MIT 科技评论
Technology Review

AI × LIFE SCIENCE
AI VIRTUAL CELL

AI 虚拟细胞（AIVC）技术趋势、产业生态与应用前景研究报告

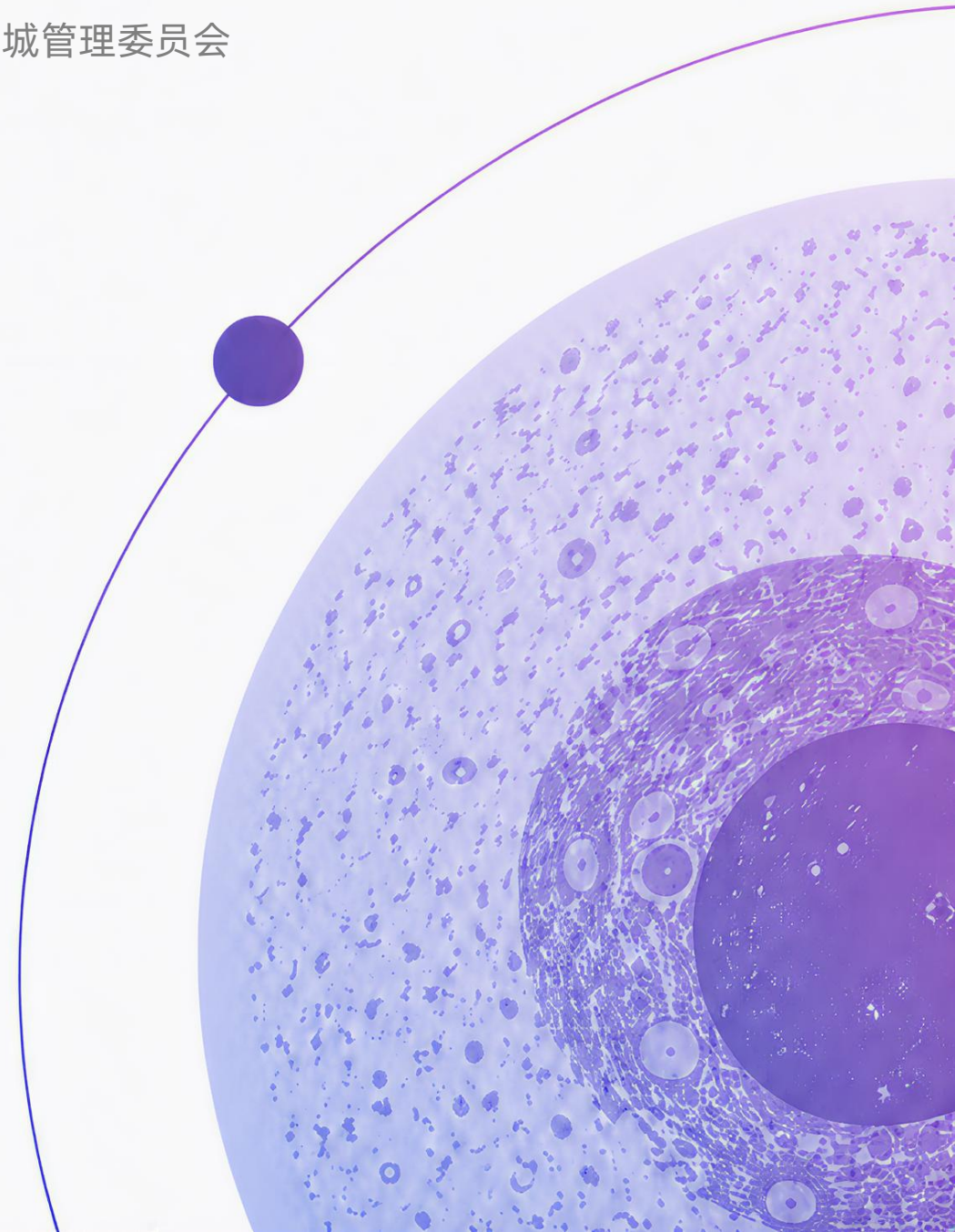
——AI 时代生命科学的新型基础设施

指导单位: 中关村科学城管理委员会

OBSERVE

SIMULATE

EXPLORE



2026

序

陈润生

中国科学院院士

中国科学院生物物理研究所
研究员

认识生命，是人类科学探索中最深刻的命题之一。过去，基因组学和生物信息学的发展使我们能够以数据读取生命；今天，人工智能与生命科学的交汇，使我们开始尝试描述和推演生命过程。

我较早开始关注人工智能在生物学研究中的应用。从 1985 到 1995 年间我曾发表多篇用神经网络预测生物大分子结构和功能的研究。三十多年来，算法、数据与算力发生了巨大变化，但一个基本问题始终存在：人工智能可以发现复杂信息之间的联系，却不能等同于对生命机制的理解。

细胞是生命活动的基本单元。能否建立具有某些生命特征、可以接受扰动并呈现动态响应的虚拟细胞，检验着我们对生命规律的理解，也可能成为生命科学研究范式演进的重要标志。

真正有价值的虚拟细胞，首先应当“活”起来。生命不是某一时刻的静态状态，而是在时间中展开、在空间中组织的连续过程。只有引入时间尺度、空间结构和分子相互作用，虚拟细胞才可能从状态描述走向过程描述。

建设虚拟细胞也不能简单等同于汇集数据。数据只有经过降噪、去重、标准化和模态对齐，才能成为高质量语料。模态对齐不是把组学、影像和生理信号机械地放在一起，而是寻找它们在同一生命过程中的联系。这既是生物医学问题，也是多学科共同面对的基础问题。这不仅是一项技术工程，也是一项需要长期积累、开放协作和审慎求证的科学事业。

虚拟细胞的发展不会一蹴而就，也不宜一开始就追求包罗万象。现实的路径，是以基本生命能力为底座，从疾病机制、药物作用等具体问题切入，发展可验证的垂类模型，并同步建立评价标准，明确精度、适用范围和能力边界。短期不应高估其成熟度，长期也不应低估其科学价值。

本白皮书对人工智能虚拟细胞的技术进展、关键问题与产业生态进行了梳理。希望它能够帮助科研与产业界既看到方向的广阔，也尊重生命系统的复杂；既积极创新，也坚持科学验证。唯有如此，我们才能让人工智能真正成为认识生命、服务健康的有力工具，并在不断逼近生命规律的过程中，拓展人类理解自身的边界。

MIT 科技评论
Technology Review

AI 虚拟细胞（AIVC）技术趋势、产业生态与应用前景研究报告

——AI 时代生命科学的新型基础设施

04 前言

05 第 1 章 AIVC 兴起：生命科学进入模型驱动的新范式

研究范式重构：从实验试错走向模型驱动

发展窗口期：政策、技术、监管与资本条件共同成熟

AIVC 将成为科学发现与产业应用的新基础设施

11 第 2 章 技术体系：如何构建理解、预测和设计生命的 AIVC

数据：AIVC 构建的数据基础

模型：从细胞表征走向扰动响应和多元范式探索

自主智能：将科研问题拆解为可执行的虚拟实验和验证规划

干湿闭环：构成 AIVC 持续迭代的长期价值与壁垒

25 第 3 章 产业发展：从模型开发步入应用验证

产业格局：新锐公司与成熟企业加速入局，产业生态初步形成

应用实践：从技术探索走向高价值应用场景

能力验证：走向真实场景中的效果评价

价值评估：从 Biotech 与 AI 双重赛道审视企业投资价值

36 第 4 章 AIVC 的关键挑战与演进趋势

关键挑战：AIVC 仍需提升模型能力与应用边界

未来趋势：从模型竞争走向产业生态协同

43 结语&致谢

45 版权声明

前言

过去十余年，人工智能的快速发展正在改变科学研究的方式。随着深度神经网络、基础模型和大规模计算能力不断进步，人工智能已不再局限于信息识别、检索和分类，而是开始从海量实验与观测数据中学习复杂系统的表征、结构关系与变化规律，并参与科学假设生成、结果预测和方案设计。这一趋势已在蛋白质结构预测、材料发现、化学反应建模和生命科学等领域逐步显现，推动人工智能从通用信息技术走向以数据、模型和实验协同为特征的 AI for Science，成为辅助科学理解、预测与发现的重要基础设施。

与此同时，单细胞测序、空间组学、多组学测量、扰动实验和自动化平台快速发展，生命科学正将细胞状态、分子变化及表型结果转化为可计算的数据。传统依赖经验和反复试错的研究方式，面对细胞这一多层级、动态变化的复杂系统，越来越需要模型提供统一表征、机制线索和可检验的预测。

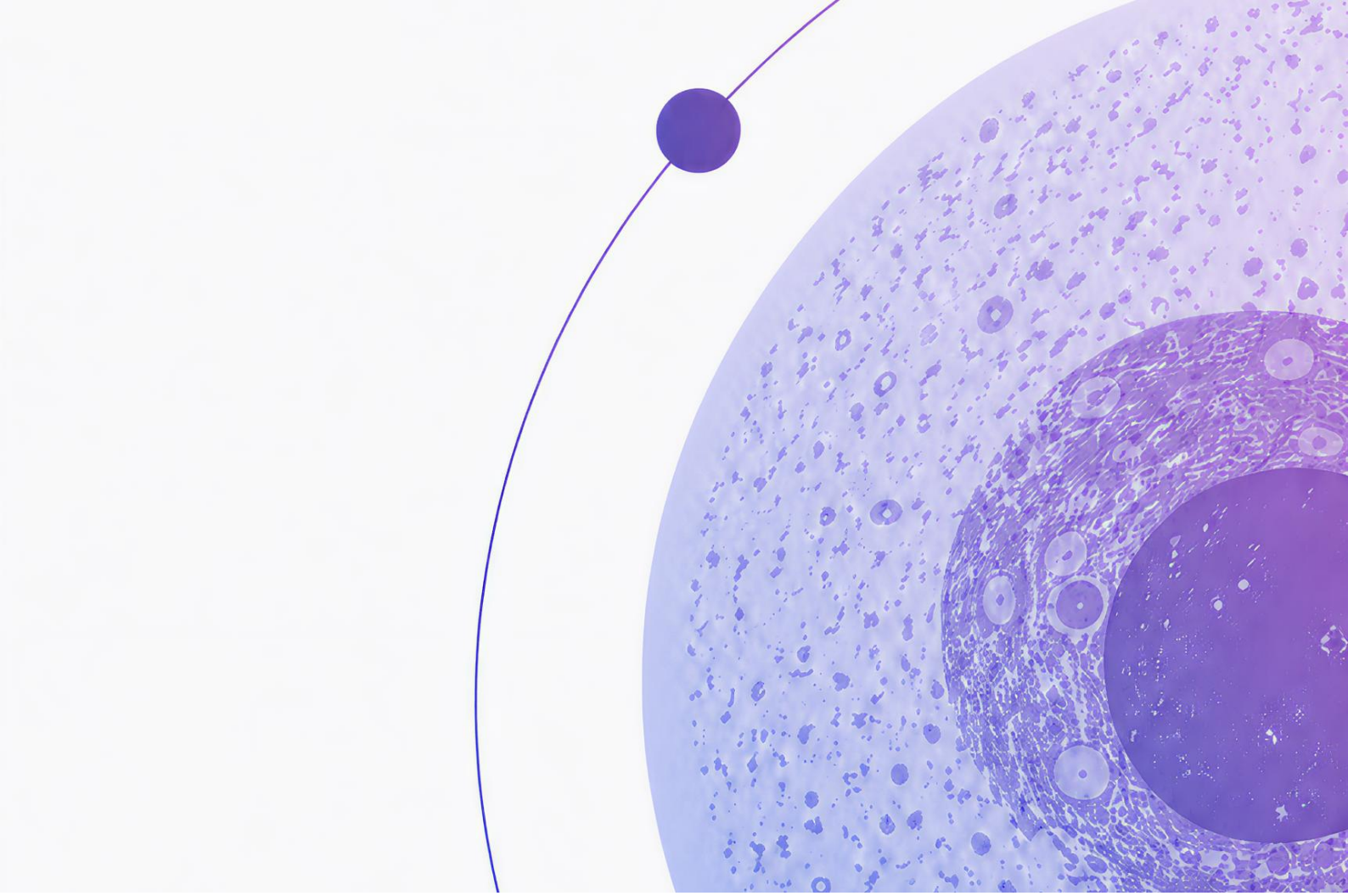
在这一背景下，AI 虚拟细胞（AIVC）受到关注。它试图在计算空间中描述细胞状态，模拟基因、药物及环境变化带来的状态迁移，并将预测结果交由实验验证，推动生命科学由观察和解释走向预测、设计与工程化。

从产业发展看，AIVC 正在由概念探索进入产业验证和早期商业化阶段。近年来，相关领域融资活动明显增多，资本市场开始关注数据、模型、实验平台与研发服务之间的协同价值。同时，药企、科技企业、科研机构和专业服务平台正在通过技术授权、数据合作、联合研发、产业孵化等方式推进布局。

基于此，本研究报告希望回答以下几个问题：

- AIVC 试图解决生命科学研究中的哪些核心问题
- 如何构建 AIVC——数据、模型、自主智能与干湿闭环
- AIVC 相关的产业生态正在如何形成
- 有何关键挑战与产业趋势

本研究报告在指导单位「中关村科学城管理委员会」的支持下，由「百曜科技」发起并提供项目支持，《麻省理工科技评论》中国团队负责策划、调研、撰写与制作。通过对这些问题的讨论，本报告希望厘清 AIVC 的技术内涵、现实价值与发展边界，构建其从技术创新走向科学基础设施和产业赋能的可能路径。



第 1 章

AIVC 兴起：生命科学进入模型驱动的新范式

研究范式重构：从实验试错走向模型驱动

发展窗口期：政策、技术、监管与资本条件共同成熟

AIVC 将成为科学发现与产业应用的新基础设施

从实验试错到模型驱动：生命科学正在重构研究范式

从 17 世纪罗伯特 · 胡克第一次在显微镜下看见细胞，到 21 世纪初人类基因组计划的完成。几个世纪以来，生命科学主要依靠实验不断逼近生命真相——研究者基于有限观测提出假设，通过控制变量和实验逐一验证，再从离散结果中归纳规律。

如今，当单细胞测序技术使我们在全基因组尺度上观测数以百万计的细胞成为可能，多组学、扰动实验和高通量自动化正在把细胞转化为数据。深度神经网络与单细胞组学数据的结合，为研究生命提供了一种新的可能，从细胞整体状态出发，学习基因之间如何相互作用，以及这些作用如何共同决定细胞的状态与表型。这是比 AlphaFold 难度维度更大的挑战，也是下一个“AlphaFold 时刻”。

AI 虚拟细胞（AI virtual cell, AIVC）**是一种跨尺度、多模态、基于大规模神经网络的模型，能够表征并模拟细胞在多种状态下的行为和状态转换。**AIVC 能够建模细胞在不同状态之间转换的动态过程，包括由分化、发育、衰老等自然过程、遗传变异或人为设计的扰动所引起的状态变化，以及在亚细胞和多细胞尺度的生命过程预测。

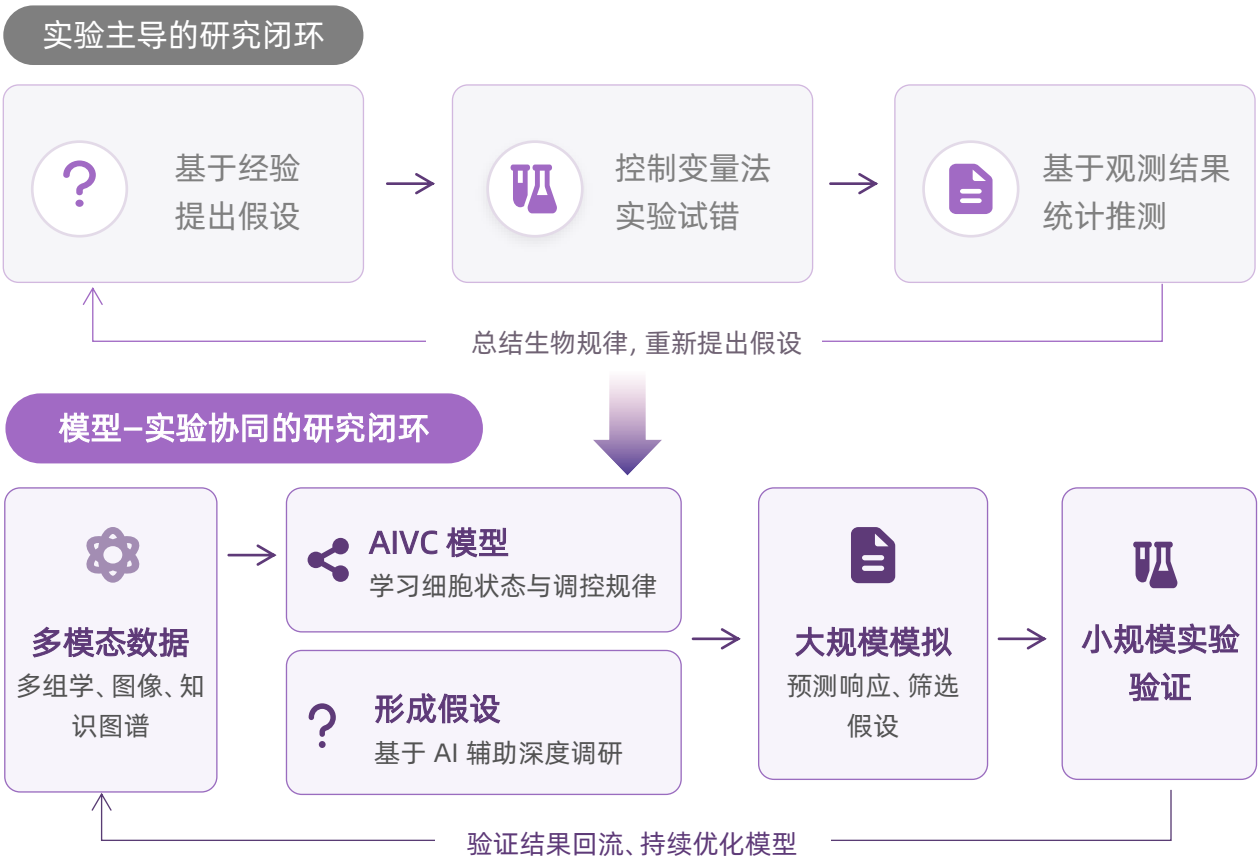


图 1 | 传统生命科学研究向基于 AIVC 的研究范式转变（来源：《麻省理工科技评论》中国、公开信息）

生命数据积累与 AI 能力演进协同催生 AIVC

1990–2003 年
生命信息开始规模化数据化
人类基因组计划启动并完成

2009 年前后
单细胞转录组测序出现
为 AIVC 提供了数据准备

2016 年
开始建立“扰动–响应”数据
为 AI 预测扰动结果创造条件

2017 年 Transformer 架构提出，为生物基础模型提供算法基础

2020 年 AlphaFold2 取得突破进展，证明 AI 可以学习复杂生命规律

2023 年 5 月
大规模预训练验证基因网络迁移能力
博德研究所发布 Geneformer 模型

2024 年 2 月
生成式预训练拓展单细胞多任务能力
多伦多大学发布 scGPT

2024 年 06 月
超大规模全基因表达表征
清华大学联合百图生科发布发布 scFoundation

2024 年 10 月
细胞基础模型走向跨物种与知识融合
中国科学院指南针联盟发布 GeneCompass

2024 年 12 月
How to build the virtual cell with artificial intelligence
发表于 *Cell*，AIVC 成为明确、系统的研究愿景。

2025 年 2 月
细胞数据进入亿级规模
AI 制药公司 Tahoe Therapeutics 开源 Tahoe-100M 数据集，一个大规模药物扰动单细胞数据，进一步扩展模型训练规模

2025 年 6 月
Arc Institute 发起首届虚拟细胞挑战赛（VCC）
评估模型的预测准确性和跨细胞情境泛化能力

2025 年 12 月
美国 FDA 现代化法案 3.0
为 AIVC 数据用于药物申报提供监管通道

2026 年 1 月
AIVC 基础模型首笔药企授权合作
Noetik 与葛兰素史克（GSK）达成人类生物学模型授权合作，用于药物靶点发现和研发决策

2026 年 7 月
中国 NMPA 启动 NAMs “先锋计划”
推进计算机模拟等新方法（NAMs）研究和应用试点

2026 年 7 月
大额药企订单验证 AIVC 服务价值
Relation Therapeutics 与葛兰素史克（GSK）达成合作，开发用于训练 AI 的大型单细胞数据库，合同规模 1.1 亿美元

图 2 | AIVC 发展历程（来源：《麻省理工科技评论》中国、公开信息）

政策、技术、监管与资本条件共同成熟，AIVC 进入发展窗口期

AIVC 领域的快速升温并非由单一技术突破驱动，是政策、技术、监管与资本在同一时期形成共振。

政策层面

中美正将 AI 与生命科学的交叉能力提升至国家战略高度

中国："十五五"将人工智能、生物科技并列纳入科技战略部署

- 推动通用大模型与行业专用模型协同发展，通过高价值场景加快模型迭代。
- 面向健康等领域建设高质量数据集，完善数据标准和质量管理体系。
- 发展合成生物技术，加强生物数据资源的存储与安全管理。

美国：通过"创世纪任务"将 AI for Science 上升为跨部门国家工程

- 统筹国家实验室、联邦科研数据、超级计算资源与自动化实验平台。
- 公布国家科学与技术挑战，进一步提出以 AI 驱动生物技术、实现生物系统按需设计。
- 2026 年 7 月宣布投入超过 50 亿美元联邦承诺资金，覆盖首批 278 个项目。

技术层面

AI 正从处理信息走向理解和预测生命系统

- **基础模型向 AI for Science 延伸：**大模型正从语言、视觉等通用领域向材料、化学、物理和生命科学等学科拓展。在生命科学领域，其应用已覆盖蛋白质、单细胞、多组学和扰动响应，并逐步从识别生物特征走向预测细胞变化。
- **生命数据的观测通量和分辨率提升：**单细胞测序、空间组学等技术不断提高，使细胞状态及扰动结果能够进入大规模模型训练。

监管层面

计算证据开始获得进入药物评价体系的制度接口

- 中国 NMPA 发布"推进新方法（NAMs）研究和应用试点计划（先锋计划）"的征求意见，计算机模拟等新方法，正在从实验室中的技术探索，逐步进入方法验证、标准建立和药品监管应用的新阶段。
- 美国 FDA 推进新方法学（NAMs），分阶段减少临床前安全性研究中的动物试验，将人源体外系统、类器官、器官芯片、计算模型和真实世界数据纳入新的证据体系。

资本层面

市场开始为完整技术闭环买单

- Xaira Therapeutics：融资 10 亿美元，2026 年 3 月发布首款 AIVC 模型 X-Cell。
- 2026年 Recursion 与罗氏及其旗下基因泰克推进了累计 2.16 亿美元合作。
- 百曜科技等多家国内初创陆续获得融资，国内虚拟细胞加速布局。

AIVC 将成为科学发现与产业应用的新型基础设施

加速基础科学发现。 AIVC 能够整合单细胞、多组学、空间成像和扰动实验等多源数据，形成对细胞状态及其变化规律的统一表征。通过模拟基因、药物及环境扰动后的细胞响应，AIVC 可辅助识别关键调控因子与潜在作用机制，为实验验证提供更聚焦的候选假设，加快从数据分析走向机制发现。

重构生命机制研究范式。 AIVC 是 AI for Science 在细胞系统层的重要实现形态：它将 AI 的模式识别、假设生成和预测能力嵌入生命科学研究。推动研究走向“数据训练模型、模型生成假设、虚拟扰动预测、真实实验验证、实验数据回流”的迭代闭环。使生命机制研究逐步从经验驱动走向模型与实验协同驱动。

提升科研效率与产业创新能力。 AIVC 可作为核心系统模型被基础研究、靶点发现、细胞模拟、疾病模拟等场景调用，在海量候选方案中识别更可能成功的实验，减少低价值试错，并辅助确定下一轮数据采集和实验设计。随着模型与自动化实验平台等工具结合，科研人员能够探索过去受成本、时间和实验条件限制而难以覆盖的组合扰动与复杂生物情境，提升科研效率与创新边界。

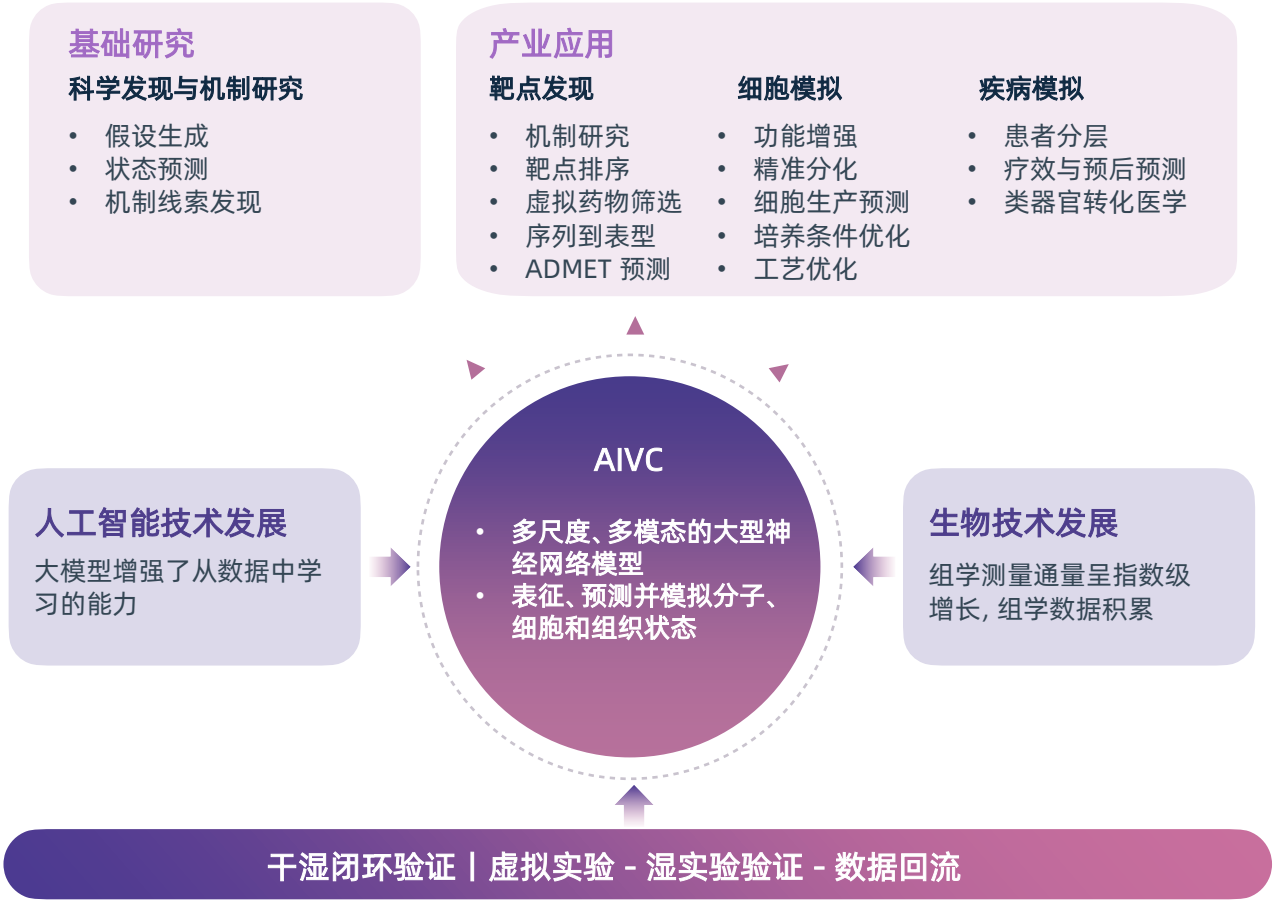


图 3 | AIVC 是 AI 时代生命科学的新基础设施（来源：《麻省理工科技评论》中国、Bunne et al., *Cell*, 2024.）

后 AlphaFold 时代，目标建立细胞尺度的新研究范式

人工智能进入生命科学后，以 AlphaFold 为代表的突破，首次证明了 AI 能够将蛋白质结构等复杂生命现象转化为可规模化预测的计算问题。由此，AI 制药（AIDD）快速发展，并在靶点发现、虚拟筛选、分子生成和候选物优化等环节提升了药物研发效率。关注重点主要是结构、性质、分子互作等物理化学问题。

但药物进入真实生物系统后，其效果还会受到细胞类型、基因调控网络、信号通路、代谢状态和疾病微环境等因素影响，这也是药物从分子设计走向体内和临床转化时仍面临的重要不确定性。

AIVC 将研究对象从“分子层面”推进到“细胞层面”。在这一新研究范式下，细胞不再只是实验观测的对象，而成为能够被表征、模拟、干预和持续校准的计算系统。

随着跨尺度数据与模型能力成熟，AIVC 构建的这套方法论，还可由细胞进一步延伸至代谢网络、组织、器官乃至个体层级，逐步形成连接微观机制与整体生理效应的生命系统模型。

多层级生命系统

宏观尺度 | 多模型调度

细胞、组织、器官与个体等模型逐级对齐；实验系统验证并校准体内效应。

AI 虚拟细胞 AIVC

介观尺度 | 细胞模型

处于连接分子与组织的介观层级。它不只考察单个分子与单个靶点的关系，而是把细胞视为由基因调控网络、信号通路、代谢过程和表型共同构成的动态系统。

AI 制药 AIDD

微观尺度 | 分子模型

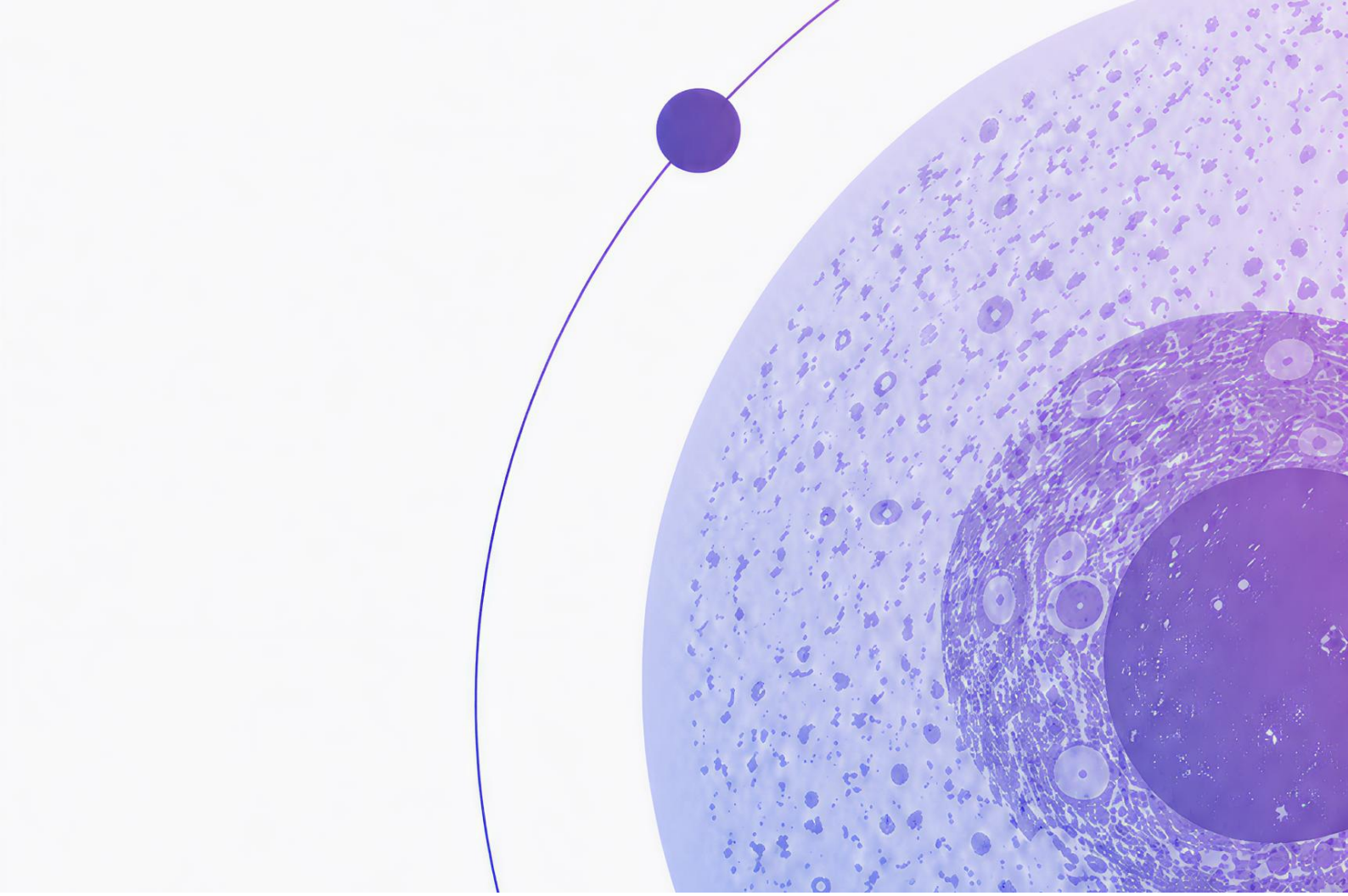
在分子尺度开展预测与设计，研究对象包括靶点蛋白、抗体、小分子及其相互作用。其典型任务包括蛋白结构预测、分子生成与筛选、结合亲和力优化。

从分子命中到细胞响应：AIVC 开启细胞状态预测

- AIDD 主要解决分子结构、性质与互作问题，提升候选分子的发现和优化效率，但难以判断一个靶点在特定疾病细胞中是否真的会带来有效响应。
- AIVC 在靶点命中之上，升维至细胞效应预测。既可预测药物分子干预后的细胞状态变化，同时也可以直接对细胞治疗进行预测和设计。
- 但细胞层结果仍不能简单外推为组织、器官或患者疗效。后续还需多尺度系统连接微环境、组织结构与个体差异，并通过实验持续验证和校准。

分子命中 → 细胞响应 → 系统效应

图 4 | 生命科学新研究范式演进（来源：《麻省理工科技评论》中国、君联资本周琼访谈、英矽智能访谈）



第 2 章

技术体系: 如何构建理解、预测和设计生命的 AIVC

数据: AIVC 构建的数据基础

模型: 从细胞表征走向扰动响应和多元范式探索

自主智能: 将科研问题拆解为可执行的虚拟实验和验证规划

干湿闭环: 构成 AIVC 持续迭代的长期价值与壁垒

AIVC 构建：数据、模型与智能体的分层架构

AIVC 的核心目标是建立能够连接不同生物状态的通用表征，预测细胞功能、行为与动态，并通过计算实验提出和检验科学假设，指导后续数据采集。

数据层面，AIVC 的构建起点是对细胞状态的数字化描述，现阶段多以单细胞转录组为主要数据基础，并逐步引入基因和药物扰动、空间组学、蛋白组、细胞形态、微环境及临床表型等信息。

其中，新生产数据是构建 AIVC 的关键数据要素之一，针对模型训练目标，通过设计实验、控制条件和统一测量形成的数据，能够更好地补足关键状态、扰动响应和多模态对应关系。

模型层面，将来源分散、形式各异的生物数据转化为可计算的细胞状态，并在这一状态空间中读取细胞特征、模拟外部干预、预测状态变化。

- 通用表征（Universal Representation, UR）：把不同物种、不同实验条件、不同测量技术下得到的细胞状态，映射到一个统一的、连续的向量空间中。
- 虚拟工具（Virtual Instruments, VI）：将通用表征中的数字状态转化为细胞类型、基因表达、形态或功能等人可理解的结果；模拟基因编辑、药物处理、细胞分化或环境变化如何推动细胞由一种状态迁移至另一种状态。

应用层面，自主智能负责理解和拆解科研问题，编排调用 AIVC 模型、虚拟工具、知识库及分析工具，完成虚拟扰动、候选方案排序和实验设计。预测结果经过湿实验验证后形成新数据，再反向用于模型迭代。

图 5 | AIVC 构建的数据、模型、应用分层（来源：《麻省理工科技评论》中国、百曜科技访谈）

静态多组学数据为 AIVC 建立统一的细胞状态空间

细胞是一个不断变化的复杂系统，模型构建一个条件化的细胞状态转移规律："初始细胞状态 + 所处环境 + 外部干预 + 时间"如何共同决定新的分子状态、细胞表型和组织结果。

因此，AIVC 至少要回答三类问题：这是什么细胞，它现在处于什么状态？敲除一个基因、加入一种药物或改变环境，它会怎么变化？这种分子层面的变化，最终是否会改变细胞功能、组织行为或患者疗效？

为了解决以上问题，AIVC 需要细胞静态数据、扰动–响应数据、临床数据以及生物学先验知识。

01 细胞静态数据

用于建立细胞状态空间，让模型理解不同细胞类型、组织、疾病和发育阶段的分子特征。

主要包括

- **基因组：**全基因组测序
- **转录组：**以单细胞转录组数据为主，反映不同基因在特定时间和空间下的活跃程度
- **表观组：**染色质状态、染色质可及性、组蛋白修饰等，揭示基因的开放、关闭或易被激活的状态
- **蛋白组：**刻画细胞中蛋白质的组成及状态，反映基因表达最终转化为哪些功能分子
- 未来还可以加入空间转录组、空间多组学、DNA 甲基化、染色质三维结构等信息

作用：帮助模型学习细胞之间的共性、差异和潜在调控关系，是 AIVC 的基础表示层。

局限：静态数据仅能描述当前状态，**不能单独说明细胞受到干预后的变化**。

当前构建 AIVC 主要以单细胞转录组数据为主。其主要反映特定时间点细胞的基因表达状态，能够刻画细胞的类型、功能及所处状态。当模型学习大量不同细胞类型、疾病阶段和实验条件下的转录组后，便可以识别基因之间的协同关系和调控模式。但本质上单细胞转录组属于对细胞状态的静态观测，难以完整呈现细胞随时间发生的连续变化及其动态调控过程。

细胞状态的形成和变化是多层级因素共同作用的结果，因此，仅依赖单细胞转录组数据，难以进一步解释细胞状态如何形成、向何种方向变化，以及受到特定干预后会产生怎样的响应。要实现对细胞动态过程的理解和预测，还需要结合表观组、蛋白组、空间信息以及不同时间点和扰动条件下的数据，从单一状态描述进一步走向对细胞状态变化规律的建模。

扰动数据让细胞变化成为可学习、可预测的对象

扰动-响应数据记录细胞在遗传、化学或环境干预前后的状态变化，能够为模型学习干预因素与分子及表型响应之间的关系提供直接依据。

02 扰动–响应数据

这是 AIVC 学习因果关系和细胞状态转移的核心数据，即**细胞受到干预后如何变化。**

扰动条件

- 基因敲除、敲低、激活和碱基编辑
- 小分子、抗体、细胞因子和其他刺激
- 剂量、作用时间和给药顺序
- 单基因、多基因和药物组合
- 缺氧、营养、感染、温度和细胞间相互作用等环境条件

细胞分子与功能响应

- RNA、染色质、蛋白质、磷酸化和代谢变化
- 细胞存活、增殖、死亡和迁移
- 细胞形态、分化、免疫杀伤和电生理活动
- 药物敏感性、耐药性、毒性和组织功能

细胞动态、时空与谱系响应

- 多时间点单细胞测量
- 活细胞成像和细胞分裂追踪
- 空间转录组、空间蛋白组
- 细胞邻域、细胞通讯和微环境
- 谱系追踪、分化轨迹和细胞命运

作用：让模型学习干预如何改变细胞，并把分子变化连接到结果。

关键要求：扰动数据必须有匹配的**基线、对照、重复、扰动效率、时间和剂量信息**，并尽量覆盖不同细胞类型、供体和生物环境。

扰动-响应数据进一步提供特定基因、药物或环境干预条件下，细胞由初始状态向响应状态转变的观测样本。通过联合初始细胞状态、扰动类型、剂量、作用时间、细胞背景和干预后多维读出，模型可以学习条件依赖的细胞状态转移规律，并预测未观测扰动或未见细胞背景下的转录响应和细胞命运变化。随着多时间点、剂量梯度、组合扰动和多模态响应数据不断积累，AIVC 可以由静态状态表征进一步扩展至细胞状态变化轨迹的预测。

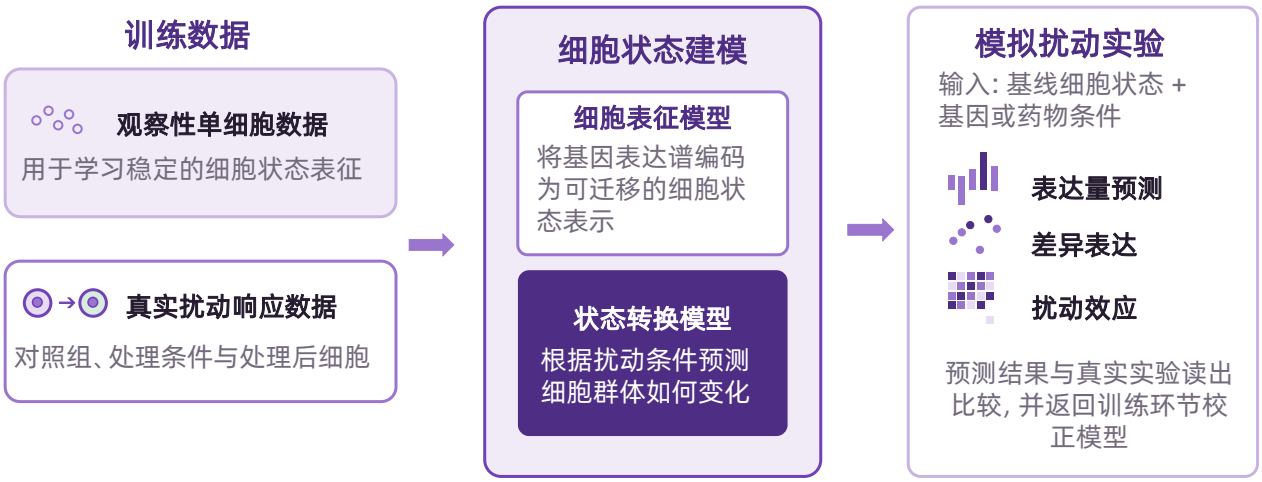


图 6 | 扰动响应数据推动 AIVC 从状态识别走向状态变化预测（来源：《麻省理工科技评论》中国）

临床数据和先验知识共同让 AIVC 更接近真实生命系统

与此同时，临床肿瘤样本、空间微环境等难以规模化获取的数据，能够帮助模型缩小标准实验体系与真实人体环境之间的差距。临床数据用于检验模型预测能否对齐患者与临床预后；生物学先验知识则通过调控网络、信号通路等已知规律，为模型学习提供结构化约束，提升其外推与解释能力。

03 临床数据

细胞层面的预测能否转化为真实人体结果，用于把体外细胞模型逐步对齐到患者和临床任务。

主要包括

- 患者遗传背景、突变和疾病分型
- 治疗前后的单细胞和多组学数据
- 病理图像、空间微环境和组织状态
- 用药方案、剂量、联合治疗和治疗时间
- 疗效、复发、生存、不良反应和耐药
- 临床成功药物与临床失败药物
- 多中心、不同人群和独立患者队列

作用：帮助模型学习体外细胞反应与人体药效之间的差异，支持患者分层、药物敏感性预测和临床决策。

关键难点：① 跨模态数据难以实现有效的对齐。② 患者数据规模较小、变量复杂、干预难以严格控制。

04 生物学先验知识

生物学先验知识是人类长期实验与研究的积累，作为知识层，为模型理解数据、建立生物学联系和约束预测结果提供支持。

主要包括

- DNA、RNA 和蛋白质的序列与结构信息
- 基因调控网络和转录因子–靶基因关系
- 蛋白质互作网络及蛋白复合物
- 信号通路、代谢通路和基因功能注释
- 基因家族、同源关系和跨物种保守信息
- 药物结构、药物–靶点和药物–通路关系
- 基因–疾病、基因–表型及疾病机制知识
- 论文和专业数据库中积累的生物学知识

作用：① 帮助模型识别实验数据中难以直接观察的生物学联系，在扰动数据有限时提升知识迁移与外推能力。② 约束模型，使预测更符合已知的基因调控、信号传导和分子作用规律，并增强可解释性。

（来源：《麻省理工科技评论》中国，西湖大学裴唯珂访谈）

去噪、去重、标准化与跨模态对齐，数据才能成为有效训练语料

去噪 | 剔除技术偏差，保留真实生物信号

AIVC 模型的性能不仅取决于数据规模，更取决于数据是否准确、稳定并具有明确的生物学意义。单细胞数据在进入分析和模型训练前往往需要投入大量精力进行清洗，低质量数据会引入噪声、批次效应和错误标注，增加后续处理难度。去噪不能仅依赖统一阈值机械过滤，还需要结合统计指标与生物学先验，避免将稀有细胞状态或特定条件下的弱信号一并删除。

去重 | 去除冗余，避免重复信息被过度学习

通过识别并控制同一序列或高度相似信息的冗余，训练语料能够保持更合理的信息分布，使不同信息以适当比例参与训练，帮助模型形成更稳定的参数。去重可以避免重复内容被过度强化，使模型学习更接近真实的生物规律。

标准化 | 统一数据表达，补全实验语境

通过统一数据的记录、处理和表达规则，并对测序深度、覆盖范围及质量指标设定明确标准，不同来源、平台与批次的数据才能形成一致、可比较的质量基础。充足且稳定的测序深度有助于识别低丰度基因、稀有细胞亚群和较弱的扰动响应。每一批数据都应先达到既定标准，再转化为可直接用于模型训练的语料。

跨模态对齐 | 让不同模态生物数据建立可学习的对应关系

对齐是数据加工中最核心、难度最高的要求，只有让不同来源的信息产生内在关联，才能突破单一模态数据的局限。

跨模态对齐不仅覆盖组学数据的整合，还涉及影像（病理切片、医学影像等）、生理生化指标（心电、脑电等）等差异极大的模态，目标是找到这些独立数据向量之间的底层逻辑关联，将基因与功能、序列与表型、分子变化与病理状态联系起来。

为 AIVC 训练提供可靠数据

去噪
保留真实生物信号
- 剔除技术噪声
- 控制批次效应
- 保护稀有状态

去重
保持合理信息分布
- 识别重复序列
- 控制冗余比例
- 避免过度学习

标准化
形成一致质量基础
- 统一记录与表达
- 明确测序质量标准
- 保证数据可比较

跨模态对齐
建立跨模态对应关系
- 突破单一模态局限
- 对齐组学、影像与表型
- 支持多模态融合

图 7 | 如何构建可靠的高质量数据（来源：《麻省理工科技评论》中国、陈润生院士访谈、耶鲁大学张迪访谈）

自监督预训练将细胞数据转化为可迁移的通用表征

AIVC 模型的构建始于对细胞状态及其影响因素的系统化数据建模，目标是建立能够统一表征分子、细胞与多细胞组织状态，并在特定生物学情境和干预条件下预测状态变化及可观测表型的人工智能系统。当前的模型训练以大规模单细胞数据为基础，融合跨物种同源基因信息、生物学先验知识等，通过自监督预训练学习基因之间的关联及细胞状态表征，经过微调和解码后，模型可用于细胞类型注释、扰动预测、剂量反应分析和基因调控网络推断。

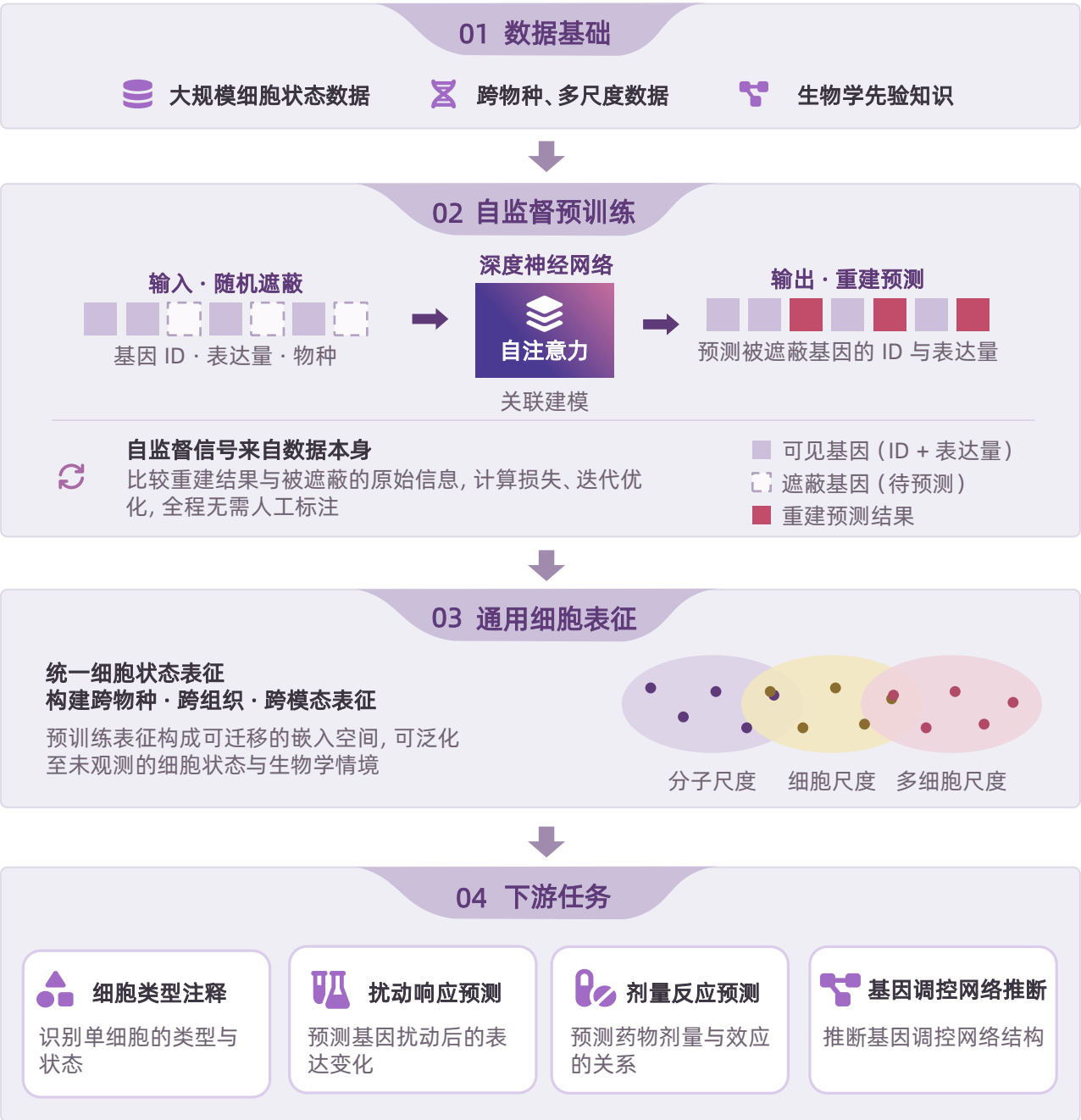


图 8 | AIVC 模型预训练（来源：《麻省理工科技评论》中国、Yang X, et al. *Cell Research*, 2024、Bunne et al., *Cell*, 2024 、中国科学院动物研究所李鑫访谈）

细胞建模正从静态表征推进到状态迁移与连续演化

AIVC 模型也正在经历快速的迭代，其发展不仅体现在模型规模和数据规模的提升，也是建模对象与建模目标的重新定义。以 Transformer 为代表的自监督预训练，主要通过遮蔽训练等方式学习细胞数据中的统计规律。当前，领域内正在向更高阶的动态建模探索：包括潜在状态预测、因果机制建模、连续动力学生成、多模态多尺度融合等。

	① 细胞状态建模	② 细胞扰动建模	③ 细胞世界模型探索
核心目标	对细胞状态进行表征，回答“细胞是什么”。	给定初始状态与干预条件，预测扰动后的细胞状态变化。	可干预的动态细胞世界模型探索，在潜在空间中学习状态转移规律。
核心问题	细胞是什么	细胞如何变化	细胞如何演化
建模对象	细胞状态	状态轨迹	状态转移与环境交互
数据特征	静态单细胞数据	时序 / 扰动数据	多模态 + 时空 + 干预
预测目标	当前状态	未来状态 / 轨迹	状态转移过程
技术范式	Transformer	Transformer	多路线
核心能力	表征与生成	动态预测	模拟、干预与推演

当前探索技术路线：

潜在状态预测：不直接预测每个基因或每个观测值，而是在潜在空间中预测细胞状态在时间和干预下如何变化。

因果机制建模：以因果图模型、结构因果模型等因果建模方法为代表，学习基因、通路、蛋白、表型之间的因果关系和机制链条。

连续动力学生成：以 Neural ODE、Flow Matching 等连续动力学方法为代表，把细胞变化建模成分布或状态空间中的连续演化过程。

多模态多尺度融合：以多模态 Foundation Model、跨模态表征学习等方法为代表，把转录组、表观组、蛋白组、代谢组、空间组学、形态学、细胞邻域和组织环境纳入统一表示。

其中，以 JEPA 为代表的世界模型是更适合推理的范式之一，进一步尝试在潜在空间中预测系统未来的状态变化，将预测对象从数据本身进一步抽象到潜在状态，为动态系统建模提供了新的技术思路。世界模型 JEPA 架构嵌入,突破传统 Transformer 架构局限，多视角，降噪音，可以提升时空动态数据处理能力、提升多模态数据处理能力、能够进行更高效的特征学习。

代表性模型图谱: 从单模态表征走向扰动表征

目前, AIVC 领域尚未形成单一主导的技术范式, 不同模型主要围绕预测对象、数据类型与状态表达方式展开。整体来看, 相关技术路线已由单模态模型, 逐步延伸至扰动模型、LLM 基础、多模态模型以及向动态建模的新兴范式演进。

01 单模态表征模型

代表模型	年份	核心团队 / 机构	主要信息
TranscriptFormer	2026	陈·扎克伯格生物中心–斯坦福联合团队	基于最多 1.12 亿个细胞、覆盖 12 个物种和约 15.3 亿年进化跨度的单细胞转录组训练生成式细胞图谱模型。
Cell-GraphCompass /GeneCompass	2025/ 2024	李鑫 团队/指南针联盟 中国科学院动物研究所/ 中国科学院	以图结构学习将基因调控关系显式编码为模型归纳偏置, 探索结构感知型预训练路径。/整合逾 1.01 亿质控后的人源与鼠源细胞, 融合基因调控网络与跨物种先验, 构建知识增强型跨物种基础模型。
scGPT	2024	Bo Wang 团队 多伦多大学	基于 3300 万人类细胞 (主要源自 CELLxGENE) 生成式预训练, 是公开生态中数据覆盖较完整的人类单细胞基础模型之一。
scFoundation	2024	百图生科 × 清华大学	基于逾 5000 万人类单细胞转录组、覆盖约 2 万个基因, 验证超大规模预训练在单细胞表征学习中的可扩展性。
Geneformer	2023	Theodoris 团队	基于约 2990 万人类单细胞转录组预训练, 构建静态细胞表征基座, 支撑基因调控网络推断与疾病靶点发现。

02 扰动模型

代表模型	年份	核心团队 / 机构	主要信息
UniPert-G2CP	2026	李敏 团队/姚建华 团队 中南大学/腾讯生命科学实验室	统一基因与药物扰动表征, 将遗传筛选知识迁移至不同细胞背景下的药物响应预测; 覆盖 4994 个基因、7860 个化合物及 5 种癌症细胞系.
X-Cell	2026	Xaira Therapeutics	基于 X-Atlas/Pisces 约 2560 万 CRISPRi 扰动细胞、16 种细胞上下文, 融合扩散语言模型与因果先验, 构建因果扰动预测框架。
State / Stack	2025/ 2026	Arc Institute	基于 1.67 亿细胞与逾 1 亿扰动态细胞混合语料预训练。/基于约 1.49 亿预训练与 5500 万后训练细胞构建上下文学习框架, 探索大规模语料中的少样本生物学推理能力。
GEARS	2022	Jure Leskovec 团队 斯坦福大学	基于 Perturb-seq 基因扰动数据训练, 以人类 K562 和 RPE-1 细胞系为实验体系, 建立单细胞基因扰动效应预测的经典基线方法。
CPA / chemCPA	2021/ 2022	Theis 团队 德国亥姆霍兹慕尼黑研究中心	以条件生成模型建模药物处理数据 (含剂量与组合信息), 开创单细胞药物反应外推与剂量-效应建模的计算范式。

* 基于公开信息不完全统计

代表性模型图谱: LLM 基础与新兴范式探索

03 LLM 基础的细胞模型

代表模型	年份	核心团队 / 机构	主要信息
CellWhisperer	2025	Christoph Bock 团队 维也纳医科大学	通过对比学习对齐约 100 万份 RNA 测序数据及其文本描述，支持自由文本检索、无参考零样本注释和对话式图谱探索，是语义对齐/交互模型。
VCWorld	2025	郑双佳 团队 上海交通大学	融合大语言模型与生物知识图谱，通过显式知识检索与推理预测药物扰动下的细胞响应，强调预测结果的可解释性与可追溯性。
LangCell	2024	聂再清 团队 清华大学	包含约 2750 万条数据，覆盖了细胞类型、发育阶段、组织器官、疾病等 8 个维度的信息。
Cell2Sentence / C2S-Scale	2024/ 2025	David van Dijk 团队 耶鲁大学	将转录组表征转化为自然语言序列并进行生成式预训练，开创“细胞即文本”的跨模态表征范式，为 LLM-JEPA 混合路线提供方法论来源。
GenePT	2023	Chen Y / Zou J 斯坦福大学	利用预训练大语言模型对基因文本描述进行语义嵌入，将外部生物医学知识显式注入表征空间，探索知识驱动型零样本细胞分析。

04 多模态模型

代表模型	年份	核心团队 / 机构	主要信息
Nicheformer	2025	Theis 团队 德国亥姆霍兹慕尼黑研究中心	同时使用人类和小鼠的解离单细胞转录组及空间转录组数据，学习细胞状态与组织微环境的联合表示，可预测空间组成、空间标签。
HoloCell	2026	鄞启进、刘海广、江瑞等 中关村学院/清华大学等	面向表观组、转录组和蛋白质组的统一生成式基础模型，基于约 4.68 亿单细胞多组学数据预训练，支持多组学表征、跨模态对齐与生成，探索统一建模细胞状态。
scGPT-spatial	2025	Bo Wang 团队 多伦多大学	使用约 3000 万个细胞/空间点数据，引入空间感知采样、邻域建模和 MoE 解码，强化对组织空间关系与细胞邻域状态的建模。

05 新兴范式: 从潜在状态预测到跨扰动状态建模

代表模型	年份	核心团队 / 机构	主要信息
CellOS / AURA CellOS	2026	百曜科技	基于 3.905 亿单细胞转录组与 1.2 亿扰动细胞混合语料，以 JEPA 架构构建细胞世界模型，探索异质数据下的潜在动态预测。
AlphaCell	2026	刘琦 团队 同济大学	一种具备完备细胞状态表征、连续几何流形塑造、状态转移建模以及真实空间重构的虚拟细胞世界模型原型系统（POC study）。
AIDO Cell	2026	GenBio AI	多尺度、有状态人类细胞世界模型,强调 DNA、RNA、蛋白、调控网络、影像等多尺度读出。

* 基于公开信息不完全统计

后训练把通用基模转化为垂直场景模型

通用基模学习的是跨细胞类型、跨条件可迁移的共性表征与基础扰动规律，面向应用端是更具体的靶点发现、细胞模拟或疾病模拟。若直接调用通用能力，模型往往缺少对关键条件的刻画。后训练因此成为连接通用基模与场景能力的关键环节。通过引入垂类数据和明确任务目标，使模型学习特定场景下的状态变化与响应规律，并接受实验、临床或生产结果的持续校准。

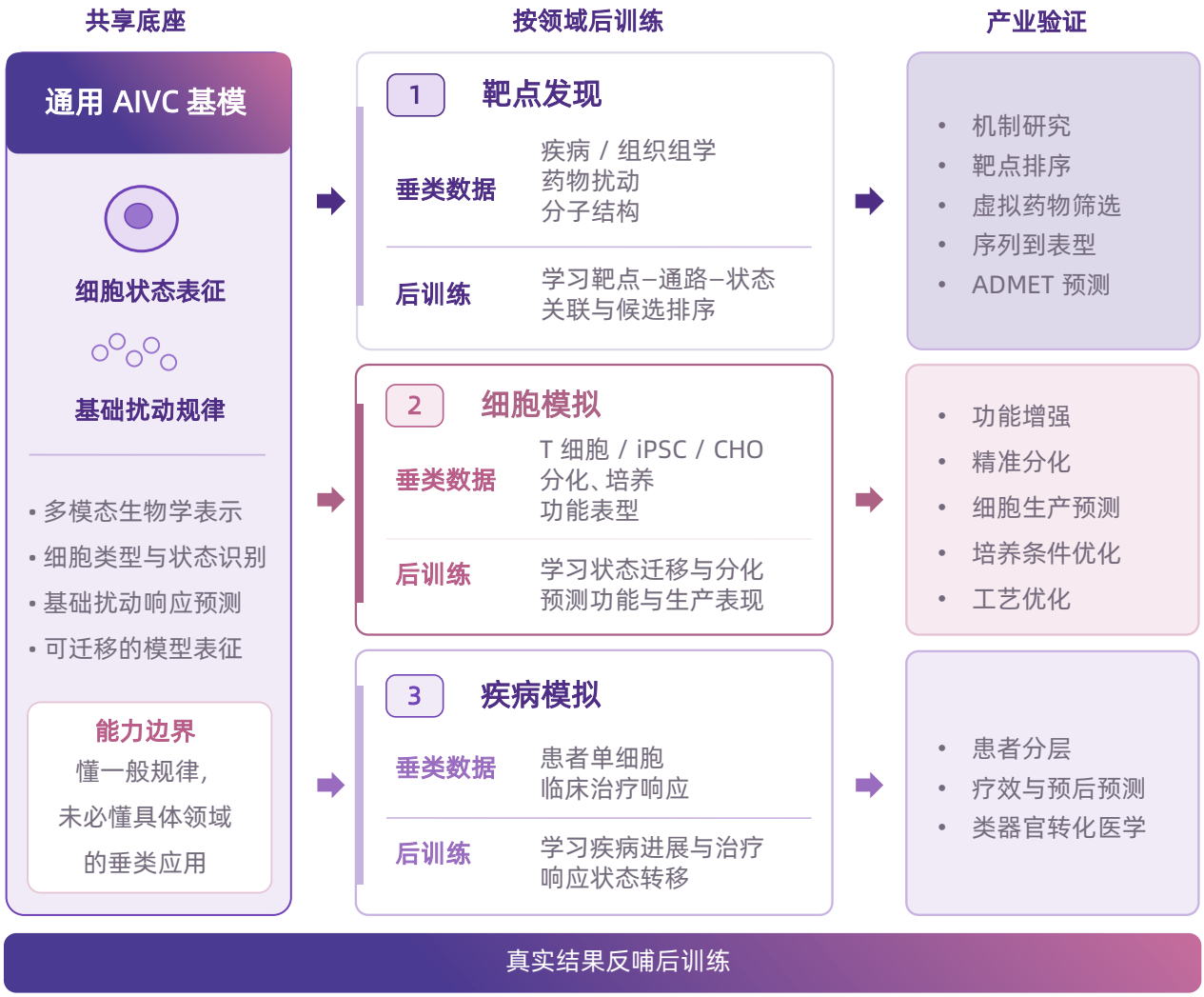


图 9 | 后训练让通用基模形成垂直模型能力（来源：《麻省理工科技评论》中国、Feng et al., *Advanced Science*, 2026.、Tejada-Lapuerta et al., *Nature Methods*, 2025）

从通用基模走向垂直模型的最优路径仍处于探索阶段。共享基模的领域后训练有利于保留迁移能力和泛化能力；对于高价值、数据密集、目标清晰且具备实验、临床或生产验证闭环的场景，也可以从头训练专用模型，以牺牲通用性换取更深的场景能力和更快的产业落地。两条路径的共同前提，是模型能力边界与实际场景需求相匹配，并由真实结果持续校准。

自主智能将科研问题拆解为可执行的虚拟实验和验证规划

随着基础模型能力不断提升，AIVC 也正从单一任务的辅助工具，走向能够理解科研目标、自主规划任务、调用专业工具并根据实验结果持续迭代的自主智能系统，将进一步具备组织和执行虚拟实验、连接模型与湿实验的能力。

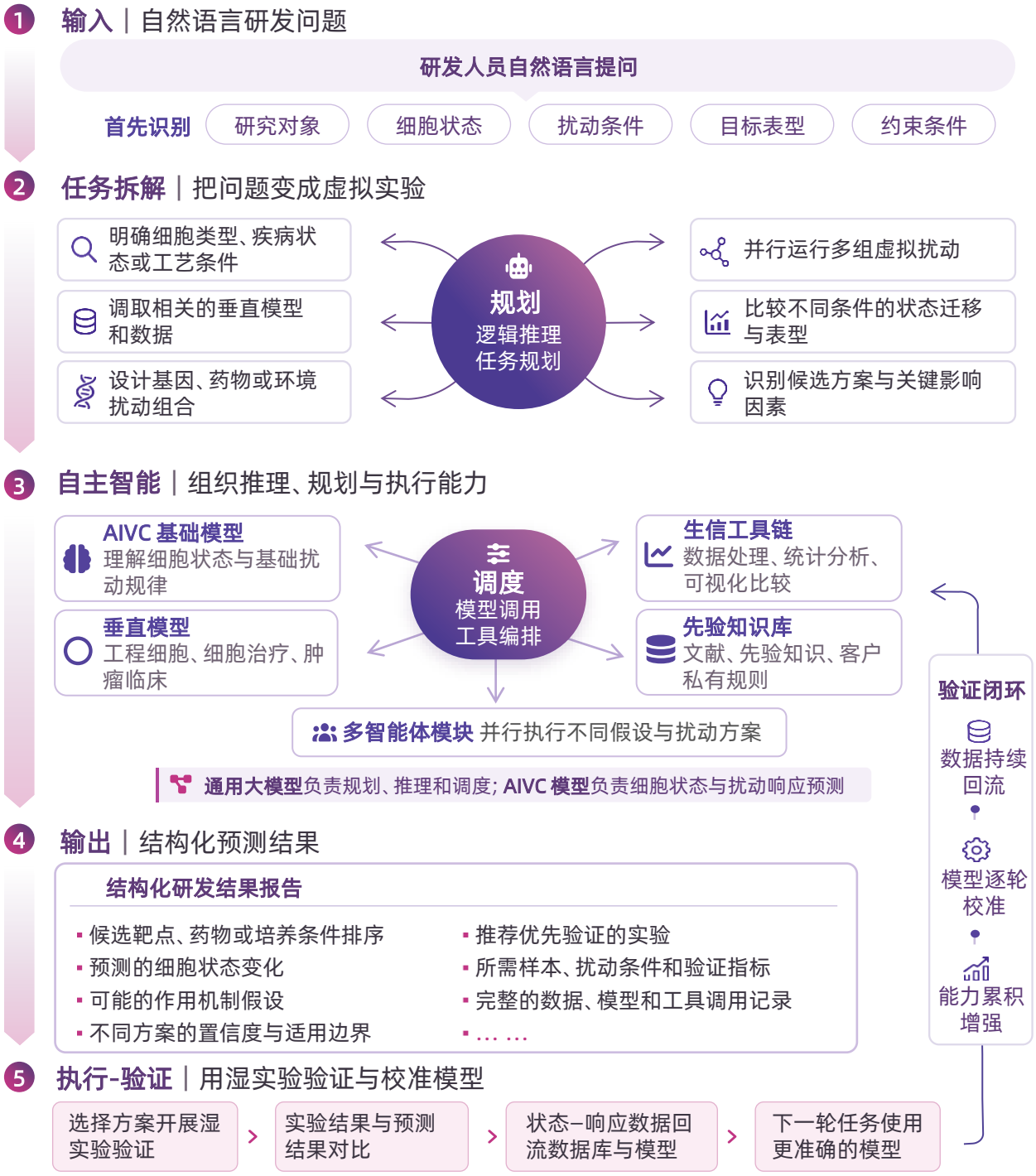


图 10 ｜ 自主智能工作流（来源：《麻省理工科技评论》中国）

AIVC 智能体项目图谱: 市场供给仍处于早期阶段

在生命科学研究中，模型通常只解决一个局部问题，根据细胞状态和扰动条件，预测细胞可能发生的变化，而真正的科研任务还需要将模型适配到不同问题和任务流中，并按需调用知识库、分析软件和实验工具。智能体因此成为连接模型与科研流程的关键环节，它理解科学家的问题，组织和调度多种能力，将分散的模型输出转化为可执行、可验证的研究方案。

代表性 AIVC 智能体:

产品名称	开发方	专注方向
CellAura	百曜科技	面向应用的 AIVC 任务编排。调度 AIVC 模型、生信工具、知识库和实验反馈，完成虚拟扰动预测与智能体。
CellForge	耶鲁大学等	面向模型研发的 AIVC 方法设计。根据单细胞多组学数据和任务描述，自动设计、训练、验证 AIVC 模型。
Virtual Aging Cell（VAC）	英矽智能	基于多智能体群架构的虚拟衰老细胞（VAC）平台，将“年龄/时间”贯穿于持久共享的机体全局环境，动态模拟细胞随时间推移的分化与损伤累积过程，并模拟对细胞命运的干预与重塑。
CellHermes	同济大学等	探索解决异源异构组学数据与LLM对齐的挑战，基于强化学习使LLM内化生物学机制、推理，进而支持生物学发现。
VCHarness	GenBio AI	面向虚拟细胞扰动响应模型的自主研发，结合生物基础模型、编码代理与搜索策略，迭代完成模型设计、代码生成、训练和评估。
TPP（The Popper Project）	无界进化	面向药物研发场景，将模型预测、实验设计、实验验证与结果反馈纳入统一工作流。

* 排序无先后顺序，基于公开信息不完全统计

AIVC 智能体的市场供给仍处于早期阶段。面向 AGI 的人工智能技术、面向细胞状态与扰动响应的预测模型，以及负责模型调用和科研流程协同的智能体系统，均处于持续探索与验证阶段。多层技术的成熟度与集成难度共同制约产品落地，经过充分验证、可稳定复用的成熟供给仍较为有限。率先开展探索的参与者，有望通过模型与工具集成、实验反馈及应用经验的持续积累，形成先发优势。

干湿闭环构成 AIVC 持续迭代的长期价值与壁垒

自动化平台同时连接两端：向后为新的模型训练生产数据，向前承接模型预测后的实验验证。结果一方面用于判断预测是否成立，另一方面沉淀为新的训练数据，回流数据–模型平台进行误差校准和场景优化。由此，自动化实验成为连接预测、验证与持续产数的关键节点，推动系统形成可迭代的干湿闭环。

图 11 | 数据–模型–实验的循环飞轮（来源：《麻省理工科技评论》中国、百曜科技访谈）

数据来源：外部数据共计与自建数据生产结合。公开数据库、商业化数据库为模型提供广泛的细胞状态先验和基础训练信息。自建实验数据则形成面向具体场景的核心壁垒。

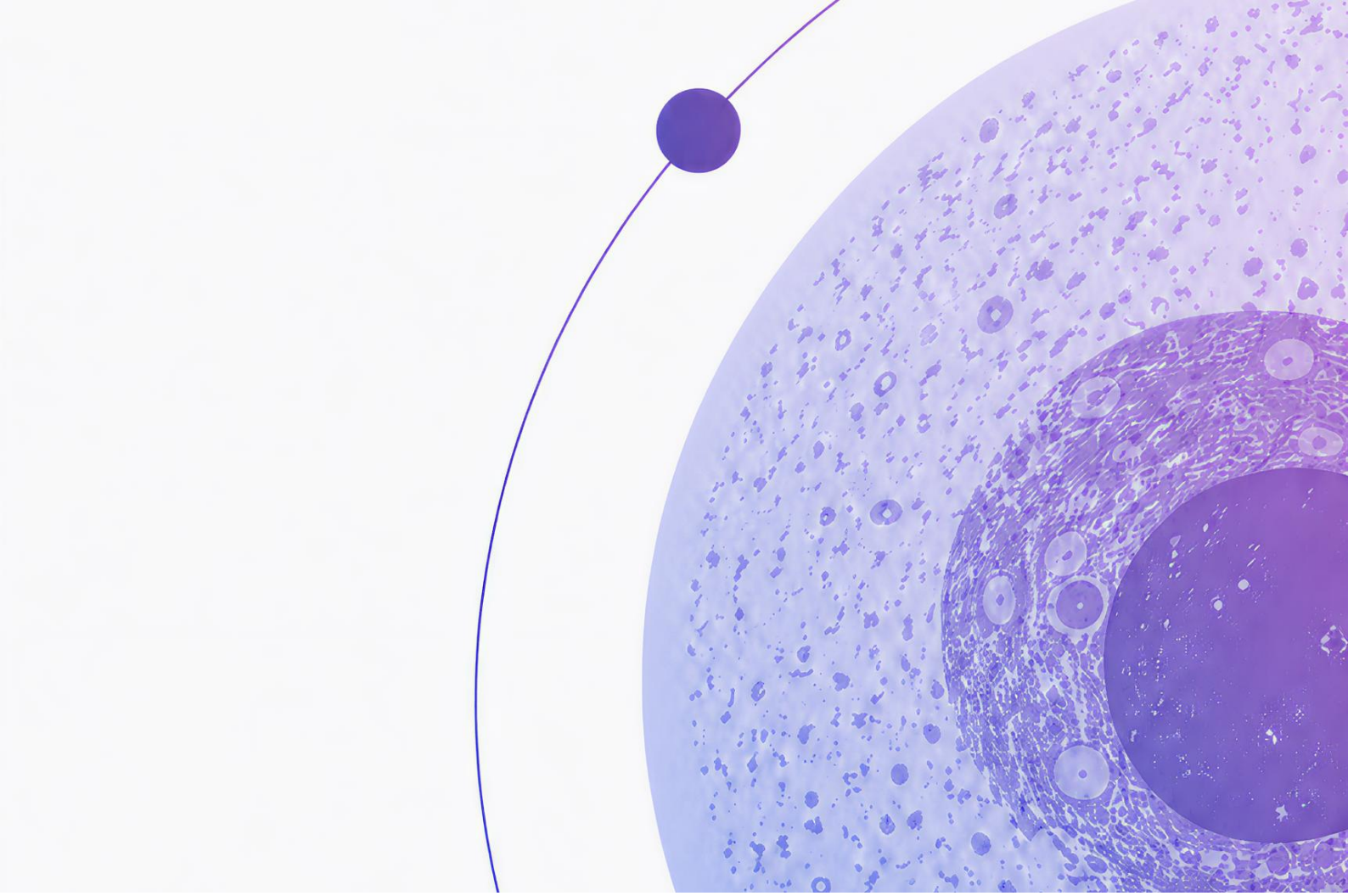
外部数据：

- 覆盖广、启动快
- 适合建立基础模型
- 数据质量不可控
- 数据无法独占
- 数据信息与开发目标不完全匹配

自建数据生产：

- 成本更高、周期更长
- 质量更可靠
- 覆盖关键扰动数据
- 更直接收集多组学数据
- 形成场景化能力和数据资产

干湿闭环：数据工厂对外部数据库、自有实验和临床合作等多源数据进行清洗、标注、质控与多模态对齐，形成可用于训练的细胞状态和扰动数据。模型平台基于这些数据学习细胞状态变化规律，输出基因、药物或培养条件等扰动预测及实验方案；自动化平台随后执行实验并采集真实响应。每一轮实验同时产生两类结果：一是对模型预测的验证与误差反馈，二是可进入下一轮训练的新数据。经过回流、校准和迭代，系统逐步提升在特定细胞类型、疾病或研发场景中的预测可靠性，形成“数据生产–模型预测–实验验证–数据回流”的干湿闭环。



第 3 章

产业发展: 从模型开发步入应用验证

产业格局: 新锐公司与成熟企业加速入局, 产业生态初步形成

应用实践: 从技术探索走向高价值应用场景

能力验证: 走向真实场景中的效果评价

价值评估: 从 Biotech 与 AI 双重赛道审视企业投资价值

AIVC 价值链：上游基础提升上限，下游场景兑现价值

上游决定 AIVC 获得的数据获取、数据质量和算力条件；中游将这些资源转化为细胞表征、扰动预测和机制分析能力；下游通过研发、临床与生产场景，兑现模型与服务价值。

上游 | 数据、实验与算力基础设施

测序与单细胞技术平台
- 提供样本测序、单细胞/空间组学与多组学数据采集能力

开放生物图谱与训练数据资源
- 提供细胞图谱、基因功能、扰动、疾病与临床数据
- 如 Human Cell Atlas、CZ CELLxGENE 等

生命科学 AI 开发基础设施
- 生命科学研究基础设施方
- 如 NVIDIA BioNeMo、Microsoft Discovery 等

CRO 与实验数据服务
- 提供样本制备、药理药效、扰动实验与结果验证服务。

AI 芯片与算力基础设施
- 提供 AI 芯片、云计算、数据存储、模型训练与部署环境
- 如 NVIDIA、AWS 等

↓

中游 | AIVC 模型与服务

把细胞数据和生物学知识转化为可调用的细胞表征、扰动预测与机制分析能力

↓

下游 | 需求方与典型交付

药企、Biotech 与 CXO

药物研发

核心任务：
- 靶点发现
- 药物响应预测
- 虚拟筛选
- 虚拟动物实验

典型交付：
- 候选靶点
- 药物响应模型
- 虚拟筛选结果
- 模型授权

细胞治疗、干细胞

核心任务：
- 细胞功能优化
- 干细胞定向分化
- 管线优化

典型交付：
- 分化预测模型
- 细胞产品优化方案
- 工艺参数
- 管线/IP 授权

工程细胞、生物制造

核心任务：
- 细胞改造
- 产量提升
- 工艺优化

典型交付：
- 工程细胞模型
- 高产细胞株
- 培养工艺优化
- 细胞株授权

临床研究

核心任务：
- 患者分层
- 标志物发现
- 药物响应研究
- 联合用药
- 适应症拓展

典型交付：
- 候选标志物
- 响应分析
- 数据共研

消费健康

核心任务：
- 功效评价
- 成分筛选
- 配方开发

典型交付：
- 抗衰老护肤品
- 健康食品
- 农业产品
- 功效预测
- 配方迭代
- 产品验证

高校与科研机构

核心任务：
- 细胞机制解析
- 扰动预测
- 实验筛选

典型交付：
- 模型/API
- 虚拟实验结果
- 机制假设
- 联合研究

AIVC 产业呈现原生创新与延伸布局双线并进格局

AIVC 原生创新企业

vitaura 百曜科技
元示科技 YUANSHI TECHNOLOGY
INFevo
华源智因

AIVC 延伸布局企业

Insilico Medicine HKEX: 3696
华大BGI
Tencent 腾讯
阿里巴巴 Alibaba.com
ByteDance 字节跳动
BioMap 百图生科
DP Technology 深势科技
西湖欧米 WESTLAKE OMICS
ONEX INTELLIGENCE 观壹智能
EonArk
BioCosmos X
水木分子 PHARMOLIX
耀速科技 Xellar
iORGAN 淇嘉科技
衍因科技
联川生物
OxTium 津渡生科
瑞普晨创 REPROGENIX
BioGeometry
数因智科
REVIR Therapeutics
Nanomics 人工智能数字化蛋白组学平台
璨辰科技 CANCHEN TECH
哲源科技 PHILRIVERS
旦白龙 DEEPKINASE 嘉华药锐
贝芸 BioAI-global
ZYBio

* 排序无先后顺序，基于公开信息不完全统计，持续更新中

海外 AIVC 相关公司

Google DeepMind
Meta
CZI
Microsoft
NVIDIA
NOETIK
xaira
Recursion
Cellarity
CytoReason
Algocell
OMPHALOS LIFESCIENCES
GenBio
turbine
BIOPTIMUS
NeoLife
CELLULAR INTELLIGENCE
TAHOE
Shift bioscience
DeepLife.
Cytocast
ChemT Biotech
biomy
CellType
RELATION
tolemy bio
QurieGen
CARTABIO
Panacent BIO
Infinitusbio.AI
reflector.bio

* 排序无先后顺序，基于公开信息不完全统计，持续更新中

融资与项目并进，AIVC 进入多路径商业探索新阶段

国内代表性融资事件及进展

企业	时间	事件	关键进展
百曜科技	2026.07	多轮融资	总部位于北京，近期完成数轮累计过亿元融资
华源智因	2026.07	种子轮	总部位于北京，完成千万级人民币种子轮融资
元示科技	2026.07	新一轮	总部位于上海，完成数千万最新一轮融资
无界进化	2026.06	天使轮	总部位于北京，由晶泰科技孵化，完成数千万元天使轮融资
深度细胞	2026.08	种子轮	总部位于北京，完成数千万种子轮融资
英矽智能	2026.08	产业进展	中国香港上市公司，发布多智能体驱动的多模态虚拟衰老细胞生成平台
百图生科	2026.03	提交 IPO	总部位于北京，以保密形式向港交所提交上市申请
华大集团	2026.09	产业进展	总部位于深圳，华大智造与百图生科达成合作、华大基因研究院参与 AIVC 成果发表
深势科技	2026.09	产业进展	总部位于北京，与北京智能研究院等单位合作，发表 AIVC 相关成果
西湖欧米	2026.07	新一轮融资	总部位于杭州，以蛋白质组学+AI虚拟细胞双核心为驱动
观壹智能	2026.09	产业进展	总部位于上海，与赛陆医疗达成战略合作
无尽方舟	2026.07	种子+轮	总部位于深圳，完成数千万元种子+轮融资
津渡生科	2026.06	A轮	总部位于深圳，完成近亿元A轮融资
璨辰科技	2026.08	天使系列	总部位于深圳，完成数千万天使系列融资
生寰未来	2026.09	产业进展	总部位于上海，近期正在开展首轮融资
耀速科技	2026.09	产业进展	总部位于深圳，与 M20 Genomics 达成战略合作
淇嘉科技	2026.07	产业进展	总部位于苏州，与赛陆医疗达成战略合作
衍因科技	2026.01	产业进展	总部位于上海，与神拓生物达成战略合作
联川生物	2026.08	产业进展	总部位于杭州，与珞米科技达成战略合作
珞米科技	2026.08	产业进展	总部位于杭州，与联川生物达成战略合作
百奥几何	2026.06	战略融资	总部位于北京，完成数亿元战略融资
数因智科	2026.06	Pre-A 轮	总部位于苏州，完 Pre-A 轮融资
溪砾科技	2026.04	产业进展	总部位于深圳，与 CureCMT 达成战略合作
嘉华药锐	2026.09	产业进展	总部位于北京，一家 AI+ 蛋白质组学技术平台公司
哲源科技	2026.08	A2 轮	总部位于北京，完成近 2 亿元融资
瑞普晨创	2025.11	A+ 轮	总部位于杭州，完成 5 亿元 A+ 轮融资
水木分子	2025.07	产业进展	总部位于北京，与清华大学联合发表 AIVC 相关成果
贝芸科技	2026.09	产业进展	总部位于天津，发布虚拟细胞智能体服务平台
正雅生物	2026.05	产业进展	总部位于常州，专注合成菌群理性设计

* 排序无先后顺序，基于公开信息不完全统计，持续更新中

融资与项目并进，AIVC 进入多路径商业探索新阶段

海外代表性融资事件及进展

企业	时间	事件	关键进展
GenBio AI	2026.08	产业进展	总部位于美国，发布新模型，由胡祖六、邢波、David Baker、宋乐等创立
ChemT Biotechnology	2026.06	种子轮	总部位于新加坡，由孙杰和吴凌联合创立
DeepLife	2026.06	产业进展	总部位于法国，与国内公司分销合作进入中国市场
NeoLife AI	2026.06	产业进展	总部位于美国，与上海交通大学联合发布 VCWorld 模型
Biomy	2026.06	产业进展	总部位于日本，与 10x Genomics 达成数据合作
Recursion	2026.05	产业进展	总部位于美国，与罗氏、基因泰克的合作，加速创新靶点转化
Tolemy	2026.05	种子前轮	总部位于英国，完成 140 万欧元种子前融资
CellType	2026.03	产业进展	总部位于美国，与 Senhwa Biosciences 合作加速临床开发
Turbine	2026.02	B 轮	总部位于匈牙利，完成 2500 万美元 B 轮融资
infinitusbio.AI	2026.01	产业进展	总部位于美国，与 Oncovita 合作开展癌症免疫疗法
Omphalos Lifesciences	2025.12	产业进展	总部位于美国，获得 BARDA 支持，开展抗菌策略模拟
Relation Therapeutics	2025.12	产业进展	总部位于英国，与诺华达成多项目合作，新靶点发现与验证
QurieGen	2025.04	种子前轮	总部位于荷兰，完成 220 万欧元种子前轮融资
Reflector Bio	2025.08	种子前轮	总部位于美国，获得 22 万美元种子前轮融资
Tahoe Therapeutics	2025.08	A 轮	总部位于美国，完成 3000 万美元新一轮融资
Algocell	2025.07	种子前轮	总部位于以色列，获得 280 万美元种子前融资
Cellular Intelligence	2025.05	A 轮	总部位于美国，完成 4700 万美元融资，曾获扎克伯格等投资
Bioptimus	2025.01	A 轮	总部位于法国，累计融资超 7600 万美元
Shift Bioscience	2024.10	种子轮	总部位于英国，累计融资 1600 万美元
CartaBio	2024.04	天使轮	总部位于美国，完成数千万元天使轮融资，创始人Robert Yang
Noetik	2024.08	A 轮	总部位于美国，完成 4000 万美元 A 轮融资
Panacent Bio	2024.07	新一轮	总部位于美国，完成 1300 万美元融资
CytoReason	2024.07	新一轮	总部位于以色列，完成 8000 万美元新一轮融资
Xaira Therapeutics	2024.04	创立融资	总部位于美国，获超 10 亿美元承诺资本
Cytocast	2023.09	种子前轮	总部位于匈牙利，获得 190 万欧元的种子轮前融资
Cellarity	2022.10	C 轮	总部位于美国，获 1.21 亿美元 C 轮融资

* 排序无先后顺序，基于公开信息不完全统计，持续更新中

AIVC 大玩家持续增加，生命科学成为 AI 大厂布局的新焦点

随着模型能力逐步走向产品化，AIVC 正在吸引越来越多产业玩家。Google DeepMind、Microsoft、NVIDIA，以及药企和新锐生物科技公司，正分别从细胞模型、扰动预测、算力平台和联合研发等环节切入，推动 AIVC 由前沿研究走向产业探索。

与此同时，大型 AI 公司的布局正在向更广泛的生命科学领域扩展。OpenAI、Anthropic 目前并未直接进入 AIVC 赛道，而是从科研智能体、生命科学工作台和企业服务等方向切入。这些布局说明生命科学正在成为通用 AI 能力落地的重要产业场景，也将为 AIVC 带来更成熟的模型、工具和研发基础设施。

Google

以 AlphaFold、AlphaGenome 推进蛋白结构与基因组调控预测，并参与 C2S-Scale 等单细胞语言模型研究。

Microsoft

依托 Azure、Project Ex Vivo 与 CausCell，探索单细胞模型评估、因果建模和反事实细胞状态预测。

NVIDIA

以 BioNeMo、NIM 提供生物 AI 训练与推理基础设施，并参与 CZI Virtual Cells Platform 生态。

Meta / CZI

Meta 以 ESM 系列推进蛋白质语言模型；CZI 建设 Virtual Cells Platform、CELLxGENE 等细胞科研基础设施。

腾讯

参与 UniPert-G2CP 统一遗传与化学扰动表征，探索跨细胞背景的表型预测。

阿里巴巴

发布灵枢细胞模型，全转录组尺度的细胞动态建模，预测基因扰动和药物反应。

字节跳动

AML AI4Science 团队持续加码 AI 生物，围绕蛋白质结构预测、蛋白设计、药物研发等方向推进布局。通过 Anew Labs 进一步开展商业探索，并完成 2.9 亿美元首轮融资。

以下为非 AIVC 相关生命科学领域布局

OpenAI

发布 Rosalind Workbench，覆盖蛋白质与小分子设计、结构与序列分析、基因组学等生命科学任务。

Anthropic

以 Claude for Life Sciences、Claude Science 等研究工作台切入，强调文献、实验设计、代码与知识库的协同处理。

* 排序无先后顺序，基于公开信息不完全统计

产业合作推动 AIVC 走向产业研发实践

合作形态已呈现从投资/战略合作、联合研发到模型许可的梯度，AIVC 正由能力建设逐步进入研发流程。大药企掌握管线、样本和验证体系，新锐公司补足细胞状态预测、多模态数据和扰动建模能力。AIVC 领域的新锐公司，正通过与药企、科研机构等开展合作，从技术展示和模型验证，逐步进入产业研发实践。

★ 模型许可 / 权利获取　● 技术研发 / 长期合作　○ 投资 / 战略合作

合作方	自研	Noetik	Relation	Cellular Intelligence	Turbine	CytoReason	Recursion
Novartis 诺华	●		●				
GSK 葛兰素史克		★	●				
AstraZeneca 阿斯利康					●		
Novo 诺和诺德				●○			
MSD / Merck 默沙东					●○		
Pfizer 辉瑞						●○	
Sanofi 赛诺菲						●	
Ono 小野制药					●		
Roche / Genentech 罗氏/基因泰克							●
Agenus 艾吉尼斯		●					
Beiersdorf AG 拜尔斯道夫					○		
NVIDIA 英伟达						○	●○

* 排序无先后顺序，基于公开信息不完全统计

整体来看，全球 AIVC 产业化进程已呈现出较为明显的阶段差异。海外市场起步更早、推进速度更快，已逐步形成由技术公司、制药企业共同参与的协作体系，合作内容也从早期的数据共享、模型研发和技术授权，进一步延伸至药物靶点发现、候选分子筛选及临床前研究等研发环节，开始进入以实际效果检验技术价值的产业验证阶段。相比之下，国内 AIVC 仍处于发展早期，相关探索更多集中在技术研发与能力建设层面，未来需要进一步加强产学研协同，推动模型在更多研发场景中的测试、反馈与迭代，加快完成从技术可行性验证到产业价值验证的跨越。

AIVC 价值正在场景落地中兑现

从模型授权、战略投资到联合研发，AIVC 已开始进入产业。药企通过大额合作为技术和管线提供资源支持，并将其嵌入疾病机制研究、候选筛选与功效评价等环节，推动 AIVC 从概念验证走向持续迭代和规模化应用。

场景一 · 肿瘤与免疫

患者组织多模态数据开始进入药企模型许可与研发合作

Noetik　葛兰素史克

2026 年 GSK 与 Noetik 达成 AI 基础模型授权合作，GSK 将支付约 5000 万美元的前期及近期款项，并另行支付年度许可费用，用于获得 Noetik 肿瘤生物学模型在药物研发中的使用权。

Noetik 构建能够同时描述肿瘤细胞状态、空间微环境和遗传特征的人类多模态数据底座，并在此基础上训练模型。

Relation Therapeutics　葛兰素史克

2024 年 Relation 与 GSK 开展合作，涉及纤维化疾病和骨关节炎，并披露约 4500 万美元首付款；2026 年双方进一步扩展生物数据合作，公开报道的潜在规模最高约为 1.1 亿美元。

Relation 将患者组织、多组学测量、基因和药物扰动实验，以及自动化实验平台结合起来，训练和细胞生物学模型。

场景二 · 活性成分

虚拟实验从靶点发现延伸至消费健康活性成分评价

Turbine　拜尔斯道夫

2026 年 Turbine 获得拜尔斯道夫（Beiersdorf）投资，期望助力研究人员评估活性成分与皮肤生理、皮肤病症及安全性。

此前，Turbine 已与默沙东、阿斯利康和拜耳等公司合作，已在肿瘤和免疫中应用其虚拟检测技术，已超过 30 个新药发现项目。

Turbine 的虚拟检测技术能够模拟数千种生物样本和数百种药物的生物反应，从而生成预测性见解，用于识别和优先筛选靶点、疗法和联合用药方案。

通过与 Crown Bioscience（JSR 生命科学旗下公司）合作，可以使用肿瘤类器官检测对选定的假设进行实验验证，从而简化实验设计、降低成本并缩短开发时间。

场景三 · 细胞管线

细胞状态预测开始连接细胞工程、药物发现与自有管线

Cellular Intelligence　诺和诺德

2026 年 Cellular Intelligence 宣布收购诺和诺德的一款帕金森（STEM-PD）细胞疗法项目。该疗法目前已获 FDA IND 批准，进入二期临床试验准备阶段。此外，还拥有六项临床前资产。

诺和诺德还成为 Cellular Intelligence 的股权投资者，并有资格获得未来的里程碑和版税。

Cellular Intelligence 曾宣布已完成超过 100 万种组合实验条件，并展示了生成多种细胞类型的能力。

Recursion　罗氏基因泰克

2026 年罗氏与旗下基因泰克推进了与 Recursion 合作的首个项目，共计 2.16 亿美元的预付款和里程碑付款。Recursion 用 AI 找到了一个此前无人探索的神经科学新靶点。

模型测评正从学术指标走向多维评价与产业验证

AIVC 的能力边界仍在形成，行业目前把“扰动后的表达谱预测”作为共同评测起点。统一数据、细胞背景和评价指标可以减少选择性报告，但学术 SOTA 并不等于产业有效。模型可能只在熟悉数据集上表现突出，也可能依赖特定预处理和任务设定，难以迁移到新的细胞情境。

当前学术评价	问题	从	未来产业验证	方向
评价标准	在特定数据集上刷 SOTA 易受数据选择影响	学术指标	评价标准	预测是否被湿实验验证
能力展示	挑选优势场景 跨数据集泛化不清	到	能力展示	统一赛道同台竞技
概念边界	Benchmark 碎片化 能力边界模糊	产业指标	概念边界	聚焦扰动预测与泛化 能力边界逐步收敛
成果形态	论文 真实 POC 稀缺		成果形态	产业 POC 、商业合作

虚拟细胞挑战赛（Virtual Cell Challenge，VCC）

2025 年 Arc Institute 发起的 VCC 是目前较具代表性的扰动响应类 Benchmark 之一。由英伟达、10x Genomics 和 Ultima Genomics 赞助，吸引来自一百多个国家的一千多只队伍参赛。

VCC 挑战赛难度和考察方向的演进：

VCC 2025	→	VCC 2026
少量目标细胞数据		6 个匿名细胞系
300 个 CRISPRi 基因扰动		无新增扰动训练集
少样本泛化		零样本预测

图 12 丨两届 VCC 挑战赛的任务与数据（来源：Virtual Cell Challenge 官网）

两届 VCC 的目标变化，体现了评测的重要趋势变迁：2025 年主要考察模型能否利用少量目标细胞数据，完成对新细胞背景的扰动响应预测；2026 年则进一步要求模型在缺少当届扰动训练数据的情况下，将已有规律迁移到多个未见细胞背景。

能力测评走向零样本预测，模型泛化面临更严格检验

Benchmark 测试的指标不能只衡量预测表达谱与真实表达谱的整体距离，还需要同时检验三个层面：模型是否找对了扰动后发生变化的基因，是否能够将正确的响应归因于正确的扰动，以及预测的变化方向和效应强度是否与真实实验一致。只有同时覆盖整体表达准确性、扰动特异性和生物学响应信号，评测结果才更接近模型能否支持候选筛选和后续实验验证的实际价值。

VCC-2025 | 3+4 项指标

PDS | 扰动区分能力
衡量预测的扰动效应是否仍然彼此可区分

DES | 差异表达恢复
评估模型是否正确识别出上调和下调基因集

MAE | 全基因表达误差
量化整个转录组中基因水平的预测准确性

三项指标之外
进一步考察**表达变化一致性、基因变化排序一致性、差异基因识别能力和以及扰动效应大小一致性**判断是否接近真实实验结果

VCC-2026 | 6 项指标

指标	检测内容与含义
PDS	模型能否区分不同扰动的响应
MSE	预测表达谱是否接近真实细胞状态
nMAE	模型对变化幅度的估计是否可靠
FID	模型是否准确预测上调或下调方向
REACH	模型是否覆盖了较完整的细胞响应
JAC	模型是否准确识别受影响的基因

对应到产业研发，这意味着模型评测不再只关注能否服务于一个具体项目，，逐步适应多项目、多场景并行推进的研发节奏。

2025 年 VCC 挑战赛结果排名

排名	通用性排名	主榜排名	验证集排名
1	Cleopatra/Altos Labs	BM_xTVC/Biomap Research*	x.Compass
2	xBio/INSTITUTE OF ZOOLOGY *	XLearning Lab*	Outlier*
3	Mean Predictors	Outlier*	WIND

（来源：Virtual Cell Challenge 官网，*为中国团队或中国参与团队）

价值评估：从 Biotech 与 AI 双重赛道审视企业投资价值

从投资视角看，AIVC 长期价值取决于它被定义为**服务单一场景的研发工具**，还是可复用的**生命科学基础设施**。医药研发是当前最容易形成收入和验证的应用场景，但模型、数据资产与平台复用能力，才决定其长期估值上限。AIVC 应以 AI 科技公司的逻辑评估长期价值，同时以实验与临床验证的逻辑检验近期兑现。

按 Biotech 型公司评价

价值单位	单项目、单管线、单适应症或单一场景
商业模式	研发服务费、里程碑付费、管线授权、联合开发、按部署计费
关键指标	项目交付进展、实验验证结果、场景空间、现金消耗、项目成功概率
价值锚点	风险调整后的项目与管线价值，以及未来销售或许可现金流
主要风险	项目失败、验证周期长、资本消耗大、单项目依赖程度高

VS

按 AI 科技公司评价

价值单位	基础模型、专有数据、生命状态表示、扰动预测与实验反馈系统
商业模式	模型许可、API 调用、研发平台订阅、联合研发、数据与实验基础设施
关键指标	数据壁垒、模型泛化与迭代速度、平台复用率、场景扩展能力、实验验证
估值锚点	平台能力、可规模化收入、跨场景复用、持续学习与生态网络效应
主要风险	技术兑现速度、数据质量、模型外推边界、真实场景验证不足

无论采用哪种评价框架，投资人都应重点考察六个维度：

1 模型能力

是否拥有泛化的细胞表示与扰动预测能力，而非针对单项目的定制分析

2 数据资产

数据是否私有、AI-ready 与具备生产新数据能力

3 实验闭环

模型预测能否转化为实验设计，实验结果能否反哺模型，形成干湿迭代

4 平台复用能力

能力能否跨疾病、跨细胞类型、跨药物与跨研发阶段迁移扩展

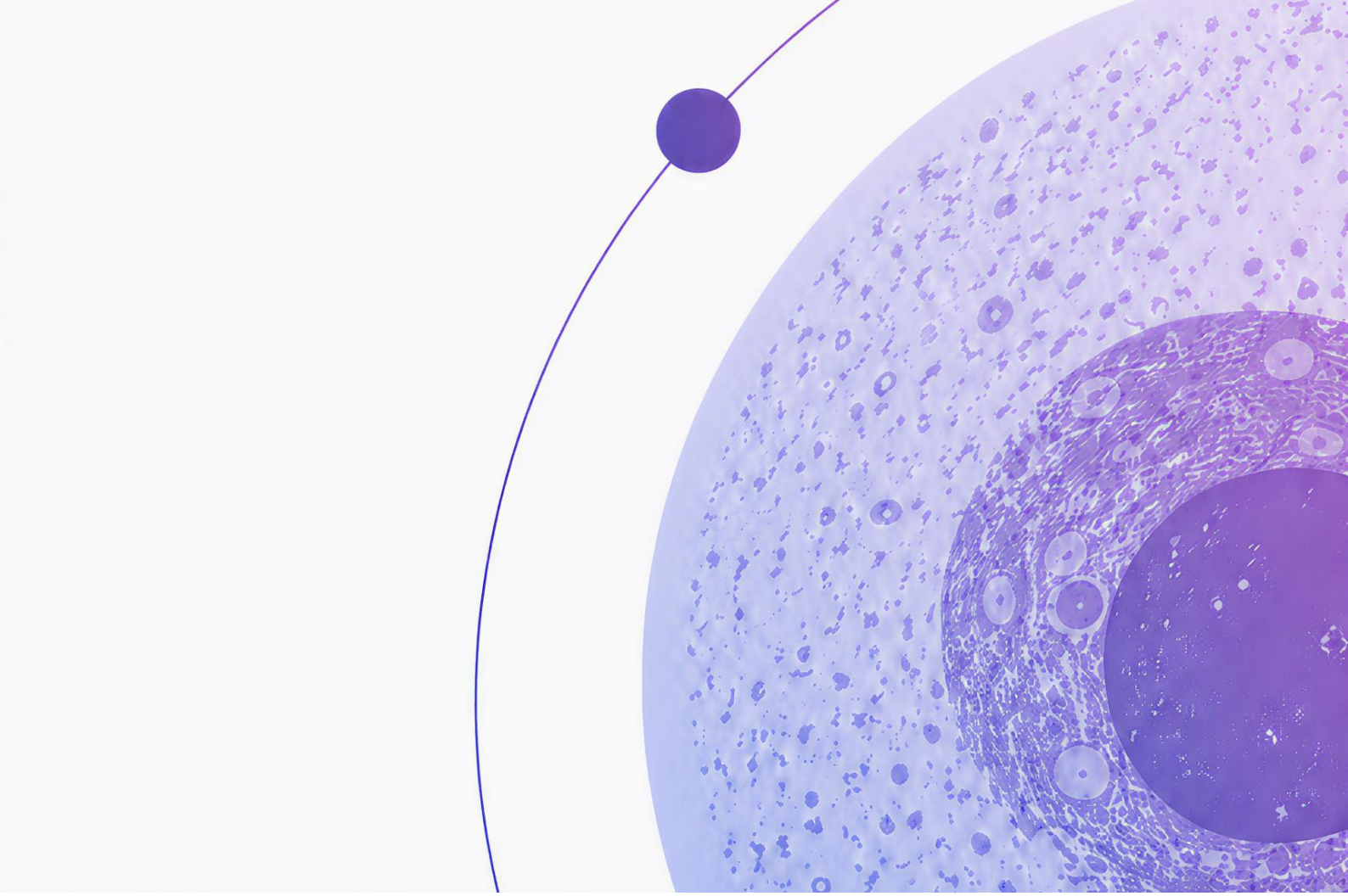
5 商业化质量

收入来自一次性项目，还是可重复销售的模型、平台与数据能力，以及商业化落地时间

6 真实验证结果

能否以实验、临床或客户部署结果，证明模型预测的真实有效性

（来源：《麻省理工科技评论》中国、君联资本周琼访谈）



第 4 章

AIVC 的关键挑战与演进趋势

关键挑战: AIVC 仍需提升模型能力与应用边界

未来趋势: 从模型竞争走向产业生态协同

AIVC 已具备限定场景预测能力，跨情境泛化仍待验证

AIVC 已能识别细胞状态，并在部分已知细胞和扰动条件下预测响应；但跨细胞、跨疾病、跨药物和跨实验条件的稳定泛化，以及预测结果对机制和实验决策的支持，仍需进一步验证。

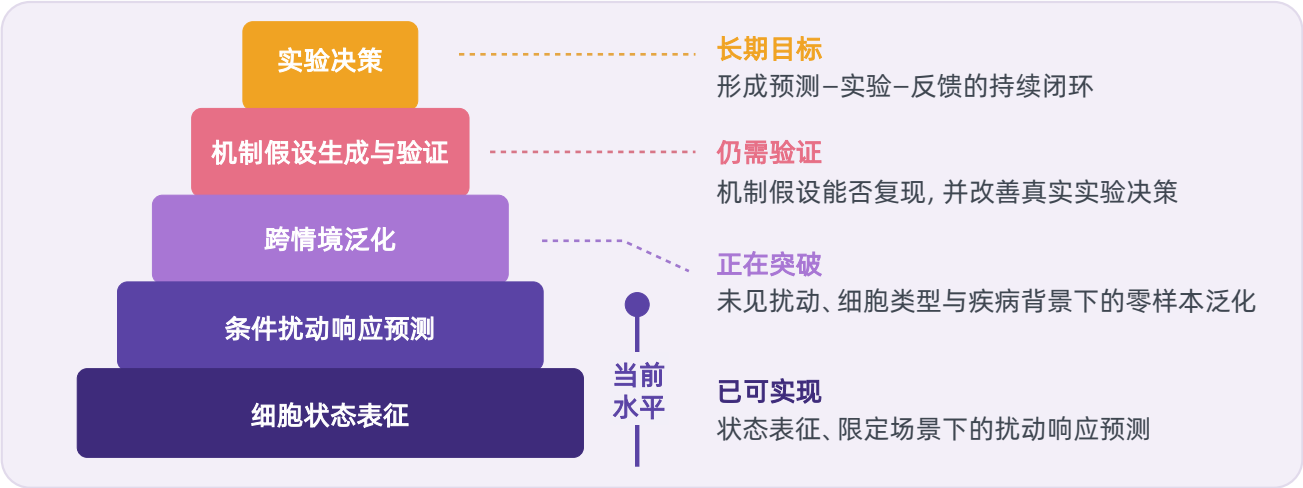


图 13 | AIVC 能力与成熟度阶梯（来源：《麻省理工科技评论》中国、中国科学院动物研究所李鑫访谈、同济大学刘琦访谈）

更高层能力取决于模型能否在未见扰动、细胞类型、疾病背景与实验条件的零样本泛化中保持稳健，并由独立湿实验复现，任何一层能力的薄弱，都会限制模型进入上层能力。

当前阶段，在特定场景 AIVC 模型已可开展表征和预测工作。围绕特定细胞类型、特定基因或药物干预和明确研发终点，AIVC 模型已经具备实践条件。可以预测细胞状态、分子表达和功能表型的变化，支持药物候选筛选、靶点及作用机制判断、细胞工程与分化方案优化，以及药物响应评估和患者分层等任务。其价值主要体现在减少无效搜索、辅助实验设计和提升研发决策效率。

再进一步，泛化能力不断提升，走向多场景。在可预见的未来，随着多模态扰动数据持续积累、模型表征能力提升以及预测–实验–反馈闭环逐步建立，AIVC 将获得更好的跨情境泛化能力。将逐步从已知细胞和已知条件外推至更多细胞类型、疾病背景、药物和实验环境，覆盖多领域、多场景的研发任务。届时，AIVC 将进一步发展为支撑生命科学研究与药物研发的通用基础能力。

将来，AIVC 将走向可计算、可设计的动态生命系统建模。AIVC 最终希望建立一个具有生命特征的动态系统模型，在时间和空间尺度上具有追踪细胞连续演化的能力。随着多模态、跨尺度数据不断积累，模型有望逐步连接基因信息、调控网络、细胞表型以及组织和个体效应，形成对生命过程更完整的计算表达。

从概念验证走向产业应用，AIVC 仍需跨过关键挑战

当前 AIVC 面临的核心问题，是从模型研发到产业验证的能力尚未形成。行业尚未建立起将数据生产、模型训练、实验反馈、应用决策与商业回报有效贯通的完整机制。由此，AIVC 从概念探索走向规模化应用，仍需补足研发体系、验证体系与产业化机制之间的连接。

数据基础仍不成熟，多模态新生产数据仍然稀缺

生命科学数据存在质量不一、标准不统一、实验条件差异大、不同模态缺乏对应关系等问题；数据规模的增长并不直接转化为高价值训练语料，失败实验、负样本和重复实验数据尚未得到充分利用。AIVC 需从单纯扩大数据规模，转向围绕模型训练目标持续获得“新生产数据”，但其生产成本高、获取周期长，目前仍然十分宝贵。

动态模型能力尚需提升

现有组学数据难以完整刻画细胞分化、药物作用和疾病进展等动态过程；扰动数据也仍为扰动前后两个静态切片的变化关系。AIVC 需要进一步引入多时间点采样、谱系追踪、活细胞成像、空间组学和细胞互作数据，才能建立更加完整的动态状态表示。

干湿闭环仍缺乏有效的产业场景

干湿闭环是 AIVC 持续成长的基础，目前许多 AIVC 项目仍停留在一次性验证、单个任务或局部合作阶段，尚未形成能够长期运行的闭环场景。有效的落地场景，需要同时满足：① 具有持续产生新数据的实验过程；② 模型能够对实验设计或研发决策产生实际影响；③ 客户愿意为模型或研发结果付费。缺少其中任何一项，都无法转化为持续的数据资产和模型能力。

AIVC 的价值仍受成果与管线叙事主导，服务价值尚未充分显现

当前市场对 AIVC 的价值判断仍较多集中于候选药物和研发管线，对其服务价值关注不足。相比之下，海外企业已更多探索模型服务、虚拟实验、实验设计和临床转化等服务形态。AIVC 不仅可以形成具体研发成果，也能够作为生命科学研究的基础能力，帮助不同机构降低成本与风险。当这种服务能够被量化和验证，服务价值就能形成更加稳定的商业模式。

从模型规模竞争走向数据–模型共同进化

AIVC 的产业化单纯构建技术模型尚无法满足产业需求，需要数据、模型、场景协同发展的系统性能力建设,并逐步形成核心壁垒与更长期的商业价值。

01 趋势一: 模型规模跃迁，但参数量不再是唯一胜负标准

单细胞基础模型的参数规模已经从早期的百万级、千万级，逐步进入数亿级，部分路线开始进入十亿级。参数扩展带来扰动预测性能持续改善，表明该领域开始出现类似大模型的规模效应。

但与自然语言模型不同，单纯扩大数据规模或参数规模并不必然带来更好的零样本泛化能力。模型效果还取决于训练目标、扰动覆盖、数据多样性、细胞上下文、任务类型。

02 趋势二: AI-ready 且私有的新生产数据将成为关键竞争要素

数据库不等于 AI-ready 数据，用于模型训练的数据是需要经过专门加工、能够被模型直接学习、能够复现实验条件并支持外部验证的数据。“新生产数据”是围绕模型训练目标专门设计实验、控制变量而获得的数据。未来，企业能否持续生产、验证并沉淀这类数据，将成为模型能力和产业竞争力的重要分水岭。

维度	数据内容	对产业的意义
数据质量	噪声、重复、批次效应和 QC 是否可控	能否稳定训练和复现实验
数据关系	是否记录初始状态、干预条件、剂量、时间、对照和真实响应	能否支持扰动预测，而不只是描述细胞
数据对齐	不同组学、影像、空间和时间信息能否对应到同一细胞或同一实验	能否训练多模态和动态模型
数据资产化	场景覆盖、正负样本、失败实验、数据版本和合规管理是否完整	能否形成长期壁垒并持续回流

公开数据可以被所有团队获取，单纯依赖公共数据库容易陷入同质化竞争，且公开数据之间存在显著的批次差异。更有价值的是企业自己生产、统一标准和可持续更新的数据。

因此，未来的护城河可能按以下顺序形成：① 公开数据基础；② 自有实验平台生产的标准化数据；③ 数据形成专有资产；④ 通过干湿闭环持续更新模型；⑤ 通过实际研发结果证明模型的决策价值。

AIVC 产业化需要模型与场景两端并进

03 趋势三：干湿闭环将从“加分项”变成产业化门槛

干湿闭环不只是自动化实验室能力，而是能否建设并持续运转一套“资金投入–体系建设–服务交付–模型迭代”的干湿闭环。这是一项重资产、长周期能力，主要包括四个方面：

资金投入
持续投入实验设备、算力、样本、人员和研发周期

体系能力
建立从问题定义、模型开发、实验验证到结果回流的标准流程

服务能力
把模型和实验能力转化为可交付服务

资源配置
明确哪些能力自建、哪些环节合作，以及平台、服务、联合研发和管线之间的关系

这是湿实验平台的高效运转成为连接模型研发与产业客户的中心关键环节。未来真正有竞争力的企业，除拥有模型、实验室或客户资源外，还应能够将三者组织成一个持续运转的系统：模型为客户提供预测和决策服务，客户需求推动实验验证，实验结果反过来提升模型能力，模型能力提升又带来更多客户和更深的合作。

04 趋势四：模型能力与垂直落地，快慢有别但缺一不可

AIVC 企业未来都必须同时发展两种能力：一是构建具有跨场景预测能力的模型；二是将模型垂直落地到具体疾病、药物、管线或各类相关产业领域中。

不同企业的发展节奏可能不同。有的企业模型能力提升更快，先通过通用模型积累数据和客户；有的企业管线落地更快，先在特定疾病或细胞场景中取得验证，再反向推动模型升级。两条路径没有统一的先后顺序，但长期来看，两者缺一不可。

- **高泛化性的模型：能够跨细胞类型、疾病和干预场景复用，减少每个客户重新采集数据、重新训练模型的成本。**
- **管线的垂直落地：进入具体应用场景，通过具体研发问题、实验验证和客户付费证明模型价值。**

以数据、模型资产为壁垒　　以场景验证积累信任　　以管线落地驱动迭代

产业协同与价值评估共同重塑 AIVC 价值

05 趋势五：产业协同与政策推动，或将成为 AIVC 的重要锚点与放大器

随着 AIVC 从技术探索进入研发场景落地，产业合作的形成将越来越依赖多种能力的组合。

大型产业方：将从单纯的客户或投资方，逐步转向产业生态的共同建设者。可能影响 AIVC 的技术路线和评价标准。谁来定义有价值的研发问题，什么样的预测可以进入决策流程，哪些数据和验证结果能够被行业认可，都可能在长期合作中逐步形成共识。合作方式也将从项目合作扩展到联合实验室、战略投资、平台授权、数据合作、联合管线等。

政府侧：主管部门可通过顶层规划、制定行业政策等方式助力产业发展。通过建设大科学装置、相关国家级平台，建设共享数据、算力和实验验证设施，提升 AIVC 在数据生产、资源共享与实验验证等方面的整体支撑能力。

06 趋势六：AIVC 的投资逻辑将转向“平台价值与研发兑现”双重定价

AIVC 产业既不完全等同于传统 Biotech，也不能仅按 AI 大模型公司进行估值。其长期价值来自 AIVC 模型所构建的预测能力；其短期价值则需要通过具体研发场景、客户收入和产业验证逐步兑现。

因此，可能形成两套相互关联的评价逻辑：一方面，考察模型能否跨场景复用、数据是否形成壁垒、平台能否持续迭代，以及客户和研发流程能否不断复用；另一方面，考察模型是否真正改善候选筛选、响应预测、患者分层等实际结果。

这意味着，模型平台决定企业价值的上限，研发结果决定价值兑现的程度。

（来源：《麻省理工科技评论》中国，峰瑞资本谢达访谈）

> AIVC正在构建一套全新生命科学计算基础设施，在其之上，未来可能形成从数据资源、基础模型到专业模型、实验验证和应用服务的完整产业生态。
>
> 唯有既看到方向的广阔，也尊重生命系统的复杂；既积极创新，也坚持科学验证，才能让人工智能真正成为认识生命、服务健康的有力工具，在不断逼近生命规律的过程中，拓展人类理解自身的边界。

陈润生

中国科学院院士
中国科学院生物物理研究所研究员

数据和语料是两个概念，不能简单把数据库等同于模型训练的语料；只有经过降噪、去重、标准化和跨模态对齐，才能真正成为人工智能可用的语料。

李鑫

中国科学院动物研究所研究员

虚拟细胞统一模型与垂类模型并行，统一模型追求多模态融合与强泛化；垂类模型在特定场景下更容易形成独特能力并率先落地。

刘琦

同济大学
长聘特聘教授

"Scaling Law" 在生命科学领域是否适配仍存在争议。虚拟细胞研究需要找到其合适的生物“尺度法则”，发展生命组学适配的AI技术路径。同时明确应用场景的边界条件，并不断拓展模型在有限训练样本下的泛化能力。

周璟

君联资本
董事总经理

做新药研发就像攀登珠峰。AlphaFold 可以更快、更低成本地从走到珠峰大本营，但继续向登顶迈进，还需要理解药物进入动物乃至人体后与复杂生命系统之间的相互作用。

Insilico Medicine
HKEX: 3696

当前的 AI 药物研发已取得显著突破，临床转化仍面临巨大挑战。为了攻克生物学的动态复杂性难题，只有将时间维度、多尺度推理与个体异质性引入生物学模型，才能真正提升候选药物在真实患者群体中的临床成功率。

vitaura
百 曜 科 技

推动虚拟细胞走向产业化，首先要理解模型的能力边界，找到与之匹配的应用场景，再以数据和模型能力共同支撑；一旦能力与场景错配，要么失败率极高，要么投入成本极高。

结语

AIVC 的意义，是尝试让机器学习从海量生命科学数据中学习细胞运行的规律。当模型开始具备对细胞状态、动态变化和动态建模能力时，AIVC 模型就开始成为理解和探索生命的新型基础设施。

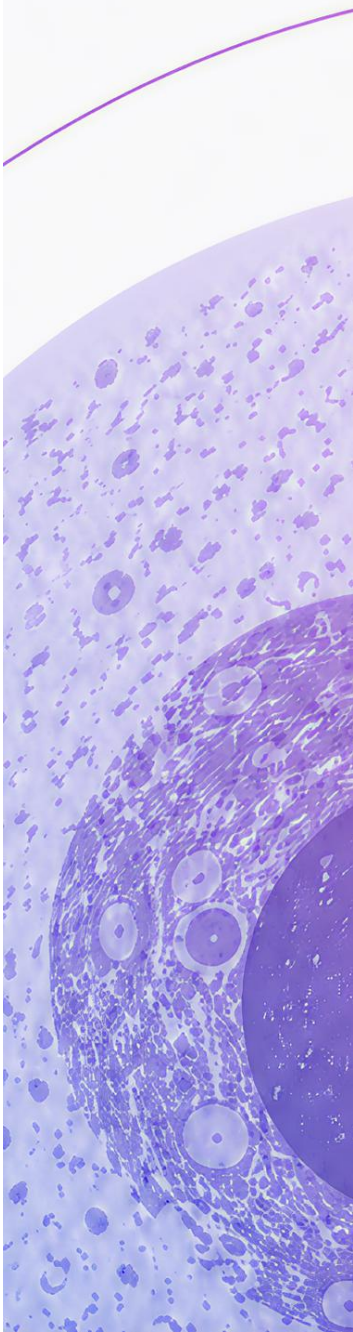
从技术发展阶段来看，当前 AIVC 仍处于由静态表征向动态建模、由单一任务模型向通用生命基础模型演进的早期阶段。模型规模、数据质量、多模态融合、跨尺度建模等方面仍存在诸多技术挑战。

因此，AIVC 未来的发展更依赖于数据、模型、干湿闭环和应用场景之间形成持续迭代的飞轮。随着自主智能、自动化实验平台等技术的发展，模型预测、实验验证与数据反馈之间的连接也有望进一步加强。

与此同时，AIVC 的产业价值也将随着模型能力和应用成熟度的提升逐步释放。其更深层的产业意义，在于推动生命科学研发从以实验为中心的线性流程，逐步向以模型为支撑、以数据为驱动、以实验验证为闭环的研发范式转变。推动生命科学产业的知识生产、研发决策和资源配置方式发生变化。

从更长周期来看，AIVC 所构建的是一套新的生命科学计算基础设施。围绕这一基础设施，未来可能形成从数据资源、基础模型到专业模型、实验验证和应用服务的完整产业生态。

相信随着人工智能基础模型能力持续提升，生命科学数据体系不断完善，以及计算模型与实验系统进一步融合，AIVC 将逐步从早期的技术探索走向更加成熟的应用阶段，并在生命科学研究与产业研发中发挥越来越重要的作用。



致谢

感谢指导单位**「中关村科学城管理委员会」**对报告研究工作的关心与支持，为报告的顺利开展提供了重要指导与帮助。

本报告的完成，受到了来自科研、产业、投资等领域专家学者及相关人士的支持与帮助。在研究过程中，相关专家就 AIVC 的技术发展、研究进展、产业应用及未来趋势等内容提供了宝贵的意见与建议，为本报告的研究与撰写提供了重要参考。在此谨致以诚挚感谢。

以下为致谢名单（排名不分先后）：

陈润生	中国科学院院士、中国科学院生物物理研究所研究员
李鑫	中国科学院动物研究所研究员
刘琦	同济大学长聘特聘教授
裴唯珂	西湖大学特聘研究员、助理教授
张迪	耶鲁大学副研究科学家
周琅	君联资本董事总经理
谢达	峰瑞资本副总裁
英矽智能	Insilico Medicine
百曜科技	Vitaura

vitaura
百 曜 科 技

本研究报告由**「百曜科技」**发起并提供项目支持，《麻省理工科技评论》中国团队负责策划、调研、撰写与制作。感谢百曜科技对本项目的研究推进、资料整理与成果呈现提供支持。

版权声明

本研究报告由百曜科技发起并提供项目支持，《麻省理工科技评论》中国发布。如需转载，请完整注明出处；如需引用，请完整注明来源，不得歪曲、篡改报告原意。未经书面许可，不得将本报告用于商业用途。任何未经授权的商业使用行为，将依据中华人民共和国相关法律、法规追究其法律责任。

免责声明

本报告所载数据和观点仅反映《麻省理工科技评论》中国截至本报告发布之日的判断。《麻省理工科技评论》中国对报告所载信息的准确性、完整性或可靠性做尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或表述不构成任何投资建议。《麻省理工科技评论》中国对该报告的数据和观点不承担法律责任。不同时期，《麻省理工科技评论》中国可能会发布其他与本报告所载资料、结论不一致的报告。同时《麻省理工科技评论》中国对本报告所载信息，可在不另行通知的情况下进行修改，读者应自行关注。

本报告最近更新于 2026 年 9 月 28 日。

关于《麻省理工科技评论》中国

《麻省理工科技评论》*（MIT Technology Review）*于 1899 年在美国麻省理工学院创刊，是全球历史悠久、具有广泛影响力的技术商业类杂志。内容覆盖广泛，涉及生命科学、人工智能、能源、新材料等领域。

2016 年，《麻省理工科技评论》与 DeepTech 合作落地中国，并以《麻省理工科技评论》中国为主体发掘改变世界的新兴科技，以全球新兴科技创新服务平台，影响那些关注科技及其影响的人。

联系我们

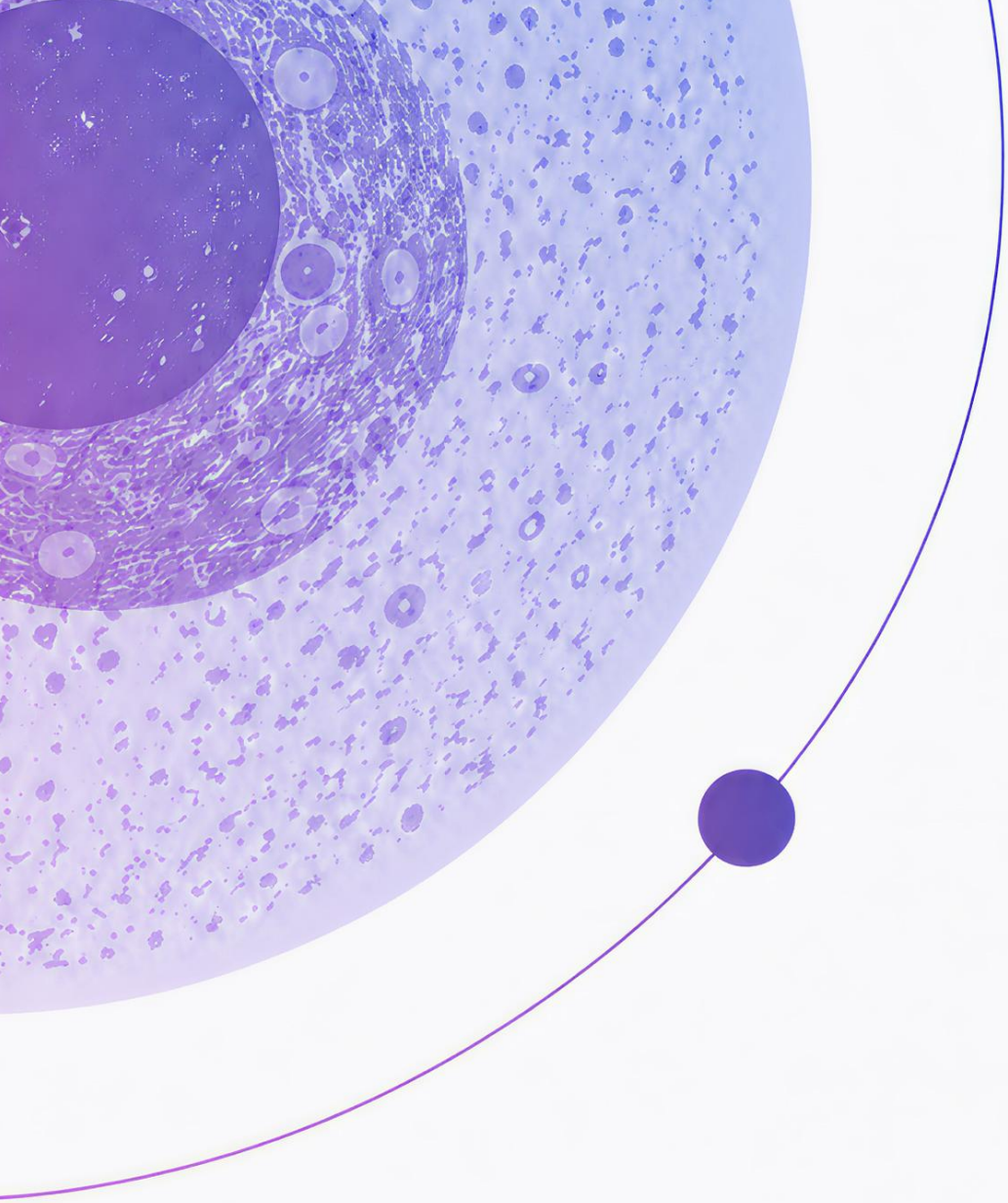
songbingran@deeptechchina.com M: +86 18811152209

jing.zhang@vitaura.ai M: +86 15083545825

了解更多

https://www.mittrchina.com

https://www.vitaura.cn



百曜科技

DeepTech