测等行业需求，搭建了MegaPlant AI算法平台，融合了包括小样本学习、强化学习和图计算等技术在内的CV与NLP算法。
科亿科技	- 为各工业领域提供视觉成像、算法算力及自动控制的AI系统解决方案。 - 基于深度学习的AI图像视觉检测融合了AI算法云平台、边缘计算技术、嵌入式处理平台、定制专用相机、光学成像系统、定制化的人机交互系统、远程运维平台等技术。 - 可为药企提供产品的缺陷检测，如：黑点、裂纹、异物、破损等。

此外，太仓中科信息技术研究院携手华为推出智晶系列智能视觉检测产品。该产品基于昇腾AI基础软硬件平台，融合迁移学习、数据增强、弱监督检测等AI模型算法，解决了对泡罩药品跨产品、跨型号、跨类别的通用检测难题，为制药行业打造药品封装缺陷检测的数智化解决方案。

智晶系列智能视觉检测核心产品包括：

A.智晶 ZMAI 视觉检测软件：基于昇思MindSpore AI 框架和MindX深度学习视觉SDK；B.智晶工业AI视觉检测平台：可实现大规模数据的快速训练和迭代，建立全面的缺陷检测模型库；C.智晶药品检测设备及模组：基于昇腾Atlas 300 AI推理卡，独立检测设备ZMX-100可直接对药板的正反面包装缺陷及药品缺陷进行全方位检测；独立视觉检测模组ZMX-320可加装在药品泡罩机上，在高节拍快速检测的同时实现识别精度的有效提升。

图表1 昇思MindSpore总体架构



图源：MindSpore官网

■ AI赋能药物市场开拓与商业化

■ 最新技术举例

AI营销是指利用人工智能技术和算法来辅助和改进市场营销活动的过程。它结合了大数据分析、机器学习和自然语言处理等技术，通过对海量数据的处理和分析，提供个性化、精准和智能化的营销解决方案。在AI营销中，人工智能可以帮助企业实现的目标包括：市场分析和预测、目标受众定位、个性化营销和推广、自动化营销、客户关系管理，以及营销决策支持等^[1]，这些应用可以帮助药企更好地理解市场和消费者需求，以提供个性化的营销和推广。

此外，AI在药物市场开拓中，一个重要的作用为帮助产品定价，确定价格的方法为利用其可模仿人类专家思维的能力，对产品生产后的定价控制因素进行评估^[2]。决定创新和仿制药价格的因素包括：药品研发过程中的支出、相关国家的价格监管制度、独占期的长短、专利到期一年后创新药的市场份额、参照产品的价格以及定价政策等^[3]。

目前已有企业将AI应用到药品市场开拓与商业化中，如 Intelligence Node、Veeva Systems、Aktana、DeepIntent等等，Node推出的 In competitor 是一个零售竞争情报平台，可分析竞争对手的定价数据，帮助零售商和品牌商监控竞争对手。

在AI药物市场开拓中，用到的AI技术主要包括：NLP、ML、数据挖掘和大数据分析等。在 ML 中，软件会分析大量统计数据，如产品开发成本、产品市场需求、库存成本、制造成本和竞争对手的产品价格等，然后开发出预测产品价格的算法。

推荐系统

- 推荐系统是一种常见的算法模型，用于根据用户的历史行为和偏好，为其提供个性化的药物推荐。这些推荐系统可以基于协同过滤、内容过滤和深度学习等技术进行建模和优化。

文本分类和情感分析

- 在药物营销中，对医学文献、患者反馈和社交媒体数据进行文本分类和情感分析是常见的任务。这些任务可以利用机器学习算法，如支持向量机（SVM）、朴素贝叶斯分类器和深度学习模型（如卷积神经网络和循环神经网络）来实现。

预测模型

- 药物营销中的预测模型可以用于预测市场需求、药物销售量和患者行为等。这些模型可以基于回归分析、时间序列分析和机器学习算法（如决策树、随机森林和梯度提升树）来构建。

聚类分析

- 多角度应用于药物营销中的市场分析，为市场营销战略和策略的制定提供科学合理的参考。

深度学习模型

- 深度学习模型在药物营销中也得到广泛应用。例如，利用卷积神经网络（CNN）进行图像识别，可以用于识别和分类药物图像。另外，循环神经网络（RNN）和变换器模型（Transformer）等也可用于处理序列数据和自然语言处理任务。

[1] doi: 10.3386/w24282
[2] doi.org/10.1016/j.eswa.2008.09.014
[3] doi.: 10.1016/j.jval.2016.03.816

国内企业布局

2023年全球经济不确定性增强，市场承压、竞争加剧，传统医药营销模式下，成本高、效率低、转化差、手段单一所带来的问题更加凸显，这也倒逼企业加速转变营销模式与技术策略，开辟第二增长曲线。而通过人工智能技术可以更准确地了解患者需求，从而确定更适宜的药品营销策略，基于人工智能的分析，也能帮助制药公司缩短市场反馈周期，从而优化药品推广策略等。

在此背景下，国内在AI赋能药物商业化上迎来一定进展，据药融数据统计，目前国内赋能药物市场开拓及商业化企业共六家。通过对其服务内容和应用场景进行分类，我们发现，企业提供服务内容大致可分为：与KOL的沟通交流、市场准入策略制定、销售预测、竞品对比以及话术合规。

企业	服务内容及应用场景	
百奥知	MedAI eMAF解决方案 知识图谱 Knowledge graph 学术网络 Academic Network	AI赋能市场策略制定，学术推广。 - 竞品主题对比分析，产品优势主题发现，产品研究热点发现，客户沟通材料制作。 - KOL动态跟踪和管理。
零氪科技	Data Insights：数据洞察 Digital Marketing：数字化营销	- 剖析市场状况和模式，开发定制化数据分析包，阐明患者在疾病发病率高或对该特定疗法认识较高的地区之间的分布。优先选择新药发布区域。 - 发现不同地区的趋势，监控特定营销策略的执行情况。
开心生活科技	AI赋能真实世界证据的产出	为企业进行药物经济学研究，制定市场准入策略。
云势软件 SaaS服务商	大数据和人工智能套件 GeniusDATA 智能数据 GeniusSALES 销售预测 GeniusMKT 智能营销 GeniusAIDE 智能助理	为制药、医疗器械、动物健康、作物科学、营养保健等行业公司及CRO、医院等临床相关机构提供信息自动化管理云端解决方案。 销售市场（一站式SFE管理平台）为其核心业务环节之一。
火石数智	Chatbot：医学对话机器人 Social Listening——医/患诉求洞察服务	通过聆听社交媒体收集的指标可帮助企业了解其与竞争对手相比的位置，市场份额和差异特征。 亦可通过研究目标药品与其他药物的联用情况，了解药品使用情况，为延长药品成熟期周期提供重要指标参考
沃丰科技	智能知识库 智能沟通 智能SCRM 智能合规	虚拟代表解决方案 话术合规流程、增强话术传达、数据可视化

■ AI逐步应用于药物警戒领域

药物警戒(Pharmacovigilance-PV)：药物警戒是与发现、评价、理解和预防不良反应或其他任何可能与药物有关问题的科学研究与活动。

目前，AI已在PV管理中发挥了非常重要的作用。AI，包括ML方法，如自然语言处理和深度学习，可以检测和提取药物不良事件的信息，从而使药物警戒过程自动化，改善对已知和记录的药物不良事件的监测。此外，随着对远程医疗服务需求的增加，在管理急性和慢性疾病方面，AI可以在检测和预防药物不良事件方面发挥作用。以数据智能技术为支撑，健全药品安全监管体系，是完善健康保障的重要一环。

哈佛大学对基于ML的AI的使用进行了范围审查，检索了PubMed、Embase、Web of Science和IEEE Xplore数据库，以识别2000年至2021年9月发表的与ML在药物警戒中的使用相关的文章，结果显示，大多数研究（53%）侧重于使用传统统计方法检测安全信号。在使用较新的ML方法的研究中，61%的研究使用了现成的技术并稍作修改。时间分析显示，近年来，深度学习等较新的方法使用越来越多。AI的进步尚未完全渗透到药物警戒领域（尽管最近的研究表明这可能正在改变）。在整个药物警戒管道中，存在着实施ML方法的机会。

FDA将AI应用于PV以提高IVSRs分析的效率和科学价值时间相对较早，除了每年从制药行业收到的近200万份FAERS报告外，FDA还将公众直接提交给其的数十万份报告处理后传输至FAERS数据库。从对FDA工作进展的描述中可以看出，AI用于PV时，大多数工作都涉及到使用NLP（自然语言处理）从ICSRs叙述中自动提取涉及因果关联性评价的关键性特征，少数工作试图开发可预测的机器学习（ML）算法，试图将ICSRs关键信息要素提取、整合和分析的人类认知过程自动化。

FDA近年来将人工智能应用于药物警戒活动的主要进展见表：

图表1 FDA将人工智能应用于药物警戒的主要进展

应用主题	具体进展
ICSRs关键信息提取和分析	用NLP提取个例报告中的时间信息
	人口统计学信息和临床概念的提取
	应用NLP和ML总结ICSRs的关键特征
ICSRs关联性评价	发了用于提取临床特征的LP和用于定义不良反应判定评价标准（如过敏反应）的机器学习分类
	用ML预测哪些ICSRs对因果关联性评价最有价值
	NLP通过提取临床数据并使其可视化（如时间性），以支持因果关联性评价的病例系列分析
	ML算法，用于识别无法评价的病例报告（例如，包含不充分信息的ICSRs,因此无法支持因果关联性评价)
其他	用NLP和统计聚类算法/树络分析来识别相似疾病的报告
	药品标签/包装插页中提取不良事件并对其进行编码
	对ICSRs开发了去除重复报告的算法

[1] 《中国药物警戒》2023年

国内企业布局

根据已发布的2022年国家药品不良反应监测年度报告，2022年，全国药品不良反应监测网络收到不良反应报告202.3万份；1999年至2022年，全国药品不良反应监测网络累计收到不良反应报告2085.6万份^[1]。鉴于每年接受的庞大ICSRs（个例安全性报告）数量，如何处理、分析、评价和利用这些庞大而异构的海量数据，如何更好地利用ICSRs为上市后药品监管工作服务，是我国药物警戒（PV）工作的重大挑战之一。PV系统的两个主要构成部分包括单个案例处理：即个人不良事件的收集、解读和报告，以及信号检测，对AE数据进行持续性的趋势监测以判断是否有未知的安全信息会对风险获益评估结果产生影响。

我国在人工智能用于PV上起步较欧美国家晚，但当前，基于真实世界数据和数据智能技术的药物警戒系统正在重塑我国药品监管格局。

目前国内利用AI技术提供药物警戒服务的企业约七家左右，这些企业大多集中在北京，且类型均为Saas 服务。除洞悉网络外，其他公司除药物警戒外还提供药物研发生命周期的其他服务。而成立于2016年的洞悉网络专注医药安全，采用“全生命全产业链医药安全产业数字化SaaS+安全重度垂直服务”的模式，面向医药行业，提供覆盖临床试验到上市后全生命周期的药物警戒及风险管理服务。

图表1 国内AI+药物警戒企业清单

企业	地址	技术平台及提供服务
深度智耀	北京	Deep-PV 依靠自然语言处理技术，实现：个例药物不良反应/不良事件报告；可疑安全信号检测；报告自动生成。
开心生物科技	北京	- HLT eMediSafe药物警戒管理系统 - 收集、处理、上报与分析药品安全相关数据的管理系统 - 核心功能：个例报告处理、各类表格导出、上报与反馈、报告翻译、年度报告撰写
云势软件	北京	VirgoPV 药物警戒系统 为企业建立药物安全性数据库，整合各阶段药物的数据，为企业的数据分析做基础数据体系。
零氪科技	北京	真实世界研究服务 为药品上市后的安全性和有效性提供等提供真实世界证据
爱特曼	北京	SUSAR报告端到端解决方案 针对药物警戒SUSAR报告处理中涉及的CIOMS翻译、提取、整理和分发等流程中的问题进行优化，服务药企应对SUSAR报告site分发环节。
北京洞悉网络 SaaS	北京	Holox®Clinical、Holox®Pharma、Holox®Vaccin、Holox®RWE
神州医疗	北京	为医药企业/CRO公司提供采集、管理、上报药品安全性数据
太美医疗科技	杭州	eSafety药物警戒系统

[1] 《中国药物警戒》2023年
[2]来源于公开信息，药融咨询整理

06

第六章

AI制药行业总结及未来展望

■ 现状分析及展望

■ AI新药研发技术迎来成果检阅，多款药物进入临床，但尚无药物成功上市

尽管全球已有多款AI研发药物进入临床，且最高进展已到临床三期，与此同时，有很多利用AI技术研发的药物进入临床后失败，目前尚无AI研发的药物成功上市。如：2023年 7 月 31 日，日本住友制药和大冢制药宣布，其合作在研的 ulotaront 药物的两项 III 期研究精神分裂症临床试验未能达到主要终点。Ulotaront 是一款利用人工智能技术发现，具有 5-HT1A 激动剂活性的TAAR1激动剂，目前正在研究用于治疗精神分裂症、广泛性焦虑症和重度抑郁症的辅助治疗，并正在考虑其他适应症。由Exscientia和日本住友制药合作开发DSP-1181的药物，是一种用于治疗强迫症(OCD)的长效血清素5-HT1A受体激动剂，这是全球首个由AI设计的分子，但因临床I期研究未达标而停止研发。

英国头部AI药企Benevolent AI公布其治疗特应性皮炎的局部泛Trk抑制剂BEN-2293的IIa期临床没有达到次要疗效重点，Relay Therapeutics在AACR2023会议上披露了选择性PI3K α 抑制剂RLY-2608的临床数据有效性不佳。

Recursion（Nasdaq：RXXR）开发的REC-3599用于治疗一种罕见病—GM2神经节苷脂沉积症临床一期结束后下阶段的临床试验迟到2年也已停止研发。

图表1 国内借助AI进入临床阶段药物及进展

企业	名称/代码	临床阶段	药物类型
英矽智能	ISM001-055	Phase 2 Clinical	小分子药物
红云生物	H002	Phase 2 Clinical	小分子药物
锐格医药	RGT-075	Phase 2 Clinical	小分子药物
Exscientia;湃隆生物	GTAEXS617	Phase 2 Clinical	小分子药物
冰洲石生物	AC0682	Phase 1 Clinical	小分子药物 (PROTAC)
未知君	XBI302	Phase 1 Clinical	微生物疗法
未知君	LBP01	Phase 1 Clinical	微生物疗法
未知君	LBP02	Phase 1 Clinical	微生物疗法
剂泰医药	MTS004	Phase 1 Clinical	小分子药物
宇耀生物	YY001	Phase 1 Clinical	小分子药物
睿健医药	NouvNeu001	Phase 1 Clinical	细胞治疗
锐格医药	RGT-028	Phase 1 Clinical	小分子药物
锐格医药	RGT-419B	Phase 1 Clinical	小分子药物
锐格医药	RGT-264	Phase 1 Clinical	小分子药物
费米子	FZ002	Phase 1 Clinical	小分子药物
中以海德	HDM-8421034	Phase 1 Clinical	小分子药物

■ AI制药存在“数据困境”，但并非无法破局

数据是AI的三大核心要素之一，一方面，AI能够从海量的生物医药数据中挖掘出价值点；另一方面，AI药物研发又普遍受到数据之苦。AI/ML模型本质上是数据驱动的，对于训练、测试和验证的数据的属性或特征特别敏感，然而数据需要考虑偏见、完整性、隐私和安全、出处、相关性等各种问题。鉴于AI/ML模型本质上是数据驱动的，基于获取的各种来源的训练数据不断升级，这使得模型很容易纳入人为偏见，所以，在开发各种AI/ML模型时，应尽一切努力获取平衡、客观和高质量的训练数据集。

现阶段，大多数公开数据库来源于文献和实验室数据以及通用数据库。但目前存在诸多问题，如：由于生成数据的实验条件和记录数据的格式或注释往往不相同，存在数值不一致、数据缺失、误差等问题，难以被计算机统一识别。且医药领域普遍存在“数据孤岛”，大多数高质量和核心数据掌握在少数企业中，保密性极强，公开可能性较小，制约了行业的整体向前发展。大型通用数据库能够对初始生成模型进行预训练，用来评估其模型能力。但公开数据集一般质量参差不齐，需要自主进行筛选过滤和标准化。针对不同的场景，基于不同的数据维度与颗粒度进行数据清洗处理，还需要较大的投入。

与此同时，从偏差的数据中，亦可挖掘出有价值的信息。2023年8月31日，发表在NEJM综述表示：开发医学人工智能工具的时候，“偏差的数据”，比如人群代表性差的数据、信息缺失的数据等等，也具有其价值，不应该只简单丢弃或者技术上的“矫正”（特别是有些“矫正”缺乏依据，加剧偏差）。这些“偏差的数据”蕴含着丰富的信息，比如，它们就像“文物/遗迹”一样，可以反映当时或现在的社会、制度状态；将来结合创新的人工智能工具，就可以很好的总结这种“数据偏差模式”，帮助人们进一步分析偏差成因，并促进更公平的医疗。

■ AI赋能药物开发多集中在药物发现阶段，临床试验阶段应用或为AI带来新方向

全球范围内，AI制药均主要集中在药物发现环节。可能原因在于，该环节以化学过程为主，研究人员对候选化合物数据的完整性及可重复性、化学稳定性、理论认知度等都有较好的把握，有利于AI建模。

但临床试验阶段有较大不同，是以生物学过程为主，其复杂性在数据和AI建模两方面带来巨大挑战。在数据方面，需要将临床数据加以结构化处理，而诸如病历、随访记录目前还很难标准化、数字化；此外，由于涉及患者隐私，目前尚无能充分保障数据安全的有效措施，这也限制了临床数据的灵活运用。在AI建模方面，化合物与人体靶点的反应过程非常复杂，但目前理论认知还不足，受环境影响因素很大，数据稳定性和可重复性差，所以不利于AI建模，对疗效和安全性的影响也难以把控。

因此，对于真实可信、覆盖临床各环节的完整临床数据库，还有较大发展空间。

■ “黑盒”算法带来安全风险，突破其不可解释难题意义显著

AI算法“黑盒”问题是指由于广泛应用在AI产品上的深度学习等主流算法模型内部结构复杂、运行过程自主性较强且人工无法干预等因素，在数据输入、模型训练、结果输出等方面出现**运行机制难以解释**的情况，人工智能的内部工作原理对最终用户来说是不可见的，导致运行结果无法完全掌控。

当人工智能模型是“黑盒”时，很难理解模型是如何得出预测以及建议的。人工智能的可解释性提供了对人工智能模型内部工作的深入了解，例如有助于预测的潜在因素、特征或模式，并解释了这些模型的决策过程和输出。目前深度学习作为AI技术的主要算法之一，其特点是通过大量的训练数据对模型训练，最终确保在特定输入数据下通过“黑盒”运行，得到更加智能、精准的输出结果。

在生命科学行业，人工智能越来越多地被用于在可能危及生命的地方做出具有深远影响的关键决策和预测，如疾病诊断、治疗计划或药物开发，**可解释性对于确保透明度、可信度和监管合规性至关重要**。如：浙江大学指出，由于历史数据的偏差，曾发生过算法对肺炎患者诊断出错等问题。从产业实践上看，目前各家科技公司基本利用自身所搜集的有限数据进行测试，各公司宣称“安全”的AI产品往往具有很多较难发现的安全漏洞。因此，一旦未经充分安全测试的相关产品大范围应用于医疗/制药领域，将暗藏错误推理、错误决策等较大安全隐患。

对“黑盒”算法进行更多可解释性的分析研究，使人工智能模型的输出更具透明度，如：对于小分子，如果结合自由能的预测，或是基于靶点口袋的分子生成，或是对先导化合物进行优化改造，将微观物理的信息融入到深度学习框架之中，这个模型就能够更好地学习到真实的因果关系或物理规律，模型的鲁棒性和可靠性也会得到大幅提升。或许会成为未来的主要发展趋势之一。

■ 集采、医疗反腐背景冲击下，AI+市场营销迎来新机遇

2023年7月21日，国家卫健委同教育部、公安部、审计署等十部门联合召开视频会议，部署开展为期1年的全国医药领域腐败问题集中整治工作。随后多个省份已发文，跟进集中整治。此次反腐行动力度大，覆盖范围广，强调针对医药领域生产、供应、销售、使用、报销等重点环节和“关键少数”，深入开展医药行业全领域、全链条、全覆盖的系统治理。

在此背景下，医药行业对于药品学术推广及市场营销等的合规性要求愈加严格，而传统质检难以全面覆盖销售推介和客户服务交流的内容，存在极大的合规风险。但智能质检可提供通话记录全量质检，不遗漏任何违规项，如过度承诺、询问药量等，在医疗反腐大背景下可能迎来一定的市场机会。此外，在药物市场开拓中，AI技术辅助营销可以及时向医生提供专业医学内容，帮助医生提高专业知识储备、满足研究需求等。

■ 上市AI制药企业商业化模式

纵观全球上市的AI制药公司，从业绩上讲，盈利能力较低，年度营收过亿的企业仅有2家。鉴于目前尚未有真正意义上的AI药物上市，AI制药行业目前商业模式主要分为：

- AIDD相关软件销售/技术平台授权：该模式毛利率极高，但技术门槛高且集聚效应强，市场规模优先且少数头部企业几乎实现垄断，通过该模式实现盈利企业数量较少，如Schrödinger；
- 转移研发风险的CRO模式：帮助企业完成药物发现工作，收取授权费。相较于创新药物研发，变现难度相对低，回款周期较短，是多数企业的选择，如AbCellera Biologics；
- 自行承担研发失败的风险的合作/项目转让：通常采取里程碑式付款，目前首付款比例有下降的趋势；
- “财务投资+项目管理”模式：低价买入药企闲置的临床管线，减少研发成本，后续自行研发药物，如Roivant Sciences、Erasca

图表1 全球AI制药上市企业信息汇总

企业名称	上市时间	技术领域	上市地区	临床管线数量	临床管线治疗领域	2022年营收（亿美元）
AbCellera Biologics	2020年12月	抗体药物发现、药物筛选	美国	AI-CRO		4.85
Absci Corp	2021年7月	分子生成与优化（SoluPro）	美国	AI-CRO		0.06
Benevolent AI	2022年4月	靶点发现、分子生成与优化（Benevolent Platform）	荷兰	1	特异性皮炎（已宣布失败）	0.21
C4X Discovery	2007年1月	基于基因组的靶点发现（Taxonomy3）、分子生成与优化（Conformetrix）	英国	1	神经系统疾病	0.04
Exscientia	2021年10月	靶点发现、分子生成与优化、ADMET性质预测、临床预测（CentaurAI）	美国	5	肿瘤 神经系统疾病 炎症	0.27
Erasca	2021年7月	药物发现（OPRA）	美国	4	肿瘤 （临床管线为引进）	0.09
Evaxion Biotech	2021年2月	癌症个性化治疗靶点发现（ObsERV）	美国	2	肿瘤	--
Icosavax	2021年7月	蛋白质分子设计	美国	3	疫苗	0.01
Landos Biopharma	2021年2月	靶点发现、精准医疗（LANCE）	美国	1	克罗恩病	--
Recursion Pharmaceuticals	2021年4月	靶点发现、表型筛选（Recursion Operating System）	美国	4	肿瘤 微生物感染	0.40
Relay Therapeutics	2020年7月	分子动力学模拟、药物筛选（ZebiAI）	美国	3	肿瘤	0.01
Roivant Sciences	2021年10月	分子动力学模拟、蛋白质降解深度学习模型	美国	1	炎症	0.55
Schrödinger	2020年2月	分子动力学模拟、分子生成与优化、性质预测	美国	1	肿瘤	1.81
Syntekabio	2019年12月	药物发现（DeepMatcher）	韩国	0	-	0.36
Valo Health	2021年6月	靶点（生物标记物）发现、药物筛选、分子优化、临床预测（Opal Platform）	美国	3	肿瘤、心血管疾病	--

■ 全球主要AI制药上市企业

AbCellera Biologics

AbCellera Biologics，主营抗体发现和开发平台AbCelera，专门设计用于解决传统抗体发现的障碍，以更高的精度和速度找到最佳的临床候选药物，更快地到达诊所和患者。借助平台，公司搜索范围广泛的不同抗体，快速选择最有效和可开发的先导物，并通过授权、交易先导化合物及建立合作伙伴，开发优化的解决方案实现盈利。公司与40个合作伙伴签订超174个项目合同，其中149个项目包括里程碑和版税。

Exscientia

Exscientia是全球具有代表性AI驱动的制药企业，采取和外部企业合作形式，共同推进研发管线，利用广泛合作沉淀更多的数据支持其算法模型进行迭代和优化。Exscientia利用已开发的人工智能平台进行自动化药物的研发指导，利用大数据和机器学习方法，根据已有药物研发数据自动设计的小分子化合物，并根据药效、ADMET等条件对化合物进行评估和筛选，对筛选出来的化合物进行实验检测，并反馈到AI系统中进行筛选。Exscientia内部管线专注于肿瘤领域的药物开发，而合作管线则侧重于其他治疗领域。

Schrödinger

Schrödinger 成立于 1990 年，是 AI 制药领域中的明星企业。Schrödinger的核心技术平台是物理计算平台，即基于物理学的预测方法和机器学习技术，为预测模型、数据分析、合作等提供整合差异化解决方案，以加速药物发现。Schrödinger采取“软件营收+新药研发”的布局，与预营收模式的药企相比，具备客户稳定且毛利率极高的分子模拟技术及相关软件销售兜底（全球TOP 20药企均为其客户，且使用时间超10年），而与AI平台公司相比，其新药研发业务更具潜力。

Valo Health

Valo Health成立于2019年。Valo的Opal平台能够通过分析人类数据，发现专有的新分子、遗传标记和疾病之间的未知关联。Opal是一个完全集成的、组件化的、端到端的药物开发平台，采用云计算和人类数据。这个计算平台由于减少了药物开发过程的成本、持续时间和失败次数而使其更加高效。

C4X Discovery

C4X Discovery是一家药物发现公司，将AI技术与尖端的药物发现技术相结合，有效地发现世界领先的创新药物。公司致力于发现和开发小分子药物，治疗免疫炎症性疾病，以高效地发现世界领先药物。同时，C4X Discovery与国内外制药公司合作（如阿斯利康、Horizon、成都先导等），致力于开发出更好、更安全、更有效的药物。

Recursion Pharmaceuticals

Recursion Pharmaceuticals，是一家利用计算机视觉技术处理细胞图像并分析细胞特征以评估患病细胞的药物后反应结果的公司。利用成像技术和AI技术进行高通量的细胞模型实验。公司致力于在数百种疾病的细胞模型中测试数千种候选药物，最终找出不同疾病对应的新药。

Relay Therapeutics

Relay Therapeutics是一家以分子动力学模拟见长的AI药物发现公司，旨在将蛋白质结构和运动的深刻理解应用于药物发现。该公司将基因组数据、计算科学以及实验科学深度融合，致力于改变药物发现的过程。

Roivant Sciences

Roivant Sciences成立于2014年，专注孵化创新药企。Roivant旗下现拥有12家子公司涉猎AI制药、数据整合与分析、免疫疗法等新兴技术。Roivant Sciences成立AI+CRO子公司VantAI，主要为大药企提供从头药物设计、靶点预测、蛋白降解剂的计算设计和优化等服务。收购蛋白降解公司Oncopia Therapeutics，形成子公司之间的协同布局。同时创办新公司Psivant Therapeutics搭建计算物理驱动的药物设计平台。

Icosavax

Icosavax成立于2017年，主要利用华盛顿大学蛋白质设计研究所授权的类病毒颗粒技术，搭建技术平台用于新型高效疫苗的发现。VLP平台技术旨在实现复杂病毒抗原的多价、基于粒子的呈递，这将诱发对特定病毒的保护。Icosavax的产品线包括针对呼吸道合胞病毒（RSV）、人类偏肺病毒（hMPV）和严重急性呼吸道综合征冠状病毒2（SARS-CoV-2）的候选疫苗。

Erasca

Erasca是一家开发RAS/MAPK途径驱动的癌症疗法的生物技术公司，其目标是开发新一代抗癌药物，彻底治愈癌症。Erasca采用了一种模式诊断方法，通过小分子治疗药物、大分子治疗药物和蛋白质降解剂，选择性地、有效地抑制或降解RAS/MAPK途径中的关键信号节点。OPRA（肿瘤模式识别算法）是Erasca专有的人工智能药物发现平台，它利用机器学习等先进的计算工具，通过阐明新颖的肿瘤生物学和创新策略来加速药物发现。

Evaxion Biotech

Evaxion成立于2008年，公司使用免疫学、生物信息学和机器学习方法，来发现和开发针对感染和癌症的有效疫苗。Evaxion的两个平台PIONEER和EDEN可以预测、排列和优化表位和抗原。EDEN能快速准确地揭示新的抗原，而PIONEER则能识别引发自然保护性免疫反应的变异蛋白质表位。

图表1 全球AI制药企业临床管线汇总^[1]

药物名称	公司名称	临床阶段	药物类型	治疗领域
REC-2282	CereXis	Phase 3 Clinical	小分子药物	肿瘤
RLY-4008	Relay Therapeutics	Phase 2 Clinical	小分子药物	肿瘤
BEN-2293	BenevolentAI	Phase 2 Clinical	小分子药物	特应性皮炎
OPL-0401	Valo Health	Phase 2 Clinical	小分子药物	视网膜病变
BPM31510	BERG	Phase 2 Clinical	小分子药物	肿瘤
ISM001-055	英矽智能	Phase 2 Clinical	小分子药物	特发性肺纤维化
BMF-219	Biomea Fusion	Phase 2 Clinical	小分子药物	肿瘤
H002	红云生物	Phase 2 Clinical	小分子药物	肿瘤
RGT-075	锐格医药	Phase 2 Clinical	小分子药物	糖尿病
GTAEXS617	Exscientia;湃隆生物	Phase 2 Clinical	小分子药物	肿瘤
gaboxadol	Healx	Phase 2 Clinical	小分子药物	神经系统疾病
REC 4881	Recursion Pharmaceuticals	Phase 2 Clinical	小分子药物	肿瘤
REC-994	Recursion Pharmaceuticals	Phase 2 Clinical	小分子药物	肿瘤
NDI-034858	Nimbus Therapeutics (交易至武田制药)	Phase 2 Clinical	小分子药物	银屑病
NDI-101150	Nimbus Therapeutics	Phase 2 Clinical	小分子药物	肿瘤
IVX-A12	Icosavax	Phase 2 Clinical	疫苗	微生物感染
OPL-0301	Valo Health	Phase 1 Clinical	小分子药物	心血管疾病/肾损伤
Valo-C1	Valo Health	Phase 1 Clinical	小分子药物	肿瘤
INDV-2000	Indivior	Phase 1 Clinical	小分子药物	药物成瘾
EXS-21546	Exscientia	Phase 1 Clinical	小分子药物	肿瘤
PHI-101	Pharos iBio	Phase 1 Clinical	小分子药物	肿瘤
RLY-1971	Relay Therapeutics	Phase 1 Clinical	小分子药物	肿瘤
SNX-281	Silicon therapeutics	Phase 1 Clinical	小分子药物	肿瘤
SGR-1505	Schrödinger	Phase 1 Clinical	小分子药物	肿瘤
RLY-2608	Relay Therapeutics	Phase 1 Clinical	小分子药物	肿瘤
REC-163964	Recursion Pharmaceuticals	Phase 1 Clinical	小分子药物	微生物感染
DSP-0038	Exscientia	Phase 1 Clinical	小分子药物	神经系统疾病
BPM31543	BERG	Phase 1 Clinical	小分子药物	药物成瘾
AC0682	冰洲石生物	Phase 1 Clinical	小分子药物 (PROTAC)	肿瘤
XBI302	未知君	Phase 1 Clinical	微生物疗法	肿瘤
LBP01	未知君	Phase 1 Clinical	微生物疗法	溃疡性结肠炎
LBP02	未知君	Phase 1 Clinical	微生物疗法	糖尿病
MTS004	剂泰医药	Phase 1 Clinical	小分子药物	假性延髓情绪

药物名称	公司名称	临床阶段	药物类型	治疗领域
YY001	宇耀生物	Phase 1 Clinical	小分子药物	肿瘤
NouvNeu001	睿健医药	Phase 1 Clinical	细胞治疗	神经系统疾病
RGT-028	锐格医药	Phase 1 Clinical	小分子药物	内分泌与代谢
RGT-419B	锐格医药	Phase 1 Clinical	小分子药物	肿瘤
RGT-264	锐格医药	Phase 1 Clinical	小分子药物	肿瘤
FZ002	费米子	Phase 1 Clinical	小分子药物	疼痛
HDM-8421034	中以海德	Phase 1 Clinical	小分子药物	微生物感染
IVX-121	Icosavax	Phase 1 Clinical	疫苗	微生物感染

[1] 红字表明该企业为国内AI制药企业

图表1 全球AI制药企业大宗项目合作交易^[1]

企业名称	合作企业名称	交易/合作时间	交易合作内容	潜在交易金额	预付款
辉瑞	PostEra	2022年1月	合作建立AI实验室，以肿瘤药物/新冠病毒为初期研究重点	2.48亿美元	1300万美元
	CytoReason	2022年9月	获得CytoReasonAI技术平台和疾病模型的使用授权+1项商业合作协议	1.1亿美元	2000万美元股权投资
默克	Iktos	2023年3月	战略投资	1550万欧元	1550万欧元
	Exscientia	2023年9月	合作开发小分子药物项目	5.94亿美元	数百万美元
	Benevolent AI	2023年9月	合作开发小分子药物项目	6.74亿美元	2000万美元
艾伯维	Anima	2023年1月	合作开发小分子mRNA药物项目	5.82亿美元	4200万美元
罗氏	Relay Therapeutics	2020年12月	口服小分子SHP2变构抑制剂全球独家许可	7亿美元	7500万美元
	Recursion	2021年12月	约40个新药项目合作	120+亿美元	1.5亿美元
	Tempus	2022年11月	获取Tempus AI临床试验设计平台+患者数据库使用权	/	7000万美元
Celgene (2019年BMS收购)	Exscientia	2019年3月	合作开发PKC-θ药物项目	/	2500万美元
BMS	Schrödinger	2020年11月	5个新药项目合作	27亿美元	5500万美元
	Exscientia	2022年	合作开发免疫学和炎症药物项目+AI平台使用费用	/	1500万美元
	Evotec	2022年3月	利用 EVO panOmics和EVO panHunter以及基于AI/ML的药物平台开发新药	50亿+美元	2亿美元
	Owkin	2022年6月	使用AI平台设计心血管疾病临床试验	1.8亿美元	8000万美元
	Exscientia	2022年1月	合作开发15款肿瘤或免疫相关的小分子新药	52亿+美元	1亿美元
赛诺菲	Owkin	2021年11月	独家投资+合作开发4种癌症新疗法	2.7亿+美元	1.8亿美元
	天演药业	2022年3月	利用AI技术设计并开发4种单抗/双抗药物	25亿美元	1750万美元
	Atomwise	2022年8月	合作开发5个靶点相关的药物	10.2亿美元	2000万美元
	英矽智能	2022年11月	达成战略研发合作	12亿美元	2150万美元
	百图生科	2023年10月	基于百图生科生命科学大模型，合作开发生物药物发现新模型	10亿+美元	1000万美元
阿斯利康	Scorpion Therapeutics	2022年1月	合作开发针对难以靶向的癌蛋白药物	/	7500万美元
GSK	Tempus	2022年10月	续签Tempus AI临床试验设计平台+患者数据库使用权	/	7000万美元

图表1 全球AI制药企业大宗项目合作交易^[1]

企业名称	合作企业名称	交易/合作时间	交易合作内容	潜在交易金额	预付款
武田	Anima	2021年3月	合作开发针对早期亨廷顿病mRNA药物项目	10.5亿美元	1.2亿美元
	Evozyne	2022年4月	合作开发用于4种罕见病靶点的蛋白质药物	4亿美元	/
	Nimbus Therapeutics	2023年2月	收购1款AI设计用于治疗银屑病的临床阶段药物	60亿美元	40亿美元
礼来	Anima	2018年7月	合作开发针对多款未公开靶点药物	10.5亿美元	4400万美元
	Genesis	2022年5月	利用Genesis的AI药物发现平台，发现针对5个靶点的新疗法	6.7亿美元	2000万美元
	Schrödinger	2022年10月	1个未公开靶点的药物发现及分子优化	4.25亿美元	/
	晶泰科技	2023年5月	晶泰科技利用其特有的小分子药物发现平台研发首创新药，填补未满足的临床用药需求	2.5亿美元	/
阿斯泰来	Dyno Therapeutics	2021年12月	合作开发功能优化的新型AAV衣壳载体用于基因疗法	16亿美元	1800万美元
默沙东	Saama	2021年11月	收购Saama公司多数股权，支持其开发其AI和其他技术产品	4.3亿美元	/
	Absci Corp	2022年1月	生物制造酶的定制及3款药物合作开发	6.1亿美元	/
安进	Generate Biomedicines	2022年1月	合作开发5款药物	19亿美元	5000万美元
Exelixis	英矽智能	2023年9月	USP1抑制剂ISM3091全球独家许可协议	/	8000万美元
未披露	Benevolent AI	2014年6月	合作开发2个阿尔茨海默氏症新药项目	8亿美元	4亿美元
Recursion Pharmaceuticals	Cyclica	2023年6月	收购AI生物科技初创公司Cyclica	4000万美元	4000万美元
	Valence	2023年6月	收购AI生物科技初创公司Valence	4750 万美元	4750 万美元
BioNTech	InstaDeep	2023年1月	收购AI技术公司InstaDeep	3.62亿英镑	3.62亿英镑