中邮证券
CHINA POST SECURITIES

证券研究报告

AI4S（三）：因果数据——机制理解瓶颈突破口，AI+生物数据新基建正崛起

行业投资评级：强大于市|维持

盛丽华/徐智敏

中邮证券研究所　医药团队

中邮证券

发布时间：2026-11-04

投资要点

- **AIDD(AI-driven Drug Discovery，AI加速药物发现)新周期开启：此轮和上一轮有什么不一样？**1）主要经济体同步加码AI for Science，AI制药正逐步成为AI巨头公司的战略主航道。2026年以来，谷歌/Anthropic等巨头将AI制药视作公司未来的重要增长引擎；2）AI逐步深度参与MNC药物研发整个周期。3）投资新潮初现：以训练下一代AI大模型为目标，工业化大规模产出扰动干预湿实验数据，搭建AI数据基础设施。
- **AI制药新周期核心竞争力：大规模高质量有效数据。** 只依托存量公开数据迭代模型，边际收益持续下降。 **AI 驱动药物研发阶跃式进步极度依赖于高质量湿实验数据，而其绕不开的是对疾病机制的理解。**我们看好虚拟细胞（AIVC）成为揭示细胞系统的“编程语言”的接口，带动AI+生物基础设施建设。1）**需要什么样的数据？**行业存量数据大量是观测类数据只能得到相关性结论，**建立因果关系需要大量干预/扰动实验数据。**动态状态是构建具备真实推演能力模型的关键，是支撑AI虚拟细胞实现动态仿真的核心数据底座。**2）现存数据缺口大，专业化数据生产商正快速崛起。**当前可用高质量带标签扰动数据集，相比理论需要还差约1000倍，**因此自建实验平台建立干预/扰动实验数据是下一代AI模型竞争中的必要条件。**数据供给端已经形成多元参与者矩阵：**单细胞测序企业是重要数据来源、类器官与器官芯片企业**可产出贴近生理环境的细胞实验数据、**蛋白质组学企业**依托自研质谱平台，产出转录组难以覆盖的功能蛋白、完整的磷酸化修饰数据，部分生物医药企业也开始**对外输出临床数据资产**。单一模态数据很难完整还原细胞复杂运行逻辑，多组学融合的数据集合可以有效降低模型的认知偏差。
- **投资展望：1）湿实验需求的确定性。**发挥AI 在生物学领域的全部潜力，必须产出大量全新湿实验数据；为数据瓶颈环节对应的工具栈带来长期利好。**2）AI对CXO影响：** CXO公司一致认为AI制药将提升候选分子数量，增厚上游客户管线，拓宽疾病覆盖领域并提升相关临床/临床前研究复杂度，从而带来外包需求的提升 。过去经验显示，药物研发技术革新，工具效率提升，药企研发总投入持续上行。如果AI驱动研发投入能够带来更高回报，药企整体R&D总开支反而进一步扩张。
- **风险提示：**AI制药模型研发进度不及预期风险；投入制药领域的资本开支不及预期风险；市场竞争加剧风险；地缘政治风险；政策变动风险。

目录

一 AI制药新周期开启？此轮和上一轮有什么不一样

二 AI制药新周期核心竞争力：大规模高质量有效数据

三 投资展望

四 风险提示

一 AI制药新周期开启？此轮和上一轮有什么不一样

AI制药正逐步成为AI巨头公司的战略主航道

- **AI制药正逐步成为AI巨头公司的战略主航道。**谷歌和Anthropic将生命科学视为 AGI 终极落地场景，从模型延伸至临床阶段。字节跳动旗下Anew Labs成立于2021年，一直承担从基础模型研究到产业化的核心职能，拥有IL-17小分子项目等早期管线。英伟达做全栈基础设施或全链路自研，作为突破现有业务天花板的核心新增长极。微软等巨头选择推出专用模型 / 云解决方案，服务 B 端企业客户。

图表1：国内外科技巨头公司在制药领域布局

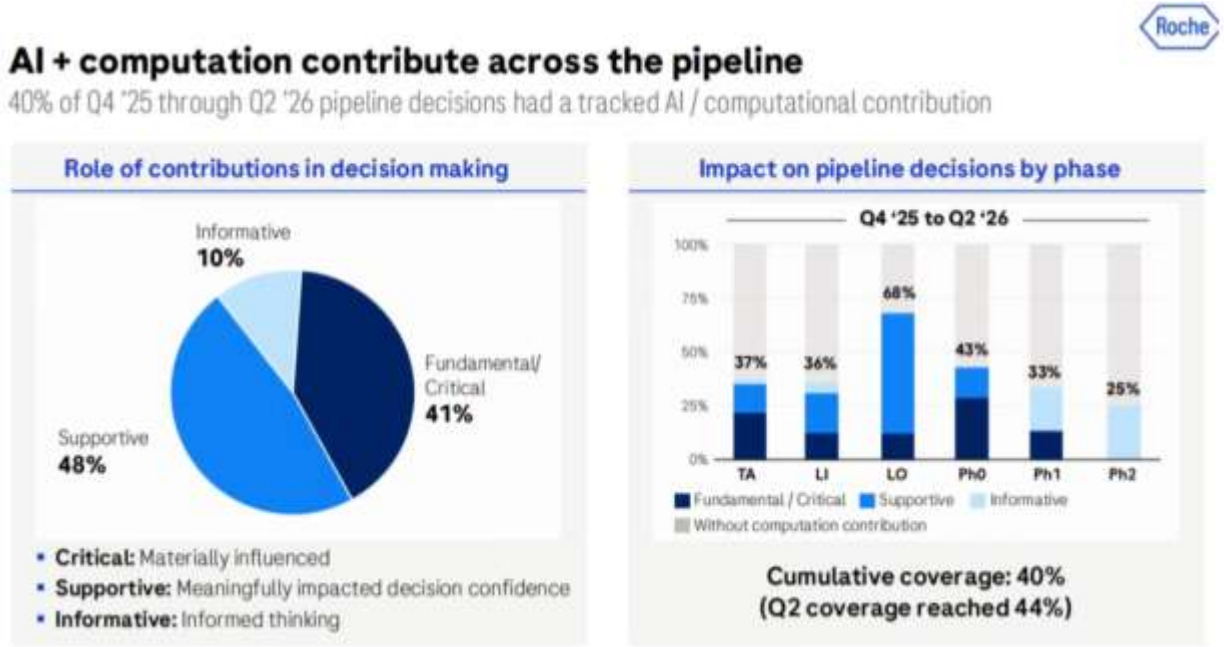
公司	可验证投入规模	投入构成说明
谷歌 / Alphabet	Isomorphic Labs 累计融资超 27 亿美元；药企合作潜在总额近 30 亿美元	2025 年 6 亿美元 B 轮、2026 年 21 亿美元新一轮融资与礼来、诺华的合作含里程碑付款；研发打包在 DeepMind 整体预算中
英伟达	与礼来 5 年合计最高 10 亿美元共建实验室	含算力基础设施、联合研发、自动化实验室；BioNeMo 平台研发年投入数亿美元级
Anthropic	4 亿美元收购 Coefficient Bio；年研发投入 3-5 亿美元级	并购 + 团队扩张 + 模型训练；生命科学团队超百人
字节跳动	AI 整体年资本开支超 2000 亿元人民币，AI 制药为重点方向	含算力、团队、管线研发；规划 60 亿元建设 AI 原生医院
OpenAI	GPT-Rosalind 相关研发投入超 10 亿美元量级	含通用模型生物能力训练；以企业合作 API 收入为主
微软	6000 万美元投入美国能源部 Genesis AI4S 项目；药企合作单家数千万至数亿美元	以云服务收入为主，不承担药物研发风险
腾讯	AI 制药领域累计投资超 50 亿元人民币	以财务投资为主，自研投入规模较小
百度（百图生科）	累计投入数十亿元；与赛诺菲合作潜在超 10 亿美元	IPO 拟募资数亿美元；合作含里程碑付款

资料来源：36氪、微软、英伟达公司等官网新闻公告，中邮证券研究所

AI逐步深度参与MNC药物研发整个周期

- **AI逐步深度参与MNC药物研发整个周期。**罗氏2026年研发日披露数据：截止到最新，100%的大分子研发和50%的小分子研发是有AI参与的。25Q4-26Q2统计下来，约40%的管线有AI参与贡献，参与的管线中AI起决定作用占比41%，近16%的管线决策中，AI起到了决定性的作用。AI在先导优化LO阶段AI贡献是最大。靶点评估AI Agent预计Q4会参与80%研究组合决策。2026 年 2 月，礼来在印第安纳波利斯启用 LillyPod，礼来与 NVIDIA 宣布五年共同投入最高 10 亿美元建设 AI 共创实验室，将药物专家与模型工程师集中合作，并探索机器人和物理 AI。

图表2： 罗氏新管线中AI参与度



资料来源：2026年罗氏研发日，中邮证券研究所

AI 制药为何具备现实意义：药物研发成本高昂，而生物学具备可编程性

- **药物研发成本居高不下：**“双10”定律：一款新药从立项到上市周期超过 10 年，成本超 20 亿美金，整体成功率仅 10%。Eroom ‘s law 难题：自 1950 年以来，每一款新药的开发成本大约每九年翻一倍。

- **生物学具备可编程特性：**

✓ 细胞不仅仅是信息存储载体，更是信息处理器。DNA 由 4 种核苷酸组成：A 腺嘌呤、T 胸腺嘧啶、C 胞嘧啶、G 鸟嘌呤。

✓ 基因是 DNA 上独立片段，编码蛋白质，蛋白质承担生物体内绝大多数生理功能。分子生物学中心法则描述 DNA 如何转化为功能蛋白：DNA 转录为信使 RNA（mRNA）；mRNA 进入细胞质，核糖体将 mRNA 翻译为氨基酸，组装为功能完整蛋白质。Moderna 称 mRNA 是 “生命软件”，将 DNA 存储信息转化为蛋白质功能。

- DNA、RNA、蛋白质的语言在生物界是通用的；结合测序、成像、多组学技术产出海量多模态数据集，生物学非常适合 AI 技术落地。

- AI 药物开发目前聚焦于分子生成，但 AI/ML 本质上是一套基于数据做决策的技术。可以应用在药物开发全链条：理解生物学、新药靶点发现、分子设计优化，临床试验设计执行全流程。

新周期开启？此轮和上一轮有什么不一样

- **阶段1:前 AI 时代｜CADD：计算辅助药物设计（1980-2010）**

✓ 分子力学、分子对接理论诞生后，计算机开始用来模拟小分子与蛋白结合；1980s分子对接软件（AutoDock 等）问世，用于虚拟筛选；1990–2010：CADD 在药企内部普及，用于分子构象、药效团建模；

✓ 局限：依赖已知蛋白晶体结构、算力有限，只能做局部优化，无法从头生成分子；预测能力弱，假阳性高。效率工具，不具备自己的创造力。

- **阶段2:AIDD 萌芽期｜机器学习初应用（2010–2018）**

✓ 传统机器学习（SVM、随机森林、浅层神经网络）进入药物筛选。2012 前后深度学习浪潮兴起，行业开始尝试用 ML 做 QSAR（定量构效关系）、毒性预测；

✓ 初创公司陆续成立：BenevolentAI、Recursion、Insitro 早期起步；

✓ 行业状态：药企小范围试点，多作为辅助筛选工具，管线几乎没有 AI 原生分子进入临床。

新周期开启？此轮和上一轮有什么不一样

- **阶段3:AIDD 第一波热潮 & 泡沫期（2019–2022）：**

✓ 标志性引爆点：AlphaFold 2（2021），DeepMind 实现高精度蛋白质结构预测，大量 AIDD 初创公司上市 / 大额融资

✓ 关键里程碑：2019：Generative 模型（GAN、VAE）用于小分子从头生成；2021：AlphaFold2 发布，可预测几乎所有人类蛋白结构；DSP-1181 首个进入临床的 AI 设计药物；

✓ 泡沫破灭：II临床失败，泡沫破裂之后，行业对 AI 驱动药物研发持有怀疑态度。

- **阶段4:第二轮 AIDD 范式转型期（2023–至今）：**

✓ **核心认知转变：湿实验室训练下一代模型，搭建自有数据库，不仅局限分子设计，逐渐前移到靶点发现和临床实验阶段。**

✓ 行业共识：AlphaFold 解决蛋白结构预测，但不等于解决疾病机制与临床结局预测；企业开始投入自建自动化湿实验室，构建「AI 预测—湿实验验证—扰动数据回流」闭环；

AI大模型公司参与 （Isomorphic、Anthropic等）、MNC参与。

✓ 技术方向拓展：全基因组 CRISPR 扰动筛选、高内涵成像、空间组学、虚拟细胞 / 虚拟人体、AI+自动化实验室等；

✓ **临床验证时间窗口：**

2026–2027：Isomorphic AI 设计药物首次进人体试验

2028–2030：AI 原生药物二期 / 三期临床数据读出，是检验 AIDD 价值的关键节点

主要经济体同步加码AI for Science

- 2025年以来，美国、英国、欧盟、日本、中国等主要经济体密集出台AI for Science政策，普遍将其上升为国家级战略工程。

图表3：2025-2026年主要经济体同步加码AI for Science



资料来源：动脉网，中邮证券研究所

新一轮投资浪潮初现：专门以训练下一代 AI 模型为目标，搭建AI数据基础设施

- 上一轮AI制药投资：算法、早期发现药物为主；
- 2025-2026投资潮出现：以训练下一代AI大模型为目标，工业化大规模产出湿实验数据，搭建AI数据基础设施。我们认为，这波AI数据基础设施或许才刚刚开始。

图表4： 已公布AI生物学数据基础设施项目

项目 / 平台	牵头机构 / 主体	核心内容	投入规模	类别
Virtual Biology Initiative（虚拟生物学倡议）	Chan-Zuckerberg Biohub，联合NIH、DOE、Google DeepMind、Isomorphic、Meta、NVIDIA	统一数据标准；大规模CRISPR、药物扰动、多模态细胞成像；产出扰动表型数据集，服务虚拟细胞/人体模型训练	合计约18亿美元	非营利联盟公共基建
Billion Cells Project（十亿细胞计划）	CZI + 10x Genomics + Ultima Genomics	十亿级别标准化单细胞多组学，遗传扰动数据集，服务虚拟细胞	大型公益项目	非营利联盟公共基建
Virtual Human™ 配套自动化湿实验工厂	insitro	大规模CRISPR人源细胞扰动、POSH全基因组光学筛选；细胞扰动数据对齐人类临床队列，产出因果数据集	未披露（大规模自建）	商业AI-Biotech私有工业化数据基建
Dataverse数据基座+内部自动化实验平台	Google DeepMind	Dataverse数据基座+内部自动化实验平台：多模态生命科学知识库，配套蛋白、小分子扰动功能湿实验数据	B轮21亿美元	商业AI-Biotech私有工业化数据基建
Lab-in-the-Loop 自主AI实验室，AI工厂	罗氏 / 基因泰克	3500+ Blackwell GPU算力集群；机器人自动化实验；TargetNexus AI智能体驱动实验决策；实验室自主化L0-L6分级演进	未披露（3500+ GPU等重资本投入）	MNC自建闭环基建
10x Genomics（AI细胞模型基建）	10x Genomics，合作Arc Institute、CZI	单细胞、空间多组学标准化测量，产出高质量原始测量数据，支撑扰动、疾病队列数据生产	未披露（合作项目）	上游工具厂商数据基建（卖铲人）

资料来源：罗氏、 10xgenomics.、 isomorphiclabs.公司等官网新闻公告，中邮证券研究所

2024-2026 虚拟细胞(AIVC)资本市场关注度显著提升

- **2024-2026 虚拟细胞资本市场关注度显著提升。**

✓ 海外创业公司中，Xaira Therapeutics 于 2024 年成立时即完成 10 亿美元承诺资本，规模在早期生物技术融资中极为罕见；Cellarity 累计融资超过 1.2 亿美元，CytoReason 的新一轮达 8000 万美元，法国的 Bioptimus 累计融资超过 7600 万美元，Noetik、Turbine、Cellular Intelligence 等公司在过去两年也相继完成千万美元级别的新一轮融资。

✓ 国内相关统计显示，仅 2026 年 4 月以来，中国 AIVC 赛道就已发生近 10 起融资事件，红杉中国、君联资本、百度风投、国新基金等主流机构相继入场，数亿资金涌入这一新兴技术的产业化进程。百曜科技完成数轮累计数亿元融资。元示科技、深度细胞、无界进化等国内初创也均在近期完成新一轮融资。

二

AI制药新周期核心竞争力：大规模高质量有效数据

AI 驱动药物研发阶跃式进步极度依赖于高质量湿实验数据

- **现阶段AI驱动药物开发并没有提高药物研发率。**2026年6月发表在《Pharmaceuticals》期刊上的一项系统性研究显示：AI辅助药物发现进入临床开发后，约90%的候选药物未获批准，这一比例与传统药物开发“基本持平”。AI加速了早期发现——靶点识别时间缩短50-80%，先导化合物优化时间缩短30-70%——但在最关键的临床转化环节，它还没有展现出超越传统方法的优势。 Tech Emergence的数据同样显示，AI技术可以将临床阶段新药研发成功率从12%提升至约14%。根据BCG在2024年发布的分析，AI发现的分子在Phase I期取得80%-90%的成功率（行业历史水平为40-65%），但Phase II期成功率约为40%，与行业历史水平接近。

- **AI 在药物研发局限性与失败来源，核心是当前 AI 资源主要集中在分子设计环节，分子设计的再好如果靶点不合适（机制不对），会倒在临床2期，无法真正提高药物研发成功率，AI提高药物研发成功率，下一步要走到上游，理解疾病机制和发现药物靶点阶段。**

- **AI 驱动药物研发进入下一阶段，极度依赖于高质量实验数据。**必须依赖于工业化大规模产出湿实验数据，产出大规模、高质量、与人体疾病相关的生物测量有效数据，湿实验数据质量决定AI模型上限。

2.1 为什么生物学AI数据基础设施现在开始做

- **紧迫性：“算法+公开数据”第一波红利已尽？行业已经撞上 “数据墙”？**

✓ 依托存量公开数据，前期模型迭代很快，但全行业共用同一套公开存量资源，继续在这套数据上扩大模型规模，边际收益持续下降。

✓ 行业已经撞上 “数据墙”，上一轮AI驱动药物研发，集中在已知靶点的分子设计方面，目前已知靶点红利和开发空间不多了，需要向上游走，走到更有价值和更难的“疾病到机制”（Disease-mechanism）阶段，去发现疾病机制和靶点，必须要拓展数据库。

- **必要条件：AI在搭建数据基础设施中起关键作用**

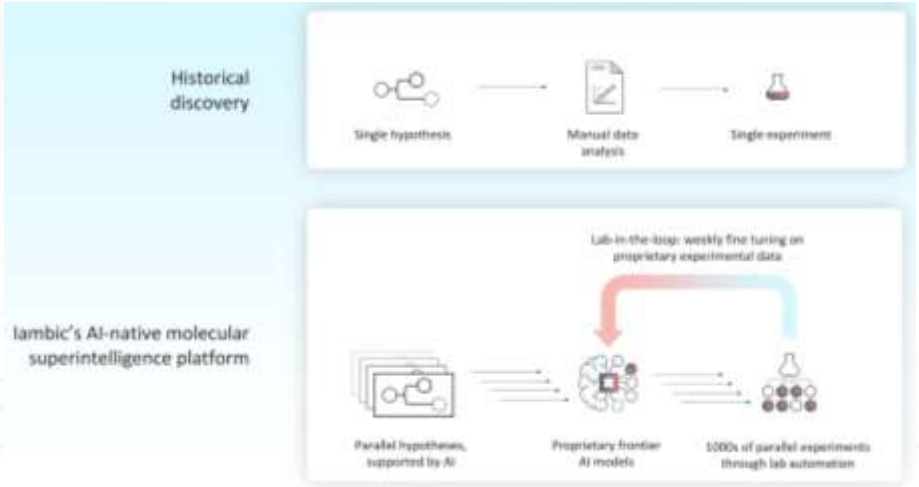
✓ AI agent+自动化实验室

✓ 虚拟细胞

2.2 药物开发中AI模型迭代存在哪些特殊性

- **AI4S为何在分子设计最先取得亮眼成果？核心在于能够实现快速迭代。**AI模型的快速迭代，高度依赖快速、准确、低成本、可获得的评价标准。具体来看在分子设计领域，拥有成熟、高通量读取的评价指标：结合亲和力、溶解度、表达量等，同时叠加早研科学服务及相关CRO行业的迅速发展，使得这些数据具备高通量、快交付、低成本等特征。对比传统的单一线性方式，AI工作范式中形成了海量假设并行的高通量的方式，可参考下图：海量并行假设（由AI支持）→自研前沿 AI 模型预测 → 实验室自动化实现上千次并行实验。参考美股Iambic（股票代码：IAM，2026年10月上市）路演材料，其AI平台在早研细分环节中的失败率对比传统大大缩减，hit-to-lead阶段由25%的失败率降至7%（参考基数15个项目）preclinical阶段的失败率由40%降至0%（参考基数3个项目）。
- 目前，蛋白药物的分子设计对比小分子领域仍处发展初期，而随着AI+公司在蛋白设计中逐渐取得阶段性成果，我们看好相关产业链公司受益于模型训练迭代所需的分子验证需求的增长。

图表5：传统与 AI 药物设计工作流对比



资料来源：Iambic 路演材料，中邮证券研究所

图表6：Iambic AI 平台提升分子早研环节成功率



资料来源： Iambic 路演材料，中邮证券研究所

2.2 药物开发中AI模型迭代存在哪些特殊性

- **长远来看，AI制药的发展绕不开的是疾病机制的理解。**我们认为，在药物开发中，评价药物有效性最终和唯一标准将一直是人体临床实验，要形成数据闭环必须要拿到临床数据，临床实验是整个药物开发过程中耗时最长（3-5年），耗资巨大的部分，属于慢反馈闭环，不满足AI模型快速迭代所需要的廉价高通量要求。而行业存量数据大量是观测类数据（单细胞、细胞图谱、蛋白结构）只能得到相关性结论，建立因果关系需要大量干预/扰动实验数据。当前可用高质量带标签扰动数据集，相比理论需要还差约1000倍，因此自建实验平台建立干预/扰动实验数据是下一代AI模型竞争中的必要条件。只考虑细胞生物学——探究生物学机制所必须进入的层级——需要探索的空间也已经极其庞大（10-100TB）。再考虑药物干预，这个空间就会以指数级扩大（达到100PB级别）。

图表7：药物发现链路： AI 发力点与失败源头分析



资料来源：a16z.news《Drug Discovery Has No Magic Wands》，中邮证券研究所

图表8：高质量带标签扰动数据集的数据缺口



资料来源： a16z.news《Drug Discovery Has No Magic Wands》，中邮证券研究所

2.3 关于数据缺口：需要什么样的数据?

- **什么样的数据才是有效高质量数据?**

✓ 能揭示宏观层面的人体疾病临床表现与药物所干预的细胞层面的具体机制的对应关系数据，形成以下数据闭环，从实验生成数据→训练AI模型→模型预测新分子→实验验证预测结果（人体临床实验结果）→验证数据回馈模型，上一代AIDD闭环：大多停留在分子层面闭环（分子生成-合成-活性测试回流）。下一代面向**靶点因果发现**的闭环：需要扰动细胞实验 + 人体临床结果标签对齐。

✓ 静态单细胞数据只能刻画相关性，难以支撑「干预—响应」的因果推断。Perturb-seq（CRISPR 扰动结合单细胞 RNA 测序读出）的价值正在于此：它为每个细胞提供「扰动输入—转录组输出」的配对，使模型可以学习由干预到响应的方向性映射。因此，扰动数据（尤其是基因组规模 Perturb-seq 与药物扰动图谱）被视为虚拟细胞训练数据中不可替代的一类。

2.3 关于数据缺口：需要什么样的数据?

- **虚拟细胞（AIVC）有望成为揭示细胞系统的“编程语言”的接口。**参考cell在2024刊发的文章，我们把 AIVC 理解为一种 可学习的细胞与细胞系统模拟器： 它能在不断变化的条件下和不断变化的情境中运行，能够跨生物尺度、跨越时间、跨越数据模态地工作；它还应有助于揭示细胞系统的“编程语言”，并提供一个可将其用于工程化改造的接口。其本质并非数据库/知识图谱、亦不是“预测某个任务的模型”，而是能接受新条件、并对新条件给出结果的模拟器。
- 具体而言，AIVC 需要具备以下能力：（1）构建通用表征（Universal Representation, UR）：建立跨物种、跨模态、跨数据集、跨情境的生物状态表征，涵盖细胞类型、发育阶段和外部条件；（2）预测细胞的功能、行为与动态，并揭示其底层机制；（3）开展 in silico（计算）实验，用以生成并检验新的科学假设，并指导数据采集，从而高效地扩展虚拟细胞的能力。
- 参考动脉网白皮书，AIVC从数据层来看，由**先验知识、静态架构、动态状态**组成。**先验知识**整合已发表的生物医学文献、已知的分子表达规律和多尺度成像数据，明确细胞生命活动的基础规则。**静态架构**收录细胞形态结构、分子空间分布等固定信息，还原细胞固有物理特征。**动态状态**是构建具备真实推演能力模型的关键，聚焦细胞自然发育、衰老、药物干预、基因编辑等外部扰动下的状态变化，是支撑AI虚拟细胞实现动态仿真的核心数据底座。

图表9：虚拟细胞AIVC建设图景

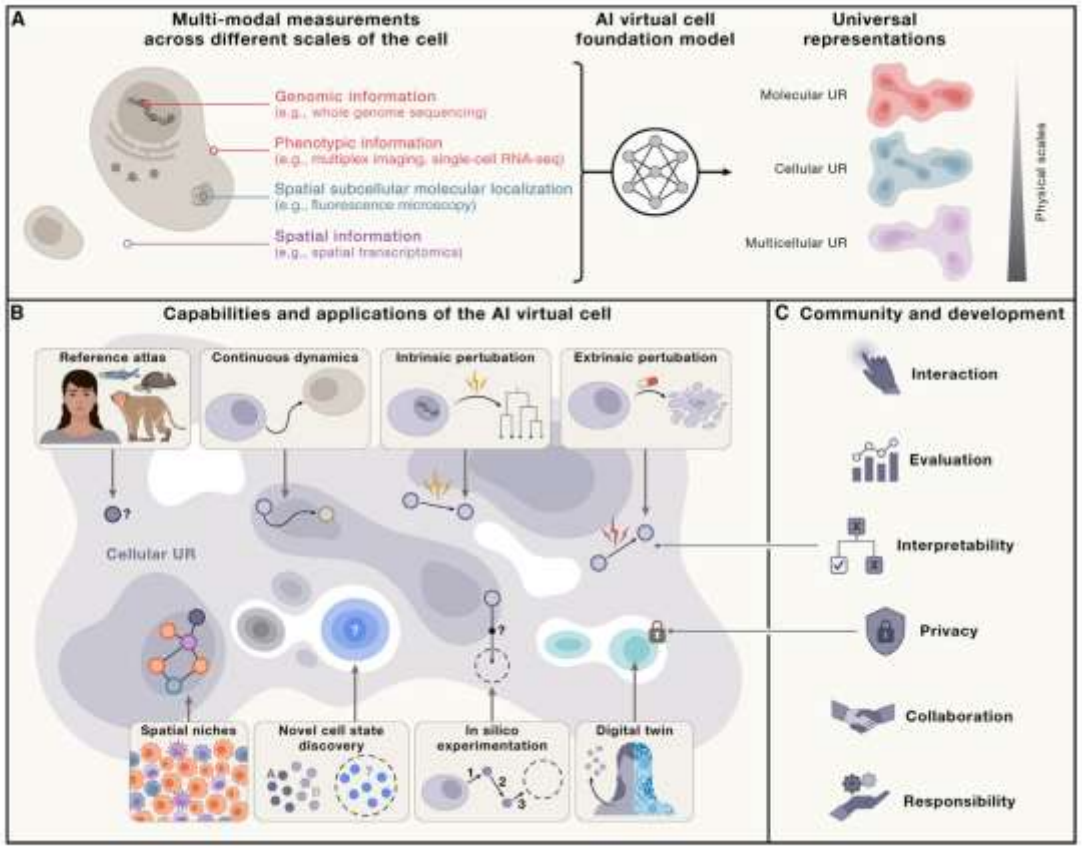


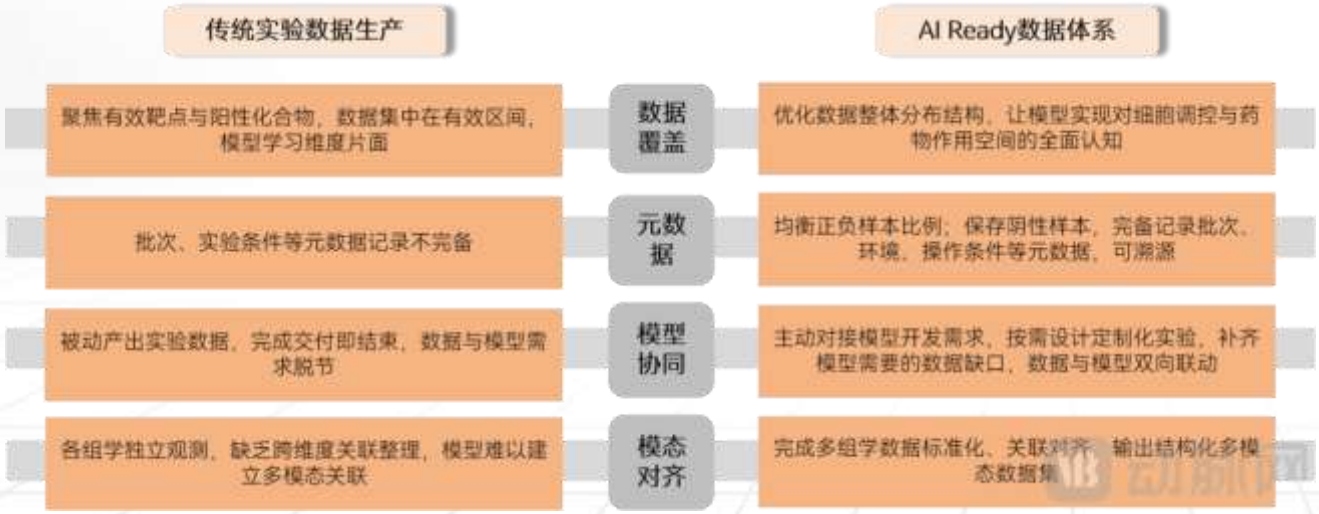
Figure 1. Capabilities of the AIVC

资料来源：Cell. 2024 December 12; 187(25): 7045–7063，中邮证券研究所

2.4 发展现状：模型规模扩张，专业化数据生产商正崛起

- AIVC数据底座已进入亿级细胞规模，基础模型与扰动预测模型快速涌现，但严格评测提示其能力边界尚未真正突破。行业整体仍存在四大硬约束：
  - ✓ 数据：稀缺、批次效应、模态不对齐、元数据缺失；
  - ✓ 评测：没有公认 benchmark，“预测得准”无法定义。
  - ✓ 因果：观察数据只能给相关性，机制只能靠扰动实验换。
  - ✓ 生态：需要开放、标准、联邦式协作。此外，从商业化落地角度，仍存在价值定位的痛点和应用落地的鸿沟等现实问题。
- 目前行业参与者类型多元、覆盖维度广泛，不同背景的企业与科研机构依据自身资源禀赋，分别在产业数据层、模型层、应用层形成差异化布局，呈现出多点开花、协同发展的繁荣格局，尚未形成高度集中的竞争态势，各类主体仍处于能力搭建与赛道卡位阶段。

图表10：传统数据生产vs AI-ready 数据体系



资料来源：动脉网，中邮证券研究所

2.4 发展现状：模型规模扩张，专业化数据生产商正崛起

- **产业合作已从概念验证进入实际交付阶段。**Recursion 与罗氏（Roche）及其旗下基因泰克的合作进展也是标志性的一笔。2026 年 7 月，基因泰克将合作管线中的首个神经科学新靶点推进至联合早期发现项目，这是该合作自 2021 年启动以来首次实现“AI 发现-药企验证-进入研发管线”的闭环，截至 2026 年第二季度，Recursion 已从该合作累计交付 2.16 亿美元预付款和里程碑款。
- **专业化数据生产商快速崛起**：数据供给端已经形成多元参与者矩阵。**单细胞测序企业是重要数据来源，**包括10x Genomics、寻因生物、墨卓生物、新格元等。**类器官与器官芯片企业**可产出贴近生理环境的细胞实验数据，代表有黑玉科学、耀速科技、淇嘉科技等；**蛋白质组学企业**依托自研质谱平台，产出转录组难以覆盖的功能蛋白、完整的磷酸化修饰数据，包括嘉华药锐、西湖欧米等；部分生物医药企业也开始**对外输出数据资产**，例如神拓生物这类体内治疗技术企业，沉淀大量真实研发过程数据与失败边界信息。单一模态数据很难完整还原细胞复杂运行逻辑，多组学融合的数据集合可以有效降低模型的认知偏差。业内数据生产商正积极通过跨主体合作补齐自身数据维度短板。
- 模型开发企业也意识到了数据底座的价值，数据生产商已经成为模型开发与迭代不可或缺的外部合作伙伴。2026年赛道诞生多起产业合作，例如衍因科技与神拓生物，百图生科与耀速科技，无界进化与墨卓生物，百曜科技与新格元，无尽方舟与寻因生物，英矽智能与墨卓生物，百图生科与华大智造，数据层与模型层的绑定日益紧密。

2.5 关注AI+生物新基建，竞赛已然开始

- **AI agent+自动化实验室：本质上就是虚拟模型的物理验证引擎。**

礼来与英伟达双方计划将能够自主执行部分实验的机器人实验室与 AI 模型结合，让实验数据生成、模型训练和实验验证形成持续循环，并探索将物理 AI 和机器人技术应用于药物制造。

- ✓ AI **agent**与实验室自动化系统形成闭环，提出假说、指挥机器人、分析结果，迭代。 礼来CEO Ricks 也曾提出，建立能够全天候运行的机器人实验系统，通过大规模自动化实验持续生成生物学数据，可以主动探索尚未被充分研究的生物过程，并将实验结果用于改进模型。
- ✓ 自动化系统在优化无细胞蛋白质合成、设计针对已知靶点的抗体等任务上，已经取得出色表现。

- 评判最终标准：是否解决临床实验结果最终闭环的问题。如果AI无法获得 “下游”环节的反馈数据，其预测结果将始终停留在理论层面。
- 自动化实验室不是今天才诞生，2026年美国橡树岭国家实验室牵头的《Toward Trustworthy Autonomous Science: A Two-Year Community Roadmap》，以一项标志性自主发现成果的更正事件作为切入点指出：产生候选发现已不再是难点，验证才是；截至目前，还没有任何系统能够端到端完成真正新颖发现，并实现实验层面的自我闭环验证。

请参阅附注免责声明

2.5 关注AI+生物新基建，竞赛已然开始

- **虚拟细胞研究**

✓ 陈-扎克伯格研究院 2025-02 宣布十亿细胞项目，多年目标产出 10 亿细胞单细胞数据集，专门服务下一代生物大模型，合作 10x Genomics、Ultima Genomics。Illumina 与 Broad 临床实验室合作，规模化单细胞工作流，目标达到数十亿细胞规模数据集。Arc 研究所搭建虚拟细胞图谱，至少 3 亿细胞，使用 10x 做单细胞，Ultima 做 NGS。

✓ 商业公司案例：Xaira 发布 X-Cell 虚拟细胞模型，训练基于 X-Atlas/Pisces 全基因组 CRISPRi Perturb-seq 数据集，包含 2560 万扰动单细胞转录组，覆盖 7 种不同细胞环境，是目前公开披露规模最大的功能基因组数据集之一。Owkin MOSAIC 项目，10x 单细胞 + 空间技术，构建泛癌症空间转录组大数据集，用于 AI 疗效预测、靶点挖掘。Bioptimus STELA 项目，联合 10x（Xenium）、Broad 研究所，计划分析多达 10 万份患者标本，规模较现有空间图谱提升约 20 倍，训练 M-Optimus 模型。

- **人类的大多数疾病，都是系统层面的功能紊乱，涉及跨越多种细胞类型、随时间演变的复杂生物过程相互作用。要理解这些过程，就需要系统层面的测量。虚拟细胞研究将大规模扰动数据与AI结合，但即使是其中规模最大的项目，也只采样了所有可能扰动中微不足道的一小部分，而且几乎完全局限于少数细胞环境。**

2.5 结论

- **AI会颠覆药物研发的逻辑吗？AI会重新定义制药产业的底层逻辑吗？可能不会，AI要走的路很漫长。**
  - ✓ 评价药物有效性最终和唯一标准将一直是人体临床实验，慢反馈闭环。
- **生物学AI数据基础设施，需求巨大类基建，行业发展基石，投入大耗费时间长，需要国家层面投入参与。**
  - ✓ *10月8日，美国政府联合Google、Meta，以及扎克伯格资助的陈·扎克伯格生物中心（Chan Zuckerberg Biohub，Biohub），正式推出一项总投入18亿美元的重大计划，目标搭建AI生物学领域的数据底层基础设施*

- **下一轮 AI 制药竞争同时发生在几个层面：**
  - ✓ 模型仍然重要
  - ✓ 大规模高质量有效数据开发能力：获取大规模、标准化的人体和实验数据；持续产生扰动数据的自动化实验平台，把细胞层面机制与人体临床结果连接起来；以上数据需要多年积累。
  - ✓ 足够的资金实力，多年临床验证周期的资金需求。

- **“平台型AI制药”有可能出现，成为最终赢家。** 最先构建有效数据，形成临床验证闭环，将具备持续输出靶点发现能力，形成平台能力。初期投入大，平台和模型稳定运行后，不断发现有效靶点，边际成本低，马太效应明显。

三

投资展望

3.1 湿实验需求的确定性和空间具有想象力

- **湿实验需求的确定性。 AI 在生物学领域的潜力释放，必须产出大量全新实验数据（也不依赖于AI是否真的最终提高新药研发成功率）；这会为数据瓶颈环节对应的工具栈带来长期利好。**

✓ 蛋白语言模型 PLM（ESM 等）、结构预测模型（AlphaFold）、新一代生成式蛋白设计（RFdiffusion、Chai-2）： 主要基于历史积累公开数据集训练（UniProt、PDB 蛋白数据库），不需要大量新增湿实验。

✓ 蛋白适应性、生物学功能预测模型，必须直接依赖实验输入。

✓ 细胞状态、扰动响应模型（例如陈-扎克伯格研究院虚拟细胞项目），需要大规模全新实验：单细胞测序、CRISPR 扰动文库、高内涵成像等。

- **湿实验室需求空间：**

✓ 模型训练数据本身；

✓ 把通用基础模型改为特定研发模型：针对特定疾病、特定场景运行闭环 DBTL 循环。

- **AI+分子验证需求：**AI模型降本增效显著，将催生更强烈分子验证需求，产出假设未必比传统方法质量更高，但可以确定产出数量更大、速度更快。

3.1 湿实验：数据瓶颈环节对应的工具栈均受益

- **数据瓶颈环节对应的工具栈**

✓ **基因/蛋白合成、表达及检测服务公司有望率先受益，**客户利用 AI 在计算机内虚拟生成、优化海量候选 DNA 序列（in silico 硅基虚拟设计），不做实体实验。**基因/蛋白合成公司承**做高通量 DNA 合成和蛋白表达（即将DNA序列变成目标蛋白），而后开展体外实验分析：结合能力、可开发性、关键成药性特征等检测，产出实验测定数据。而后将实验检测数据回传给客户，指导下一轮 AI 设计；每一轮迭代驱动客户复购 。

✓ **测序**：细胞状态、扰动响应模型实验需要单细胞测序、CRISPR 扰动文库等。

✓ **实验试剂和耗材**：分子验证的需求爆发，其试验所需的试剂、耗材等增长有望得到同步提振。

✓ **实验室自动化设备**：AI+自动化实验需求驱动

✓ **蛋白表达及纯化：**蛋白表达及纯化亦是蛋白药物分子验证、生产中必不可少的一环。采用细胞表达蛋白的方式需要培养基充当细胞之养料，进而收取及纯化蛋白的过程需要色谱（层析介质）。

✓ **动物模型体内验证：**体内药效检测目前仍然是临床前不可少的一环，通过前端的分子筛选验证需求的传导， 我们认为在体内验证的分子有望同比例的快速增加。趋势来看，行业、监管层面越来越重视非动物模型 NAMs；类器官、器官芯片、预测模型，预计将与动物体内模型形成互补。**湿实验景气观察指标**

✓ AI制药投融资情况、大模型企业投入情况、AI进展驱动

3.2 AI对CXO行业影响

- **国际头部CXO公司一致认为AI制药将提升候选分子数量，增厚上游客户管线，拓宽疾病覆盖领域并提升相关临床/临床前研究复杂度，从而带来外包需求的提升**

✓ 早期研发影响：先导分子工程与优化需求可能减少，占临床前总投入约 10%。

✓ 安评影响：更多分子推出，需求增加，中长期存在借助药效、安全性预测模型（NAMs技术），目前药物申报仍需要传统的体内体外研究，未来两者会形成互补而非替代关系。

✓ 临床实验影响：更多分子进入临床，2期成功率提升，顺利推进3期，1/2期临床数量有望增加

✓ CDMO影响：更多管线获批

3.3 相关标的

- **建议关注：**

1）蛋白合成表达：金斯瑞、义翘神州、奥浦迈、百普赛斯、近岸蛋白等；

2）动物模型验证：百奥赛图、药康生物、南模生物。

3）试剂&耗材：皓元医药、毕得医药、洁特生物、昌红科技等；

4）测序&自动化实验室：华大智造、诺禾致源、华大基因；

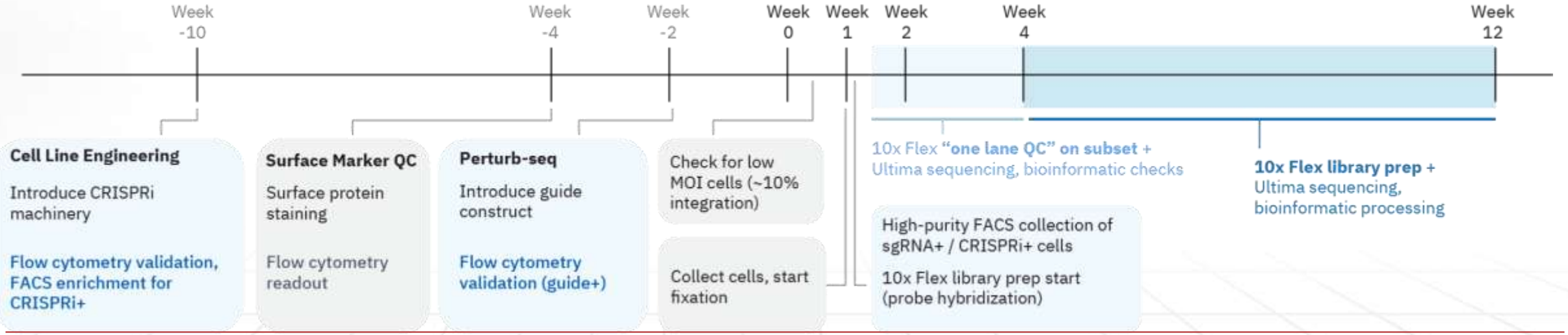
5）设备：东富龙、楚天科技、海尔生物等；

6）临床前CRO：昭衍新药、益诺思、美迪西、维亚生物、康龙化成、药明康德。

附：虚拟细胞工作流程

- Arc Institute 为 2026 年 Virtual Cell Challenge 生产的扰动响应数据，是目前公开领域 质量最高的 Perturb-seq 数据集，其生产流程标准化到了 12 周一个完整周期：前 6 周（Week -10 至 -4）给 6 种差异很大的细胞系（干细胞 H1、永生细胞 RPE1/HEK293T、癌细胞 HCT116/A549/K562）逐个转染 CRISPRi 机器 dCas9-KRAB，流式富集到 ≥95% 阳性；接着 2 周（Week -2 至 0）转导 sgRNA 文库——400 个基因各 3 条 guide，外加 100 个非靶向对照（NTC）定义基线，低 MOI（~0.3）确保每个细胞只被敲一个基因；然后 3 周（Week 1 至 4）用 10x Flex 探针法建库、Ultima UG100 测序， 每个细胞中位数 20,000 UMI （行业平均 5,000–10,000），超深测序捕获微弱扰动效应；最后 8 周（Week 4 至 12）用 Arc 自研的 cyto 管道（比 Cell Ranger 快 16.5 倍）做生信处理，以 80% on-target knockdown 为阈值严格筛选（行业一般 50% 就过关），下采样到每个扰动 400 个细胞，最终产出 3000 万+ 高质量单细胞转录组——整个流程的标准化和自动化，正是 Arc 能把 AI-in-the-loop 闭环压缩到 14 天、而传统实验室要 6–12 周的核心原因。

图表11：虚拟细胞工作流程



资料来源：Arc Institute，中邮证券研究所

附：计划资金预算项目分配

- 2026年10月7日，旧金山首届AIxBio活动上，Biohub联合Google DeepMind、Meta、Isomorphic Labs、美国能源部与NIH，围绕同一个目标凑出约18亿美元——在计算机里造出一个“足够可信”的细胞。几乎同时，近期行业白皮书给出另一个数字：全球虚拟细胞赛道累计融资突破30亿美元。
- **钱不是最难的一环，数据才是。**Biohub 首席科学家 Alex Rives 给出的量级判断是：眼下所有细胞数据集加起来只覆盖了数亿个细胞的观测，而要训练出真正能预测细胞行为的模型，可能要数十亿、甚至数万亿个细胞的数据。对照Arc Institute的 scBaseCount——5 亿个细胞、21 个物种、72 种组织，已经是行业头部，但按Rives的口径连零头都不够。Arc把数据分两类：观测数据能说明“细胞处于什么状态”，只有扰动数据才能回答“为什么变成这个状态”。

图表12：Biohub 18亿美金计划资金预算项目分配

出资方	出资口径	形式
Biohub（陈-扎克伯格生物中心）	约4–5亿美元（2026年4月）	测量工具与项目投入
美国能源部（DOE）	逾5亿美元 / 5年（Genesis Mission）	资金 + 算力 + 自动化实验室
Google DeepMind + Meta + Isomorphic Labs	合计约3亿美元	资金 + 模型能力 + 算力
NIH	协调数据集（此前逾5亿美元联邦投资）	数据与资源
合计	约18亿美元（承诺总量口径）	现金 + 数据 + 算力

资料来源：易贸医疗17talk，中邮证券研究所

四 风险提示

风险提示

风险提示：

- AI制药模型研发进度不及预期风险：新药研发本身存在较高研发失败风险，AI制药模型迭代进度存在不及预期风险。
- 投入制药领域的资本开支不及预期风险：AI制药模型的商业模式未完全成型，前期投入主要依赖资本开支及投融资等，存在不及预期风险。
- 竞争加剧风险：随着投融资等外部因素带动，行业竞争可能会趋于激烈。
- 地缘政治风险：BD合作模式部分涉及跨国跨地区，存在地缘政治因素扰动风险。
- 政策变动风险：行业易受监管、药物审批等政策扰动。

请参阅附注免责声明

中邮证券
CHINA POST SECURITIES

感谢您的信任与支持！

THANK YOU

盛丽华（首席分析师）

SAC编号：S1340525060001

徐智敏（分析师）

SAC编号： S1340525100003

免责声明

分析师声明

撰写此报告的分析师（一人或多人）承诺本机构、本人以及财产利害关系人与所评价或推荐的证券无利害关系。

本报告所采用的数据均来自我们认为可靠的目前已公开的信息，并通过独立判断并得出结论，力求独立、客观、公平，报告结论不受本公司其他部门和人员以及证券发行人、上市公司、基金公司、证券资产管理公司、特定客户等利益相关方的干涉和影响，特此声明。

免责声明

中邮证券有限责任公司（以下简称“中邮证券”）具备经中国证监会批准的开展证券投资咨询业务的资格。

本报告信息均来源于公开资料或者我们认为可靠的资料，我们力求但不保证这些信息的准确性和完整性。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价，中邮证券不对因使用本报告的内容而导致的损失承担任何责任。客户不应以本报告取代其独立判断或仅根据本报告做出决策。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，中邮证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。

中邮证券及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，也可能为这些公司提供或者计划提供投资银行、财务顾问或者其他金融产品等相关服务。

《证券期货投资者适当性管理办法》于2017年7月1日起正式实施，本报告仅供中邮证券签约客户使用，若您非中邮证券签约客户，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司不会因接收人收到、阅读或关注本报告中的内容而视其为签约客户。

本报告版权归中邮证券所有，未经书面许可，任何机构或个人不得存在对本报告以任何形式进行翻版、修改、节选、复制、发布，或对本报告进行改编、汇编等侵犯知识产权的行为，亦不得存在其他有损中邮证券商业性权益的任何情形。如经中邮证券授权后引用发布，需注明出处为中邮证券研究所，且不得对本报告进行有悖原意的引用、删节或修改。

中邮证券对于本申明具有最终解释权。

免责声明

中邮证券 CHINA POST SECURITIES

公司简介

中邮证券有限责任公司于2002年9月经中国证券监督管理委员会批准设立，公司注册资本61.68亿元人民币，是中国邮政集团有限公司绝对控股的证券类金融子公司，公司是中邮创业基金管理股份有限公司的第二大股东。

公司经营范围包括:证券经纪，证券自营，证券投资咨询，证券资产管理，融资融券，证券投资基金销售，证券承销与保荐，代理销售金融产品，与证券交易、证券投资活动有关的财务顾问，具备展业的各项资格。截至2025年10月底，公司在全国设有58家分支机构(含29家分公司、29家营业部)，1家资产管理分公司和1家另类投资子公司。

中邮证券紧密依托中国邮政集团有限公司的雄厚实力，通过强化“自营+协同”发展模式，实现快速发展，当前服务的经纪客户已超过260万人。公司始终坚持诚信经营、践行金融为民，为社会大众提供全方位专业化的证券投融资服务，努力成为员工自豪、股东放心、客户信赖、社会尊重的优秀企业，打造契合中国邮政资源禀赋和市场地位的特色精品券商。

投资评级说明

投资评级标准	类型	评级	说明
报告中投资建议的评级标准：报告发布日后的6个月内的相对市场表现，即报告发布日后的6个月内的公司股价（或行业指数、可转债价格）的涨跌幅相对同期相关证券市场基准指数的涨跌幅。 市场基准指数的选取：A股市场以沪深300指数为基准；新三板市场以三板成指为基准；可转债市场以中信标普可转债指数为基准；香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	股票评级	买入	预期个股相对同期基准指数涨幅在20%以上
		增持	预期个股相对同期基准指数涨幅在10%与20%之间
		中性	预期个股相对同期基准指数涨幅在-10%与10%之间
		回避	预期个股相对同期基准指数涨幅在-10%以下
	行业评级	强于大市	预期行业相对同期基准指数涨幅在10%以上
		中性	预期行业相对同期基准指数涨幅在-10%与10%之间
		弱于大市	预期行业相对同期基准指数涨幅在-10%以下
	可转债评级	推荐	预期可转债相对同期基准指数涨幅在10%以上
		谨慎推荐	预期可转债相对同期基准指数涨幅在5%与10%之间
		中性	预期可转债相对同期基准指数涨幅在-5%与5%之间
		回避	预期可转债相对同期基准指数涨幅在-5%以下

中邮证券研究所

北京
邮箱：yanjiusuo@cnpsec.com
地址：北京市丰台区北甲地路2号院6甲1号，玺萌大厦南塔
邮编：100050

上海
邮箱：yanjiusuo@cnpsec.com
地址：上海市虹口区东大名路1080号大厦3楼
邮编：200000

深圳
邮箱：yanjiusuo@cnpsec.com
地址：深圳市福田区滨河大道9023号国通大厦二楼
邮编：518048

中邮证券
CHINA POST SECURITIES