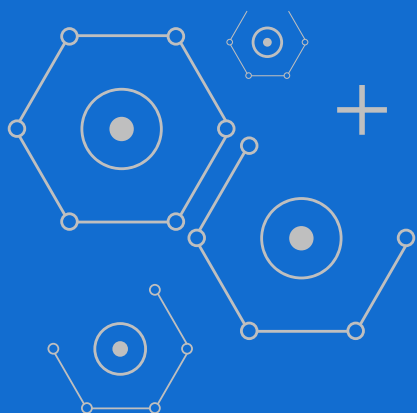
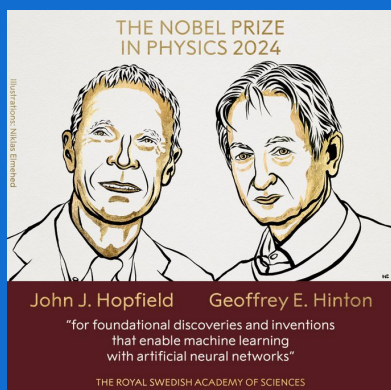


AI × SCIENCE

十大前沿观察



从未来回望，2025年将被定义为“人工智能与科学发现深度融合的元年”，这源于2024年两项诺贝尔奖均授予了AI深度参与其中的研究。物理学奖授予了科学家John J. Hopfield和Geoffrey E. Hinton，以表彰他们在使用人工神经网络进行机器学习方面的基础性发现和开创性贡献。诺贝尔化学奖则颁发给了科学家David Baker、Demis Hassabis和John M. Jumper，以表彰他们在计算机蛋白设计及蛋白质结构预测领域的杰出成就。



出品： 漆 远 吴力波 张 江

出品机构：



运营： 孟晋宇 王 婷

撰稿团队：

张 江 杨燕青 王 婷 王朝会 十三维 周 莉 梁 金 袁 冰
江千月 刘志毅

鸣谢（按姓氏拼音顺序，排名不分先后）：

曹风雷 陈小杨 程 远 杜沅岂 段 郁 方榭楷 付彦伟 高 悦
黄柯鑫 李 昊 刘圣超 谭伟敏 吴泰霖 吴艳玲 向红军 张 骥
张 艳 朱思语

目 录

Part1 序章

1.1 AI for Science	6
1.1.1 AlphaFold	6
1.1.2 气象预测大模型	6
1.1.3 AI可控核聚变	7
1.2 Science for AI	7
1.2.1 Hopfield网络与受限Boltzmann机	8
1.2.2 视觉架构启发的CNN网络	8
1.2.3 等变神经网络	9

Part2 十大前沿观察

AI for Science	2.1 基于LLM的科学研究	12
	2.1.1 背景介绍	12
	2.1.2 研究进展	13
	2.1.2.1 跨学科知识的整合与创新	13
	2.1.2.2 基于LLM的科研流程重构	13
	2.1.2.3 革新人机协同科研模式	14
	2.1.3 挑战与展望	15
	2.1.4 参考文献	16
	2.2 垂直领域科学大模型	17
	2.2.1 背景介绍	17
	2.2.2 研究进展	18
	2.2.2.1 生命科学基础模型	18
	2.2.2.2 医学基础模型	19
	2.2.2.3 气象预测基础模型	21
	2.2.2.4 物质科学基础模型	22
	2.2.2.4.1 分子动力学模拟	22
	2.2.2.4.2 物理模拟	23
	2.2.3 挑战与展望	24
	2.2.4 参考文献	24

2.3 融入先验知识的AI模型	27
2.3.1 背景介绍	27
2.3.2 研究进展	28
2.3.2.1 神经符号人工智能	28
2.3.2.2 几何图神经网络	29
2.3.2.3 物理启发的图神经网络	31
2.3.3 挑战与展望	31
2.3.4 参考文献	32
2.4 AI科学家	36
2.4.1 背景介绍	36
2.4.2 研究进展	36
2.4.2.1 AI科学家与自主科研智能体	36
2.4.2.2 科学家助手	40
2.4.3 挑战与展望	42
2.4.4 参考文献	43
2.5 复杂世界的多智能体建模	45
2.5.1 背景介绍	45
2.5.2 研究进展	46
2.5.2.1 基于LLM的多智能体系统研究宏观经济学	46
2.5.2.2 基于LLM的多智能体系统研究社会群体行为	47
2.5.2.3 基于多智能体系统的未来城市治理	49
2.5.3 挑战与展望	50
2.5.4 参考文献	50
2.6 AI 仿真器与系统模拟	53
2.6.1 背景介绍	53
2.6.2 研究进展	53
2.6.2.1 AI for PDE(人工智能求解偏微分方程)	53
2.6.2.2 因果数字孪生	54
2.6.2.3 AI驱动的多尺度数字有机体建模	55
2.6.2.4 生物-环境闭环智能仿真系统	57
2.6.2.5 具身生成式物理模拟引擎	58
2.6.3 挑战与展望	58
2.6.4 参考文献	59
2.7 物理世界的第一性原理	62
2.7.1 背景介绍	62
2.7.2 研究进展	63
2.7.2.1 自由能原理: 智能的第一性原理	63
2.7.2.2 泊松流模型: 物理启发的深度生成模型	64
2.7.2.3 范畴论: 刻画大模型的能力边界	65
2.7.3 挑战与展望	65
2.7.4 参考文献	66

AIS 基础设施

2.8 科学启发的AI新架构	68
2.8.1 背景介绍	68
2.8.2 研究进展	69
2.8.2.1 Komogorov Arnold Networks (KAN)	69
2.8.2.2 脉冲神经网络	70
2.8.2.3 储备池计算	70
2.8.2.4 机械神经网络	72
2.8.2.5 傅里叶神经算子	73
2.8.3 挑战与展望	74
2.8.4 参考文献	75
2.9 合成数据和数据基础设施	77
2.9.1 背景介绍	77
2.9.2 研究进展	78
2.9.2.1 合成数据生成: 从序列数据到复杂数据	78
2.9.2.2 科学数据表示: 从图神经网络到多模态数据表征	79
2.9.2.3 模型自我改进: 通过合成数据提升AI系统性能	80
2.9.2.4 数据基础设施: 人体蛋白质组计划	81
2.9.3 挑战与展望	82
2.9.4 参考文献	83
2.10 新型智能计算	86
2.10.1 背景介绍	86
2.10.2 研究进展	88
2.10.2.1 凡人计算	88
2.10.2.2 光学计算	89
2.10.2.3 量子计算	90
2.10.3 挑战与展望	93
2.10.4 参考文献	94

Part3 展望

3.1 特殊挑战	98
3.1.1 Science for AI: AI如何与人类共享知识	98
3.1.2 AI for Science: AI如何具备实验思维	98
3.2 科学世界模型框架	99
3.2.1 数据驱动模型	99
3.2.2 深思考	100
3.2.3 AI与现实的交互	100
3.2.4 人类专家与AI系统的对齐	101

PART 1

序章

AI × Science

序章

AI × Science，即人工智能与科学研究的深度融合，具体包括两方面：将AI技术应用于领域学科的新兴研究方法（AI for Science）；将领域学科知识用于AI算法和架构的理解和改进工作（Science for AI）。

2024年诺贝尔物理学奖授予的是将物理学用于AI系统设计的研究，即Science for AI，而诺贝尔化学奖授予的是人工智能算法用于解决化学领域重要问题的研究，即AI for Science。

1.1 AI for Science

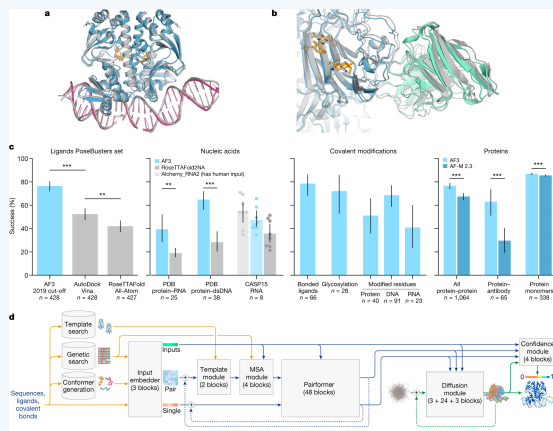
随着数据的快速积累和文献的爆炸式增长，人类科学家自身的信息处理能力已经成为了制约某些学科快速发展的瓶颈。而越来越多的科学研究领域在面对复杂问题的挑战时难以利用传统的数学和物理方法，例如蛋白质结构预测、新材料的设计、复杂流体系统的模拟等等。

另一方面，随着大数据时代的降临和算力的提升，人工智能技术正飞速发展。以ChatGPT为代表的大语言模型（Large Language Model，简称LLM）是AI发展史上的重大突破，而近期推理模型异军突起，例如GPT-o1/o3和Deepseek R1已经具备了强大的学习能力和推理能力，甚至可以成为科研人员重要的助手，辅助解决一定级别的科研问题。因此，如何将AI先进技术用于具体的科学领域，加速科学研究的速度成为了AI for Science研究领域的重要挑战。

已有的AI for Science案例包括：

1.1.1 AlphaFold

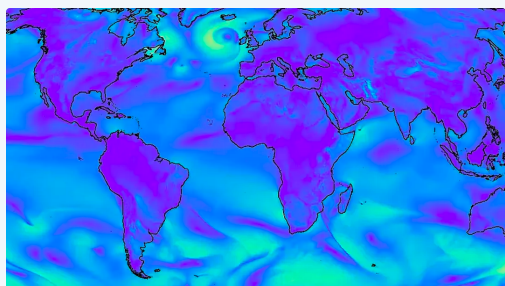
斩获2024年诺贝尔化学奖的Baker、Hassabis和Jumper的研究工作就是将深度学习技术应用于蛋白质结构预测与设计这一经典领域。蛋白质是生命的“基础建筑块”，它不仅可以催化生化反应、调节生理机制，还可以调控基因表达，运输生命必需物质以及用于抵抗病毒、细菌入侵等重要功能。Hassabis与Jumper开发的AlphaFold系列工作，可以精准预测蛋白质结构，准确度已经接近专家水平，速度则远超人类。最新版本的AlphaFold则可以预测几乎所有分子类型的蛋白质复合物结构。



用AlphaFold3预测的精准蛋白质结构 | 来源：Abramson, J., Adler, J., Dunger, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500 (2024).

1.1.2 气象预测大模型

随着全球变暖，极端天气会越来越的影响人类生活生存环境。因此，如何更加精准地对未来天气发



GraphCast: 用于更快速、更准确的全球天气预报的人工智能模型 | 来源: Remi Lam et al., Learning skillful medium-range global weather forecasting. Science 382, 1416-1421 (2023)

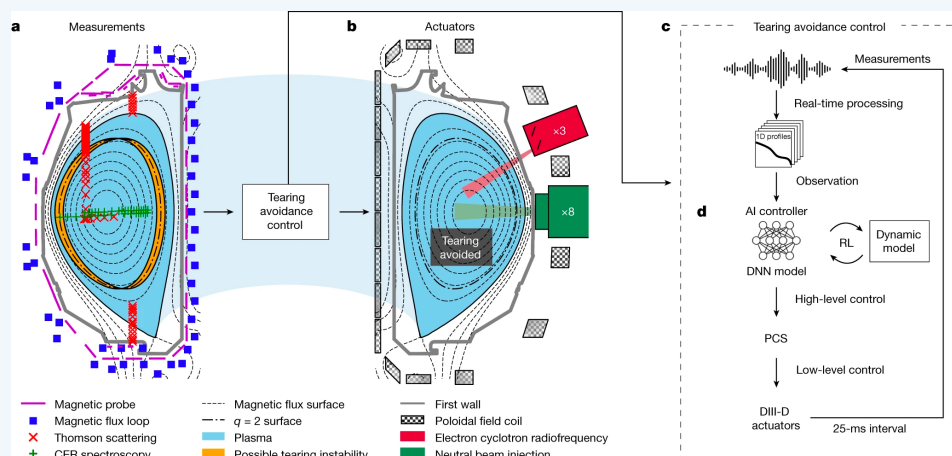
展进行准确预测，特别是防患极端天气的到来越来越成为意义重大的科学问题。然而，传统的气象学研究受限于数据的缺乏、人类知识和经验的局限以及气象系统内在的不确定性因素等影响，无法给出较长期而精准的预测。因此，如何将大模型等最新的AI技术引入气象领域，已经成为了近年来 AI for Science 研究的焦点之一。近年来，若干气象预测大模型被陆续开发出来，包括Google的GraphCast、华为的盘古大模型、英伟达的FourCastNet、复旦大学-上海科学智能研究院的伏羲气象大模型等都取得了显著进展。

1.1.3 AI可控核聚变

核聚变是一种清洁、安全、便利的新型能源。与化石燃料相比，聚变能的使用可以显著减少对环境的影响，有助于应对气候变化。核聚变反应需要在极高温度和压力下进行，因此等离子体的稳定性是实现可控核聚变的关键。AI技术可以用于实时监测和控制等离子体状态，预测不稳定性并及时调整实验参数，以保持等离子体的稳定。2024年初，普林斯顿大学和普林斯顿等离子体物理实验室的研究人员在可控核聚变领域取得突破，他们发现了一种方法，可以利用AI预测这种潜在的不稳定性，并且实时防止实验中断。相关研究成果发表在了2024年2月份的Nature刊物上。

1.2 Science for AI

即使已获重大突破，但AI技术仍然面临数据稀缺、算力稀缺、耗能过大、解释性较差等重大挑战。



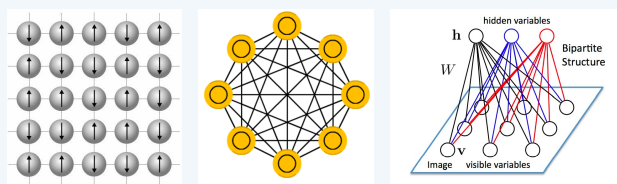
通过深度强化学习控制聚变等离子体以避免撕裂不稳定性的模型框架 | 来源: Seo, J., Kim, S., Jalalvand, A. et al. Avoiding fusion plasma tearing instability with deep reinforcement learning. Nature 626, 746-751 (2024)

另一方面，人类科学家已经积累了各个学科领域的海量知识。如何将科学家的经验和知识，甚至一些不成型的启发式想法，转化为AI系统的能力，让AI系统与人类专家能够更好地互动、协作，构成了Science for AI研究领域的重大挑战。

目前，已有的Science for AI案例包括：

1.2.1 Hopfield网络与受限Boltzmann机

本次诺贝尔物理学奖授予的两名学者，Hopfield和Hinton正是将统计物理中的经典模型的变种应用于神经网络架构设计，从而大大提升了机器学习问题的求解效率，为人类进入深度学习时代开辟了先河。Hopfield所提出的Hopfield网络模型是在统计物理的ISING模型的基础上改进而得到的，它可以通过训练学习记忆一些固定的模式，并在应用阶段激活、回忆起这些模式。在Hopfield网络模型的基础上，Hinton所提出的受限Boltzmann机模型则可以通过大量数据的训练，学习数据的压缩表征，并可以通过拼接多层的方式，获得数据在多个尺度上的深层次表征。这些表征可以进一步用来分类或预测。

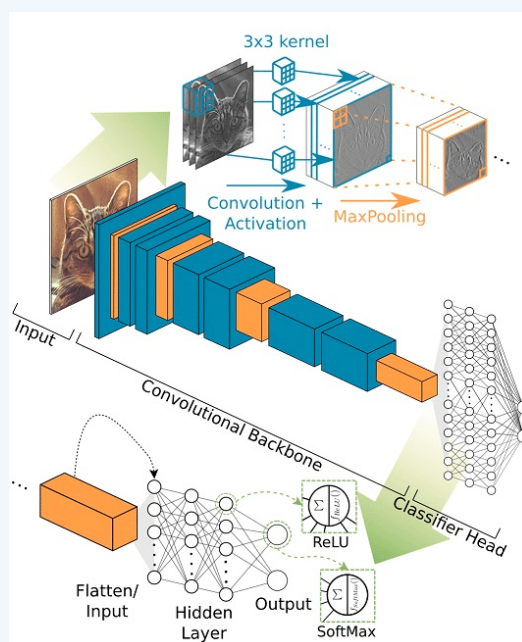


ISING模型、Hopfield网络以及受限Boltzmann机 | 来源：
<https://www.nobelprize.org/prizes/physics/2024/press-release/>

1.2.2 视觉架构启发的CNN网络

对真实生物神经网络结构的研究有助于人们提出更先进的人工智能系统架构。例如，卷积神经网络

(CNN)的提出就受到了生物神经网络研究的启发，尤其是对视觉系统的理解。早在20世纪60年代，神经科学家David Hubel和Torsten Wiesel通过对猫的视觉皮层进行实验，发现了视觉神经元的不同反应特性。他们观察到，某些神经元对特定方向的边缘或图案表现出强烈的反应，称为简单细胞；而另一些神经元则对更复杂的视觉刺激（如形状和运动）有反应，称为复杂细胞。这种对视觉神经元功能的理解为CNN的设计提供了重要的理论基础。CNN模拟了生物视觉系统中的层次结构，通过多层网络逐步提取特征。最初的卷积层相当于简单细胞，负责检测图像中的基本特征，如边缘和角落；而后续的层则类似于复杂细胞，能够识别更高层次的特征，如形状和物体。此外，CNN中的卷积操作和池化操作也受到生物神经网络的启发。卷积操作通过局部感受野的方式，模拟了生物神经元对局部区域的敏感性，而池化

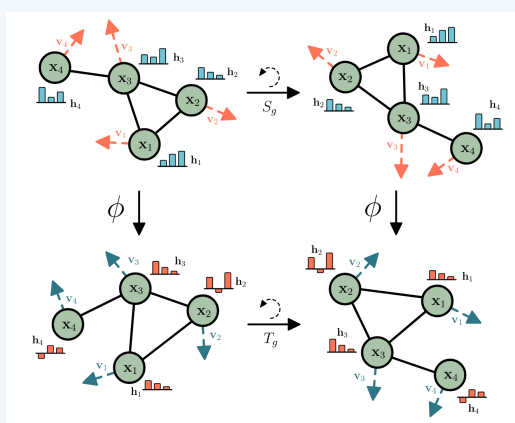


用于图像识别的卷积神经网络架构 | 来源：Hoeser, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. Remote Sens. 2020, 12, 1667.

操作则通过下采样减少特征图的维度，增强了模型对位置变化的鲁棒性。这种设计不仅提高了计算效率，还使得网络能够更好地处理图像中的变换和扭曲。

1.2.3 等变神经网络

在分子建模、蛋白质预测、计算机视觉和机器人等领域中，很多数据都具备几何空间上的对称性，如一组点云在旋转、平移或缩放变换下会保持一定的不变性。如果用普通的神经网络技术处理数据就很难捕获到这种几何对称性。于是，人们提出了等变神经网络（Equivariant Neural Networks, EGNN）技术，可以将几何上的对称性作为一个先验偏置编码进神经网络的架构中，使得网络在输入经过某种变换时，输出能够以相应的方式变化，从而保持对称性和不变性。等变神经网络在建模分子结构、模拟多体物理系统、计算机视觉、机器人等领域具有广泛的应用场景，是将几何学、群论和刚体运动等数学、物理学知识应用于神经网络设计的典型案例。



等变神经网络工作原理示意图 | 来源：Victor Garcia Satorras, Emiel Hoogeboom, Max Welling
Proceedings of the 38th International Conference on Machine Learning, PMLR 139:9323-9332, 2021.

尽管刚刚起步，人工智能和科学研究的深度融合已呈现出“井喷式”增长：不同领域的科学家纷纷将AI融入自己的研究，同时科学领域的研究者也在关注人工智能发展，力求理解其工作原理，以期实现更好的改进。为了梳理这一快速发展的全新领域，上海科学智能研究院、集智科学研究中心和阿里云共同发起了“AI×Science十大前沿观察”项目，希望能够把脉和提炼这一新兴领域的最新发展动态。

在与上海科学智能研究院、复旦大学和大量集智俱乐部社区中从事前沿研究的学者深度访谈基础上，我们做了大量的文献梳理，并最终凝练了“AI×Science十大前沿观察”的三个维度和十个具体方向。三个维度包括AI for Science、Science for AI和基础设施，其中，AI for Science的前沿方向包括基于LLM的科学研究、垂直领域科学大模型、融入先验知识的AI模型、AI科学家、复杂世界的多智能体建模，以及AI仿真与系统模拟；Science for AI的前沿方向则覆盖了物理世界的第一性原理和科学启发的AI新架构；基础设施前沿方向包括合成数据和数据基础设施、新型智能计算。在十个具体方向上，我们梳理出共计35个研究前沿。需要指出的是：这十个方向的具体研究案例未必都是共识的权威成果，但在科学问题设定、方法论建构、精度和效率的提升等方面，这些研究不仅效果显著，且令人耳目一新。展望未来，AI×Science大幕已启，正如DeepMind在2024年11月发布的研究报告所预见，一个科学发现的黄金时代正在到来。

AI × Science 十大前沿观察

AI for Science

- 基于LLM的科学研究
 - 跨学科知识的整合与创新
 - 基于LLM的科研流程重构
 - 革新人机协同科研模式
- 垂直领域科学大模型
 - 生命科学基础模型
 - 医学基础模型
 - 气象预测基础模型
 - 物质科学基础模型
- 融入先验知识的AI模型
 - 神经符号人工智能
 - 几何图神经网络
 - 物理启发的图神经网络
- AI科学家
 - AI科学家与自主科研智能体
 - 科学家助手
- 复杂世界的多智能体建模
 - 基于LLM的多智能体系统研究宏观经济学
 - 基于LLM的多智能体系统研究社会群体行为
 - 基于多智能体系统的未来城市治理
- AI仿真与系统模拟
 - AI for PDE (人工智能求解偏微分方程)
 - 因果数字孪生
 - AI驱动的多尺度数字有机体建模
 - 生物-环境闭环智能仿真系统
 - 具身生成式物理模拟引擎

Science for AI

- 物理世界的第一性原理
 - 自由能原理：智能的第一性原理
 - 泊松流模型：物理启发的深度生成模型
 - 范畴论：刻画大模型的能力边界
- 科学启发的AI新架构
 - Komogorov Arnold Networks (KAN)
 - 脉冲神经网络
 - 储备池计算
 - 机械神经网络
 - 傅里叶神经算子

AIS 基础设施

- 合成数据和数据基础设施
 - 合成数据生成：从序列数据到复杂数据
 - 科学数据表示：从图神经网络到多模态数据表征
 - 模型自我改进：通过合成数据提升AI系统性能
 - 数据基础设施：人体蛋白质组计划
- 新型智能计算
 - 凡人计算
 - 光学计算
 - 量子计算

PART 2

十大前沿观察

AI × Science

AI for Science

2.1 基于LLM的科学研究

2.1.1 背景介绍

近年来，大语言模型(Large Language Models, LLMs)正在深刻影响科学研究的方式。正如Nature Reviews Physics的评论文章[1]指出，LLM在科研中扮演的是“增强型助手”而非“替代者”的角色，这种增强作用体现在对科研活动不同维度的重构中。

从本质上看，LLM之所以能带来深远影响，在于它提供了一个统一的符号系统来重构科研活动的不同维度。具体而言，这种重构体现在三个深层次的维度上：

第一个层次是知识重构维度。传统的科学知识是分散在各个专业领域的，不同学科之间存在着“知识鸿沟”和“语言障碍”。LLM通过提供统一的知识表达框架，正在打破这些壁垒。例如，Galactica[2]通过整合海量科学文献，重新定义了科学知识的组织方式；BioGPT[3]在生物医学知识表示和生成方面取得突破。这些工作为科学发现提供了新的认知基础，使得研究者能够更容易地发现跨领域的关联和洞见。

第二个层次是过程抽象维度。科研过程传统上高度依赖研究者的专业经验和直觉判断，这种经验难以形式化和传承。LLM通过将科研流程抽象为可计算的任务序列，实现了科研活动的形式化和自动化。在化学领域，ChemCrow[4]展示了复杂实验流程的自动化执行；在生物领域，BioMedLM[5]展示了在生

物医学任务中的多任务学习能力。这些工作不仅提高了研究效率，更重要的是为科研方法的系统化和标准化提供了新的范式。

第三个层次是界面创新维度，体现为用自然语言重构人机物理系统的交互方式。传统上，科研人员需要掌握复杂的专业操作界面来使用科研设备，这种非自然的交互方式不仅限制了研究效率，也阻碍了创新思维的发挥。VISION[7]通过模块化架构实现了科研设备的自然语言控制，Med-PaLM[8]在医学图像分析中实现了直观的人机对话，这些创新不仅降低了使用专业设备的门槛，更重要的是实现了人与设备之间的认知层面协同。

这三个维度的变革是递进和互补的：知识重构提供了认知基础，过程抽象实现了方法创新，界面创新则打通了实践环节。这种多维度的系统性变革也带来了深层的挑战：

1. 知识表达：如何在知识重构中保持专业深度并确保表达的完备性？
2. 流程保障：如何在过程抽象中平衡自动化与创新空间？
3. 交互设计：如何构建更符合科研认知特点的自然交互界面？
4. 系统集成：如何实现知识、流程、界面三个维度的协同创新？

理解和应对这些挑战，将决定LLM能在多大程度上推动科学研究的进步。本报告将重点分析几个代表性工作，展示当前在应对这些挑战方面的最新进展，并探讨LLM驱动的科研新范式的未来发展方向。

2.1.2 研究进展

2.1.2.1 跨学科知识的整合与创新

推荐理由：Galactica[2]是首个专门面向科学文献训练的大规模语言模型，通过整合4800万篇科研论文、教材和知识库的内容，加上独特的训练策略和模型设计，展现了LLM在科学知识综合理解与应用方面的强大潜力。

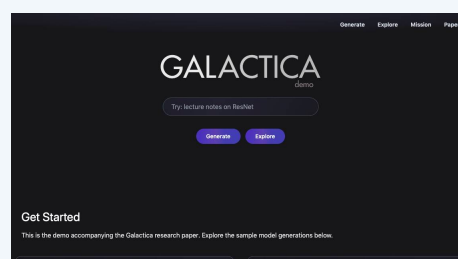
跨学科知识的整合与创新已成为推动科学进步的关键动力，然而，面对指数级增长的科研文献，研究人员往往难以全面把握不同领域的知识脉络。能否构建一个“科学知识的通用理解者”，打破学科壁垒，实现跨领域知识的有机融合？

Galactica是一个非常重要的尝试。通过创新的数据处理和训练策略，Galactica开创了专业领域大模型的新范式。它的核心突破在于提出了“高质量数据+多轮迭代”的训练方法，通过对精选的4800万篇科研文献进行4.25轮深度训练，验证了这一方法优于传统的“大规模数据+单轮训练”范式。同时，它设计了独特的Tokenization策略统一处理科研领域的多模态数据（如数学公式、化学结构、蛋白质序列），并创新性地引入Reference Token和Work Token分别用于构建知识图谱和支持多步推理。这些设计让Galactica在多个科学任务上取得了突破性进展：LaTeX公式理解准确率达68.2%（超GPT-3近20个百分点），数学推理准确率达41.3%（超Chinchilla 5.6个百分点），医学问答准确率创下77.6%的新纪录。

Galactica实现了科学知识的系统性整合。同时，由于有大量研究指出这项技术容易产生偏见和将

谎言断言为事实的倾向等缺陷，即文本生成中的幻觉（Hallucination），经过三天的激烈批评后即被迅速下线。

尽管如此，Galactica通过将领域知识系统性注入预训练过程，为构建专业领域大模型提供了重要范式，其影响已超越科学领域，为其他垂直领域大模型的发展提供了有益借鉴。其中，来自上海交通大学的团队将这一颇具潜力的模型引入到了地球科学领域，完成了300亿参数的地学大语言模型GeoGalactica的训练[9]。



Galactica的产品页面 | 来源：Galactica的官网（现已下线）

2.1.2.2 基于LLM的科研流程重构

推荐理由：科学研究的过程抽象是一个根本性挑战，它涉及如何将复杂的研究活动分解为可计算、可执行的任务序列。ChemCrow[4]通过创新性地将LLM与专业工具相结合，为这一挑战提供了一个可行的解决方案。

科学研究过程的形式化和自动化一直是一个重要而困难的课题。传统上，科研活动高度依赖研究者的经验和直觉，这种依赖不仅限制了研究效率，也使得

许多宝贵的研究经验难以传承和推广。如何将专家的经验判断转化为明确的操作流程，如何确保复杂实验过程的可重复性，如何在保持灵活性的同时实现高度自动化，这些问题长期困扰着各个科学领域。

ChemCrow为解决这些普遍性问题提供了一个创新的范式。其核心思想是构建一个“思考-行动-观察”的闭环系统，通过LLM作为协调者，将各类专业工具有机整合。具体而言，该系统包含三个关键创新：首先，它建立了一个基于自然语言的任务规划框架，能将复杂的研究目标分解为具体的操作步骤；其次，它设计了一套工具调用机制，使LLM能根据需求精确调用相应的专业工具；最后，它实现了一个动态反馈系统，能根据执行结果实时调整策略。

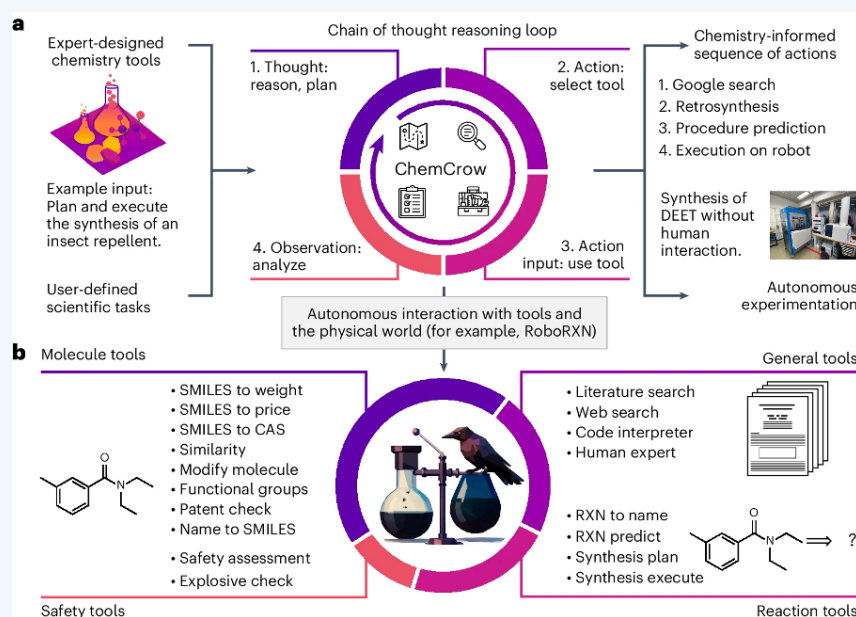
ChemCrow的成功意义远超化学领域。首先，它证明了将复杂的科研过程形式化是可行的，这为

其他领域的过程自动化提供了重要参考。其次，它展示了如何在保持系统灵活性的同时确保研究的可靠性和可重复性。最重要的是，它提供了一个可推广的框架，这个框架可以根据不同学科的特点进行调整和扩展。它代表了科研活动从“经验驱动”向“系统驱动”的重要转变，预示着一个更加自动化、规范化的科研新时代的到来。

2.1.2.3 革新人机协同科研模式

推荐理由： VISION[7]开创性地将模块化AI架构引入科学实验，通过将不同的认知模块有机整合，实现了首个全面的自然语言实验控制系统。它为人机协同科研提供了新的交互范式。

科学实验设备的操作一直是一个专业性很强的领域，传统上需要研究人员掌握复杂的专业操作界面和



ChemCrow任务解决过程概述 | 来源: M. Bran, A., Cox, S., Schilter, O. et al. Augmenting large language models with chemistry tools. Nat Mach Intell 6, 525–535 (2024). <https://doi.org/10.1038/s42256-024-00832-8>

命令。这不仅给研究人员带来很大的学习负担，也成为科研自动化的重要障碍。如何让科学家能够用自然语言与科研设备直接“对话”，让设备理解并执行研究人员的意图，这是一个具有重要意义又充满挑战的课题。

VISION的突破性在于它提出了一种创新的模块化AI架构。系统由多个功能模块（Cognitive Blocks）组成，每个模块都基于一个针对特定任务优化的大语言模型。这些模块协同工作，共同完成从语音识别、意图理解到任务规划和代码生成的全流程。

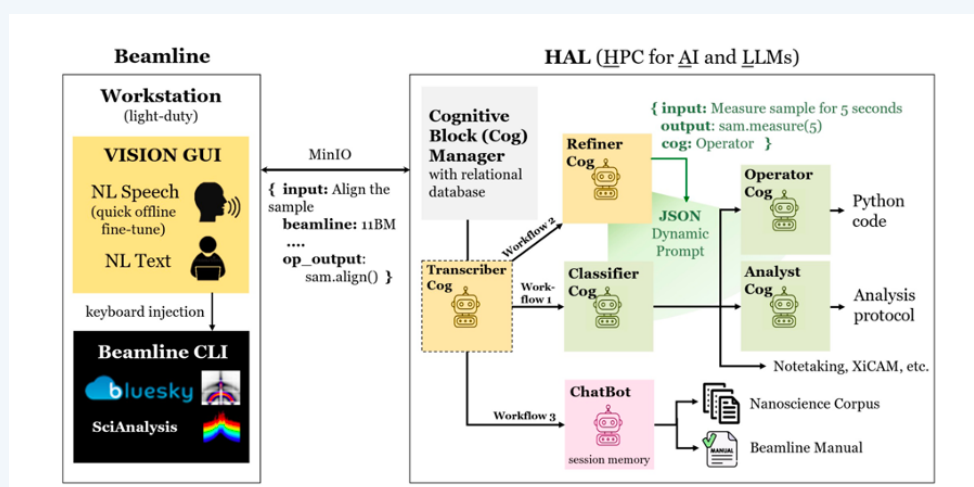
具体来说，系统包含转录器（Transcriber）、分类器（Classifier）、操作员（Operator）和分析师（Analyst）等关键模块。这些模块以工作流（Workflow）的形式进行组织和调度。当用户发出一个请求时，系统首先判断其属于数据采集、分析还是其他类型，然后调用相应的工作流处理。最后，系统将生成的代码或分析结果返回给用户确认，并在获

得许可后提交给底层的实验控制平台执行。整个过程中，用户通过自然语言与系统交互，无需了解复杂的技术细节。

VISION的意义在于它开创了一种全新的人机交互范式。通过模块化的架构设计，它充分利用了大语言模型在不同任务上的特长，实现了全面的语言理解和任务执行能力。同时，它将前沿的AI技术与传统的科研工作流程巧妙结合，极大地提高了实验效率和灵活性。更重要的是，它为实现“用语言做实验”这一科研人员的长期愿景迈出了关键一步。

2.1.3 挑战与展望

然而，基于LLM的科学研究也面临着几个根本性的挑战：幻觉问题，LLM可能生成表面上合理但实际上不准确的内容，这在科学研究中尤其危险；创新的认知边界问题，LLM的“创新”源自已有知识的重组和推理，这与真正的科学突破性发现可能存在



VISION的系统架构概览 | 来源:Mathur, S., van der Vleuten, N., Yager, K. G., & Tsai, E. (2023). VISION: A modular AI assistant for natural human-instrument interaction at scientific user facilities. arXiv preprint arXiv:2412.18161.

本质差异；可重复性与可解释性的矛盾，科学研究要求结果可重复、过程可追溯，但LLM的黑盒特性与此形成本质冲突。

面向未来，大语言模型驱动的创新亟需在以下方向实现突破：

1. 构建可信科研平台：整合跨模态知识，建立链上可追溯、过程可审计、结果可复现的科研基础设施。

2. 探索人机协同创新范式：发掘研究人员专业经验与LLM知识处理能力的最佳协同路径，实现辅助决策到联合创新的跃升。

3. 推进跨学科协同治理：计算机、伦理学、科学哲学、科研管理等领域，需协同推进LLM在科研领域的责任评估、伦理审查、规范制定等工作。

科学研究正在经历一场方法论的革新。基于LLM的科学研究不仅提供了新的研究工具，更重要的是开创了一种新的认知范式。这种范式将人类的创造力与AI的能力有机结合，有望加速科学发现的进程。随着技术的进步和应用的深入，这种新范式可能会重塑科学研究的方式，开启科学探索的新纪元。

2.1.4 参考文献

[1] Birhane, A., et al. "Science in the age of large language models." *Nature Reviews Physics* 5 (2023): 277-280.

<https://doi.org/10.1038/s42254-023-00581-4>

推荐理由：这是一篇重要的评论性文章，由四位AI伦理和政策专家撰写，全面讨论了LLM在科学研究中的潜力和风险。文章不仅提供了清晰的概念框架，还引发了关于LLM在科研中应用的深入讨论。

[2] Taylor, R., et al. "Galactica: A Large Language Model for Science." *arXiv preprint arXiv:2211.09085* (2022).

<https://arxiv.org/abs/2211.09085>

推荐理由：这是首个专门面向科学文献训练的大规模语言模型的

技术报告，详细描述了模型架构、训练策略和评估结果。虽然模型最终下线，但其技术创新和失败教训都具有重要的参考价值。

[3] Luo, R., et al. "BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining." *Briefings in Bioinformatics* 23.6 (2022): bbac409.

<https://doi.org/10.1093/bib/bbac409>

推荐理由：这篇论文介绍了专门针对生物医学领域的预训练语言模型，展示了领域特定LLM的潜力，为其他垂直领域的模型开发提供了重要参考。

[4] Bran, A.M., et al. "Augmenting large language models with chemistry tools." *Nature Machine Intelligence* 6 (2024): 525-535.

<https://doi.org/10.1038/s42256-024-00832-8>

推荐理由：这篇发表在Nature Machine Intelligence的论文展示了如何将LLM与专业工具结合，实现科研流程的自动化。文章提供了详细的系统设计和实验验证，具有很强的实践指导意义。

[5] Bolton, E., et al. "BioMedLM: A 2.7B Parameter Language Model Trained on Biomedical Text." *arXiv preprint arXiv:2403.18421* (2024).

<https://arxiv.org/abs/2403.18421>

推荐理由：这是一个针对生物医学领域的开源语言模型，论文详细描述了模型训练过程和多任务学习能力，为生物医学领域的AI应用提供了重要工具。

[6] Mathur, S., et al. "VISION: A Modular AI Assistant for Natural Human-Instrument Interaction at Scientific User Facilities." *arXiv preprint arXiv:2412.18161* (2023).

<https://arxiv.org/abs/2412.18161>

推荐理由：这篇预印本详细描述了一个创新的模块化AI系统，展示了如何通过自然语言实现科研设备控制，为提高实验效率提供了新思路。

[7] Singhal, K., et al. "Large Language Models Encode Clinical Knowledge." *Nature* 620.7972 (2023): 172-180.

<https://doi.org/10.1038/s41586-023-06291-2>

推荐理由：这篇Nature论文展示了LLM在医学领域的应用潜力，特别是在医学知识理解和图像分析方面的突破，为医学AI的发展提供了重要参考。

[8] Lin, Z., et al. "GeoGalactica: A Scientific Large Language Model in Geoscience." *arXiv preprint arXiv:2401.00434* (2024).

<https://arxiv.org/abs/2401.00434>

推荐理由：这篇论文介绍了如何将Galactica的架构应用到地球科学领域，展示了专业领域大模型的发展潜力，为其他学科的模式开发提供了借鉴。

2.2 垂直领域科学大模型

2.2.1 背景介绍

近年来，以GPT和BERT为代表的基础模型(Foundation Models)[1]在人工智能领域取得了突破性进展。这类模型通过大规模预训练获得了强大的表示学习能力和良好的迁移能力，为人工智能的发展开辟了新方向。然而，在专业科学领域的应用中，研究者们发现通用基础模型往往难以直接应用于特定科学问题的解决。这主要源于科学研究对专业知识深度整合的需求、数据特点的差异性以及任务特征的特殊性。

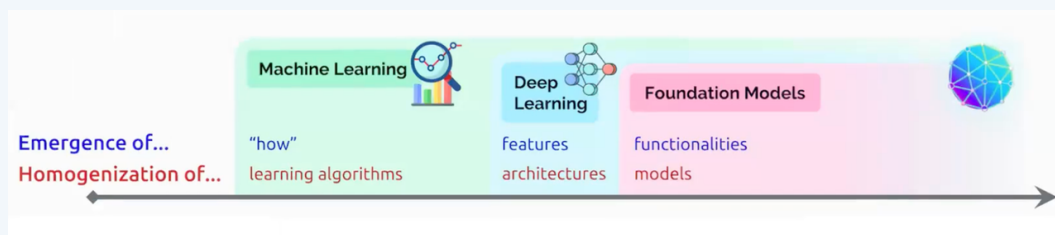
在科学研究领域，模型不仅需要具备强大的学习能力，更需要严格遵循物理规律，保持结果的可解释性，并确保预测的物理合理性。例如，在分子设计中，模型需要确保预测结构满足化学键角和键长的约束；在材料研究中，需要考虑晶体结构的周期性和对称性。这些特殊需求推动了针对单一领域的基础模型研究。以AlphaFold3[2]为代表的成功案例，展示了将基础模型思想与领域专业知识深度融合的巨大潜力。

单一领域基础模型继承了基础模型的核心特征，如大规模预训练、通用表示学习和迁移能力，同时深度整合了领域特定知识，包括物理规律、专业知识和

领域约束。与通用基础模型相比，领域基础模型在数据特点、任务特征和应用场景等方面都呈现出显著的特性。

从发展历程来看，领域基础模型的演进大致可分为三个阶段。在早期探索阶段(2017-2019)，研究主要集中在特定任务的深度学习方法上，如DeepChem[3]在分子性质预测中的应用，以及基于3D-CNN[4]的材料结构分析；快速发展阶段(2020-2021)见证了几个具有里程碑意义的突破，其中最具有代表性的是AlphaFold1[5]在蛋白质结构预测领域的重大突破。进入成熟应用阶段(2022-2023)，我们看到了更多创新性的工作：物质科学领域中针对分子动力学模拟的GPTFF(Graph-based Pre-trained Transformer Force Field)[6]以及面向高效扩展神经算子的统一框架Universal Physics Transformer (UPT)[7]，还有生命科学领域的scFoundation[8]和GMAI[9]和地球气候系统领域的Pangu-Weather[10]、FuXi[11]、FengWu[12]等工作，都越来越走向科学统一的基础模型。

这种发展过程始终都是围绕着核心关键的问题：能否将基础模型的理念推广到更广泛的科学研究领域？特别是，能否构建一个统一的科学基础模型，来加速跨学科的科学发现进程？要回答这个问题，仍面临三个根本性的挑战[13]：



人工智能的范式转变

1. 如何处理跨模态、跨尺度、跨领域的科学数据？

2. 如何将这些基本科学规律融入模型设计，使模型既保持灵活性又不违背基本物理定律？

3. 如何设计模型架构，使其能够捕捉并利用这些潜在关联，实现跨学科的知识迁移和创新发现？

虽然目前还没有出现统一的科学基础模型，但是物质科学、生命科学、气象预测等各个领域已经涌现出一些创新性的工作，出现了大量的单一领域科学基础模型，解决领域问题。以下我们将重点介绍三个最具代表性的研究方向及其最新进展。

2.2.2 研究进展

2.2.2.1 生命科学基础模型

推荐理由： scFoundation[8]代表了生命科学领域基础模型的重要突破。它实现对复杂生命系统的建模，不仅在技术上创新性地解决了单细胞测序数据的特殊挑战，更为生命科学的智能化分析开辟了新范式。

随着单细胞RNA测序技术的发展，科研人员能够开始观察和理解细胞内的基因表达模式。然而，如何从海量、异质、存在技术噪声的单细胞数据中学习到有意义的生物学知识，一直是该领域的核心挑战。传统的计算方法往往只能处理部分基因，或者需要针对特定任务单独构建模型，缺乏统一的分析框架。

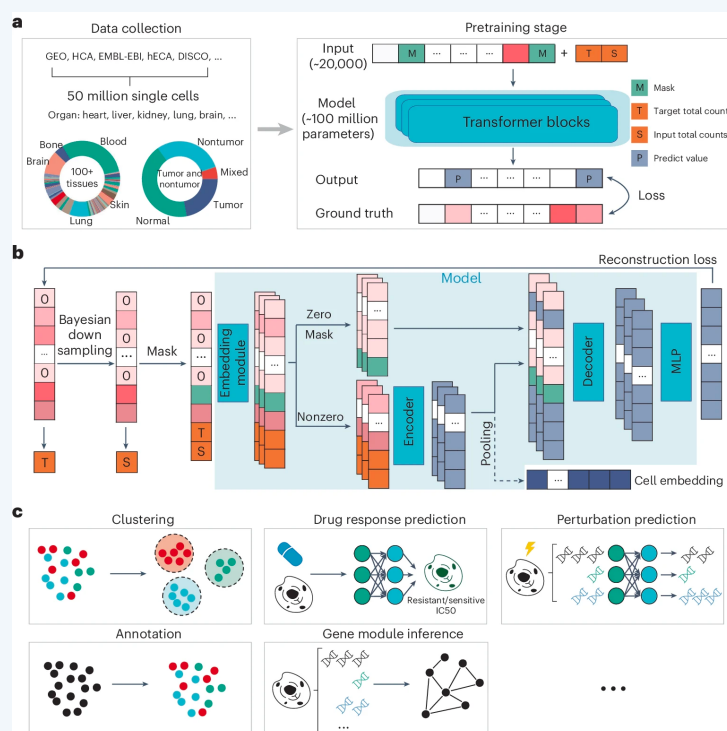
在2023年Geneformer[14]首次验证基因表达数据预训练的可行性，以及scGPT[15]探索生成式方法的基础上，scFoundation[8]通过独特的技术方案实

现了对复杂生命系统的统一建模，标志着生命科学正在进入基础模型时代。

scFoundation通过创新的模型架构设计和预训练策略，首次实现了对全基因组范围的统一建模。该模型包含1亿个参数，覆盖约2万个蛋白编码基因，在超过5000万个人类单细胞转录组数据上进行预训练。其核心创新在于：首先，设计了一个非对称的Transformer架构(xTrimoGene)，通过将编码器限制在非零表达基因上处理同时让解码器覆盖全基因组范围，巧妙解决了数据稀疏性带来的计算挑战；其次，提出了读取深度感知(RDA)的预训练任务，不仅可以捕获基因间的复杂共表达关系，还能自动处理不同测序深度带来的技术噪声，实现了对细胞转录组的统一表征。

在多个下游任务的实验中，scFoundation展现出了卓越的性能。在基因表达增强任务上，相比传统方法将平均绝对误差降低了一半；在药物响应预测中，无需针对性微调就达到了最先进的预测准确率；在细胞扰动预测任务中，模型能够准确预测基因敲除后的表达变化。值得注意的是，scFoundation通过预训练获得的细胞嵌入向量和基因上下文嵌入向量，展现出了优异的生物学意义，可以用于细胞类型注释、基因模块发现等多种分析任务。

通过大规模预训练，scFoundation学习到了普遍的生物学知识，可以迁移到各种下游任务中。这种方法避免了传统方法需要针对每个任务重新训练模型的低效率，为生命科学研究提供了一个统一的计算基础设施。



scFoundation预训练框架概览 | 来源: Hao, M., Gong, J., Zeng, X. et al. Large-scale foundation model on single-cell transcriptomics. Nat Methods 21, 1481-1491 (2024).
<https://doi.org/10.1038/s41592-024-02305-7>

此外，蛋白质结构预测在理解蛋白质的结构-功能关系、推动生物学研究以及促进制药开发和实验设计等方面具有关键意义。尽管日益成熟的深度学习技术以及丰富的实验三维蛋白质结构数据，在很大程度上加速了结构预测的发展，但蛋白质的结构动力学特性尚未得到充分关注。然而，蛋白质在功能过程中的构象变化对于精准药物设计至关重要：通过纳入动态变化的因素，能够显著优化药物与蛋白质的结合效率，从而提高靶向治疗的有效性。基于此，文献[16]提出了一种创新性4D扩散模型，该模型结合了分子动力学（MD）模拟数据，旨在学习并模拟蛋白质的动态结构。这一研究在行业内率先实现了对10纳秒级动态运动过程的生成，为蛋白质结构动力学在药物设计等实际应用场景中的进一步探索提供了重要的启示

与方法支持。

2.2.2.2 医学基础模型

推荐理由： GMAI (Generalist Medical AI)[9]代表了医疗AI领域的范式转变，从传统的任务特定模型转向了更加通用和灵活的基础模型方法。这项发表在Nature上的开创性工作不仅提出了医疗AI的新范式，更为未来医疗保健的智能化提供了全新视角。

长期以来，医疗AI面临着三大核心挑战：数据的稀缺性和异质性、预测任务的复杂性、领域知识的整合难度。传统的医疗AI系统采用“一个任务一个模型”的方式，不仅需要大量标注数据，而且难以利用最新

的医学知识。早期的一些尝试，如CheXzero[17]在放射影像解读上的探索，以及GPT-3[18]在医学执照考试上的表现，都展示了基础模型在医疗领域的潜力，但距离临床应用仍有不小差距。

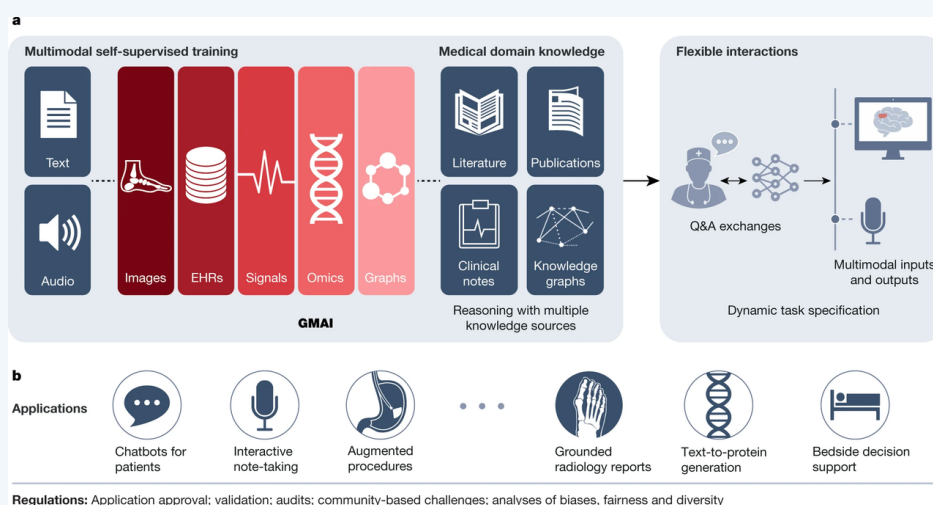
GMAI的创新之处在于，它提出了一个统一的医疗AI框架。这个框架通过自监督学习方式，在海量异构医疗数据上进行预训练，同时融入了结构化的医学知识。它可以训练来自图像、EHR信号、组学数据（基因组学和转录组学）等的不同类型的数据，能够摄取医学领域的多源异构知识（出版物、教科书、临床节点的知识图谱），还具有零样本学习能力，根据几个例子就可以学习去做模型从来没有见过的任务。

这样做的好处是，通过自由组合和交错不同的数据模式，能够实现动态任务规范（Dynamic Task Specification），即根据这些指令即时解决任务。因此基本上是零样本或少样本的上下文类型学习。这些模型具有大量的知识，因为它们允许我们用结构化

领域知识进行推理。例如，当遇到一个复杂的病例时，模型会首先分析各项检查结果，然后检索相关的医学文献和指南，最后通过逻辑推理给出诊断和治疗建议，并清晰地解释推理过程。

GMAI的出现推动了一系列后续创新。Med-Flamingo[19]基于GMAI的框架，进一步优化了多模态医疗诊断能力；RA-CM3[20]通过改进检索机制，显著提升了模型的知识获取能力；而基于GMAI的临床辅助系统已经在多家医院开展试点应用，显示出良好的实际效果。

这些领域也有很多挑战，包括验证（Validation）、确认（Verification）、社会偏见（Social Biases）以及非常重要的患者隐私问题。即便如此，这项工作的意义远超技术创新，它开启了医疗AI的新纪元，也为其他领域基础模型的发展提供了新的思路。



GMAI (generalist medical AI) 模型流程图概览 | 来源: Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, et al. Foundation models for generalist medical artificial intelligence. nature, 2023, 616(7956): 259-265

2.2.2.3 气象预测基础模型

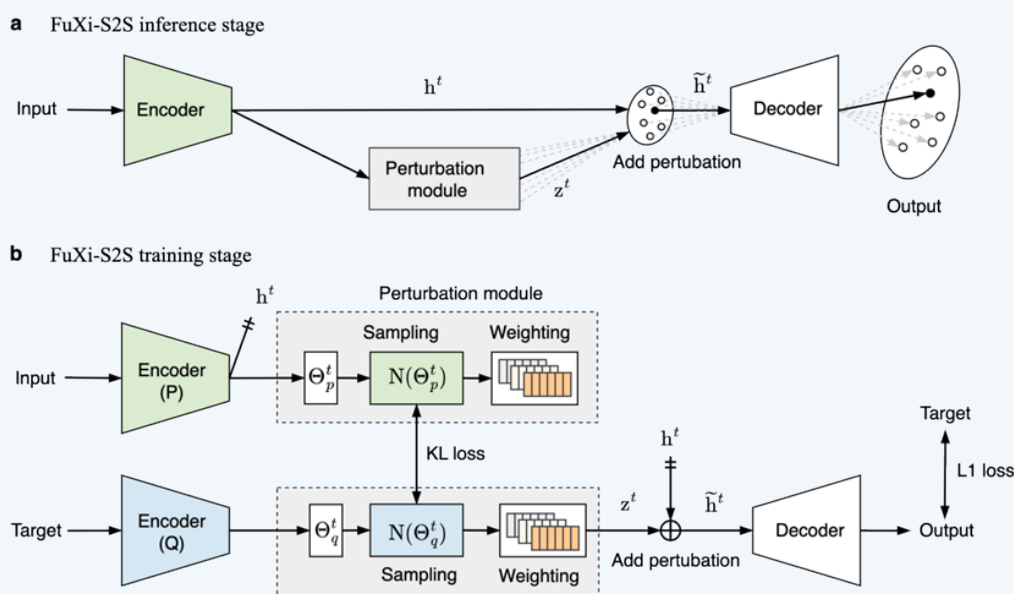
推荐理由：气象预测是最快能够在现实生活应用并对生产生活产生巨大影响的领域之一，气象预测基础模型在预测精度和效率上都实现了突破性进展，为极端天气预警和气候变化研究提供了新工具。

如何对未来天气发展进行准确预测，特别是预测极端天气，越来越成为意义重大的科学问题。然而，传统的气象学研究受限于数据的缺乏、人类知识和经验的局限以及气象系统内在的不确定性因素等影响，无法给出较长期而精准的预测。

近年来，若干气象预测大模型被陆续开发出来，包括Google的GraphCast[21]、华为的盘古大模型[10]、英伟达的FourCastNet[22]、上海科学智能研究院和复旦大学的伏羲气象大模型等[11]，都取得了显著进展。

全球首个基于人工智能的次季节气候预测大模型“伏羲”（FuXi-S2S），首次超越传统数值预报模式的标杆——欧洲中期天气预报中心（ECMWF）的S2S模式。该模型涵盖了13个气压层的5个高空大气变量和11个地面变量，实现了两个关键性技术突破：一是将热带大气季节内振荡(MJO)这一次季节最重要的可预报性来源纳入模型；二是创新性地设计了智能扰动生成模块，能够在隐空间中刻画气候系统演变的概率特征。这些创新使得模型在多个关键指标上超越了ECMWF S2S系统，特别是将MJO的有效预报期从30天延长到36天。

在实际应用中，“伏羲”仅需7秒即可完成42天的全球日平均预报，并在2023年成功预测了印度尼西亚特大暴雨事件，提前20天发出预警，为防灾减灾赢得了宝贵时间。此外，复旦大学联合上海科学智能研究院、上海创智学院于2024年11月发布了中国首个基于自主知识体系构建的气候科学大语言模型



FuXi-S2S模型的结构示意图 | Chen, L., Zhong, X., Zhang, F. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. npj Clim Atmos Sci 6, 190 (2023).

PI@Climate[23], 该模型已经实现了跨越46个学科的知识整合, 通过多智能体建模提升了对复杂气候系统的认知能力。

2.2.2.4 物质科学基础模型

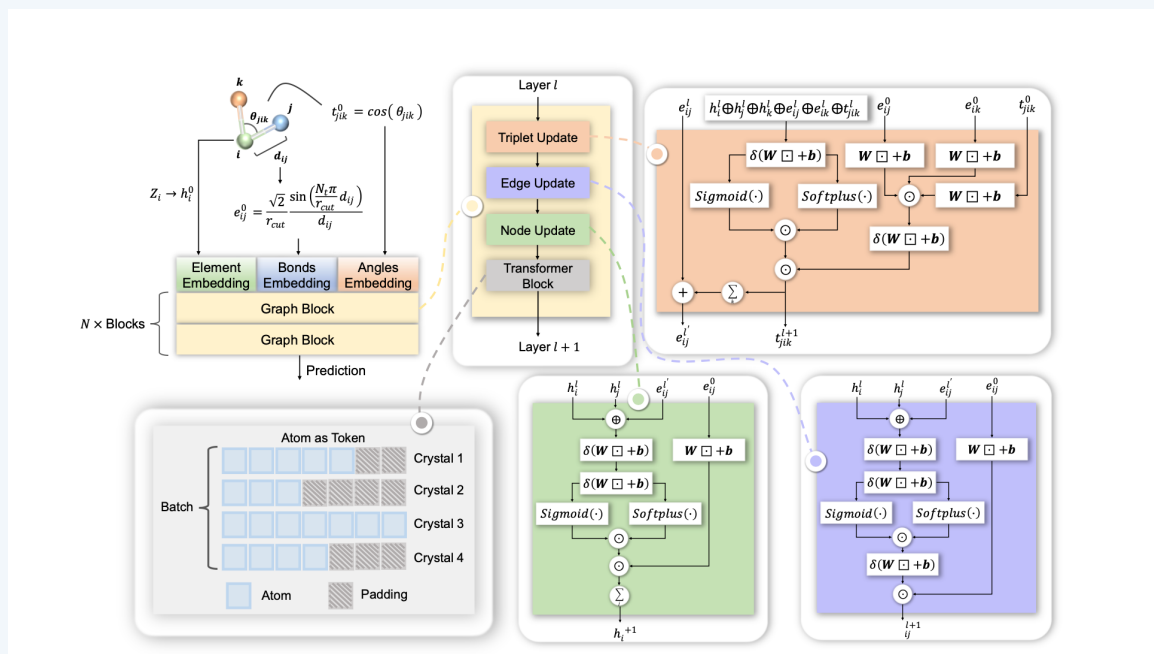
2.2.2.4.1 分子动力学模拟

推荐理由: 分子动力学模拟是材料科学研究的基础工具, 但长期受限于计算效率与精度的权衡问题。基于深度学习的新型力场模型为这一难题提供了突破口, 有望推动材料科学研究范式的革新。

分子动力学模拟是研究材料微观行为的重要方法, 但其准确性很大程度上依赖于所使用的力场模型。传统的解析力场函数形式较为简单, 难以准确描述复杂的原子间相互作用, 而密度泛函理论 (DFT)

虽然准确但计算成本过高, 这使得大规模原子体系的精确模拟面临着重大挑战。

针对这些问题, GPTFF (Graph-based Pre-trained Transformer Force Field) [6]作为一种新型的通用人工智能力场模型被提出。该模型采用图神经网络来表征晶体结构, 并引入Transformer架构来提升模型的学习能力。GPTFF模型的训练数据源自自研的Atomly材料数据库 (<https://atomly.net/>)。该数据库包含近35万个无机材料数据, 数据规模和质量位于世界顶级水准, 知识产权自主可控, 也是我国唯一可对标欧美竞品的无机材料数据库。通过利用包含3780万个单点能量、117亿个原子力和3.402亿个应力张量的大规模数据集进行训练, GPTFF在能量、力和应力预测方面都达到了极高的精度, 分别实现了32meV/atom、71meV/Å和0.365GPa的平均绝对误差, 比传统经验力场提升了



GPTFF模型的模型架构 | F. Xie, T. Lu, S. Meng, M. Liu, GPTFF: A high-accuracy out-of-the-box universal AI force field for arbitrary inorganic materials, Science Bulletin (2024), doi: 10.1016/j.scib.2024.08.039.

一个数量级，同时计算速度比第一性原理快3-4个数量级。

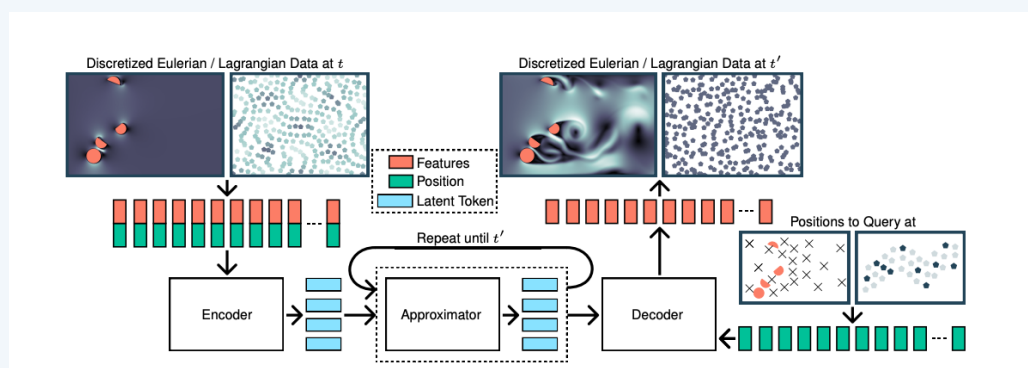
与GPTFF并驾齐驱的还有 DeepMD[24]、PaiNN [25]和Uni-Mol+[26]等代表性模型。DeepMD通过深度势能模型实现了对复杂体系的准确描述，已成功应用于二维材料的相变预测；PaiNN则创新性地引入等变性保持机制，在分子动力学模拟中展现出优异的稳定性；Uni-Mol+创新性地提出了基于双轨道Transformer的分子表示方法，通过迭代优化初始构象来逼近DFT平衡构象，并在此基础上实现量子化学性质预测，在HOMO-LUMO能隙等关键分子性质预测任务上实现了显著进展。

2.2.2.4.2 物理模拟

推荐理由：物理模拟是科学研究和工程应用的重要工具，但长期面临着不同物理系统需要不同建模方法的挑战。Universal Physics Transformer (UPT) [7]通过创新的架构设计，首次实现了对不同类型物理系统的统一处理，为发展统一的物理模拟基础模型提供了重要范式，有望推动物质科学研究方法的革新。

物理模拟对推动材料科学研究至关重要，但长期以来受困于方法的碎片化问题——不同类型的物理系统往往需要不同的模型和算法，这不仅增加了研究成本，也限制了模拟技术的广泛应用。例如，在流体动力学模拟中，常需在拉格朗日方法与欧拉方法间权衡，网格模型与粒子模型也各有优劣。这种方法的碎片化不仅增加了模型开发和维护的成本，也限制了物理模拟在更广泛场景下的应用。

针对这一问题，Universal Physics Transformer (UPT)提出了一个统一的神经算子框架。该框架创新性地采用无网格/无粒子的潜在表示，通过分层Transformer架构实现了对不同类型物理系统的统一处理。UPT的核心是其高效的三段式设计：编码器灵活处理不同类型的输入并压缩到统一的潜在空间，近似器在潜在空间中进行动力学传播，解码器则支持在任意时空点进行查询。通过Inverse Encoding/Decoding等创新训练策略，UPT实现了高效的Latent Rollout，使其能够扩展到大规模物理系统的模拟。在多个具有挑战性的任务上，如网格流体模拟、Reynolds平均Navier-Stokes模拟和拉格朗日动力学等，UPT都展现出优异的性能，同时保持了良好的可



UPT学习范式示意图架构| B. Alkin, A. Fürst, S. Schmid, L. Gruber, M. Holzleitner and J. Brandstetter, "Universal Physics Transformers: A Framework For Efficiently Scaling Neural Operators," in Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS 2024), 2024.

扩展性和泛化能力。

UPT的工作为构建通用的物理模拟基础模型提供了重要范式。它不仅展示了如何通过巧妙的架构设计实现不同物理系统的统一表示和处理，还为提升模型可扩展性提供了新思路。这对于发展更通用高效的物质科学模拟工具具有重要的启发意义。

2.2.3 挑战与展望

目前，科学家们正在各个领域探索基础模型的应用潜力，同时也在朝着构建统一科学基础模型的宏伟目标迈进。这一进程面临着异构数据融合、物理规律约束和可解释性等重要挑战，但近期的突破性进展展现了令人振奋的前景。

在气候科学领域，由复旦大学等机构联合开发的PI@Climate模型[23]已经实现了跨越46个学科的知识整合，通过多智能体建模提升了对复杂气候系统的认知能力。在生命科学领域，“女娲”系列模型[23]展示了从微观到宏观的突破：女娲基因导航大模型通过图神经网络解密了基因组“暗物质”的作用机制，女娲生命流体大模型实现了人体内流体动力学的精确模拟，而女娲生物结构大模型则在蛋白质动态结构预测方面超越了AlphaFold3的性能。燧人分子基础大模型[23]通过融合量子力学性质（如能量、力、结构）与热力学参数（包括结构分布、熵、焓及自由能变化），将分子动力学模拟过程融入大模型，实现了物质宏观与微观性质的快速预测，并在量子化学性质预测基准QM9数据集的绝大多数任务中超越现有AI模型。这些成功案例证明，将基础模型理念与领域专业知识深度融合是可行且富有成效的。

随着模型架构的创新和算力的提升，我们有望见证一个统一方法论的科学基础模型的诞生。这个模型将不再局限于单一学科，而是能够理解和运用物理、化学、生物等多个领域的基本规律，实现跨学科的知识推理和创新发现。它将成为科学家的“数字智囊”，不仅能预测复杂系统的行为，更能主动提出研究假设、设计实验方案。在材料研发、药物设计、气候预测等关乎人类福祉的重大领域，这种模型将显著加速科学发现的进程。

2.2.4 参考文献

[1] Bommasani, Rishi, et al. "On the Opportunities and Risks of Foundation Models." arXiv preprint arXiv:2108.07258 (2021).

<https://arxiv.org/abs/2108.07258>

推荐理由：这篇论文为整个报告提供了基础模型的重要背景和概述，为后续讨论奠定了基础。

[2] Abramson, J., et al. "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3." Nature 630, 493–500 (2024).

<https://www.nature.com/articles/s41586-024-07487-w>

推荐理由：这是AlphaFold系列最新的突破性成果，首次实现了统一的生物分子结构预测框架。该工作将深度学习在结构生物学领域的应用推向了新的高度。

[3] Ramsundar, Bharath, et al. "Deep Learning for the Life Sciences." O'Reilly Media (2019).

<https://www.amazon.com/Deep-Learning-Life-Sciences-Microscopy/dp/1492039837>

推荐理由：DeepChem被提及为早期将深度学习应用于分子性质预测的探索，提供了重要背景。

[4] C. Wang, "A Review on 3D Convolutional Neural Network," 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2023, pp. 1204-1208,

<https://ieeexplore.ieee.org/document/10075760>

推荐理由：三维CNN被引用为材料结构分析的早期方法，有助于说明该领域技术的发展历程。

[5] Senior, Andrew W., et al. "Improved protein structure prediction using potentials from deep learning." *Nature* 577.7792 (2020): 706-710.

<https://www.nature.com/articles/s41586-019-1923-7>

推荐理由: AlphaFold 1被指出是蛋白质结构预测的里程碑式突破, 是了解近期进展的关键基础。

[6] Xie, Feiyang, et al. "GPTFF: A high-accuracy out-of-the-box universal AI force field for arbitrary inorganic materials." *Science Bulletin* (2024).

<https://www.sciencedirect.com/science/article/pii/S2095927324006327?via%3Dihub>

推荐理由: GPTFF被强调为分子动力学模拟的创新力场模型, 代表了材料科学基础模型的最新重要进展。

[7] B. Alkin, A. Fürst, S. Schmid, L. Gruber, M. Holzleitner and J. Brandstetter, "Universal Physics Transformers: A Framework For Efficiently Scaling Neural Operators," in *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS 2024)*, 2024.

<https://openreview.net/pdf?id=oUXiNX5KRm>

推荐理由: UPT被介绍为一个统一的框架, 用于跨不同物理系统高效缩放神经算子, 是将基础模型应用于物理模拟的重要发展。

[8] Hao, Minsheng, et al. "Large-scale foundation model on single-cell transcriptomics." *Nature Methods* 21.12 (2024): 1481-1491.

<https://www.nature.com/articles/s41592-024-02305-7>

推荐理由: scFoundation被聚焦为将基础模型应用于单细胞转录组数据的突破, 克服了关键挑战, 为该领域开辟了新的可能性。

[9] Moor, Michael, et al. "Foundation models for generalist medical artificial intelligence." *Nature* 616.7956 (2023): 259-265.

<https://www.nature.com/articles/s41586-023-05881-4>

推荐理由: GMAI被呈现为开创性的通用医疗AI模型, 代表了医疗AI的范式转变, 为智能医疗的未来提供了新视角。

[10] Bi, K., Xie, L., Zhang, H. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 619, 533-538 (2023).

<https://www.nature.com/articles/s41586-023-06185-3>

推荐理由: 面向20日全球天气预报的高分辨率通用模型盘古天气预报模型是基础模型在天气预报领域的代表。

[11] Chen, L., Zhong, X., Zhang, F. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Clim Atmos Sci* 6, 190 (2023).

<https://www.nature.com/articles/s41612-023-00512-1>

推荐理由: FuXi是一个用于15天全球天气预报的级联机器学习预报系统, 展现了天气预报领域的最新进展。

[12] Kang Chen, Tao Han, Junchao Gong, et al. FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead. *arXiv:2304.02948*, 2023

<https://arxiv.org/abs/2304.02948>

推荐理由: FengWu是一种先进的AI驱动全球中期天气预报系统, 通过多模态与多任务学习方法显著提升了天气预报的精度和时效性。

[13] Chen, Kang, et al. "An Overview of Domain-Specific Foundation Models: Key Technologies, Applications, and Challenges." *arXiv*, 2024, *arXiv:2409.04267*,

<https://arxiv.org/abs/2409.04267>.

推荐理由: 这篇文章全面概述了定制化领域专用基础模型的方法论, 介绍了基础概念、通用架构, 并总结了构建领域专用模型的关键方法以及面临的挑战。

[14] Theodoris, C.V., Xiao, L., Chopra, A. et al. Transfer learning enables predictions in network biology. *Nature* 618, 616-624 (2023).

<https://www.nature.com/articles/s41586-023-06139-9>

推荐理由: 基因表达基础模型建模的先驱工作Geneformer

[15] Cui, H., Wang, C., Maan, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods* 21, 1470-1480 (2024).

<https://www.nature.com/articles/s41592-024-02201-0>

推荐理由: 基因表达基础模型建模的重要工作scGPT

[16] K. Cheng, C. Liu, Q. Su, J. Wang, L. Zhang, Y. Tang, Y. Yao, S. Zhu, Y. Qi. 4D Diffusion for Dynamic Protein Structure Prediction with Reference Guided Motion Alignment. *AAAI* 2025.

推荐理由: 行业首次实现生成10纳秒级的动态运动过程。

[17] Tiu, E., Talius, E., Patel, P. et al. Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nat. Biomed. Eng* 6, 1399-1406 (2022).

<https://www.nature.com/articles/s41551-022-00936-9>

推荐理由: CheXzero被指出是在放射学解读中应用医疗人工智能的早期探索, 有助于说明将基础模型应用于医学的潜力和挑战。

[18] Brown, Tom, et al. "Language models are few-shot learners." *Advances in neural information processing systems* 33 (2020): 1877-1901.

<https://proceedings.neurips.cc/paper/2020/hash/1457c0d-6bfc4967418bf8ac142f64a-Abstract.html>

推荐理由：GPT-3在医师资格考试中的表现，被引用为基础模型在医学领域潜力的一个指示，使其成为该部分的相关参考。

[19] Moor, Michael, et al. "Med-Flamingo: A Multimodal Medical Few-shot Learner." arXiv, July 2023, arXiv:2307.15189, <https://arxiv.org/abs/2307.15189>.

推荐理由：Med-Flamingo模型在多模态医疗诊断方面进行了优化，延续了GMAI工作的创新。

[20] Chen, Xiaoli, et al. "Retrieval-Augmented Multimodal Language Modeling" arXiv preprint arXiv:2212.06759 (2022). <https://arxiv.org/abs/2212.06759>

推荐理由：RA-CM3是另一个基于GMAI的模型，具有改进的知识检索能力，展示了近期医疗人工智能的新进展。

[21] Remi Lam et al., Learning skillful medium-range global weather forecasting. Science 382, 1416-1421 (2023) <https://www.science.org/doi/10.1126>

推荐理由：GraphCast是谷歌的一个天气预测模型，被作为人工智能天气预报系统的代表引用。

[22] Jaderberg, Max, et al. "FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators." arXiv preprint arXiv:2202.11214 (2022). <https://arxiv.org/abs/2202.11214>

推荐理由：FourCastNet是英伟达的一个天气预测模型，基于适应性傅里叶神经网络，提供了天气预报模型构建的有益参考。

[23]"上智院发布系列垂直领域科学大模型"科学网，2024年11月 <https://news.sciencenet.cn/htmlnews/2024/11/533897.shtml>

推荐理由：上海科学智能研究院发布的一系列基础模型介绍。

[24] Zhang, L., et al. "Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics." Physical Review Letters (2018). <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.120.143001>

推荐理由：这是深度学习在分子动力学领域的奠基性工作，首次实现了量子力学精度的大规模模拟。该方法为材料科学研究提供了强大的计算工具。

[25] Schütt, K., et al. "Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra." arXiv preprint arXiv:2102.03150 (2021). <https://arxiv.org/abs/2102.03150>

推荐理由：这项工作在分子性质预测中引入了等变性机制，显著提升了预测精度。该方法为分子动力学模拟提供了新的理论框架。

[26] Lu, S., Gao, Z., He, D. et al. Data-driven quantum chemical property prediction leveraging 3D conformations with Uni-Mol+. Nat Commun 15, 7104 (2024). <https://www.nature.com/articles/s41467-024-51321>

推荐理由：这项工作提出了创新的双轨道Transformer分子表示方法，实现了高精度的分子性质预测。该模型在多个化学性质预测任务上都取得了突破性进展。

2.3 融入先验知识的AI模型

2.3.1 背景介绍

在构建科学大模型的过程中，数据驱动和先验知识的融合是关键，需要AI模型具备可验证的推理能力和逻辑体系，从单独的模式识别演进为复杂认知能力。“融入先验知识的AI模型”是在数据驱动的基础上，通过融入人类积累的数学、逻辑、科学规律等系统性知识，使AI能够在复杂场景中进行可解释的推理。不同于历史上符号主义具有明确的规则编码，在现在的研究中，可以通过符号推理、知识图谱与深度学习的结合，来逐步实现这一目标。

常见的知识表示形式，包括逻辑规则、概率图、知识图谱、数学公式等。下面将分别介绍其与深度学习融合的进展。

(1) 融入逻辑规则与概率图的神经符号推理

近年来，神经符号人工智能（Neuro-symbolic AI）致力于将符号推理与连接主义学习相结合，以实现更接近人类认知的AI系统。从知识形式的角度看，逻辑规则和概率图都是符号推理的重要表现形式，它们通过显式的符号化表示来捕获和处理人类的先验知识，只是关注的侧重点有所不同[1][2][3][4]。逻辑规则主要用于表达确定性的推理关系，能够将专家知识和领域规律形式化为严格的推理规则[5]。概率图则主要用于表达不确定性知识，通过图结构描述变量间的概率依赖关系[6]。

在符号推理AI模型架构方面，研究探索了多种将知识形式与深度学习相结合的方法[7][8][9]。同时，

这种知识形式与模型架构的结合也带来了新的挑战。2024年Marconato等人[10]指出了“推理捷径”问题，即模型可能绕过预期的推理路径，利用非预期的特征达到目标。这一发现提醒我们在设计神经符号系统时，需要同时关注知识表示的语义完整性和模型推理的可控性。

逻辑规则与概率图的融合为神经符号AI注入了强大的推理能力，通过与深度学习的结合，在可解释性和性能上都取得了显著进展。随着理论框架的完善和技术的不断创新，这一方向将进一步拓展AI系统的认知边界。

(2) 图结构先验知识的融入：从图神经网络到知识图谱

图结构作为一种基础的知识表达形式，能够自然地描述实体间的关系与交互。从知识表示的角度看，图结构通过节点和边的形式提供了关系推理的基本框架。图神经网络和知识图谱作为图结构计算的两个主要范式，体现了先验知识融入的不同层次。图神经网络专注于从数据中学习结构化表示，通过网络架构本身体现关系推理的偏置[11]；而知识图谱则致力于构建显式的知识表达，通过实体关系的形式化描述来整合领域知识[12][13]。这两种范式在AI系统融入结构化先验知识方面具有一定互补性。

在图神经网络层面，关系归纳偏置为模型提供了结构化学习能力。几何图神经网络进一步强化了对物理规律的编码[14]。在知识图谱层面，图结构的表达从基础的物理几何约束提升到了高层的语义关联，并在多个领域展现出独特价值[15][16][17]。

这两种计算范式的发展揭示了图结构先验知识融入的基本规律：从隐式的结构化学习到显式的知识表达，从物理约束到语义关联，构建了AI系统获取和利用结构化知识的技术链条。

（3）数学定律与物理规律作为先验知识的融入：数理公式的形式化表达

数学定律和物理规律是人类认知自然界最严谨的先验知识。通过数理公式的形式化表达，这些知识可以精确地描述系统行为和约束条件，为AI模型提供了坚实的理论基础。近年来，如何将这种严格的数理知识有效地融入深度学习模型，成为了提升AI系统可解释性和泛化能力的重要研究方向。基于物理第一性原理的先验知识虽然具有可解释性和样本效率，但在实际应用中常常需要强假设和复杂的数值计算[18][19][20]。相比之下，纯数据驱动的深度学习虽然提供了高效的建模方案，却可能违背基本的物理规律。这一认识推动了物理引导的深度学习框架的发展，其核心是将形式化的科学知识融入数据驱动方法。

在知识融入的模型架构方面，物理引导神经网络展现出独特优势。该方法通过优化网络架构和损失函数，将物理约束引入模型训练过程[21][22][23]。作为一种重要的先验知识形式，数理公式的融入正在推动新一代AI模型的发展。这些模型既保持了科学规律的严谨性，又具备数据驱动适应能力。

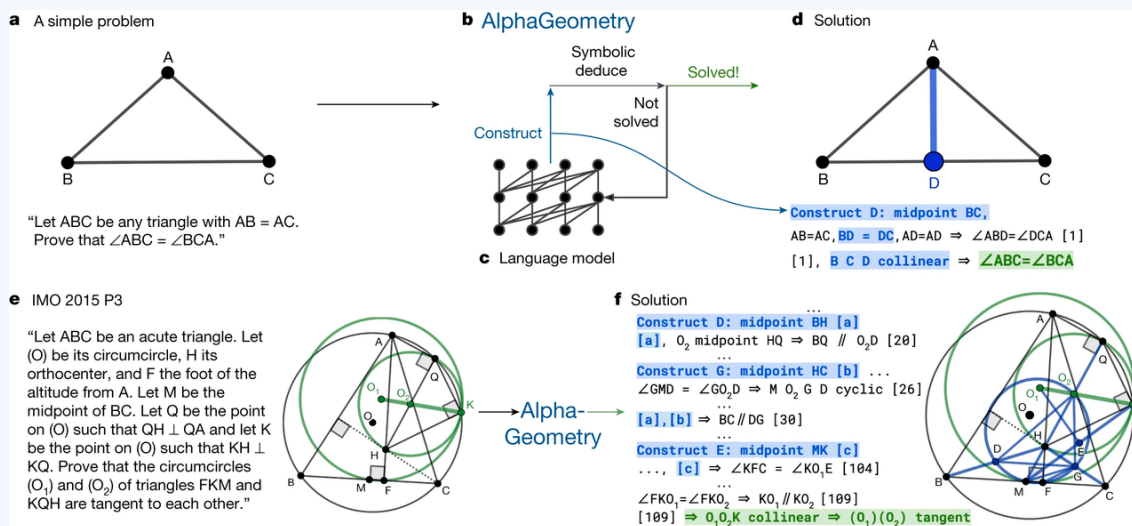
2.3.2 研究进展

2.3.2.1 神经符号人工智能

推荐理由：NSAI是一种新兴的AI范式，它整合了神经网络拓扑和符号推理的方法，来提高人工智能系统的准确性、可解释性和精度，并实现从更少的数据中学习。

NSAI模型中神经网络能从数据中提取复杂特征；符号推理模块能够结合物理世界的既定模型，增强可解释性并减少对大量训练数据的依赖。神经网络和符号推理分别对应于丹尼尔·卡尼曼（Daniel Kahneman）所说的系统1的快速思考模式和系统2的推理模式，这两者的结合是AI模型发展的一场重要革命。

具有代表性的AlphaGeometry[24]是2024年DeepMind团队在Nature文章中提出的一个神经符号系统，它由神经语言模型和符号推演引擎组成，用来寻找复杂几何定理的证明。在训练中，它使用了综合不同复杂程度的数百万个定理和证明的合成数据，通过引导符号推演引擎解决具有挑战性的问题以产生无限分支点，从而扩增数据。在包含30道最新奥林匹克级问题的测试集中，AlphaGeometry解决了25道问题，超越了之前仅解决10道问题的最佳方法，接近国际数学奥林匹克（IMO）金牌得主的平均表现。值得注意的是AlphaGeometry产生了人类可读的证明。后来模型更新到AlphaProof、AlphaGeometry2版本，并在2024 IMO大赛上进行了实战，达到了银牌得主的水平。



神经符号AlphaGeometry的概述，以及解决简单问题和IMO 2015年的第3个问题的案例 | 来源：
 Trinh, Trieu H., et al. "Solving olympiad geometry without human demonstrations." Nature
 625.7995 (2024): 476-482.

除了AlphaGeometry，从2017年到现在，人们已经提出超过 15 种的NSAI模型，包括IBM的神经向量符号架构（NVSA）[25]能够使用神经网络作为前端进行感知和语义解析；神经概率软逻辑（NeuPSL）[9]能够将高层次符号推理与深度神经网络的低级感知相结合；逻辑张量网络（LTNs）[5]可以使用逻辑公式来定义张量表示上的约束等模型；逻辑信息传递（LogicMP）[26]使用平均场变分推断，将一阶逻辑约束融合到任意深度学习网络中实现高效准确的模型训练与推理。

2.3.2.2 几何图神经网络

推荐理由：现有的图神经网络（Graph Neural Network，简称GNN）模型仅考虑研究对象的拓扑信息，而忽略了其空间几何结构特征。Geo GNN模型专注于处理和分析包含几何信息的数据，如图像、三维点云和分子结构等，具有良好的可解释性、泛化性和通用性。

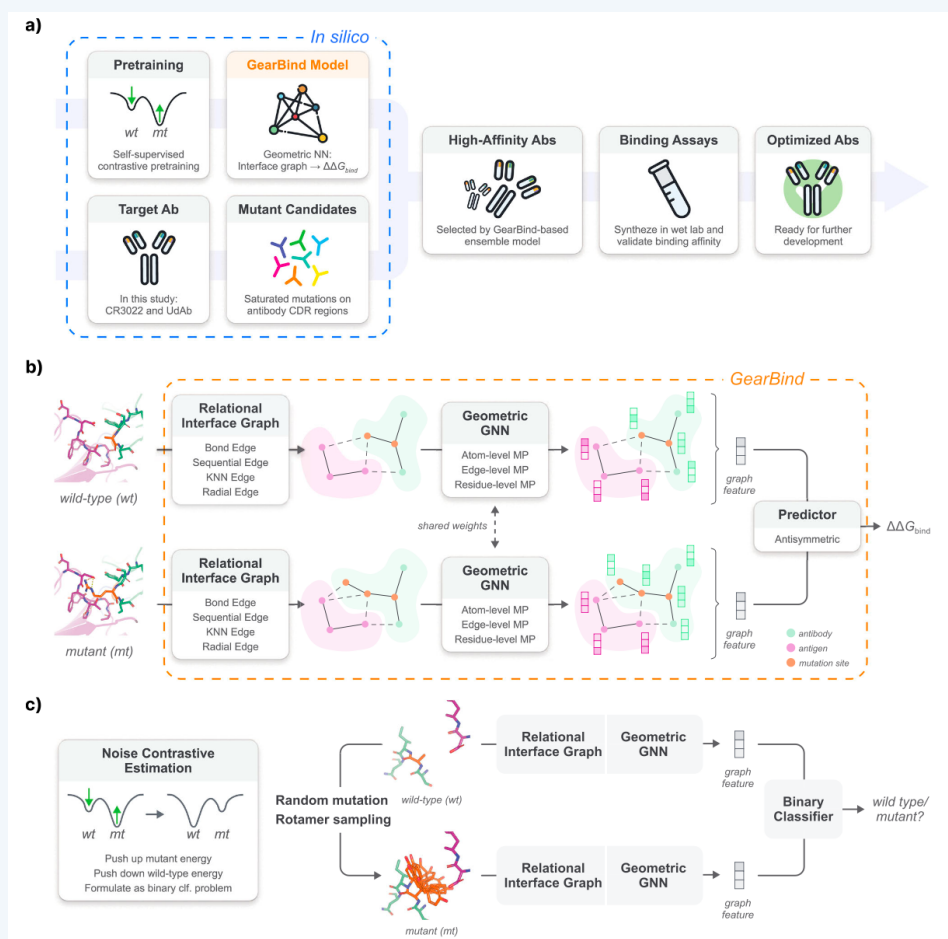
几何图神经网络在对例如分子、蛋白质、晶体、宏观世界中的物体等进行建模的时候，会根据其特定的空间几何结构，建模为几何图。其中几何图需要满足物理对称性的约束，这种几何约束便可以通过几何图神经网络嵌入到机器学习的流程中，从而更好地刻画数据的内在结构和变化规律，提高算法能力。例如在分子动力学模拟中，一个原子到另一个原子的空间方向信息在分子的经验势中发挥核心作用，而Geo GNN通过在信息传递过程中融入此类空间信息，可有效提升模型的预测能力。

在最新研究中，复旦大学和百奥几何联合推出了一种可预训练的几何图神经网络Gearbind[27]，该模型在图构建阶段，通过键边、顺序边、最近邻边、径向边等多种关系对先验知识进行嵌入；在特征提取阶段，在原子层、边层、残基层多层级进行了信息传递，来全面了解蛋白质界面上的复杂相互作用；在训练阶段，他们提出了对比预训练算法，利用CATH中

丰富的未标记单链蛋白质结构，将侧链扭转角知识提取到模型中，进一步提升模型性能。通过多关系图构建、多层次几何信息传递和大规模未标记蛋白质结构数据的对比预训练，模型提高了抗体与其靶标抗原的结合亲和力，并成功增强了两种不同格式和靶标的抗体结合能力，显著提升了抗体治疗性开发中的亲和力成熟化。

性、因果性以及拓扑等各种与结构有关的先验知识转化为图结构，从而融入到基于神经网络的机器学习框架中，并已经在AI for Science领域取得了显著进展。例如，结合物理化学约束的图神经网络模型PSICHIC[28]，与化学语言模型结合的模式DRAGONFLY[29]，等变图扩散模型Chroma[30]，晶体材料稳定性预测模型GNoME[31]，几何增强表示学习方法 GEM[32]等。

除Geo GNN外，GNN及其一些变体可以将相关



基于Gearbind的计算机抗体亲和力成熟化的流程 | 来源: Cai, Huiyu, et al. "Pretrainable geometric graph neural network for antibody affinity maturation." Nature communications 15.1 (2024): 7785.

2.3.2.3 物理启发的图神经网络

推荐理由：物理启发的图神经网络（Physics-Informed Graph Neural Networks）通过显式建模粒子或网格之间的物理相互作用关系，既保持了神经网络强大的学习能力，又融入了先验物理规律约束，在提升仿真精度和效率方面取得了突破性进展。

在科学仿真过程中，我们首先关心的问题是精确模拟。基于第一性原理求解偏微分方程的方法虽然可解释性高，但是需要已知具体的偏微分方程及其参数，而且运算速度缓慢，难以大规模求解，更重要的是难以结合大量的微观数据。

而物理启发的图神经网络技术可以弥补这些不足，这里我们以一系列基于图神经网络（Graph Neural Network, GNN）[33]的工作为例，展开论述GNN在科学仿真领域的进展。首先是DeepMind提出了一种图网络模拟器（Graph Network Simulator, GNS）[34]，模拟粒子物理系统，可以应用于流体力学或者计算图形学等领域。具体架构如下图所示。首先，研究人员可以根据先验信息建立一个图，然后基于消息传递（Message Passing）机制更新

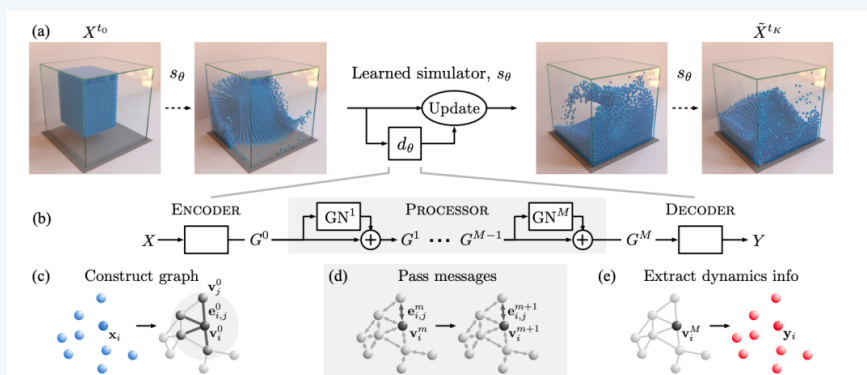
节点的特征，这里的“消息”可以代表粒子之间的相互作用关系，通过神经网络的映射，实现粒子状态的更新，最终实现模拟粒子通过相互作用发生运动的动力学过程。

鉴于GNS的成功，研究人员进一步发展了这一方法。混合图神经网络（HGNS）[35]通过引入多步预测机制显著提升了长期预测的准确性，成功应用于百万级网格规模的地下流体仿真，相比传统求解器实现了18倍的加速。最新的GraphCast模型[36]更是将这一思路扩展到全球天气预报应用上，通过巧妙设计的多尺度图结构，模型能够同时捕捉从局部（几十公里）到全球（几千公里）尺度的大气动力学，将天气预报的计算时间从小时级降低到分钟级。此外，一系列的后续工作还进一步将物理先验知识，例如对称性[37]、方程先验信息[38]等嵌入神经网络架构中，实现更高的准确度和数据应用效率。

2.3.3 挑战与展望

数据与知识

新AI模型的设计，本质是一个跨学科的科学问题。一方面需要对传统领域知识有深刻的理解，另一



使用图神经网络模拟复杂系统。| 图片来源: Sanchez-Gonzalez et al. Learning to Simulate Complex Physics with Graph Network. ICML 2020.

方面需要对AI框架和能力有系统性的认识。将领域知识融入到AI模型中，可以实现更有效率的学习和推理。

系统化整合

数据驱动和知识驱动结合的尝试仍处在初步阶段，如何通过系统化的方式，将不同类型的知识、嵌入技术进行整合仍然是一个根本性且未解决的挑战。构建统一的算法设计框架，可以将神经网络、符号推理、概率图、人类反馈等组件进行系统性整合。

底层软件框架的支持

新的AI系统通常使用底层逻辑（如模糊逻辑、参数化和可微分结构）来支持学习和推理能力。然而，当前主流AI系统多针对特定逻辑开发定制化推理工具，导致模块化架构设计与系统可扩展性受限。因此，亟需解决能够包含广泛推理逻辑功能的语法和语义扩展，同时具有快速且节省内存特性的新软件框架。此外，还需要新的编程模型、编译器和运行时，来简化和高效实现新的AI模型，并开放其全部潜力。

2.3.4 参考文献

[1] Bhuyan, Bikram Pratim, et al. "Neuro-symbolic artificial intelligence: a survey." *Neural Computing and Applications* (2024): 1-36.

<https://dl.acm.org/doi/10.1007/s00521-024-09960-z>

推荐理由：这篇文章为神经符号人工智能的综述，回顾了过去二十年文献，展示了如何将符号推理与连接主义学习相结合，以构建更具人类思维能力的AI系统。

[2] Dickens, Charles. A Unifying Mathematical Framework for Neural-Symbolic Systems. Diss. University of California, Santa Cruz, 2024.

<https://www.proquest.com/openview/9cc4f4c12499-fa9ce39f4f7890d97608/1?pq-origsite=gscholar&cbl=18750&diss>

=y

推荐理由：这篇文章提出了一个统一的数学框架，用于神经-符号系统的研究，探讨如何实现神经网络与符号方法的协同。作者定义了实现有效NeSy AI的四个里程碑，并介绍了神经符号能量基础模型（NeSy-EBMs）。

[3] Wan, Zishen, et al. "Towards cognitive ai systems: a survey and prospective on neuro-symbolic ai." *arXiv preprint arXiv:2401.01040* (2024).

<https://arxiv.org/abs/2401.01040>

推荐理由：这篇文章探讨了朝向认知AI系统的进展，强调神经符号AI在提升AI系统可解释性、鲁棒性和信任度方面的潜力。文章系统地回顾了NSAI的发展，分析了其性能特征和计算操作，提出改进的必要性，并探讨未来的挑战 and 方向。NSAI为人机协作提供了新的视角，展现了在减少数据需求下的学习能力，是发展下一代AI系统的重要途径。

[4] Marra, Giuseppe, et al. "From statistical relational to neurosymbolic artificial intelligence: A survey." *Artificial Intelligence* (2024): 104062.

<https://www.sciencedirect.com/science/article/pii/S0004370223002084>

推荐理由：这篇文章通过将神经符号人工智能（NeSy）与统计关系人工智能（StarAI）进行比较，探讨了学习与推理融合的关键挑战。

[5] Badreddine, S., Garcez, A. d., Serafini, L., and Spranger, M. Logic tensor networks. *Artificial Intelligence*, 303:103649, 2022.

<https://www.sciencedirect.com/science/article/abs/pii/S0004370221002009>

推荐理由：这篇文章提出了逻辑张量网络（LTN）一种神经符号框架，支持利用丰富的数据和关于世界的抽象知识进行查询、学习和推理。LTN提供了一种统一的语言有效地表示和完成了AI的任务，包括多标签分类、关系学习、数据分类、半监督学习、回归、嵌入学习和查询应答等。

[6] Wohlgemut, Jared M., et al. "Bayesian networks may allow better performance and usability than logistic regression." *Critical Care* 28.1 (2024): 234.

<https://link.springer.com/article/10.1186/s13054-024-05015-w>

推荐理由：这篇文章展示了在医疗预测任务中，基于贝叶斯网络的方法相比传统的逻辑回归模型具有更好的性能和可用性，特别是在处理复杂的因果关系时。

[7] Dickens, Charles, et al. "A mathematical framework, a taxonomy of modeling paradigms, and a suite of learning techniques for neural-symbolic systems." arXiv preprint arXiv:2407.09693 (2024).

<https://arxiv.org/abs/2407.09693>

推荐理由：本文提出了一种统一的数学框架神经符号能量基础模型 (NeSy-EBMs)，以促进神经符号系统 (NeSy) 的发展。作者构建了一个模型范式的分类法，聚焦于神经符号接口和推理能力，填补了 NeSy 系统中存在的理论空白。

[8] Duckham, Matt, et al. "Qualitative spatial reasoning with uncertain evidence using Markov logic networks." *International Journal of Geographical Information Science* 37.9 (2023): 2067-2100.

<https://www.tandfonline.com/doi/full/10.1080/13658816.2023.2231044>

推荐理由：这篇文章探索通过结合已有的定性空间演算与马尔可夫逻辑网络构建具有不确定性证据的定性空间推理。文章提出了一种方法论，将不确定知识库与确定的空间推理规则库相结合，旨在自动化复杂空间场景的表示和推理。

[9] Pryor, Connor, et al. "NeuPSL: neural probabilistic soft logic." *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. 2023.

<https://dl.acm.org/doi/abs/10.24963/ijcai.2023/461>

推荐理由：这篇文章介绍了神经概率软逻辑 (NeuPSL) 将先进的符号推理和深度神经网络的低级感知结合起来，比独立神经网络模型实现了30%以上的改进。

[10] Marconato, Emanuele, et al. "Not all neuro-symbolic concepts are created equal: Analysis and mitigation of reasoning shortcuts." *Advances in Neural Information Processing Systems* 36 (2024).

https://proceedings.neurips.cc/paper_files/paper/2023/hash/e560202b6e779a82478edb46c6f8f4dd-Abstract-Conference.html

推荐理由：这篇文章分析了神经符号预测模型中的推理捷径问题，即模型虽然高准确率，但可能是基于具有“意外语义”的概念，从而未能实现预期的优势。通过对推理捷径的系统性特征化，作者确定了其产生的四个关键条件，并提出了多种针对性的缓解策略。这项研究为改进 NeSy 模型的可信性提供了重要的见解，强调了解决推理捷径的重要性。

[11] Ye, Zi, et al. "A comprehensive survey of graph neural networks for knowledge graphs." *IEEE Access* 10 (2022): 75729-75741.

<https://ieeexplore.ieee.org/abstract/document/9831453>

推荐理由：这篇综述文章深入探讨了图神经网络 (GNN) 在知识图谱 (KG) 中的应用，展示了其在解决多个关键任务的潜力。知识图谱作为一种多关系图，能够有效表达各种实体之间的丰富事实信息，然而目前的知识图谱仍面临一些挑战，这些问题近年来已成为研究的热点。

[12] Rajabi, Enayat, and Kobra Etmnani. "Knowledge-graph-based explainable AI: A systematic review." *Journal of Information Science* 50.4 (2024): 1019-1029.

<https://journals.sagepub.com/doi/full/10.1177/01655515221112844>

推荐理由：这篇文章系统性回顾了知识图谱在可解释人工智能 (XAI) 中的应用，着重分析了知识图谱如何通过层次结构表示实体、属性及其关系，从而增强 AI 系统的可解释性和可理解性。

[13] Peng, Ciyuan, et al. "Knowledge graphs: Opportunities and challenges." *Artificial Intelligence Review* 56.11 (2023): 13071-13102.

<https://link.springer.com/article/10.1007/s10462-023-10465-9>

推荐理由：这篇文章对知识图谱领域进行了系统性的综述，重点讨论了知识图谱的机遇与挑战。作者指出，随着人工智能和大数据的迅猛发展，知识图谱在有效表示复杂信息方面的重要性日益突出。

[14] Han, Jiaqi, et al. "A Survey of Geometric Graph Neural Networks: Data Structures, Models and Applications." arXiv preprint arXiv:2403.00485 (2024).

<https://arxiv.org/abs/2403.00485>

推荐理由：这篇文章全面综述了几何图神经网络 (Geometric GNN) 的数据结构、模型和应用，强调了几何图在科学问题建模中的重要性。

[15] Zhong, Zhiqiang, and Davide Mottin. "Knowledge-augmented Graph Machine Learning for Drug Discovery: From Precision to Interpretability." *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023.

<https://dl.acm.org/doi/abs/10.1145/3580305.3599563>

推荐理由：这篇文章介绍了知识增强图机器学习 (KaGML) 的概念，专注于其在药物发现中的应用。作者指出传统人工智能模型在处理复杂生物医学结构时的局限，并提出通过整合外部生物医学知识来提升准确性和可解释性。

[16] Chen, Zhuo, et al. "Knowledge graphs meet multi-modal learning: A comprehensive survey." arXiv preprint arXiv:2402.05391 (2024).

<https://arxiv.org/abs/2402.05391>

推荐理由：这篇综述文章系统回顾了知识图谱与多模态学习的结合，分析了知识图谱在多模态任务中的支持作用及多模态知识图谱的

构建进展。

[17] Pan, Shirui, et al. "Unifying large language models and knowledge graphs: A roadmap." IEEE Transactions on Knowledge and Data Engineering (2024).

<https://ieeexplore.ieee.org/abstract/document/10387715>

推荐理由：这篇文章提出了将大型语言模型（LLMs）与知识图谱（KGs）结合的框架，强调两者在推理与可解释性方面的互补性。作者定义了三个框架：KG增强LLMs、LLM增强KGs以及协同作用的LLMs与KGs，展示了它们如何在不同任务中协同工作。

[18] Yu, Rose, and Rui Wang. "Learning dynamical systems from data: An introduction to physics-guided deep learning." Proceedings of the National Academy of Sciences 121.27 (2024): e2311808121.

<https://www.pnas.org/doi/abs/10.1073/pnas.2311808121>

推荐理由：这篇文章介绍了物理引导深度学习的框架，强调将物理知识融入数据驱动方法的重要性，尤其是在建模复杂动态系统方面。

[19] Faroughi, Salah A., et al. "Physics-guided, physics-informed, and physics-encoded neural networks in scientific computing." arXiv preprint arXiv:2211.07377 (2022).

<https://arxiv.org/abs/2211.07377>

推荐理由：这篇综述文章详细探讨了物理引导、物理信息驱动和物理编码的神经网络在科学计算中的应用。作者论述了这些网络如何在面临稀疏数据的科学和工程领域仍能有效地结合物理约束，提升多物理现象的数值建模能力。

[20] Han, Jiequn, and Linfeng Zhang. "Integrating machine learning with physics-based modeling." arXiv preprint arXiv:2006.02619 (2020).

<https://arxiv.org/abs/2006.02619>

推荐理由：这篇文章探讨了如何将机器学习有效结合物理基础建模，以开发可解释且可靠的物理模型。作者重点分析了施加物理约束和获取最优数据集的关键问题，并以分子动力学和动量闭合方程为案例说明相关概念。

[21] Guo, Fangzhou, et al. "Improving the out-of-sample generalization ability of data-driven chiller performance models using physics-guided neural network." Applied Energy 354 (2024): 122190.

<https://www.sciencedirect.com/science/article/abs/pii/S0306261923015544>

推荐理由：这篇文章提出了一种物理引导神经网络模型，以改善冷却器性能模型的外推能力，解决了传统数据驱动模型在有限操作条件下的通用性问题。作者通过新的网络架构和修改损失函数，将领域

知识融入数据驱动模型，从而显著提升模型的外插性能。

[22] Bolderman, Max, et al. "Physics-guided neural networks for feedforward control with input-to-state-stability guarantees." Control Engineering Practice 145 (2024): 105851.

<https://www.sciencedirect.com/science/article/pii/S096706612400011X>

推荐理由：这篇文章介绍了一种新型物理引导神经网络架构，旨在提升高精度机电控制的准确性。通过将物理基础层与黑箱神经层结合，该模型能在训练中保证输入状态稳定性，并显著改善现有物理基础前馈控制的性能。

[23] Kochkov, Dmitrii, et al. "Neural general circulation models for weather and climate." Nature 632.8027 (2024): 1060-1066.

<https://www.nature.com/articles/s41586-024-07744-y>

推荐理由：这篇文章提出了结合机器学习组件与可微分求解器的新型天气与气候模型（NeuralGCM）。该模型能在天气预报和气候模拟中实现与传统物理模型竞争的表现，尤其在多天到多周的短期气候预测上具有优势。

[24] Trinh, Trieu H., et al. "Solving olympiad geometry without human demonstrations." Nature 625.7995 (2024): 476-482.

<https://www.nature.com/articles/s41586-023-06747-5>

推荐理由：这篇文章来自谷歌 Deepmind 团队，介绍了欧几里得平面几何的定理证明器 AlphaGeometry，采用神经符号架构设计，在大规模合成数据上训练，结合符号推理引擎来证明数学定理。

[25] Hersche, Michael, et al. "A neuro-vector-symbolic architecture for solving Raven's progressive matrices." Nature Machine Intelligence 5.4 (2023): 363-375.

<https://www.nature.com/articles/s42256-023-00630-8>

推荐理由：这篇文章来自 IBM 团队，提出了神经向量符号架构（NVSA），可以利用高维分布式表示运算符作为神经网络和符号人工智能之间的通用语言，来解决神经网络无法分解联合表示来获得不同的对象，以及符号人工智能面临详尽的规则搜索等问题。

[26] Xu, Weidi, et al. "LogicMP: A Neuro-symbolic Approach for Encoding First-order Logic Constraints." arXiv preprint arXiv:2309.15458 (2023). ICLR 2024

<https://iclr.cc/virtual/2024/poster/19220>

推荐理由：通过在马尔科夫逻辑网络实现平均场变分推断，将一阶逻辑约束融合到任意神经网络中，实现高效的模型训练与推理。

[27] Cai, Huiyu, et al. "Pretrainable geometric graph neural network for antibody affinity maturation." Nature communications 15.1 (2024): 7785.

<https://www.nature.com/articles/s41467-024-51563-8>

推荐理由：这篇文章提出一种预训练的几何图神经网络Gearbind，利用多关系图构建、多级几何信息传递和对大规模未标记蛋白质结构数据的对比预训练，在SKEMPI和独立测试集上优于之前最先进的方法。并成功增强了两种具有不同形式和目标抗原的抗体结合。

[28] Koh, Huan Yee, et al. "Physicochemical graph neural network for learning protein-ligand interaction fingerprints from sequence data." *Nature Machine Intelligence* (2024): 1-15.

<https://www.nature.com/articles/s42256-024-00847-1>

推荐理由：这篇文章提出了一个结合物理化学约束的框架PSICHIC，可以直接从序列数据中解码蛋白质-配体的相互作用机制，实现更先进的准确性和可解释性。

[29] Atz, Kenneth, et al. "Prospective de novo drug design with deep interactome learning." *Nature Communications* 15.1 (2024): 3408.

<https://www.nature.com/articles/s41467-024-47613-w>

推荐理由：这篇文章提出的DRAGONFLY模型，它利用了图神经网络和化学语言模型组成的神经网络架构，用于全面研究药物-靶标相互作用组，包括蛋白质-蛋白质相互作用、药物-靶点相互作用和药物-药物关系等。

[30] Ingraham, John B., et al. "Illuminating protein space with a programmable generative model." *Nature* 623.7989 (2023): 1070-1078.

<https://www.nature.com/articles/s41586-023-06728-8>

推荐理由：这篇文章介绍了一种蛋白质和蛋白质复合物的生成模型Chroma，可以直接对新的蛋白质结构和序列进行采样，并且可以进行调节以引导生成过程实现所需的特性和功能。通过这种统一的蛋白质设计方法，来加速蛋白质物质的变成，来造福人类健康、材料科学和合成生物学。

[31] Merchant, Amil, et al. "Scaling deep learning for materials discovery." *Nature* 624.7990 (2023): 80-85.

<https://www.nature.com/articles/s41586-023-06735-9>

推荐理由：这篇文章基于图神经网络构建了一种晶体材料稳定性预测模型GNoME，发现了多达220万种理论上稳定的新型材料。

[32] Fang, Xiaomin, et al. "Geometry-enhanced molecular representation learning for property prediction." *Nature Machine Intelligence* 4.2 (2022): 127-134.

<https://www.nature.com/articles/s42256-021-00438-4>

推荐理由：这篇文章提出了一种新颖的几何增强分子表示学习方法GEM，基于几何的图神经网络架构和专用的几何级自监督学习策略来学习分子几何知识。

[33] Sanchez-Gonzalez, Alvaro, et al. "Learning to Simulate Complex Physics with Graph Networks." *International Conference on Machine Learning*, 2020.

<https://arxiv.org/abs/2002.09405>

推荐理由：这篇论文提出了基于图神经网络的物理系统仿真框架，是该领域的开创性工作。

[34] Rubanova, Yulia, et al. "Constraint-based Graph Network Simulator." *arXiv preprint arXiv:2112.09161*, 2021.

<https://arxiv.org/pdf/2112.09161.pdf>

推荐理由：该工作提出的约束图网络模拟器为复杂物理系统仿真提供了新思路。

[35] Wu, Tailin, et al. "Learning large-scale subsurface simulations with a hybrid graph network simulator." *proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2022.

<https://dl.acm.org/doi/abs/10.1145/3534678.3539045>

推荐理由：该论文提出混合图神经网络，首次将图神经网络用于3D百万网格级别的地下流体仿真，并在实际生产中部署。

[36] Lam, Remi, et al. "GraphCast: Learning Skillful Medium-range Global Weather Forecasting." *arXiv preprint arXiv:2212.12794*, 2022.

<https://arxiv.org/abs/2212.12794>

推荐理由：该工作将图神经网络应用于全球天气预报，显著提升了预报效率。

[37] Satorras, Victor Garcia, et al. "E(n) Equivariant Graph Neural Networks." *International Conference on Machine Learning*, 2021.

<https://arxiv.org/abs/2102.09844>

推荐理由：该论文将对称性约束引入图神经网络，提高了物理仿真的准确性。

[38] Li, Zongyi, et al. "Physics-informed Neural Operator for Learning Partial Differential Equations." *ACM/JMS Journal of Data Science*, vol. 1, no. 3, 2024, pp. 1-27.

<https://dl.acm.org/doi/10.1145/3590262>

推荐理由：该工作将物理知识融入神经算子，提高了偏微分方程求解的效率。

2.4 AI科学家

2.4.1 背景介绍

AI科学家（AI Scientist）是一种能够自主或半自主地执行科学研究活动的人工智能系统。这类系统试图模拟人类科学家的研究方法和思维过程，能够参与从研究构思到实验验证再到成果输出的完整科研周期。随着数据、算力驱动的AI技术快速发展，AI科学家系统经历了显著演进，可以分为两个主要阶段：

第一阶段：基于数据驱动自动发现期（2020年前）

最初的AI科学家系统主要依赖于符号回归和数据驱动方法进行科学发现，采用RNN模型、图网络模型、神经微分方程等，聚焦于从观测数据中自动提取科学规律和定律，而无需深入的领域先验知识。代表性工作包括：

- AI Feynman系统[1]：通过提高符号回归的噪声鲁棒性与方程识别精度，展示了在复杂数据中发现科学公式的能力。该系统采用Pareto优化的符号回归方法，利用图模块性来提升性能。

- AI Poincare系统[2]：专注于在动态数据中发现守恒定律。该系统能够从未知动力学系统的轨迹数据中自动识别和提取守恒量，为物理定律的发现提供了新的途径。

- AI Physicist系统[3]：通过模仿物理学家的四种关键策略——分而治之、奥卡姆剃刀、统一化和终身学习，实现了对物理规律的无监督学习。

第二阶段：基于智能体的全流程自动化探索期（2020-2024）

随着大语言模型出现，AI科学家系统开始成为自动科研探索的AI智能体（Agent）：不再局限于单一数据分析任务，而是尝试参与从初始构思到实验验证再到论文撰写的完整科研周期。

通过对LLM在科学文献上的预训练和针对性微调，以及思维链（CoT）、检索增强生成（RAG）等技术应用，使AI科学家能够自主感知和决策，模拟人类科学家的推理过程，并通过工具扩展进行各种科研操作。再加上与传统的复杂系统建模方法深度融合，使得AI科学家具备全流程的科研能力。

其中The AI Scientist[4]和Coscientist[8] 是两篇比较重要的工作，前者实现了「规划-执行-反思」范式的Agent框架，后者则实现了从化学实验设计到硬件控制的深度集成，以及对AI科学家的优化改进，例如西湖大学开发的Nova系统[9]通过创新的技术框架将创意生成效率提升了2.5倍。下面将重点介绍这几项工作。

2.4.2 研究进展

2.4.2.1 AI科学家与自主科研智能体

推荐理由：The AI Scientist是第一个完全自动化的科学发现系统，能够独立完成从研究构想到实验设计、结果分析、论文撰写到模拟评审的全过程[4]。

The AI Scientist能够从广泛的研究方向和一个简单的初始代码库开始，无缝地执行构思、文献搜

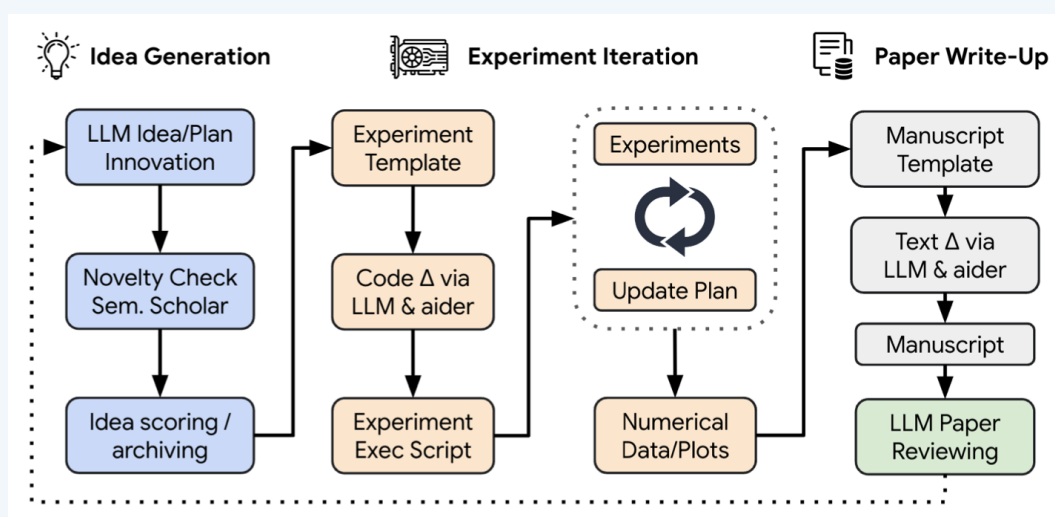
索、实验规划、实验迭代、撰写论文和同行评审，以产生有洞察力的论文。The AI Scientist自动化了整个研究生命周期，包括生成新的研究想法、编写必要的代码、执行实验、总结实验结果、可视化结果，并在完整的科学论文中展示其发现。

The AI Scientist框架能够以一个开放的循环的方式运行，这意味着它可以不断地迭代和改进。在每次循环中，系统都会基于前一次的研究成果来生成新的想法和假设，然后进行实验验证和论文撰写。这个过程可以无限重复，每次迭代都会在前一次的基础上进行改进。通过自动化的同行评审和反馈机制，The AI Scientist还能够识别和改进其研究中的不足之处。这些评审和反馈会被用来指导下一代研究想法的生成，使得新的研究能够建立在之前研究的基础上，从而实现持续的改进和创新。

想法的生成、实验的执行、结果的分析、论文的撰写以及同行评审。这种模仿不仅包括科学研究的技术层面，还包括了科学社区中知识共享和反馈的社交层面。通过这种方式，The AI Scientist能够模拟人类科学家的合作和竞争，推动科学发现的进程。

在此基础上，西湖大学蓝振忠团队研发的Nova系统带来了突破性进展[9]。Nova专注于科研创意生成环节，通过创新的技术框架，将创意生成效率提升了2.5倍。当The AI Scientist生成2个创意时，Nova已能产出5个高质量的创新想法。Nova采用迭代规划、外部知识检索、检索增强生成等先进技术，确保生成想法的质量、多样性和新颖性。实验评估显示，Nova生成的想法超过80%都是不重复的，在人工评估中获得了最高的整体质量和新颖性评分。

The AI Scientist框架以开放循环的方式运行，整个框架模仿了人类科学社区的工作方式，包括通过不断迭代和改进来推进科学发现。而Nova则通

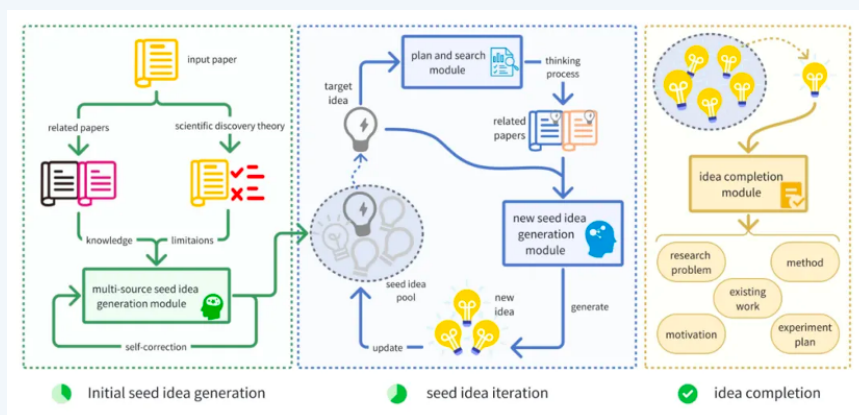


The AI Scientist: 一个LLM驱动的端到端科学发现流程 | 来源: "Lu, C., Lu, C., Lange, R.T., Foerster, J., Clune, J., & Ha, D. (2024). The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery."

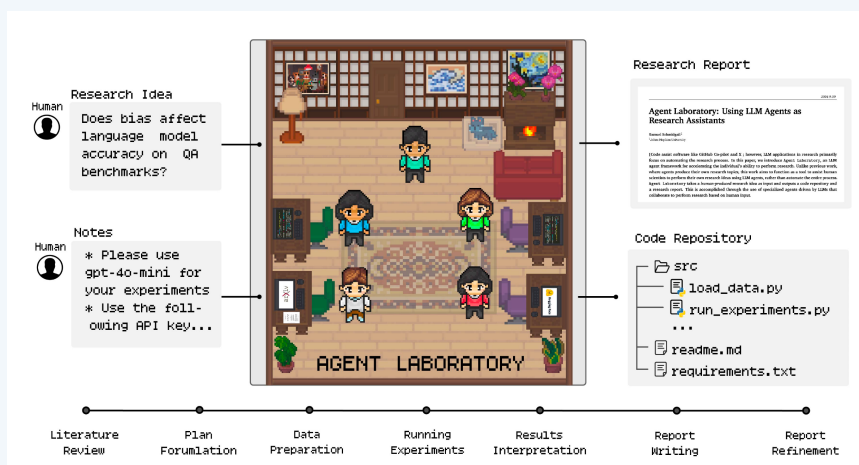
过其独特的技术优势，极大地提升了这一过程中创意生成的效率和质量。Nova的发展规划分为三步：首先在单一领域实现创新并验证效果，然后拓展到多学科领域，最终目标是实现自主的科学探索和商业创新。

此外，AMD推出的由GPT-o1-preview驱动的Agent Laboratory系统[10]，则通过整合生成式AI

与半导体芯片协同设计能力，通过文献综述、实验、撰写报告三个阶段，每一阶段分配都有不同的任务、工具和AI Agent角色，将实验数据处理效率提升40%，平均每年为科研团队节省84%的经费支出。该框架接受人类提供的研究想法，允许用户在每个阶段提供反馈和指导，可以产生全面的研究成果，包括代码库和研究报告，并达到了最先进的性能。



Nova模型：通过迭代的方式制定搜索识别新颖性和多样性的文献,使得模型能够不断优化生成的想法 |来源: "Nova: An Iterative Planning and Search Approach to Enhance Novelty and Diversity of LLM Generated Id."



Agent Laboratory：以人类的研究想法和笔记为输入，分任务和角色喂给专门的研究智能体，最终产出研究报告和代码库 |来源: "Agent Laboratory: Using LLM Agents as Research Assistants"

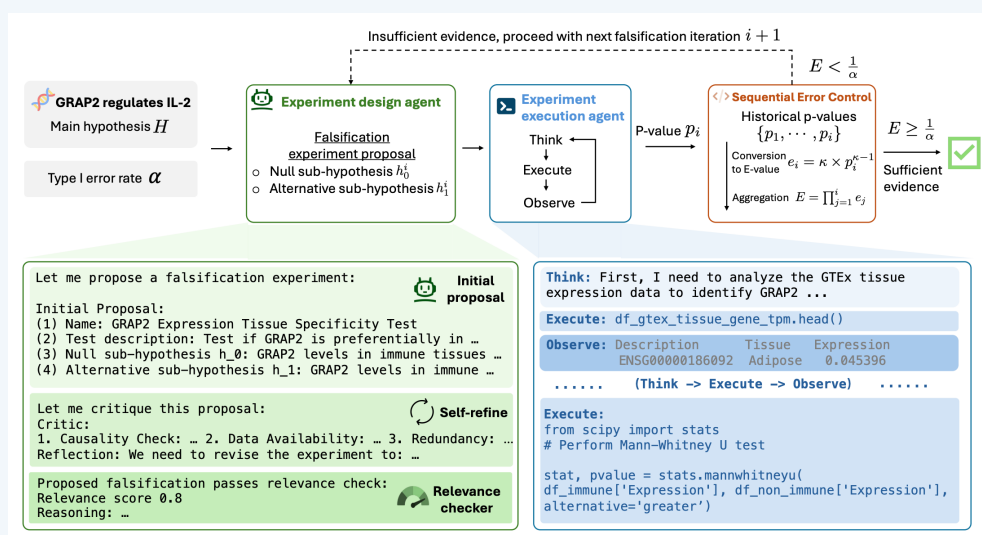
同时，斯坦福与哈佛大学的研究人员提出了一个将严格的统计原则与基于LLM的智能体相结合的自动化假设验证框架POPPER[11]。该框架受卡尔·波普尔（Karl Popper）的证伪原则启发，通过设计和执行针对假设可测量含义的证伪实验来验证假设。POPPER使用两个专门的LLM智能体：实验设计智能体和实验执行智能体，前者负责识别可测量的子假设并设计证伪实验，后者则负责实施实验并生成p值以总结实验结果。此外，POPPER引入了一种新颖的序贯测试框架，通过将多个潜在相关的LLM生成测试的证据聚合在一起，同时严格控制第一类错误率，从而构建动态化决策机制并保障统计严格性，决定是否拒绝假设、进行进一步实验或终止验证过程。

POPPER在六个领域进行了验证，包括生物学、经济学和社会学等，展示了强大的错误控制能力、高功效和可扩展性。与人类科学家相比，POPPER在验证复杂生物假设时表现出了相当的性能，同时将时间缩短了10倍，为假设验证提供了一个可扩展且严格的解决方案。该框架的提出为科学

发现提供了一种新的工具，能够有效地从大量假设中筛选出有价值的假设，减少手动验证的工作量，提高研究效率。

这几个系统的出现标志着AI辅助科研进入新阶段。The AI Scientist提供了完整的科研流程自动化框架，Nova大幅提升了创意生成的效率和质量，Agent Laboratory则大幅度降低了科研成本预算，POPPER可以进行自动化假设验证。另外，Open AI推出的基于GPT-o3模型的Deep Research自主研究智能体[12]，也展现出强大的独立文献检索、数据分析和研究综述工作能力。这些系统的结合将极大推动科研效率的提升，为未来全自动科学探索开辟新的可能。

除此之外，随着AI技术和大模型的迅猛发展，2024年还涌现出多项科学使用LLM辅助科研的研究成果，如[5][6][7]等。这些成果不仅改变了科学研究的方式，提升了科研效率，未来有望在各个领域实现更具规模的全自动科学探索。



POPPER架构概览 | 来源: "Automated Hypothesis Validation with Agentic Sequential Falsifications"

2.4.2.2 科学家助手

推荐理由：本篇论文试图解决如何将大型语言模型（LLMs）应用于自动化化学研究全流程的问题[8]。亮点在于开发了一个名为Coscientist的AI智能体系统，能够自动设计、规划和执行复杂的科学实验，并具有和信息检索、代码管理、自动化实验装置等模块相互交互的能力。

Coscientist展示了与多个模块（如网络搜索、代码执行和文档搜索）交互的能力，从而获取和处理解决复杂化学问题所需的知识。此外，Coscientist在化学合成规划、实验室硬件控制和多变量实验设计中也表现卓越，在优化实验和化学推理能力方面拥有特别的应用潜力。

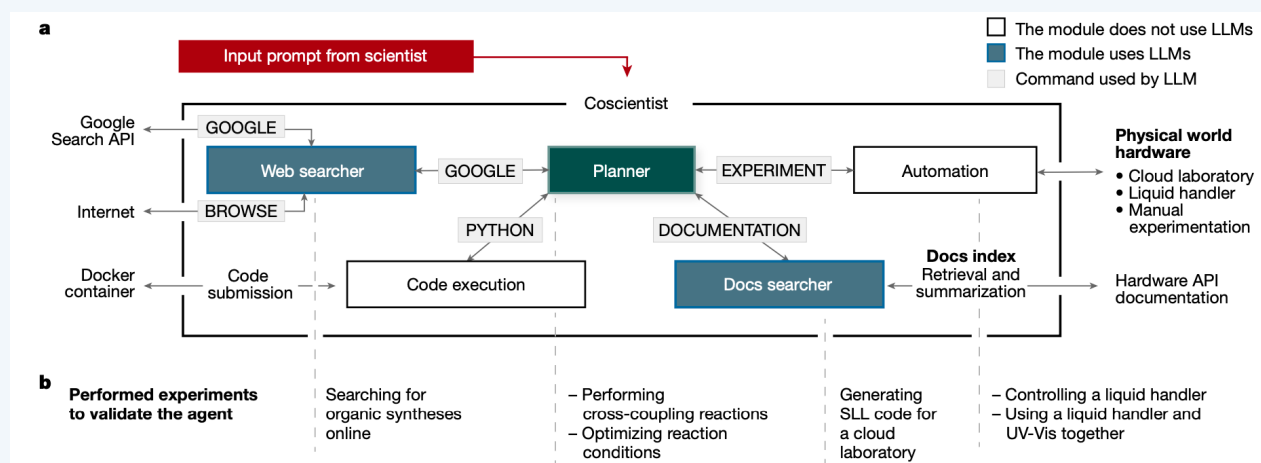
Coscientist在化学合成规划上表现卓越，特别是在采用GPT-4模型时，能够提供详尽且准确的合成步骤。系统还能通过文档搜索模块理解并运用技

术文档，比如Opentrons Python API和Emerald Cloud Lab的SLL，增强了实验的精确度和可靠性。

在硬件控制方面，Coscientist能精准操控液体处理设备，完成复杂的实验任务，显示了其在实际应用中的适应性和灵活性。它还能利用网络数据进行计算，并编写代码以规划和执行催化交叉偶联实验，展现了解决实际化学问题的能力。

Coscientist的化学推理能力同样出色，能够获取信息、处理复杂问题，并设计实验方案，这些都是衡量智能代理在化学领域应用的关键指标。这些成果不仅证实了LLMs在化学研究中的应用潜力，也为自动化和智能化科学实验的未来指明了方向。

Coscientist是在化学领域专用的科学家助手，在通用领域，Google推出的AI Co-scientist则是一个强大的跨学科科研助手系统[13]。



Coscientist系统架构图 | 来源: Boiko, D.A., MacKnight, R., Kline, B. et al. Autonomous chemical research with large language models. Nature 624, 570–578 (2023)

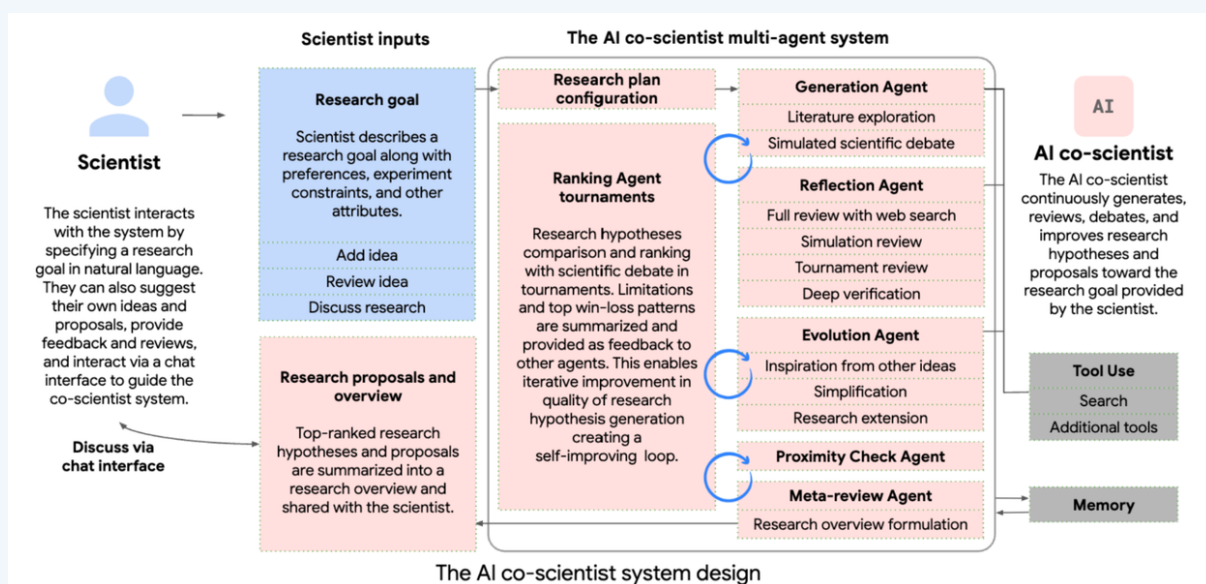
a. 组成Coscientist的多个交互模块。b. 使用单模块或多模块时可组合出不同的实验类型。

AI Co-scientist是基于Gemini 2.0构建的多智能体人工智能系统，定位为科研助手而非替代者，通过人机协作提升研究效率。它采用了六大核心智能体协同工作的架构：

- Generation Agent (生成智能体)：负责初步假设和研究方案的生成
- Reflection Agent (反思智能体)：评估假设和方案，提供改进建议
- Ranking Agent (排序智能体)：对不同假设进行排序和评估
- Evolution Agent (进化智能体)：基于现有假设生成新的研究思路
- Proximity Agent (邻近性智能体)：计算假设间的相似度，避免重复
- Meta-review Agent (元审查智能体)：总结分析所有假设和评估结果

系统采用了两项关键技术创新。**测试时计算扩展**：不同于传统AI模型在训练阶段消耗大量计算资源，Co-scientist在推理阶段投入更多算力，实现更深入的科学推理；**科学辩论机制**：通过多智能体间的“科学辩论”，互相质疑、补充，最终达成高质量的研究假设。

AI Co-scientist一经推出，就带来了亮眼的科研成果。例如，在英国帝国理工学院的研究中，Co-scientist仅用48小时就复现了科研团队耗时10年的发现。研究聚焦于cf-PICls（衣壳形成噬菌体诱导性染色体岛）在细菌耐药性传播中的作用机制。系统不仅独立提出了正确的假设，还额外提出了4个创新性理论方向。还有，在药物再利用研究中，提出了包括Binimetinib、Pacritinib等候选药物，准确预测了药物作用机制，实验验证显示多个预测结果具有显著疗效。



AI co-scientist 多智能体系统中的不同组件以及系统与科学家之间的交互范式的图示 | 来源 Google Research Team. "Accelerating scientific breakthroughs with an AI co-scientist" Google Research Blog, 2024.

Google Co-scientist的出现标志着AI辅助科研进入新阶段、扩展到了更多科学领域，它不仅能够加速科学发现过程，更重要的是实现了一种有效的新型人机协作模式。随着技术持续进步，我们有理由期待它能够帮助科研人员突破更多科学前沿，推动人类知识边界的扩展。

2.4.3 挑战与展望

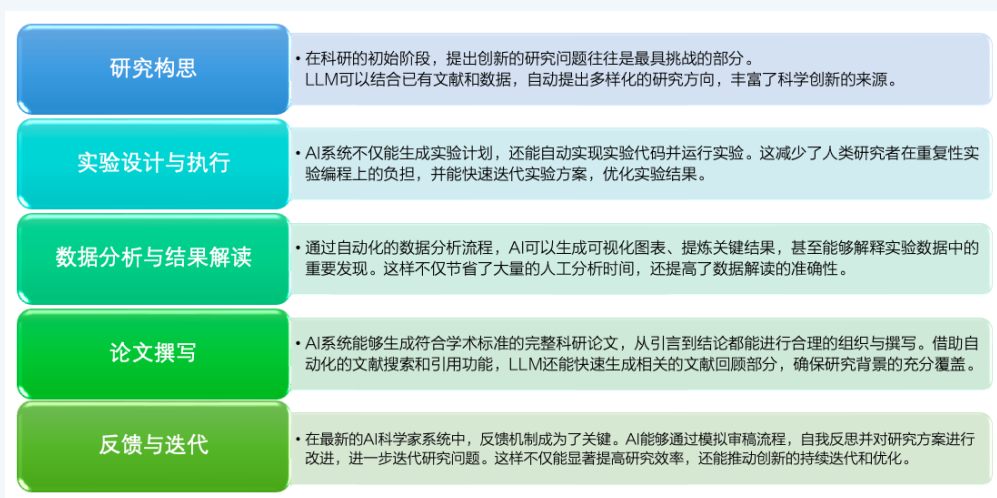
随着AI科学家、Nova以及OpenAI基于强大基础模型的Deep Research[12]等系统和功能的出现，AI科学家的发展展现出令人振奋的前景。一项[14]研究通过招募100多名NLP研究人员进行大规模人类评估实验，首次系统性地比较了人工智能系统与专家研究人员在研究创意生成方面的能力。研究发现，AI生成的研究创意在新颖性方面显著优于人类专家 ($p < 0.05$)，但在可行性方面略逊一筹。

因此我们可能开发更强大的AI科学家系统，从创新能力的跃升到全流程自动化的实现，在多个维度上推动科学研究范式的革新：

具体来说，AI智能体在未来将可以自动提出研究想法，设计实验，生成代码，运行实验，收集结果，并通过自动生成的可视化工具和语言总结实验结果，最终撰写完整的科研论文（见上图总结）。这种全流程能力大大提升了AI系统在科学研究中的自主性，减少了人类科学家在繁琐实验和数据分析上的投入。

虽然AI智能体在科学研究中的应用取得了显著进展，尤其是在仿真和数据驱动模型验证方面，但它们的自主自动化能力仍然存在不少局限性，特别是在涉及与真实世界交互的实验和科学探索领域。这些不足主要体现在以下几个方面：

1. 与真实世界的交互能力有限：当前的LLM智能体主要依赖于仿真数据或现有数据集进行科学推理和实验设计。然而，真实世界中的实验涉及复杂的物理交互，要求系统能够实时感知、操控物理对象、处理环境不确定性。这是现有数据驱动模型所无法完全胜任的。LLM在这类场景中表现较弱，因为它们缺乏与实际物理系统的直接连接。



未来科学AI智能体参与科学研究全流程的设想

2. 硬件支持不足：尽管LLM在软件层面表现出强大的计算能力，但与物理世界的交互需要更加智能化的硬件支持。具身智能（Embodied Intelligence）将成为解决这一问题的关键，它要求AI模型不仅在虚拟环境中进行推理，还要能直接控制物理设备并适应环境变化。现有的LLM代理还未能有效整合物理硬件和AI模型之间的协同工作。

3. 实验自主性的局限：虽然LLM代理可以自动生成代码、设计仿真实验并撰写论文，但其“自主性”还不够强。例如，AI在面对新兴的、未被数据集覆盖的实验问题时，往往难以进行创新的实验设计。这限制了AI在基础科学研究中的探索能力，特别是在需要物理实验的领域（如材料科学、生物技术等），LLM往往依赖于预先设定的框架或模型，而缺乏真正的探索创新。

4. 无法处理真实实验中的复杂变量：真实实验往往涉及许多复杂且难以预测的变量，如环境噪音、实验设备的差异以及意外情况。现有的LLM系统在面对这类不确定性时表现不佳，它们更适合处理理想化或简化的问题模型，而缺乏应对现实世界中随机性和复杂性的能力。

尽管当前基于LLM的AI智能体展示了其在科研自动化中的巨大潜力，特别是在数据驱动的仿真和分析方面，但它们在真实世界实验中的能力仍有待提升。要实现真正的自主科学发现，需要更强的AI模型、更智能的反馈迭代机制以及与硬件系统的深度结合。随着具身智能和更先进的AI技术的融合发展，我们有望在未来看到完全自主的AI科学家系统，不仅能够虚拟世界中推理，还能够物理世界进行真正的科学探索，为人类的科学进步提供前所未有的动力。

2.4.4 参考文献

[1] Udrescu, S.-M., Tan, A., Feng, J., Neto, O., Wu, T., & Tegmark, M. (2020). AI Feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. *NeurIPS* 2020.

<https://www.nature.com/articles/s42254-023-00581-4>

推荐理由：这篇论文介绍了AI Feynman 2.0，这是一个利用图模块化的帕累托最优符号回归算法，旨在从数据中自动发现物理定律。

[2] Liu, Z., & Tegmark, M. (2021). Machine Learning Conservation Laws from Trajectories *Phys. Rev. Lett.* 126, 180604.

<https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.126.180604>

推荐理由：这篇论文介绍了AI Poincaré算法，这是一个机器学习工具，能够从未知动力学系统的轨迹数据中自动发现守恒量，包括在五个哈密顿系统中发现所有精确守恒量以及周期轨道、相变和近似守恒定律的破坏时间尺度。

[3] Wu, T., & Tegmark, M. (2019). Toward an artificial intelligence physicist for unsupervised learning.

<https://journals.aps.org/pre/abstract/10.1103/PhysRevE.100.033311>

推荐理由：这篇论文提出了一个名为“AI Physicist”的无监督学习框架，该框架通过模仿物理学中的四种策略——分而治之、奥卡姆剃刀、统一化和终身学习——来提高机器学习性能。

[4] Lu, C., Lu, C., Lange, R.T., Foerster, J., Clune, J., & Ha, D. (2024). The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery.

<https://arxiv.org/abs/2408.06292>

推荐理由：这篇论文介绍了一个名为“The AI Scientist”的系统，旨在实现完全自动化的开放性科学发现，通过自我迭代的实验设计、执行和结果分析来推动科学进步。

[5] Guo, S., Deng, C., Wen, Y., Chen, H., Chang, Y., & Wang, J. (2024). DS-Agent: Automated Data Science by Empowering Large Language Models with Case-Based Reasoning. *Proceedings of the 41st International Conference on Machine Learning (ICML)*.

<https://arxiv.org/abs/2402.17453>

推荐理由：这篇论文介绍了一个名为“DS-Agent”的自动化数据科学系统，通过将基于案例的推理赋予大型语言模型，以实现完全自动化的开放性科学发现。

[6] Ma, Y., Gou, Z., Hao, J., Xu, R., Wang, S., Pan, L., Yang, V., Cao, Y., Sun, A., Awadalla, H., & Chen, W. (2024). SCIAGENT: Tool-augmented Language Models for Scientific Reasoning. Retrieved from arXiv:2402.11451v2.

<https://arxiv.org/abs/2402.11451>

推荐理由：这篇论文介绍了一个名为“SCIAGENT”工具增强型的大型语言模型，旨在通过结合特定领域的工具集来提升科学推理能力，使模型能够在多个科学领域中进行自动化的科学发现和问题解决。

[7] Chen, H., Shen, X., Ye, Z., Feng, W., Wang, H., Yang, X., Yang, X., Liu, W., & Bian, J. (2024). Towards Data-Centric Automatic R&D. Retrieved from arXiv:2404.11276v2.

<https://arxiv.org/abs/2404.11276>

推荐理由：这篇论文探讨了如何实现以数据为中心的自动化研发，旨在通过自动化的方法来加速科学研究和开发过程。

[8] Boiko, D.A., MacKnight, R., Kline, B. et al. Autonomous chemical research with large language models. Nature 624, 570–578 (2023).

<https://www.nature.com/articles/s41586-023-06792-0>

推荐理由：这篇论文介绍了一个由GPT-4驱动的人工智能系统，它能够自主设计、规划和执行复杂的化学实验，展示了在内的六项不同任务中的研究加速潜力，体现了大型语言模型在科学研究中的多功能性、有效性和可解释性。

[9] Hu, Xiang, et al. "Nova: An iterative planning and search approach to enhance novelty and diversity of Llm generated ideas." arXiv preprint arXiv:2410.14255 (2024).

<https://arxiv.org/abs/2410.14255>

推荐理由：提出了一种创新的迭代规划和搜索方法，专门用于提升大语言模型生成内容的新颖性和多样性。所提出的Nova框架通过结构化的迭代探索过程，有效克服了LLM生成内容单一和重复的问题。

[10] Schmidgall, Samuel, et al. "Agent Laboratory: Using LLM Agents as Research Assistants." arXiv preprint arXiv:2501.04227 (2025).

<https://arxiv.org/abs/2501.04227>

推荐理由：这篇论文提出了一个创新的LLM驱动的自主研究框架“Agent Laboratory”，通过文献回顾、实验和报告写作三个阶段实现端到端的研究自动化，实验显示该系统不仅能达到与现有方法相当的性能，还实现了84%的成本降低，为AI辅助科研提供了新范式。

[11] Huang, Kexin, et al. "Automated Hypothesis Validation with Agentic Sequential Falsifications." arXiv preprint arXiv:2502.09858 (2025).

<https://arxiv.org/abs/2502.09858>

推荐理由：这篇论文提出了一个将严格的统计原则与基于LLM的智能体相结合的自动化假设验证框架POPPER，能够有效地从大量假设中筛选出有价值的假设，减少手动验证的工作量，提高研究效率。

[12] OpenAI Team. "Introducing Deep Research: An Autonomous Research Agent." OpenAI Technical Report, 2024.

<https://openai.com/index/introducing-deep-research/>

推荐理由：该报告介绍了OpenAI开发的Deep Research系统，这是一个基于大语言模型的自主研究代理，能够独立完成文献检索、数据分析和研究综述工作。系统由最新的o3模型驱动，具有增强的推理能力、网络浏览功能和数据分析能力，可以显著提高研究效率，为科研工作者提供强大的智能辅助工具。该系统标志着AI辅助科研进入新阶段，有望加速科学发现过程。

[13] Google Research Team. "Accelerating scientific breakthroughs with an AI co-scientist" Google Research Blog, 2024.

<https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/>

推荐理由：该文介绍了Google开发的AI co-scientist 科研助手系统，展示其作为“AI联合科学家”在材料发现、实验设计等领域的实际应用。系统具备多模态理解、数据分析和知识整合能力，能协助研究人员完成文献分析、实验优化等任务。文章通过具体案例展示了AI助手如何加速科研突破，体现了大语言模型在科学发现中的实用价值，代表了AI辅助科研的重要进展。

[14] Si, Chenglei, Diyi Yang, and Tatsunori Hashimoto. "Can LLMs generate novel research ideas? a large-scale human study with 100+ nlp researchers." arXiv preprint arXiv:2409.04109 (2024).

<https://arxiv.org/pdf/2409.04109>

推荐理由：这项大规模人机对比研究首次证实了LLM在研究创意生成方面的潜力，通过严谨的实验设计和统计分析，不仅量化了AI在研究想法新颖性上的优势，也指出了当前系统的局限性。

2.5 复杂世界的多智能体建模

2.5.1 背景介绍

过去几十年里，社会科学家和相关领域的研究者，一直致力于通过实证数据与模型揭示人类行为和智能社会运行的基本规律，试图找出隐藏在各种社会现象和治理痛点背后的因果机制，从而回答“是什么”、“为什么”、“如何治”等一系列问题。相对于传统实证主义方法的局限，基于多智能体的建模作为一种科学方法论在上世纪八九十年代被提出。科学家们围绕经济学等社会科学及工程领域广泛存在的复杂系统和复杂现象所开展的探索工作，如Joshua Epstein等开发的糖域模型、Brain Arthur领导开发的人工股票市场模型、Thomas Schelling的居住隔离模型、Christopher Langton的人工生命模型[1-4]等等，几乎都是在这一时期提出和发展起来的。其核心思想是借助于计算机平台，在一个人工搭建的虚拟环境中创建若干彼此之间以及与环境之间能够交互的主体（Agent），对现实个体行为与环境进行精细刻画，进而辅助研究者的直觉推理，实现从微观个体行动到宏观群体行为与决策效应/模式的低成本、高可控的探索性研究，揭示复杂系统的非线性、动态性和不确定性等重要特征[5-6]。

多智能体系统的建模方法经历了知识驱动的早期阶段和数据驱动的近期阶段[6-7]。前者倾向于依靠精心设计的带有规划、推理、决策过程的内部世界模型的框架来对现实世界进行仿真，导致很难对其内部世界模型和规划推理模块进行优化，也因此促进了数据驱动的“感知-反馈”式架构的发展。这种方法主要依赖于直接的感知-行动循环，而不是复杂的内部世界模型或深层推理过程来做出决策。虽然以深度学习

为代表的人工智能技术促进了多智能体建模技术的发展，但仍然面临以下局限[7]：

- 单个智能体不具备处理长期复杂任务的能力。尽管在数据驱动的“感知-反馈”式架构支持下，智能体能够适应不同的环境变化，但在处理复杂特别是不确定的任务或需要长期规划的情况时可能受到限制。
- 智能体不具有通用能力。真实世界的复杂性、动态性与不确定性是多变的，目前在某一种环境中表现优异的智能体可能在另一种环境中完全无效。
- 多智能体建模方法无法支持对真实环境的综合模拟。一个实用的多智能体模型应该可以描述系统在已知条件下如何运行并对出现的现象具有解释能力，同时可以对开放环境进行观察并预测系统未来的状态。目前的建模方法只能实现二者其一的能力，不能兼而有之。

以大语言模型为代表的人工智能技术的出现，不仅增强了单个智能体在感知、决策、记忆、工具使用和自适应方面的能力，而且从群体层面提高了智能体的异质性、智能体间及与环境交互的场景复杂性与不确定性，以更好表征个体-群体的决策行为。这些能力共同促进了多智能体建模方法的跨越式发展[7-8]。

下边我们介绍三个用于社会科学领域的研究：

- 1、EconAgent，构建了基于LLM的宏观经济复杂系统[9]；
- 2、AgentSociety 1.0，精确模拟了社会舆论传播、认知观点极化、公众政策响应等[11]；
- 3、RLLNC，提出了一种基于多智能体强化学习的模型驱动的去中心化优化方法，并应用于城市治理中的若干问题，为大规模复杂的真实复杂系统的逼真模拟和优化提供了可行技术手段[12]。

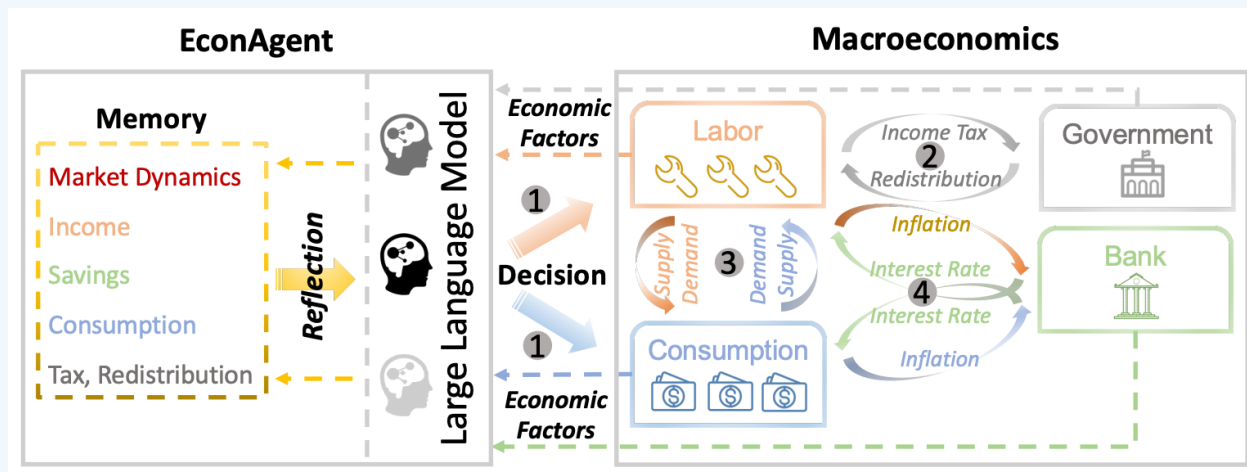
2.5.2 研究进展

2.5.2.1 基于LLM的多智能体系统研究宏观经济学

推荐理由：相较于以往的智能体，如今基于LLM的智能体能够表现出类人的决策模式。因而基于LLM的多智能体系统能模拟更符合实际的宏观经济系统，反映更真实的人类复杂经济行为。

现有的智能体建模宏观经济系统通常采用预定规则或基于学习的神经网络进行决策。前者面临的问题是不能对智能体进行个性化设计；后者则需要大规模行为数据对模型进行训练。此外，现有模型通常关注当前时期的个别宏观经济影响因子，多期市场动态和多方面宏观经济因素的影响在决策过程中往往被忽视。

为了解决上述问题，清华大学李勇团队首次提出一个具有人类特征的、由LLM赋能的智能体EconAgent，用于宏观经济模拟。他们搭建了一个包括100个智能体在长达20年时间内（模型以年为单位进行迭代）的工作与消费行为组成的动态市场仿真环境。其中，仿真环境随着智能体的行为以及财政与货币政策等宏观经济因子的变化而演化。每个智能体都是基于真实世界经济状况的目标画像进行建模的，因而系统中的智能体具有异构性，能够自动展现不同的决策机制。除了感知模块和行动模块外，模型还引入了记忆模块使智能体能够反思过去的个人经历和市场动态，从而反映宏观经济趋势变化对个体的影响。实验表明，基于大语言模型的多智能体系统能够更合理、稳定地模拟宏观经济系统，再现消费市场的通货膨胀、劳动力市场的失业率等经典宏观经济现象[9]。



基于LLM的多智能体系统模拟宏观经济系统，智能体能够依据财政、货币等宏观经济因子进行劳动和消费决策，并且能够基于记忆模块反思自己的行为。| 来源：Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024. EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 15523–15536, Bangkok, Thailand. Association for Computational Linguistics.

2.5.2.2 基于LLM的多智能体系统研究社会群体行为

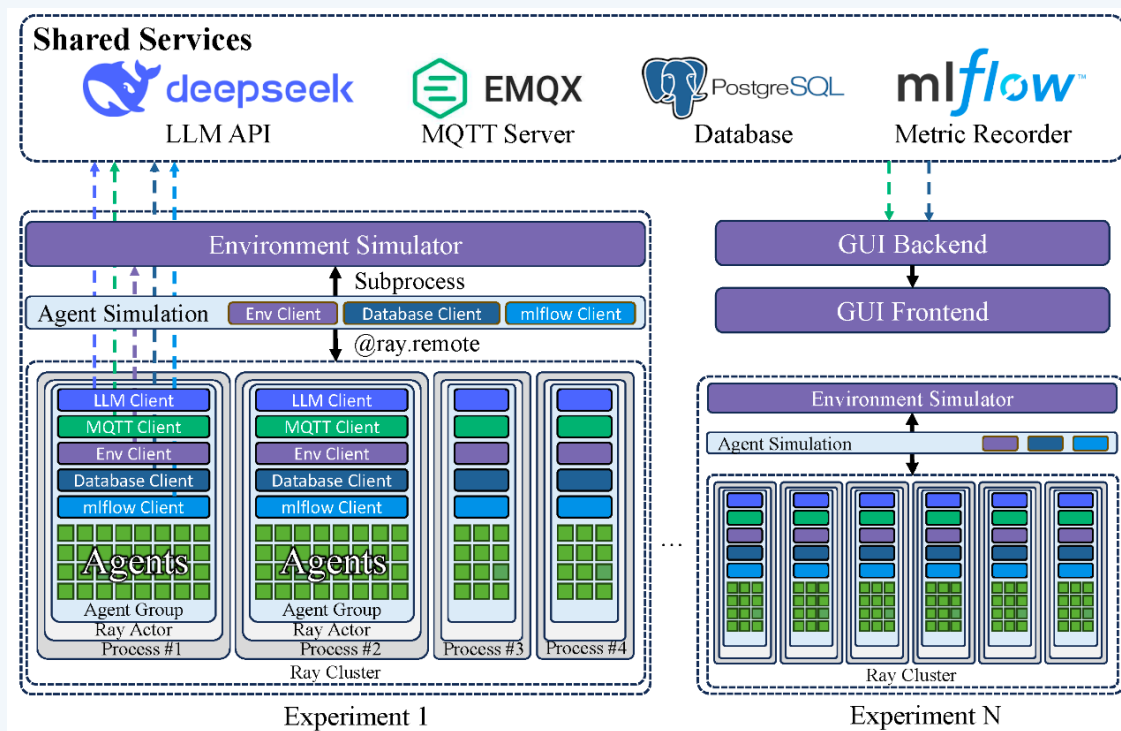
推荐理由：社会系统的本质在于其构成元素间的非线性涌现。相较于以往，基于智能体的社会群体行为模拟可以充分发挥大模型智能体自适应性、真实社会环境仿真能力与大规模模拟加速框架的技术优势，实现对社会复杂性的更真实还原。

基于多智能体的模拟在社会领域的研究中取得了一系列的成果[1-3, 9-10, 14-15, 17, 25]。但是，如何在保证模拟规模足够庞大以支持复杂性研究的同时，提升模拟的真实性，使其更贴近现实社会的运行逻辑，仍是亟待突破的重要方向。

为了解决上述问题，清华大学李勇团队充分发挥大模型智能体、真实社会环境仿真与大规模模拟加速框架的技术优势，结合智能社会治理的前沿理论与实践探索，成功构建了基于大模型的大型社会模拟器 AgentSociety1.0版本，可精确模拟社会舆论传播、认知观点极化、公众政策响应等[11]。为了让智能体的行为更加贴近现实世界，团队将智能体设计分为三个层面：心智、心智-行为耦合和行为，并构建了一个高度真实、可交互的城市社会环境，支持移动、社交和经济活动等核心行为的模拟。与仅依赖大模型推理不同，团队构建的环境结合物理约束、社会规则和资源限制，以确保智能体的行为符合现实逻辑，避免大模型生成的“幻觉”影响。



基于LLM的AgentSociety 1.0整体框架，可精确模拟社会舆论传播、认知观点极化、公众政策响应等。| 来源：AgentSociety: Large-Scale Simulation of LLM-Driven Generative Agents Advances Understanding of Human Behaviors and Society. arXiv preprint arXiv:2502.08691



基于LLM的AgentSociety 1.0大规模社会模拟引擎，结合LLM驱动的多智能体系统和真实城市社会环境，并系统架构上进行了深度优化，以确保模拟结果既具备现实合理性，又能大规模、高效运行。| 来源：AgentSociety: Large-Scale Simulation of LLM-Driven Generative Agents Advances Understanding of Human Behaviors and Society. arXiv preprint arXiv:2502.08691

大型社会模拟器通过数字化和虚拟化的方式，使研究者能够在无需进行实际实验的情况下，模拟并观察社会现象的变化过程，显著降低了研究的风险与成本。同时，模拟器能够在不同情境下进行多维度的数据采集和实验设计，克服传统研究因受限于时间、空间和资源等因素所带来的局限。团队提供可视化交互工具和智能社会治理研究工具箱，支持实时监测、主动干预和数据收集，助力社会研究和治理实验。

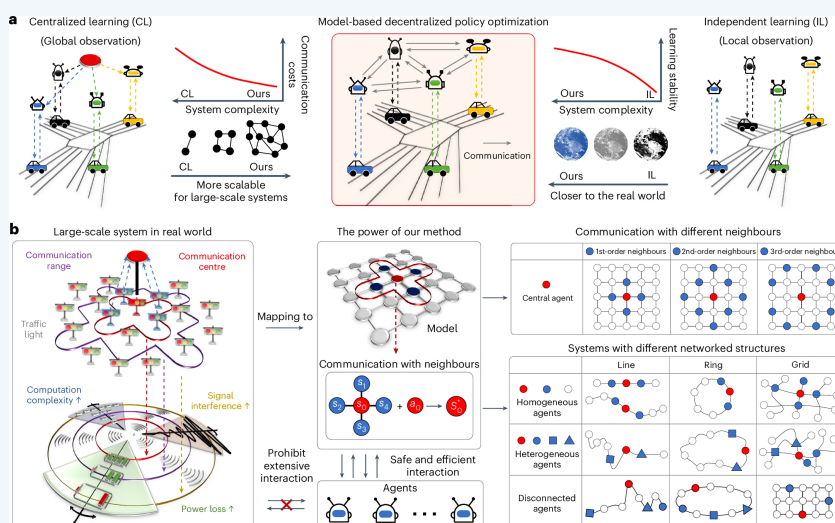
2.5.2.3 基于多智能体系统的未来城市治理

推荐理由：作为融合了人类活动、交通、医疗、电力、建筑环境等因素的复杂动态系统课题，城市科学的研究需要融合多领域、跨学科的方法。随着物联网技术和人工智能技术的发展，未来城市将会配备越来越多的传感器以及能够自主决策的边缘智能系统。因此，未来城市将是典型的大规模多智能体系统。

大规模多智能体系统面临的主要挑战在于其可扩展性和性能的平衡。传统的集中式多智能体方法在处理复杂任务时，往往需要大量的通信和数据采样，这不仅增加了系统的复杂性和成本，还可能导致性能的下降。此外随着系统规模的扩大，通信延迟和数据传输的瓶颈也变得更加明显。多智能体强化学习（MAREL）通过将复杂任务分解并分配给多个协作节点，可以显著提高系统的可扩展性。但是，其在实际部署中的仍面临横向扩展能力受限、计算效率衰减等瓶颈。

为了解决上述问题，北京大学杨耀东团队结合多智能体强化学习方法，提出了一种模型驱动的去中心化策略优化框架：RLLNC，对大规模多智能体系统进行了以智能体为单位的建模，将智能体之间的关系描述为一种拓扑连接结构下的网络化关系。其中包括线状、环状、网状等各种同构/异构节点，减少了多智能体系统对全局通信和大量交互数据的依赖，降低了系统决策复杂性[12]。此外，研究者还在框架中引入了分支回滚技术，通过采样先前策略的状态分布进行固定步数的回滚。实验表明，研究提出的去中心优化策略在车辆控制、疫情网络控制、电力控制等复杂系统的控制中都表现出良好的有效性和可扩展性。

目前，在交通[13]、城市规划[14]、公共卫生[15]等领域，如何利用多智能体强化学习技术实现未来城市中自主决策的边缘智能体更好协同，从而构建更高效、可持续的城市运行体系，已成为重点研究方向。未来，基于多智能体系统的未来城市研究与治理有望为人们提供更良好的生活环境。



大规模网络控制的强化学习（Reinforcement Learning for Large-scale Network Control, 简称RLLNC）框架及其工作原理，以及传统的中心化学习和基于模型的去中心化控制等框架的对比 | 来源：Ma, C., Li, A., Du, Y. et al. Efficient and scalable reinforcement learning for large-scale network control. Nat Mach Intell 6, 1006–1020 (2024).

2.5.3 挑战与展望

如果多智能体系统能够对世界中的真实复杂系统进行模拟，那么，它不仅能够赋能城市规划[14]、公共卫生[15]、群体博弈[16-18]、教育[19]、未来战场[20]等领域的复杂系统研究，而且能够进一步推动多领域融合的跨学科研究，适应真实世界中系统的复杂性，满足人类可持续发展的重大需求。

因此，建议从以下几方面关注并推动未来多智能体系统的发展。首先，我们需要继续探索智能体的能力边界（特别是在一些不确定的场景与任务背景下[21-23]），包括实时感知与处理外部多模态信息并学习的能力、高效的信息整合与决策能力、任务拆解与调用工具的能力，甚至是具身执行任务与操作的能力。其次，在目前强化学习方法的基础上，在多智能体系统中引入博弈论，可以实现协同策略的动态优化，提升智能体在复杂博弈环境中的决策能力，优化协同作业的系统效率[6, 24]。最后，我们需要支持开源。通过研究能够兼容异构智能体架构、支持实时扩展且具备安全合规保障的大规模多智能体系统架构，以支持多智能体之间以及多智能体与环境的复杂交互，从而实现真正的“斯坦福小镇”[25]。

2.5.4 参考文献

[1] Epstein, J. M., & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Institution Press.

<https://direct.mit.edu/books/monograph/2503/Growing-Artificial-SocietiesSocial-Science-from>

推荐理由：这是来自MIT出版社的一本书，是圣塔菲研究所（Santa Fe Institute）、世界资源研究所（World Resources Institute）和布鲁金斯学会（the Brookings Institution）的面向2050的联合项目。主要利用计算机模拟技术研究社会结构和群体行为如何从个体的相互作用中产生。其中“糖域模型”作为一种自下而上的社会科学范式，吸引了广泛关注。

[2] Arthur, W. B., Holland, J. H., LeBaron, B., Palmer, R., & Tayler, P. (1997). Asset pricing under endogenous expectations in an artificial stock market. *The Economy as an Evolving Complex System II*, 15, 15-44.

<https://www.taylorfrancis.com/chapters/edit/10.1201/9780429496639-2/asset-pricing-endogenous-expectations-artificial-stock-market-brian-arthur-john-holland-blake-lebaron-richard-palmer-paul-taylor>

推荐理由：这是来自CRC出版社的书籍《The Economy as an Evolving Complex System》的一部分，提出了一种基于异质智能体的资产定价理论。研究使用圣塔菲研究所（Santa Fe Institute）提出的人工股票市场模型。

[3] Schelling, Thomas C. "Dynamic models of segregation." *Journal of mathematical sociology* 1.2 (1971): 143-186.

推荐理由：谢林提出的居住隔离模型。

[4] Langton, Christopher G. "Studying artificial life with cellular automata." *Physica D: nonlinear phenomena* 22.1-3 (1986): 120-149.

推荐理由：朗顿提出的人工生命模型。

[5] Berry, Brian JL, L. Douglas Kiel, and Euel Elliott. "Adaptive agents, intelligence, and emergent human organization: Capturing complexity through agent-based modeling." *Proceedings of the National Academy of Sciences* 99.suppl_3 (2002): 7187-7188.

推荐理由：2002年召开的主题为“适应性Agent，智能和人类组织的涌现：通过ABM刻画复杂性”的座谈会，在PNAS发表了专刊，此文为总结性文章。

[6] Fatima, Shaheen, Nicholas R. Jennings, and Michael Wooldridge. "Learning to resolve social dilemmas: a survey." *Journal of Artificial Intelligence Research* 79 (2024): 895-969.

推荐理由：这是三位研究多智能体系统的计算机科学家撰写的关于社会困境的一篇综述文章。系统综述了面临困境的个人如何根据过去的经验调整自己的行为来学会合作，并概述了三种学习方法。

[7] Gao, C., Lan, X., Li, N. et al. Large language models empowered agent-based modeling and simulation: a survey and perspectives. *Humanit Soc Sci Commun* 11, 1259 (2024).

推荐理由：清华大学利用课题组在以大语言模型为基础的多智能体建模研究中十分活跃，先后对基于多智能体的经济系统、城市系统进行了研究，并对领域发展进行了综述，将多智能体仿真分为物理世界、社会网络、赛博空间以及混合现实四种类型。

[8] Large Language Model based Multi-Agents: A Survey of Progress and Challenges, arXiv:2402.01680

推荐理由：这篇文献从多智能体建模框架、基准、任务解决能力、能力进化、世界仿真能力几个角度对基于大语言模型的多智能体建模相关的研究进行了系统的调研。

[9] Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024. EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 15523–15536, Bangkok, Thailand. Association for Computational Linguistics.

<https://aclanthology.org/2024.acl-long.829/>

推荐理由：这篇来自ACL会议的文章提出了搭建了一个包括100个智能体在长达20年时间内（模型以年为单位进行迭代）的工作与消费行为组成的动态市场仿真环境对宏观经济行为进行模拟。

[10] Gao, Chen, et al. S³: Social-network Simulation System with Large Language Model-Empowered Agents." arXiv preprint arXiv:2307.14984 (2023).

<https://arxiv.org/abs/2307.14984>

推荐理由：这篇来自arxiv的预印文章基于实现了基于LLM的社会网络模拟系统S³。使用现实世界的数据实现了个体和群体两个模拟级别的评估。

[11] AgentSociety: Large-Scale Simulation of LLM-Driven Generative Agents Advances Understanding of Human Behaviors and Society. arXiv, 2025

<https://arxiv.org/abs/2502.08691>

推荐理由：这篇来自arxiv的预印文章实现了基于LLM的AgentSociety 1.0，可精确模拟社会舆论传播、认知观点极化、公众政策响应等。

[12] Ma, Chengdong, et al. "Efficient and scalable reinforcement learning for large-scale network control." Nature Machine Intelligence (2024): 1-15.

<https://www.nature.com/articles/s42256-024-00879-7>

推荐理由：这篇来自Nature Machine Intelligence文章提出了一种模型驱动的去中心化策略优化框架：RLLNC。框架减少了智能体系统决策的复杂性，提升了系统的可扩展性。

[13] Noaeen, Mohammad, et al. "Reinforcement learning in urban network traffic signal control: A systematic literature review." Expert Systems with Applications 199 (2022): 116830.

<https://www.sciencedirect.com/science/article/pii/S0957417422002858>

推荐理由：这篇来自Expert Systems with Application的文章综述了城市交通信号控制中强化学习的应用，涵盖了从1994到2020年的160篇同行评审的文章。

[14] Zheng, Yu, et al. "Spatial planning of urban communities via deep reinforcement learning." Nature Computational Science 3.9 (2023): 748-762.

<https://www.nature.com/articles/s43588-023-00503-5>

推荐理由：这篇来自Nature Computational Science的文章基于15分钟城市的概念提出了一个深度强化学习算法模型，可以结合人工输入、机器学习辅助土地和道路空间规划的表现，进行复杂的城市空间规划，超越其他算法和专业人类设计师。

[15] Williams, Ross, et al. "Epidemic modeling with generative agents." arXiv preprint arXiv:2307.04986 (2023).

<https://arxiv.org/abs/2307.04986>

推荐理由：这篇来自arxiv的预印文章基于大语言模型的多智能体系统模拟了传染病预警与传播，为公共卫生治理提供了更有效的解决方案。

[16] Kramár, János, et al. "Negotiation and honesty in artificial intelligence methods for the board game of Diplomacy." Nature Communications 13.1 (2022): 7214.

<https://www.nature.com/articles/s41467-022-34473-5>

推荐理由：这篇来自Nature Communication的文章给智能体系统中的个体引入了协商机制，从而研究个体间合作与背叛的条件。

[17] Koster, Raphael, et al. "Human-centred mechanism design with Democratic AI." Nature Human Behaviour 6.10 (2022): 1398-1407.

<https://www.nature.com/articles/s41562-022-01383-x>

推荐理由：这篇来自Nature Human Behaviour的文章开发了民主AI模型，用于设计与验证符合大多数人利益的社会机制。

[18] Ashery, Ariel Flint, Luca Maria Aiello, and Andrea Baronchelli. "The Dynamics of Social Conventions in LLM populations: Spontaneous Emergence, Collective Biases and Tipping Points." arXiv preprint arXiv:2410.08948 (2024).

推荐理由：这篇来自arxiv的预印文章构建了基于LLM的多智能体系统研究了社会中观点的产生、传播机制，包括自发出现、集体偏见和临界点。

[19] Jiang, Yuan-Hao, et al. "AI Agent for Education: Von Neumann Multi-Agent System Framework." Proceedings of the 28th Global Chinese Conference on Computers in Education (GCCCE 2024).

推荐理由：这篇来自GCCCE的会议文章提出了面向教育的冯·诺伊曼多智能体系统框架。

[20] Huang, Jen-tse, et al. "On the Resilience of Multi-Agent Systems with Malicious Agents." arXiv preprint arXiv:2408.00989 (2024).

<https://arxiv.org/abs/2408.00989>

推荐理由：这篇来自arxiv的预印文章提出了AUTOTRANSFORMER和AUTOINJECTION两种方法为评估多智能体系统的安全性，以及提高面对恶意智能体时的弹性提供了方法。

[21] Feng, Yu, et al. "BIRD: A Trustworthy Bayesian Inference Framework for Large Language Models." arXiv preprint arXiv:2404.12494 (2024).

<https://arxiv.org/abs/2404.12494>

推荐理由：这篇来自arxiv的预印文章提出了基于贝叶斯推断的LLM-based智能体框架BIRD，提高智能体在面对不完整信息进行规划与决策时的能力。

[22] Liu, Ollie, et al. "DeLLMa: Decision Making Under Uncertainty with Large Language Models." arXiv preprint arXiv:2402.02392 (2024).

<https://arxiv.org/abs/2402.02392>

推荐理由：这篇来自arxiv的预印文章提出了基于人类的决策理论与效用理论优化LLM-based智能体框架DeLLMa，通过优化智能体的推理过程在提高智能体在不确定环境中决策准确性。

[23] Han, Jiuzhou, Wray Buntine, and Ehsan Shareghi. "Towards Uncertainty-Aware Language Agent." arXiv preprint arXiv:2401.14016 (2024).

<https://arxiv.org/abs/2401.14016>

推荐理由：这篇来自arxiv的预印文章在LLM-based智能体框架中加入了量化评估不确定性的模块，并通过寻求外部解决方案或者人类帮助来帮助智能体不断进行学习应对不确定性的能力。

[24] Hua W, Liu O, Li L, et al. Game-theoretic LLM: Agent Workflow for Negotiation Games[J]

推荐理由：这篇发表在的文章提出了博弈论工作流：将博弈论作为基础，以理性和帕累托最优性作为两大基本评估指标--即个体是否理性，以及是否基于个体理性得出全局最优解，旨在观察并提升智能体在以自我利益最大化为指导的交互中的表现。

[25] Park, Joon Sung, et al. "Generative agents: Interactive simulacra of human behavior." Proceedings of the 36th annual acm symposium on user interface software and technology. 2023.

<https://dl.acm.org/doi/abs/10.1145/3586183.3606763>

推荐理由：这篇来自ACM会议的文章介绍了生成式智能体模拟人类行为，形成了一个由25个智能体组成的基于自然语言交互的小镇。初步探讨了基于大语言模型对人类行为模拟的架构和交互模式。

2.6 AI 仿真与系统模拟

2.6.1 背景介绍

在复杂系统研究中，从微观的量子粒子，如可控核聚变中的等离子体演化，中观的物理流体动力学、生物智能运动，到大尺度的全球气候变化等宏观现象，都需要通过科学仿真工具来预测系统的行为和演化。

从历史发展来看，科学仿真经历了三个重要阶段：最初基于第一性原理的解析求解，然后主要发展的数值模拟方法。然而由于这些真实系统往往表现出多尺度、多物理场耦合的特征，其复杂性对传统解析和数值计算方法提出了巨大挑战，一方面不能保证足够的精度，一方面又很难从数据中发掘系统的内在机制和因果关系[2]。

所幸的是，在融合物理约束与数据驱动的基础上，基于AI驱动的新一代科学仿真技术应运而生。AI仿真器和数字孪生[3]通过深度学习和AI for PDE科学模拟方法，或引入因果推理框架[4]，结合物理先验知识构建新一代仿真系统，为系统行为理解和预测开辟了新途径。在生物系统方面，以秀丽线虫为代表的模型生物为研究复杂生命系统提供了理想平台，而基于AI驱动的数字生命体(AIDO)，则构建了一个多尺度系统的基础模型，实现了从分子到整体生命的全方位预测与模拟。

2024年12月发布的Genesis项目更是将物理仿真推向了新高度。这个面向机器人和人工智能的生成式物理引擎，不仅能高效模拟从刚体到软体的各类物理交互，还首次实现了通过自然语言控制的场景生

成，为研究具身智能提供了强大工具。

这些进展标志着科学仿真正在进入一个新时代：从单一物理场景扩展到多尺度、多学科交叉，从被动观察转向主动探索，从现象预测迈向机理理解。下面我们将详细介绍几项代表性工作。

2.6.2 研究进展

2.6.2.1 AI for PDE（人工智能求解偏微分方程）

推荐理由：偏微分方程（PDEs）是现代科学和工程中不可或缺的数学工具，用于精确描述从流体力学到量子力学等各种物理现象。但传统的有限元（FEM）和有限差分方法（FDM）等数值方法，面临着计算成本高昂、难以扩展到复杂几何形状和高维问题以及处理大规模数据集的局限。由此，结合人工智能算法的AI for PDE科学模拟方法应运而生，通过将数据与PDE信息相结合，能够为特定问题提供近似解，并显著加速传统的数值求解算法。

AI for PDE领域涌现出多种创新方法，这些方法利用机器学习和深度学习技术来求解、分析和发现偏微分方程。以下是对几种主要的AI驱动PDE求解方法的简单介绍。

物理信息神经网络（PINNs） [5]是一种通过将PDE的控制方程融入神经网络的损失函数中，来训练神经网络求解PDE的强大框架。PINNs的核心思想是利用神经网络强大的函数逼近能力来表示PDE的解，并通过最小化一个包含PDE残差项和边界/初始条件项的损失函数来优化网络参数。近年来，研究人员不断探索改进PINNs在网络架构（如KAN）、优化技术

以及处理复杂问题方面的能力，以解决其在处理高频和多尺度问题时面临的挑战。Pinn-jax等开源库的出现进一步简化了PINNs的开发和训练过程。值得一提的是，PINNs不仅可以用于求解已知PDE的正问题，还可以应用于根据观测数据反演PDE参数或发现未知物理规律的反问题。

深度能量方法 (DEM) [6]是另一种基于物理原理的AI驱动PDE求解方法，它利用最小势能原理来预测静态载荷条件下的变形，可以看作是有限元分析的一种替代方案。与传统的基于配点的PINNs不同，DEM主要依赖于二阶微分和精确的数值积分来训练神经网络。为了弥补传统DEM缺乏最小余能形式的不足，研究人员提出了深度互补能量方法 (DCEM)。DCEM以应力函数作为输出，能够天然地满足平衡方程。此外，为了提高DEM的精度，研究人员还引入了随机傅里叶特征映射等技术，并提出了用于超参数优化的双循环架构。DEM的一个显著优点是其网格无关性，这使得它在处理复杂材料（如超弹性材料和粘弹性材料）的变形问题时，无需生成大量的训练数据。

算子学习则旨在训练神经网络直接学习PDE解算子，即从输入函数空间（如初始条件或边界条件）到输出函数空间（PDE的解）的映射。DeepONet是一种典型的算子学习方法，它基于算子的泛逼近定理，通过学习分支网络和干网络来逼近非线性算子。傅里叶神经算子 (FNOs) 则利用傅里叶变换在频域进行卷积操作，能够高效地学习PDE的解算子，并且具有分辨率无关性。物理信息神经算子 (PINOs) 是一种混合方法，它结合了算子学习和物理信息优化框架，旨在提高收敛速度和精度。最近的研究还对DeepONet[7]进行了改进，允许其Trunk-branch结构中的

Branch网对Trunk网非线性干扰，提出了一种物理信息增强的扩展版本，特别适用于解决复杂的非线性偏微分方程 (PDEs)，显著降低了物理损失。

微分同胚映射算子学习 (DIMON) [8]是一种新兴的AI框架，它能够比超级计算机更快地在普通个人电脑上求解复杂的PDE。DIMON的关键能力在于，它能够从单个已求解的形状中学习模式，并将该解映射到多个新的形状上，从而避免了为每个新形状重复进行网格划分和计算。这种方法在需要对多个复杂几何形状求解PDE的科学和工程领域具有广泛的应用潜力，包括医疗保健（如心律失常预测）、航空航天和材料科学。目前 DIMON 的代码库也已公开。

2.6.2.2 因果数字孪生

推荐理由: 因果数字孪生 (Causal Digital Twin) [4]将因果推理引入仿真系统，不仅能模拟系统行为，还能揭示系统内在的因果机制，为智能决策和干预提供理论基础。

传统的数字孪生技术主要关注系统行为的模拟与预测，而因果数字孪生则通过引入因果推理框架，通过概率编程与因果推理的深度融合，赋予了系统理解“为什么”的能力。

CausalCity[3]就是应用因果数字孪生技术的典型案例，它提供了一个高保真的仿真环境，使其能突破简单的“模式识别”，理解因果关系，这对自动驾驶等复杂系统非常重要。其核心创新在于引入了“主体性” (Agency) 的概念——环境中的每个实体（如车辆）都具有基本的智能性，能够自主决策低层级行为。这使得研究者可以用简单的高层级配置来定义复杂场



CausalCity 仿真环境 | 来源: McDuff, Daniel, et al. "CausalCity: Complex Simulations with Agency for Causal Discovery and Reasoning." Proceedings of the First Conference on Causal Learning and Reasoning, PMLR 177, 2022, pp. 559-575.

景，而不需要指定每个具体的低层级动作。

CausalCity与三个主流的推理模型进行了对比实验，包括：NRI，使用变分自编码器框架，通过节点间的消息传递机制来发现系统中的关系结构并重建轨迹；NS-DR，基于传播网络，将场景中的车辆建模为图中的节点，通过分析节点间的状态传播来预测车辆运动；V-CDN，采用一个三阶段架构，包含感知模块（提取视觉特征）、推理模块（构建因果图谱）及动态模块（预测未来状态）。这些方法在简单场景下表现良好，但在引入混杂因素（如交通信号灯）的复杂场景中性能会显著下降，而CausalCity则能表现良好。这揭示了传统推理方法在处理真实世界复杂性时的局限，以及引入因果推理的重要性。

2.6.2.3 AI驱动的多尺度数字有机体建模

推荐理由：GenBio AI提出的AI驱动数字有机体（AIDO）框架[9]，代表了生物计算仿真领域的一项里程碑式进展。该研究首次构建了一个整合分子、细胞、组织到个体水平的多尺度基础模型系统，通过模块化设计和分层整合，实现了生物学现象的预测、模拟与编程。相较于传统单任务模型（如AlphaFold、ESMFold），AIDO在多个维度实现了范式突破。

研究团队开发了覆盖全生物尺度的基础模型组件。针对分子层级，提出了混合分词器（Hybrid Tokenizer）架构，通过几何深度学习编码蛋白质三维结构[10]，并结合序列模型实现DNA-RNA-蛋白质的中央法则统一建模[11]；在细胞层级，设计了非对称编码器-解码器框架，有效处理单细胞转录组数据的高维稀疏特；在表型层级，则利用自监督对比学习

构建了连续生理信号（如血糖、心电图）的生成式模型[12]。

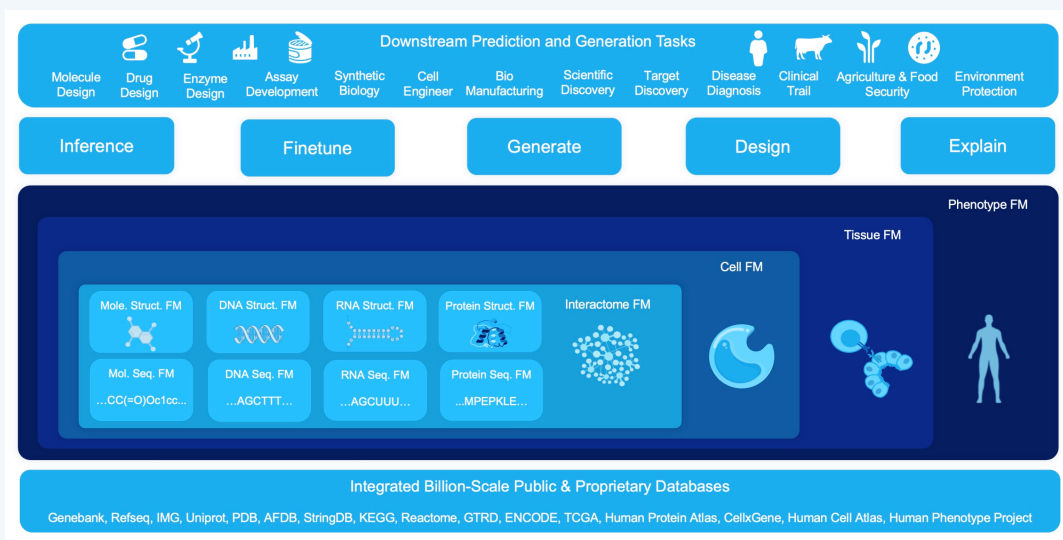
AIDO的创新性体现在其系统性整合能力。通过可微分计算图（DCG）技术，将分子交互网络、基因调控网络与细胞状态模型动态连接，形成闭环反馈机制。例如，在药物设计场景中，蛋白质生成模型可基于结构解码器生成候选序列，再通过分子动力学评分模型反向优化生成策略，最终通过湿实验验证形成迭代优化循环。

该框架的计算效率同样突出。通过混合架构（如CNN-Transformer融合模型）和高效并行化策略，实现了从基因序列（长达百万碱基）到器官级表型（如 UK Biobank 队列数据）的多尺度建模。系统支持30+种生物数据模态，涵盖NCBI基因组库、结

构数据库等超10亿级数据点，模型参数量突破千亿级别。

AIDO的开源生态体系（GitHub:genbio-ai/AIDO）为社区提供了模块化扩展能力。研究人员可独立调整DNA序列模型（AIDO.DNA[9]）、蛋白质逆折叠模型（AIDO.ProteinIF[11]）等组件，或通过跨尺度对齐优化实现全系统升级。目前，该系统已在300+生物任务中达到SOTA性能[11][13]。

随着生物数据量的指数增长和计算架构的持续优化，AIDO有望推动药物研发、合成生物学、精准医学等领域的范式变革。未来，通过整合更多实验模态（如空间转录组、代谢组）和因果推理机制[14]，该系统或将成为解码生命复杂性的核心计算平台。



AIDO的多层级架构：数据层、基础模型系统层、下游实用层、生物工程应用层 | 来源：“Toward AI-Driven Digital Organism: A System of Multiscale Foundation Models for Predicting, Simulating and Programming Biology at All Levels.”

2.6.2.4 生物-环境闭环智能仿真系统

推荐理由：BAAIWorm天宝项目代表了生物智能仿真研究领域的一项重要进展。这项发表在Nature Computational Science并登上封面的研究[15]，首次实现了将秀丽线虫的神经系统、身体和环境整合到一个完整的闭环仿真系统中，为探索“大脑-身体-环境”交互机制提供了全新的研究平台。

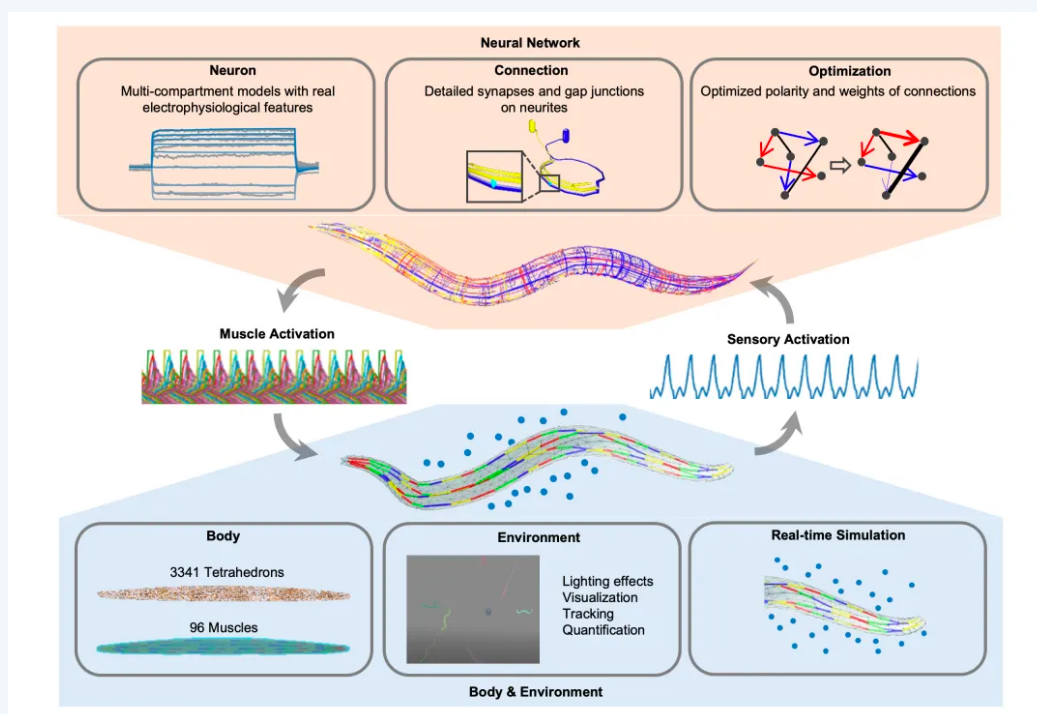
相比于原有的仿真系统（如OpenWorm[16]），BAAIWorm天宝在多个方面实现了显著突破。

研究团队构建了迄今为止最精确的线虫神经网络模型。该模型包含136个参与感知和运动功能的神经元[17]，每个神经元都采用多舱室模型进行建模，

精确还原了真实神经元的电生理特性。模型不仅在神经元层面实现了高度仿真，还通过创新的优化算法，使神经网络整体呈现出与真实线虫神经系统相似的动态特性。

此外，BAAIWorm还是一个高性能的“身体-环境”模型，创新性地采用投影动力学作为形变求解器[18]，实现了30帧/秒的实时仿真，使环境尺度提升了两个数量级，既大大提升了计算效率，又保持了生物力学特性的准确性。

更重要的是，研究首次实现了线虫神经网络与身体环境系统的持续闭环交互。例如在完整的系统仿真中，环境中的食物浓度等刺激会影响线虫感觉神经元，进而通过神经网络驱动运动神经元产生协调的行



线虫天宝概览 | 来源: An integrative data-driven model simulating *C. elegans* brain, body and environment interactions

为输出[19]，并改变环境的局部的流体动力学特性反过来再影响线虫。这种闭环设计使得系统能够模拟真实线虫的行为，如趋化性运动等复杂行为模式。系统还开发了创新的行为分析方法，可以稳定地量化线虫在三维空间中的运动轨迹。

BAAIWorm天宝展示了如何通过数字化重建来研究复杂生命系统，为理解生物智能本质提供了重要指引。目前该系统已开源，研究人员可以根据需要独立修改和扩展各个模块。未来，随着实验数据的不断积累和算法的进一步优化，BAAIWorm天宝有望模拟更多行为模式，涵盖更多神经元和环境因素，从而在生命科学和人工智能领域发挥更大的推动作用。

2.6.2.5 具身生成式物理模拟引擎

推荐理由：Genesis项目[20]由于其高精准的建模能力，开创了一种面向机器人和人工智能研究的新范式。这个由卡内基梅隆大学、清华等20多家顶尖机构历时24个月合作开发的生成式物理引擎，不仅展示了AI在物理世界模拟上的突破性进展，更为实现具身智能构建了技术路径。

Genesis的出现代表了模拟仿真平台一个新的高度：不再局限于单一的物理模拟或机器人控制，而是构建了一个统一的框架，能够同时处理物理规律、机器人动作、视觉渲染和人机交互等多个维度。这种整体性的方法使得它能够更真实地模拟复杂的物理世界，为机器人研究、AI应用和训练提供更可靠的仿真环境。

Genesis的技术创新体现在多个方面。其具有超高效的物理模拟性能，模拟速度比现实世界快约43

万倍，大幅领先现有解决方案；具备精确的物理模型，支持从刚体到软体，从单一物体到复杂交互的全方位物理模拟；此外系统采用生成式架构，通过自然语言描述就能生成复杂的物理场景和交互行为。



提示：A mobile franka arm heats the corn with the bowl and the microwave | 来源：Genesis: A Generative and Universal Physics Engine for Robotics and Beyond

Genesis在机器人研究领域展现出独特优势，不仅具备覆盖从人形机器人到软体机器人的全类型模拟能力，更在单块RTX4090显卡上实现了26秒完成可迁移至真实世界的动作策略训练，其效率突破将显著加速机器人研发进程。

随着未来Genesis的开源，可能会持续在多个领域引发连锁反应：为计算机图形学提供更真实的物理模拟，在机器人研究中加速算法验证和策略优化，在人工智能研究中推进具身智能的发展。Genesis不仅是一个技术工具，也代表了一种新的仿真研究范式，将为实现更智能、更自然的人工智能系统和探索世界的复杂系统规律持续贡献力量。

2.6.3 挑战与展望

尽管基于AI的科学仿真技术在近年来取得了显著进展，但仍面临着一些重要挑战。

多尺度与多物理场耦合的建模依然是难题。以核聚变等离子体为例，从粒子的量子效应到宏观的流体动力学行为，跨越了数十个数量级的时空尺度。现有的仿真方法难以在保持计算效率的同时实现如此大跨度的精确描述。同样，在全球天气预报中，大气、海洋、陆地等系统之间复杂的相互作用也给建模带来了巨大挑战。

仿真系统的可解释性问题依然需要探索。虽然基于深度学习的方法显著提升了预测精度，但其内部机制往往难以理解，这限制了科研人员对系统本质规律的认知。尽管有CausalCity等项目尝试引入因果推理框架，但要在复杂系统中建立可靠的因果关系图谱仍有很长的路要走。如何在保持高精度预测的同时提升模型的可解释性，成为了当前研究的重要方向。

此外，仿真系统与真实世界的对齐同样具有挑战性。虽然BAAIWorm等项目展现了在特定领域实现高保真仿真的可能，但要将这种成功推广到更复杂的生命系统和工程应用中仍面临诸多困难。如何确保仿真结果能够准确反映现实世界的物理规律，并与实验数据保持一致，仍需要更深入的研究。

要解决以上问题，除了提升计算能力和革新算法，还需要结合各种方法创造新的平台设计方案。一个可能的发展方向是将物理规律、数据驱动和因果推理等方法进行有机结合，发展新一代混合建模方法。

Genesis项目已经展示了这种融合潜力。同时，未来的仿真系统可能会更加智能和自主化，通过主动学习和迭代优化，系统能够自主设计和执行虚拟实验，提升模型精度和泛化能力，验证科学假设，甚至发现新的物理规律[19]。随着这些技术的不断成

熟，我们有望看到支持多学科、多尺度的统一仿真框架逐步建立，这将为解决生命科学、能源、气候等重大科学问题，研究复杂系统本质规律开辟的新途径和提供关键支撑。

2.6.4 参考文献

[1] Cornelio, Cristina, et al. "Combining Data and Theory for Derivable Scientific Discovery with AI-Descartes." *Nature Communications*, vol. 14, no. 1, 2023, pp. 1777.

<https://www.nature.com/articles/s41467-023-37543-4>

推荐理由：该论文提出了一个将数据驱动和理论驱动相结合的科学发展框架。

[2] Schölkopf, Bernhard, et al. "Toward Causal Representation Learning." *Proceedings of the IEEE*, vol. 109, 2021, pp. 612-634.

<https://ieeexplore.ieee.org/document/9363924>

推荐理由：该论文系统地探讨了机器学习中的因果表示问题，为因果推理与机器学习的结合提供了理论基础。

[3] Lavin, Alexander, et al. "Simulation Intelligence: Towards a New Generation of Scientific Methods." *arXiv preprint arXiv:2112.03235*, 2021.

<https://arxiv.org/abs/2112.03235>

推荐理由：该论文提出了新一代科学仿真方法的愿景，系统地总结了AI仿真器和数字孪生的关键技术。

[4] McDuff, Daniel, et al. "CausalCity: Complex Simulations with Agency for Causal Discovery and Reasoning." *Proceedings of the First Conference on Causal Learning and Reasoning*, PMLR 177, 2022, pp. 559-575.

<https://proceedings.mlr.press/v177/mcduff22a.html>

推荐理由：该工作提出了一个支持因果发现与推理的复杂仿真环境，为因果学习研究提供了重要平台。

[5] Yuntian Chen et al. "From PINNs to PIKANs: Recent Advances in Physics-Informed Machine Learning." *arXiv preprint*, 2024.

<https://arxiv.org/abs/2410.13228>

推荐理由：该论文全面回顾了物理信息神经网络(PINNs)的最新进展，重点关注网络设计和特征扩展的改进，展示了物理信息机器学习在科学计算中的发展趋势和应用价值。

[6] Sergii Mykhailovskiy, Dmytro Malynovskyi. "Model informed approach to estimate treatment effect in placebo-controlled clinical trials using an artificial intelligence based propensity weighting." *Ukrainian Journal of Medicine, Biology and Sport*, 2023.

<https://ouci.dntb.gov.ua/en/works/7qXkGYw9/>

推荐理由：该研究提出了一种基于人工智能的倾向性加权方法来评估安慰剂对照临床试验中的治疗效果，为医学临床研究提供了更准确的效果评估工具。

[7] Niyazi Sergin, Rui Duarte et al. "An architectural analysis of DeepOnet and a general framework." *Neurocomputing*, 2024.

<https://www.sciencedirect.com/science/article/pii/S0925231224014462>

推荐理由：论文分析了DeepOnet模型的神经网络设计，特别探讨了其主干-分支结构在算子学习任务中的有效性，为深度算子网络的设计提供了理论基础和改进方向。

[8] Qianxiao Li et al. "DIMON: Learning Solution Operators of Partial Differential Equations on a Diffeomorphic Family of Domains." *arXiv preprint*, 2024.

<https://arxiv.org/abs/2402.07250>

推荐理由：该研究提出了一个称为DIMON的通用算子学习框架，用于在不同形态域上学习偏微分方程的近似解，为复杂几何形状上的PDE求解提供了创新方法。

[9] Le Song, Eran Segal, Eric Xing. "Toward AI-Driven Digital Organism: A System of Multiscale Foundation Models for Predicting, Simulating and Programming Biology at All Levels." *arXiv preprint arXiv:2412.06993v1*, 2024.

<https://arxiv.org/abs/2412.06993>

推荐理由：该论文提出了整合分子至个体水平的多尺度生物建模框架，通过模块化基础模型系统实现了跨层级生物现象的预测与编程，为合成生物学和精准医学提供了通用计算范式。

[10] Jiayou Zhang et al. "Balancing locality and reconstruction in protein structure tokenizer." *NeurIPS Workshop*, 2024.

<https://www.biorxiv.org/content/10.1101/2024.12.02.626366v2.abstract>

推荐理由：该研究开发了一种基于几何深度学习的蛋白质结构分词器，通过平衡局部特征保留与全局结构重建，显著提升了三维分子表征在逆折叠和功能预测任务中的性能。

[11] Ning Sun et al. "Mixture of experts enable efficient protein understanding and design." *NeurIPS Workshop*, 2024.

<https://www.biorxiv.org/content/10.1101/2024.11.29.625425v1.abstract>

推荐理由：该论文提出专家混合架构的蛋白质语言模型，通过动

态路由机制实现序列-结构联合建模，在保证计算效率的同时将蛋白质生成多样性提升2.7倍。

[12] Levine, Zachary, et al. "Genetic underpinnings of predicted changes in cardiovascular function using self supervised learning." *bioRxiv* (2024): 2024-08.

<https://www.biorxiv.org/content/10.1101/2024.08.15.608061.abstract>

推荐理由：该工作利用自监督对比学习解析连续生理信号（如心电图）的动态特征，首次建立了心血管表型与PBMC转录组数据的跨尺度关联网络。

[13] Caleb N. Ellington et al. "Accurate DNA representations emerge from genome foundation models." *NeurIPS Workshop*, 2024.

<https://www.biorxiv.org/content/10.1101/2024.12.01.625444.abstract>

推荐理由：该研究构建了千亿参数规模的基因组基础模型，通过混合注意力机制有效捕捉长程序列依赖，在调控元件预测任务中达到98.3%的准确率。

[14] Benjamin Lengerich et al. "Contextualized Machine Learning." *arXiv:2310.11340*, 2023.

<https://arxiv.org/abs/2310.11340>

推荐理由：该理论框架系统阐述了上下文感知的机器学习方法，为多模态生物数据建模提供了可解释的因果推理机制，显著提升了模型在复杂生物系统中的应用可靠性。

[15] Zhao, Mengdi, et al. "An integrative data-driven model simulating *C. elegans* brain, body and environment interactions." *Nature Computational Science* 4.12 (2024): 978-990.

<https://www.nature.com/articles/s43588-024-00738-w>

推荐理由：首次构建了线虫完整的脑-身体-环境交互模拟系统。通过整合神经、肌肉和物理数据，实现了精确的行为预测。为理解生物神经系统控制行为提供了新方法。

[16] Sarma, Gopal P., et al. "OpenWorm: overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*." *Philosophical Transactions of the Royal Society B* 373.1758 (2018): 20170382.

<https://royalsocietypublishing.org/doi/full/10.1098/rstb.2017.0382>

推荐理由：介绍了OpenWorm项目的进展，这是一个开源的线虫整体仿真平台。整合了多层次的生物学数据。为系统生物学研究提供了重要工具。

[17] Cook, Steven J., et al. "Whole-animal connectomes of both *Caenorhabditis elegans* sexes." *Nature* 571.7763 (2019): 63-71.

<https://www.nature.com/articles/s41586-019-1352-7>

推荐理由：首次完整绘制了雌雄两种线虫的全身神经连接组。揭示了性别特异性的神经环路差异。为理解神经系统的性别差异提供了重要参考。

[18] Bouaziz, Sofien, et al. "Projective dynamics: Fusing constraint projections for fast simulation." *Seminal Graphics Papers: Pushing the Boundaries*, Volume 2. 2023. 787-797.

<https://dl.acm.org/doi/abs/10.1145/3596711.3596794>

推荐理由：提出了一种新的物理模拟算法投影动力学。通过融合约束投影实现快速仿真。在计算效率和稳定性方面取得重要突破。

[19] Wen, Quan, et al. "Proprioceptive coupling within motor neurons drives *C. elegans* forward locomotion." *neuron* 76.4 (2012): 750-761.

[https://www.cell.com/neuron/fulltext/S0896-6273\(12\)00805-7](https://www.cell.com/neuron/fulltext/S0896-6273(12)00805-7)

推荐理由：发现了线虫运动神经元内的本体感受耦合机制。阐明了前向运动的神经控制原理。为理解动物运动控制提供了新见解。

[20] Xian Z, Qiao Y, Xu Z, Wang TH, Chen Z, Zheng J, Xiong Z, Wang Y, Zhang M, Ma P, Wang Y, Dou Z, et al. *Genesis: A Generative and Universal Physics Engine for Robotics and Beyond*. arXiv preprint. 2024.

<https://genesis-embodied-ai.github.io/>

推荐理由：Genesis是一个突破性的生成式物理引擎，由CMU、清华等20多家机构联合开发。它实现了超高效的物理模拟、精确的物理建模和生成式架构，为机器人研究和具身智能发展开创新范式。

Science for AI

2.7 物理世界的第一性原理

2.7.1 背景介绍

目前的人工智能领域发展迅猛，展现出强大的能力。然而，这些基于深度学习和大模型的人工智能系统存在一个共同的弱点：它们大多是通过堆积数据和算力的手段构建的，缺乏对人类和自然智能系统本质的深刻洞察。例如，现有人工智能系统能耗巨大，而人类大脑却能以远低于此的能耗完成各种复杂任务。由人类设计的人工智能，和自然演化下的生物智能，包括我们人类的大脑，是否存在统一的运作机制？智能的本质是什么？我们是否能设计出更有效的智能系统？要回答这些问题，我们迫切需要寻找复杂智能系统的第一性原理。

物理学的思想常常被用来指导AI模型的构建。2024年诺贝尔物理学奖授予John Hopfield和Geoffrey Hinton，表彰他们“通过人工神经网络实现机器学习的基础性发现和发明”。这两位研究者将统计物理中的经典模型应用于神经网络架构设计，从而大大提升了机器学习问题的求解效率，为人类进入深度学习时代开辟了先河。

统计物理是连接微观与宏观的桥梁，提供了研究物质世界宏观性质的强有力的理论工具。在机器学习中，统计物理的思想常被用来设计处理具有随机性的模型，统计物理学中的许多概念（比如熵、混沌）和一些经典物理模型（比如伊辛模型）常被借鉴到机器学习中。在神经网络领域，基于能量的模型（Energy

-Based Models, EBM）通过引入物理学的能量概念来构建学习框架。这类模型通过最小化系统能量状态的方式进行网络训练，将能量函数与损失函数相关联，从而提升网络性能和泛化能力。典型的应用案例包括受限玻尔兹曼机（RBM）和深度信念网络（DBN），其核心架构正是基于能量模型的原理设计而成。[6] 生成模型（Generative Model）旨在学习数据样本背后的概率分布，并据此随机采样生成新的样本。目前主要的生成模型包括扩散模型、自回归模型、流模型、变分自编码器等。[5]

本部分将从物理的第一性原理和数学的第一性原理两个方面，介绍关于复杂智能系统的第一性原理的三种可能理论方向：

- **自由能原理（Free Energy Principle, FEP）** 根源于统计物理和热力学，基于自由能原理的主动推理被认为是智能的第一性原理的备选方案，可能成为新时代复杂系统的大统一理论。通过自由能原理，我们可以更深入地理解智能系统如何运作，为未来的人工智能研究和发展提供更加深刻和基础的理论支持；
- 为了展示物理的普适原理如何帮助我们改进AI系统，我们介绍了**泊松流模型（Poisson Flow Generative Model, PFGM）**这一结合了物理原理和深度学习的生成模型。

- **范畴论（Category theory）** 被称为是“数学的数学”，为现代数学提供了一套基础语言。范畴论被用来刻画大模型的能力边界，理解智能的本质，并为自我意识提供数学定义。

从物理或数学的第一性原理出发，我们希望站在复杂智能系统的普适视角更好地理解 AI 大模型的工作原理，从而为改进模型以及提高模型的可解释性提

供更高维度的洞察。

2.7.2 研究进展

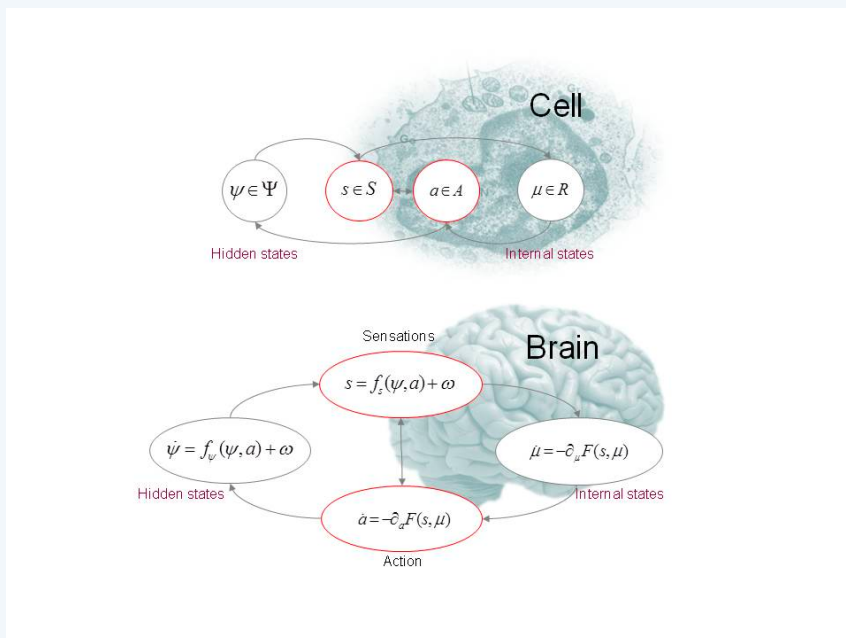
2.7.2.1 自由能原理：智能的第一性原理

推荐理由：自由能原理被认为是“自达尔文自然选择理论后最包罗万象的思想”[1]，从第一性原理出发解释智能体更新认知、探索和改变世界的机制，被认为有可能成为智能的第一性原理的重要候选方案，并有望成为新时代复杂系统的大统一理论。

生命何以存在？从微小的细菌到复杂的多细胞动物，再到具有高度智能的人类，各种形式智能背后是否存在统一的普适原理？著名理论神经科学家 Karl Friston 提出了自由能原理（Free Energy Principle, FEP），尝试从第一性原理出发解释智能体更新认知、探索和改变世界的机制。自由能原理基于统

计物理和热力学，认为任何与环境处于平衡的自组织系统都必须使其自由能最小化，即最小化内部状态与外部环境预期之间的差异。该原理本质上是对自适应系统（即生物智能体，如动物或大脑）如何抵御自然无序趋势的数学表述。基于自由能原理的主动推理（Active Inference）提供了一个建模感知、学习和决策过程的统一框架。[2]

人工智能领域中的强化学习与以自由能原理基础的主动推理框架存在紧密联系。例如在强化学习世界模型中，智能体首先通过观测数据推断隐状态的动力学模型，然后基于学习到的世界模型进行决策规划和探索。在面对复杂环境时，智能体往往需要学习一个多尺度的世界模型。这一发现启示我们，不论是自然演化下的生物智能，还是由人类设计的人工智能，在运作的底层逻辑上或许殊途同归。[3]



自由能原理尝试从第一性原理出发为生命和智能提供统一的解释 | 来源：Karl Friston

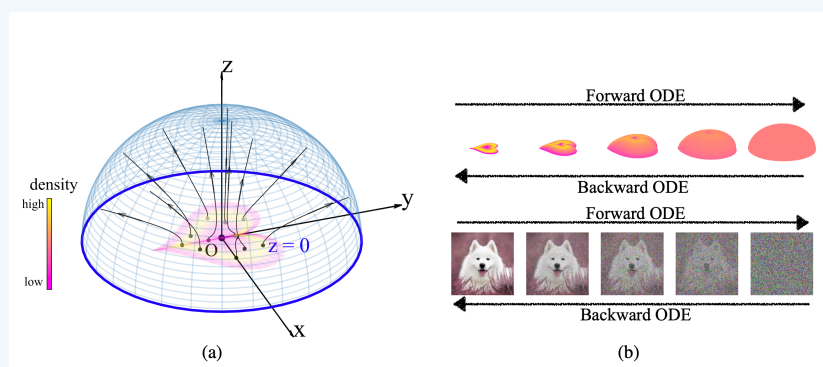
Karl Friston等人在2024年发表的论文Designing Ecosystems of Intelligence from First Principles中，展望了未来人工智能的发展愿景，认为人工智能的设计不应局限于模仿大脑，而应该拥抱各式各样的自然智能。在自然界，从粘菌到鱼群再到整个森林，集体智能无处不在，生物体共同协作以适应和生存，人类则通过网络分享想法。然而，大多数AI系统无法与人类或其他AI共享信息，也无法表达它们如何实现目标。基于此，超级人工智能更可能的形式是智能生态系统的分布式网络，一种可能的实现路径是通过主动推理框架构建集体智能（Collective Intelligence）。[4]

2.7.2.2 泊松流模型：物理启发的深度生成模型

推荐理由：对物理第一性原理的理解可以反过来帮助我们提出或改进AI算法。作为案例，由基本物理规律启发的深度生成模型，基于热力学的扩散模型[8]和基于静电学的泊松流模型[9]，已在高清图像、3D模型、蛋白质结构等生成任务中大放异彩。其中，泊松流模型由集智社区成员许逸伦、刘子鸣等人提出。

泊松流模型（Poisson Flow Generative Model, PFGM）能够学习数据分布的电场，并使用这个电场来生成新的数据点，从而实现图像生成。具体而言，模型将数据点（例如图像）解释为在 $z=0$ 平面上的正电荷，这些电荷在空间中产生一个高维电场，电场是泊松方程解的梯度，模型通过求解由电场引起的一对正向和反向常微分方程来实现泊松流。相比于扩散模型，泊松流模型更为鲁棒，能够生成同等或更高质量的图像，且速度快10至20倍。泊松流模型利用了电场这种物理结构，这个过程结合了物理原理和深度学习，提供了一种新的视角来理解和改进生成模型。

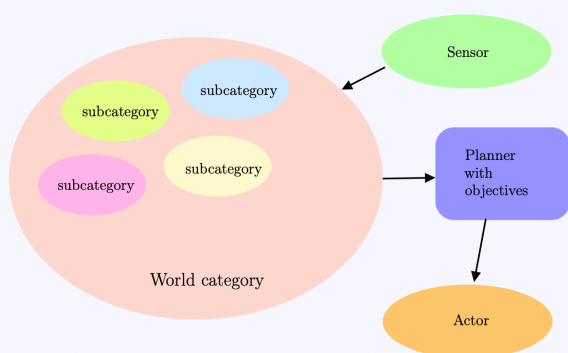
进一步，物理启发的生成模型还在原子尺度物质结构的生成与设计展现出巨大的潜力。基于和统计物理的内在联系，生成模型对于优化“大自然的损失函数”——变分自由能具有独特优势，这为求解困难的统计物理和量子多体问题提供了新的可能。同时，物理学的洞察也在推动生成模型的发展和创新。通过借鉴物理学原理和方法，还可以设计出更加高效、更加统一的生成模型，以应对人工智能领域的挑战。[5]



泊松流模型示意图。| 来源：Xu, Y., Liu Z. et al (2023). Poisson Flow Generative Models. Neural Information Processing Systems (NeurIPS)

2.7.2.3 范畴论：刻画大模型的能力边界

推荐理由：大模型的能力边界在哪里？已有的机器学习理论无法回答这个问题，因为这些理论研究的对象在这个问题中不存在了，范畴论可以为此提供一个解答。范畴论被用来刻画大模型的能力边界，理解智能的本质，为自我意识提供数学定义。[10, 11]本研究工作由集智社区成员、清华大学交叉信息学院袁洋完成。



理解智能系统的一种范畴论框架 | 来源：Yuan, Yang. A Categorical Framework of General Intelligence. arXiv preprint arXiv:2303.04571 (2023)

复杂智能系统的第一性原理，也可能包含在最新的数学理论分支之中——范畴论。假如我们有无限的资源，比如有无穷多的数据、无穷大的算力、无穷大的模型、完美的优化算法与泛化表现，由此得到的预训练模型是否可以用来解决一切问题？集智社区成员、清华大学交叉信息学院助理教授袁洋在On the Power of Foundation Models这篇论文中，通过范畴论视角，对基础模型的能力进行了理论分析，提出了限制和可能性，并提供了一个用于监督和自监督学习的范畴论框架。[10]

具体而言，范畴论提供了一种鸟瞰视角来理解监督学习。它本身不会替代原有的更具体的监督学习框架，或者用来产生更好的监督学习算法，而是将监督学习框架作为“子模块”。它更关注范畴与函子的结构，并尝试理解某个函子是否可学习。这些问题在传统监督学习框架中极为困难，但是在范畴论视角下变得简单。从机器学习的角度来看，自监督学习也是一种监督学习，只是获取标签的方式更巧妙一些而已。而从范畴论的角度来看，自监督学习定义了范畴内部的结构，监督学习定义了范畴之间的关系。因此，它们处于人工智能地图的不同板块，在做完全不同的事情。

2.7.3 挑战与展望

通过自由能原理，我们可以更深入地理解智能系统如何运作，为未来的人工智能研究和发展提供更加深刻和基础的理论支持。例如，人工智能领域中的强化学习与以自由能原理为基础的主动推理框架存在紧密联系。在强化学习领域，新算法的设计大多依赖启发式的想法，缺乏第一性原理，自由能原理则为不同强化学习算法提供了一个统一的视角。[3]范畴论则有可能帮助我们从小学的角度理解大模型的局限性[10]，并有望提炼、表述智能系统的第一性原理。

PNAS杂志近期发表的“物理学遇见机器学习”（Physics Meets Machine Learning）主题特刊引言指出，统计物理学在人工神经网络发展的初期推动了诸多关键进展，但最近深度学习神经网络追求快速的应用驱动型发展，忽视了理论基础，深度学习的下一个突破可能来自基于统计物理学概念和方法建立的坚实理论基础。[7]随着更多物理模型被深度挖掘，人工神经网络中何以涌现出复杂智

能、深度学习神经网络究竟学到了什么等人工智能谜题将逐渐清晰。

除此之外，通过借鉴量子力学的基本特性，如量子叠加或量子纠缠，人们提出了各种量子AI算法。量子机器学习结合了量子计算的速度和机器学习的学习能力，通过模拟量子力学的基本原理，在处理数据时展现出了传统算法无法比拟的潜力。量子深度学习使用量子计算代替传统的冯·诺依曼机计算，使深度学习算法实现了量子化，达到了显著提高算法并行性和降低计算复杂度的目的。量子演化算法模仿生物进化过程，大大提高了算法的并行性和收敛速度。[6]

拥有大规模参数和复杂网络结构的AI大模型是一个典型的复杂系统，展现出一系列复杂系统常见的现象规律，如涌现（Emergence）、标度律（Scaling Law）等。采用复杂科学相关的研究手段，如统计物理、复杂适应系统、复杂网络、非线性动力学、范畴论等工具，有助于我们理解和改进AI大模型，从第一性原理出发揭示智能系统底层的统一规律。

2.7.4 参考文献

[1] The Genius Neuroscientist Who Might Hold the Key to True AI
<https://www.wired.com/story/karl-friston-free-energy-principle-artificial-intelligence/>

推荐理由：《连线》杂志的这篇文章主要介绍了神经科学家Karl Friston的工作，尤其是自由能原理，这可能是通向真正人工智能的关键。

[2] Friston, Karl. The free-energy principle: a unified brain theory?. *Nature reviews neuroscience* 11.2 (2010): 127-138.
<https://www.nature.com/articles/nrn2787>

推荐理由：该综述从自由能原理视角审视了生物和物理科学中一些重要的大脑理论，包括贝叶斯大脑、预测编码和信息最大化原则、层次推理和注意力理论、神经达尔文主义、信息论和最优控制理论等，表明几种全局性大脑理论可能在自由能框架内得到统一。

[3] Mazzaglia, Pietro, et al. The free energy principle for perception and action: A deep learning perspective. *Entropy* 24.2 (2022): 301.

<https://www.mdpi.com/1099-4300/24/2/301>

推荐理由：这篇论文提供了自由能原理的概述，并从深度学习的角度探讨了感知和行动的免费原理。论文讨论了深度主动推断主体如何学习环境的生成模型，并与之互动。

[4] Friston, Karl J., et al. Designing ecosystems of intelligence from first principles. *Collective Intelligence* 3.1 (2024): 26339137231222481.

<https://journals.sagepub.com/doi/10.1177/26339137231222481>

推荐理由：文章探讨了如何从第一性原理出发设计智能生态系统，提出使用主动推理作为研究和开发人工智能的一种方法，旨在发展自然和人工智能的生态系统。尽管神经科学是AI研究的关键灵感来源，但我们必须超越大脑，拥抱自然智能的主动和层级特性，这些特性存在于生物体中，也可能在更广泛的物理系统中实现。

[5] 写给物理学家的生成模型。王磊，张潘，《物理》2024年第6期。

<https://wuli.iphy.ac.cn/article/doi/10.7693/wl20240602>

推荐理由：文章从物理的视角介绍扩散模型、自回归模型、流模型、变分自编码器等现代生成模型。

[6] Jiao, Licheng, et al. AI meets physics: a comprehensive survey. *Artificial Intelligence Review* 57.9 (2024): 256.

<https://link.springer.com/article/10.1007/s10462-024-10874-4>

推荐理由：这篇综述深入探讨了物理学与人工智能发展的相互促进。重点讨论了物理学如何启发AI深度学习范式，介绍了物理学的四大领域（经典力学、电磁学、统计物理、量子力学）如何启发现代机器学习模型的构建，并简要介绍了AI如何解决物理学问题的相关工作。

[7] Levine, Herbert, and Yuhai Tu. Machine learning meets physics: A two-way street. *Proceedings of the National Academy of Sciences* 121.27 (2024): e2403580121.

<https://www.pnas.org/doi/10.1073/pnas.2403580121>

推荐理由：文章关注机器学习领域与物理学之间的双向互动，一方面是机器学习能为物理学做些什么，另一方面则是物理学能为机器学习做些什么。文章认为，人工神经网络受益于统计物理学和神经科学这两个自然科学学科，深度学习的下一个突破可能来自基于统计物理学概念和方法建立的坚实理论基础。

[8] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020): 6840-6851.

<https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html>

推荐理由：本文介绍了去噪扩散概率模型（Denoising Diffusion Probabilistic Model, DDPM）这种生成模型。DDPM 是一种受非平衡热力学启发的潜变量模型，用于高质量图像合成，可以说是 AI 绘画的基石。DDPM 的训练过程包括前向过程（不断加噪声）和反向过程（从噪声生成图像）。

[9] Xu, Y., Liu Z. et al (2023). Poisson Flow Generative Models. Neural Information Processing Systems (NeurIPS)

https://proceedings.neurips.cc/paper_files/paper/2022/hash/6ad68a54eaa8f9bf6ac698b02ec05048-Abstract-Conference.html

推荐理由：这篇论文提出了一个由静电力学启发的生成模型 PFGM。该模型比扩散模型更为鲁棒；且在有扩散模型相同或者更优的生成质量情况下，生成速度更快。

[10] Yuan, Yang. On the power of foundation models. International Conference on Machine Learning. PMLR, 2023.

<https://proceedings.mlr.press/v202/yuan23b.html>

推荐理由：这篇论文通过范畴论视角刻画了大模型的能力边界，并提供了一个用于监督和自监督学习的范畴论框架。

[11] Yuan, Yang. A Categorical Framework of General Intelligence. arXiv preprint arXiv:2303.04571 (2023).

<https://arxiv.org/abs/2303.04571>

推荐理由：机器是否能思考？由于缺乏通用智能的坚实数学基础，至今无人能直接回答这一问题。本文提出一个关于通用智能的范畴论框架，利用范畴论，论文中许多通用智能的重要概念可以被严格定义和分析。

2.8 科学启发的 AI 新架构

2.8.1 背景介绍

随着大型语言模型的快速发展，人工智能技术在展现巨大潜力的同时，也暴露出一系列问题。一方面，以算力积累为核心的AI发展模式正面临能源消耗和硬件架构可扩展性等制约因素；另一方面，AI架构的可解释性问题成为AI发展方向和关键应用场景中的主要障碍。人类科学家积累的大量领域知识，尤其是数学和物理的基本原理，已被应用于AI模型架构中，这不仅增强了AI模型的可解释性，也提升了其性能。

基于神经网络的传统AI模型往往依赖于大量数据，缺乏对数据背后科学规律的理解，导致在物理一致性要求高的科学领域表现不佳。在AI架构改进方面，将神经网络置于数学或物理框架内，并运用前沿的数学物理工具来理解其工作机理，是至关重要的。

这一过程常涉及非凸优化、表征能力和泛化能力等前沿理论研究，从而推动新模型的提出。例如，ResNet内部的多个残差网络模块的工作机制可以解释为常微分方程的离散数值求解器，这一理念启发了ODE-Net的发明。此外，通过在损失函数中引入正则项，可以强化模型的解耦性、单调性、非负性和稀疏性等已知的先验属性。还有一些工作致力于改造神经网络本身，如KAN模型，它在激活层引入了参数化的多项式函数，替代了传统的Sigmoid和ReLU等固定激活函数，以更高效地拟合具有多项式形式的未知目标函数。

传统的数字神经网络依赖于电子计算，这涉及到大量的算术运算和数据传输，这些都是能耗密集型

的。因此颠覆传统的神经网络架构是突破传统人工智能系统能效瓶颈的重要方法。

神经科学的研究成果为AI新架构的启发提供了重要源泉。类脑计算（Brain-inspired Computing），或称神经形态计算（Neuromorphic Computing），是指模仿生物神经系统的计算范式，旨在为深入探索大脑功能及推进人工智能系统的发展提供创新途径。在该领域的初期研究中，主要目标是在硬件层面上复现或验证生物学发现，并利用神经科学的研究成果指导相关系统的设计和开发。近期，随着大数据、人工智能以及物联网等技术的迅猛发展，对计算能力提出了更高的要求。在此背景下，科研人员开始探索脑启发式计算作为一种可能的解决方案，以期构建更为高效的计算系统。以脉冲神经网络（SNNs）为代表的类脑计算模仿了生物大脑的工作方式，信息通过脉冲传递，这种事件驱动的计算方式使得SNNs在处理稀疏数据时极为高效，显示出在能效上的潜在显著优势。

机械神经网络（Mechanical Neural Networks, MNN），以及储备池计算（Reservoir Computing）都强调系统的非线性动力学特征，利用复杂的动态响应来处理信息，并且都试图通过独特的网络连接和动态来实现计算。机械神经网络采用全新的物理架构作为信息处理器，在提升AI系统的能效比方面具有巨大潜力。储备池计算结合了循环神经网络和前馈神经网络，其中所谓的“储备池计算”模拟了循环神经网络的工作机理，但其无需训练的特性，使得利用特别设计的物理系统实现储备池计算成为可能，从而大大降低系统工作能耗。

下面介绍的是五个代表性的工作：

· Komogorov Arnold Networks (KAN) 作为一种深度学习架构，受数学理论启发，能模拟复杂函数，提升模型的泛化能力和可理解性。

· 脉冲神经网络 (SNN) 通过模仿生物大脑的神经系统工作机理，有效降低了计算能耗，并适合于实时处理任务。

· 储备池计算 (RC) 则利用非线性动力系统对输入数据进行高维映射，特别适合处理时间序列数据。RC的固定的随机非线性动力系统非常适合于特殊设计的高能效硬件架构，这使得RC在实现高效信息处理方面具有巨大潜力。

· 机械神经网络 (MNN) 是一种模仿生物神经系统结构和功能的物理系统，通过特定材料和几何结构设计，能够像神经网络一样处理和传递信息，具有自适应和响应外部刺激的能力，主要应用于柔性机器人、智能材料和仿生工程领域。

· 傅里叶神经算子 (FNO) 开创了一种全新的科学AI范式，通过在频域空间学习算子映射，实现了分辨率无关的快速仿真，特别适合处理由偏微分方程描述的物理系统。

2.8.2 研究进展

2.8.2.1 Komogorov Arnold Networks (KAN)

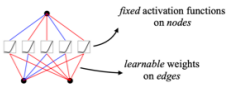
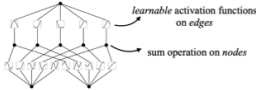
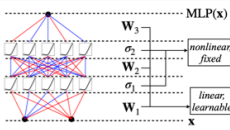
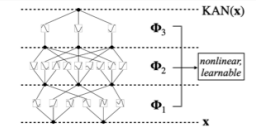
推荐理由：KAN是一种受数学理论启发的全新深度学习架构，可以替代一般的前馈神经网络，具有表示能力更强、更容易进行复杂函数拟合、更适科学计算等优点；由集智社区成员、集智科学家

刘子鸣为首的麻省理工学院科研团队开发。

Komogorov Arnold Networks (KAN)[1][2]的核心思路源自一个数学理论，即科尔莫戈罗夫-阿诺德 (Kolmogorov-Arnold) 定理。该定理表明，任何复杂的多变量函数都可以通过一系列简单函数的组合来表示。这意味着，在理论上，即使是高度复杂的关系，我们也可以将其拆解成一组简单函数叠加的形式，从而进行建模和预测。

KAN将上述定理描述的单层网络叠加为多层的深度网络。与传统多层前馈神经网络 (MLP) 相比，KAN的核心创新在于将MLP中相对固定的激活函数替换为更加丰富的数学函数集。这种设计使其能够以更简洁的网络结构建模复杂函数关系，显著提升了模型表达能力。这不仅能够提高模型的区间外插值能力，在特定问题上增强了分布外泛化能力，同时还增加了模型的易理解性，更适于在科学计算中应用。

KAN的架构通过一个称为“隐藏函数”的层将输入数据转化，然后通过多个简单函数的组合输出结果。这种方式使得它比传统神经网络更具解释性和可控性，因为我们能够清晰地看到每个简单函数在整体架构中的作用，而不需要依赖大量的参数训练。

Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(c)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) 	(b) 
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\text{KAN}(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$
Model (Deep)	(c) 	(d) 

传统多层感知机 (MLP) 结构和KAN结构的对比[1]

2.8.2.2 脉冲神经网络

推荐理由：相比当前主流AI的技术实现，人脑在高效、低功耗和实时处理能力等方面都表现出无以伦比的优势。脉冲神经网络（Spiking Neural Networks, SNN）是一种模仿生物大脑神经系统结构和功能的计算方法，有望解除AI在耗能、实时高效信息处理等方面的发展制约。

脉冲神经网络被称为第三代人工神经网络[3]，模仿了生物神经元之间的脉冲发放方式。相比传统的人工神经网络（ANN），SNN通过离散的脉冲（Spike）进行信息传递，从而有效降低计算能耗，能更高效地模拟生物神经系统的时空特性。SNN通常采用事件驱动机制，仅在事件触发时激活神经元，可有效降低能耗。此外，SNN还可模拟突触可塑性，通过模拟大脑的学习机制，使得网络能够在动态环境中实现实时学习。

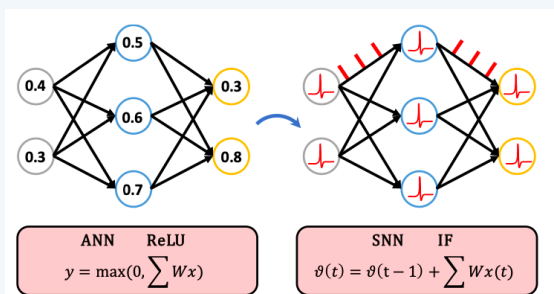


Fig. 1. ANN-to-SNN conversion, a mapping between real-valued activation neurons and spiking neurons.

从人工神经网络（ANN）转换到脉冲神经网络（SNN）[3]

相比传统的人工神经网络，采用脉冲（Spike）传递信息，事件驱动的SNN在能效比和计算资源消耗方面具有明显优势。SNN能够在数据发生变化时才进行计算，这种计算模式对于实时处理和响应是非

常有价值的。SNN可以与专门为SNN设计的神经形态硬件（如IBM的TrueNorth、Intel的Loihi）更紧密地集成，实现更高的能效和并行处理能力。此外，SNN模仿生物大脑的工作方式，使得它们在需要与生物系统接口的应用中（如脑机接口）更为有效。

2.8.2.3 储备池计算

推荐理由：储备池计算（Reservoir Computing, RC）是一种可以替代传统循环神经网络（RNN）的新型AI模型，更擅长于从时间序列中学习复杂的动力学。相比较普通的RNN，储备池计算具有训练效率更高，更容易用低成本硬件实现等特点。

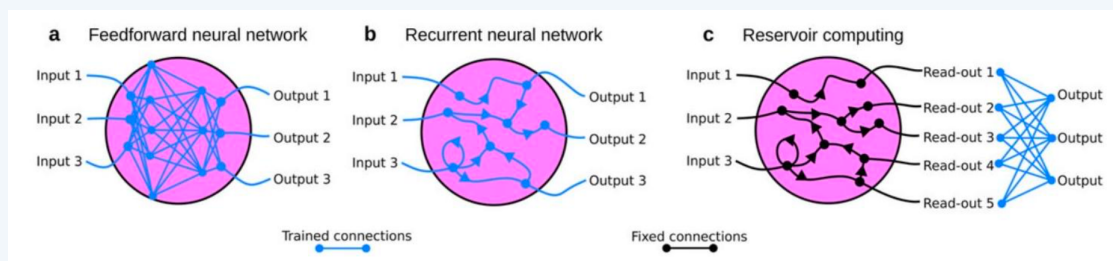
储备池计算的核心思想是通过非线性动力系统（即“储备池”）对输入数据进行高维非线性映射，使得数据的特征更容易被线性分离。它是一种特殊的RNN形式，通常使用结构固定的随机连接网络作为“储备池”（Reservoir），储备池后面连接有可训练的输出层，从而将内部状态线性映射回低维的输出变量。由于储备池网络结构和参数均固定，因此无需对其训练，而只需要训练输出层即可。这不仅可以大大提高储备池模型的训练效率，更为研发相应的专有硬件提供了便利。储备池计算通过输入驱动储备池内部的神经活动，并利用输出的线性或非线性的组合映射到低维的输出变量来完成任务[4][5]。因为储备池本身就是一个高维的动力系统，因此它更适合拟合时间序列中的复杂高维动力学。

储备池计算的发展分为三个阶段：1) 2001年提出原型：回声状态网络（Echo State Network, ESO），前十年主要研究全数字算法；2) 随后十年探索数模混合实现，包括不同架构和模拟域的物理节

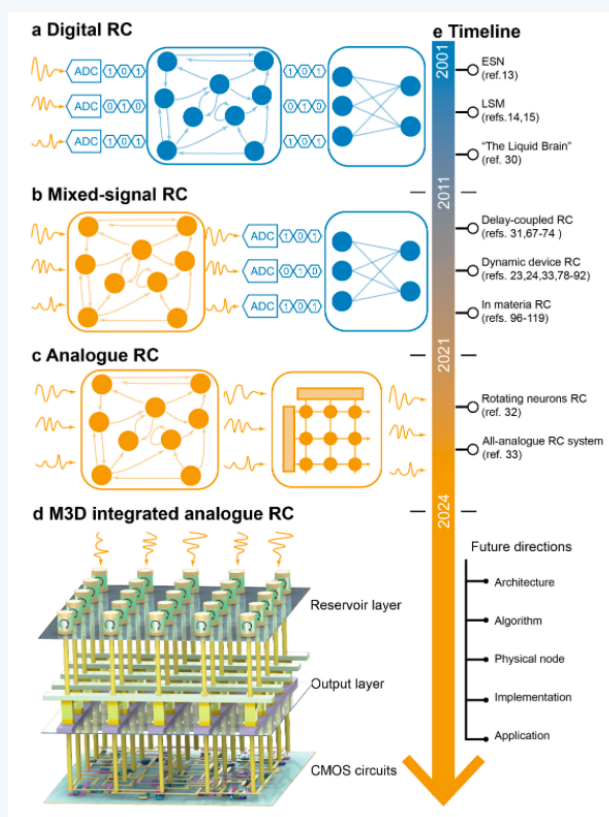
点；3) 最近几年转向全模拟系统，使用模拟型忆阻器阵列降低功耗和成本[6]。

储备池计算主要用于处理时间序列数据的任务，如语音识别、金融预测、机器人控制等。其结构简

单、训练效率高，尤其适合实现低成本的嵌入式系统。目前，物理储备池计算正快速发展，其原理和可行性已被验证，未来在多个方面有巨大潜力，性能和应用领域将进一步扩展。



前馈神经网络，循环神经网络，以及储备池计算的关系[10]



储备池计算的发展脉络与机遇 [6]

2.8.2.4 机械神经网络

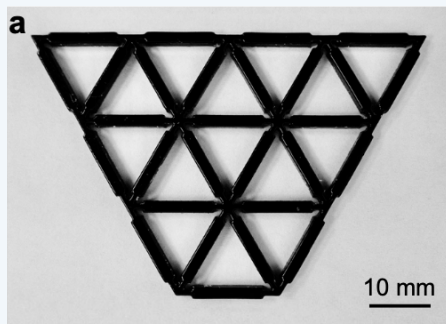
推荐理由：机械神经网络（Mechanical Neural Networks, MNN）能够在不依赖电子系统的情况下，通过材料本身的物理特性实现信息处理和智能响应，能够实现信息处理、自适应和智能响应，同时具有比传统电子神经网络更高的柔性、生物相容性和环境适应性。

机械神经网络是一种基于物理学原理的机器学习架构，通过节点之间的机械连接实现信息的处理和存储。其核心设计思路是将计算操作映射到物理行为上，以节点之间的弹簧连接为基础，将力与位移之间的响应关系模拟为神经元之间的激活与传播过程。节点的空间位置在施加外力时发生变化，通过物理连接进行信号传输，输出节点的位移代表网络的输出结果。这种机制与传统的计算机神经网络类似，但依赖于物理过程而非电子计算。

MNNs的主要架构由节点（机械神经元）、弹簧（连接权重）和输入输出节点（功能类似于感知器的输入和输出层）组成。施加到输入节点的力被视为输入信号，节点在力的作用下发生位移，表示激活程度。这些力通过弹簧传递到其他节点，决定了输出节点的最终位移。输入节点的力和输出节点的位移都是矢量，使得MNNs能够在高维空间中并行处理复杂任务。

与传统神经网络不同，MNNs具有多种独特优势：

1. 高效能与节能性：由于计算过程基于物理运动，MNNs的计算能耗极低，尤其适合在能源受限的环境中运行。



使用Polyjet橡胶类材料Agilus30的3D打印机械神经网络[11]

2. 结构简单，硬件实现成本低：MNNs依赖于简单的机械结构，减少了对复杂电路和集成芯片的需求。

3. 鲁棒性与自适应性：由于力和位移的物理特性，MNNs在电磁环境复杂或噪声干扰强的环境中具有天然优势。

4. 可扩展性与模块化：不同结构和节点布局的MNNs可以通过调整弹簧常数和连接拓扑来适配多种任务需求。

这些特点使MNNs成为未来物理神经网络硬件平台的潜在核心技术，有望在智能材料、机器人和自适应设备等应用领域中发挥重要作用。

一篇最新的工作[11]提出了一种基于伴随变量法（Adjoint Variable Method）的“原位反向传播”（In-situ Backpropagation）方法，开创性地解决了MNNs的梯度计算与参数更新难题。传统的神经网络训练依赖于计算机上的梯度下降算法，而物理系统中的参数更新一直受到精度和操作复杂度的限制。该研究提出了一种能够在物理环境中直接计算梯度的方法，大大提升了MNNs的训练效率和可扩展性。研究者成功解决了仿真与实际物理结构之间的误差问题，

为未来的机械神经网络研究提供了可靠的技术路径。

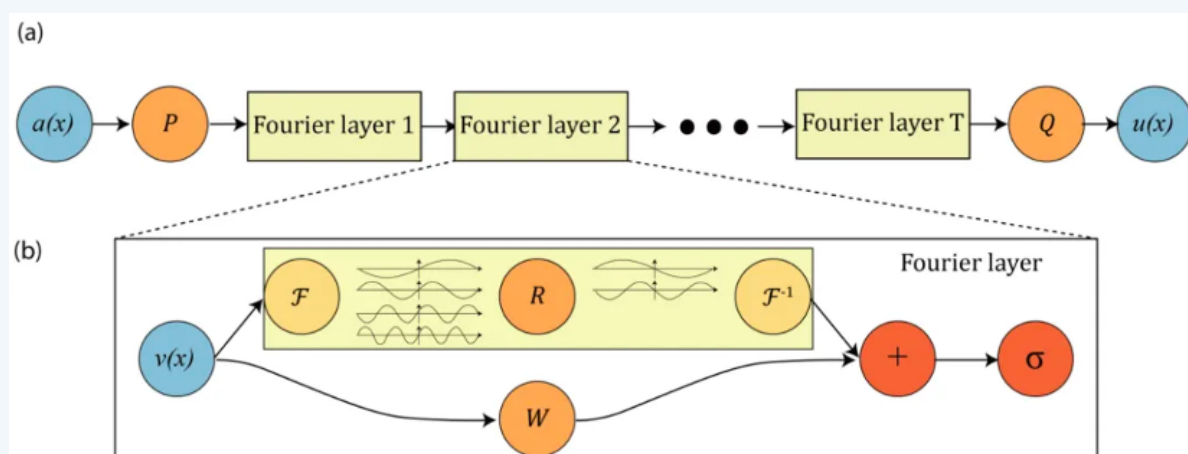
2.8.2.5 傅里叶神经算子

推荐理由:傅里叶神经算子 (Fourier neural operator, FNO) [12]开创了一种全新的科学AI范式, 通过在频域空间学习算子映射, 实现了分辨率无关的快速仿真, 特别适合处理由偏微分方程描述的物理系统。本方法由集智社区成员、加州理工学院博士生李宗宜等人提出。

傅里叶神经算子(FNO)的核心创新在于它彻底改变了神经网络处理物理问题的方式。这是一种新的深度学习架构, 能够学习无限维函数空间中的映射。如下图所示:

FNO的处理流程主要分为三步: 1. 通过线性层进行升维; 2. 应用四层积分算子和激活函数; 3. 再将输出特征投影回目标维度。其中傅立叶层对特征进行傅里叶变换, 然后对低频部分进行线性变换再应用反傅立叶变换; 而对高频部分, 则进行局部线性变换, 并将高频信息和低频信息相加输入激活函数, 再进行反傅里叶变换。这种设计既保持了物理量的守恒性, 又能显著提升计算效率。

FNO已经被应用于许多重要的科学领域, 其中在天气预报领域, 基于FNO提出的FourCastNet [13]使用Transformer模型, 搭建了人工智能和物理学之间的桥梁, 并取得了突破性的成果。它比传统的数值天气预报 (NWP) 模型快约45000倍。在训练时, 与NWP模型中的标杆——欧洲综合预报系统相比, FourCastNet预测天气所需的能耗减少了12000倍。更令人兴奋的是其在碳捕集与封存(CCS)领域的



傅立叶神经算子的网络架构 | 来源: Li Z, Kovachki N B, Azizzadenesheli K, et al. Fourier Neural Operator for Parametric Partial Differential Equations. International Conference on Learning Representations.

应用——Nested FNO[14]将复杂的地质流动预测加速了近70万倍，为大规模CCS技术的工业部署扫清了计算瓶颈。

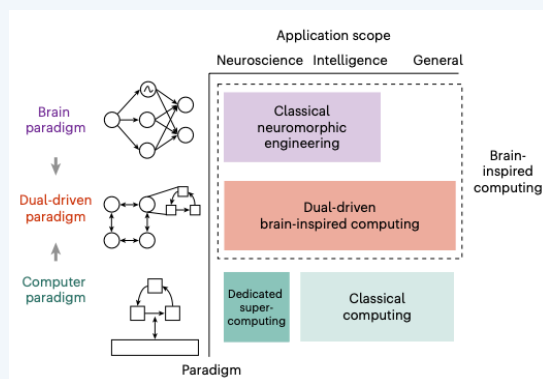
这些成功案例表明，FNO不仅是一个理论突破，更是一个能够解决实际工程问题的强大工具，也在流体动力学、材料科学、地球科学、生物医学工程等领域发挥了重要的作用[15]。

2.8.3 挑战与展望

KAN可以被归类为物理引导的机器学习（Physics-Informed Machine Learning, PIML）。作为正在发展的新型AI架构，在气候模拟、材料科学和生物医学中具有广阔前景[7]。PIML的主要优势在于可以减少对大量数据的依赖，尤其适用于数据稀缺或昂贵的领域。PIML面临的挑战主要在于如何有效嵌入物理约束，提升复杂系统中的计算效率和模型泛化能力。此外，PIML模型的有效性和精度往往依赖于物理模型的准确性，因此在物理模型不够完善时可能影响结果的可靠性。

PIML主要聚焦于AI算法层面的优化，而类脑计算则通过硬件架构革新（如神经形态芯片）与软件栈重构（涵盖操作系统、算法设计等），构建起完全不同于传统体系的计算范式。类脑计算以大脑范式为指导，复制生物神经系统的特征或功能，用于探索神经科学和加速智能任务。此外，融合类脑计算和传统计算机固有的计算范式，可以高效、灵活地处理更广泛的通用任务。

类脑计算的挑战在于实现计算的通用性，有三种可能的实现策略：应用级模式通用化、硬件级结构通用化和软件级系统通用化。类似SNN等技术最初是为



高效运行特定算法而设计的，但越来越多的系统被设计为可同时支持ANN和SNN计算，扩大了系统可支持的应用范围，从而为最终达到通用目的提供了可能[8]。

储备池计算（RC）展现出利用新兴电子技术实现高效信息处理的潜力，其未来发展将拓展至电子、光学及机械系统的多物理实现，从而构建能效突破传统架构的新型计算平台。这类系统可实现低功耗和实时数据处理，尤其适合神经形态计算。RC面临的挑战主要包括系统的稳定性、适应性以及如何提高其在复杂任务下的泛化能力和计算精度等问题。

关于机械神经网络（MNN）的最新研究[11]为其开发和部署铺平了道路，提出了机械网络在智能材料、软体机器人和自适应机械设备中的广泛应用前景。研究表明，未来可通过改进弹簧常数的动态调整机制进一步扩展该技术，设计具有高度自动化和自主学习能力的材料与设备。此外，通过研究非线性材料的响应，探索更复杂的数据集处理和决策任务，将成为未来的研究重点。这些创新为将机器学习与物理材料科学融合开辟了新的研究方向，或将在新一代节能计算硬件和智能系统的开发中发挥关键作用。

2.8.4 参考文献

[1] Liu, Ziming, et al. "Kan: Kolmogorov-arnold networks." arXiv preprint arXiv:2404.19756 (2024).

<https://arxiv.org/abs/2404.19756>

推荐理由：这篇论文提出了多层感知器（MLPs）的替代方案 KAN，通过在边缘（而非节点）上设置可学习的激活函数，并用样条参数化的单变量函数替换传统的线性权重，从而在准确性和可解释性方面超越MLPs。

[2] Liu, Ziming, et al. "Kan 2.0: Kolmogorov-arnold networks meet science." arXiv preprint arXiv:2408.10205 (2024).

<https://arxiv.org/abs/2408.10205>

推荐理由：这篇论文提出了KAN 2.0，在KAN的基础上引入了乘法节点、kanpiler编译器和树形Transformer架构，增强了模型的表达力、实用性和可解释性，并拓展了其在科学发现领域的应用范围。

[3] Hu, Yangfan, et al. "Toward Large-scale Spiking Neural Networks: A Comprehensive Survey and Future Directions." arXiv preprint arXiv:2409.02111 (2024).

<https://arxiv.org/abs/2409.02111>

推荐理由：这篇论文提供了对大规模尖峰神经网络（SNN）的全面综述，并探讨了未来的发展方向。

[4] Jaeger, H. (2001). "The "echo state" approach to analysing and training recurrent neural networks." Technical report, GMD German National Research Institute for Computer Science.

<https://www.ai.rug.nl/minds/uploads/EchoStatesTechRep.pdf>

推荐理由：这篇论文介绍了“回声状态”方法，这是一种分析和训练递归神经网络（RNNs）的技术，它通过保持网络的状态来捕捉时间动态，并将学习任务简化为输出权重的调整。

[5] Maass, W., Natschlager, T., & Markram, H. (2002). "Real-time computing without stable states: A new framework for neural computation based on perturbations." Neural Computation.

<https://ieeexplore.ieee.org/abstract/document/6789852>

推荐理由：这篇论文提出了一种新的神经计算框架，即“实时计算无需稳定状态”，它基于高维动态系统的原理和统计学习理论，可以在通用的递归电路上实现实时处理时间变化的输入，而不需要稳定的内部状态。

[6] Liang, Xiangpeng, et al. "Physical reservoir computing with emerging electronics." Nature Electronics 7.3 (2024): 193-206.

<https://www.nature.com/articles/s41928-024-01133-z>

推荐理由：这篇论文系统梳理了基于新型电子器件的物理储备池计算的起源、发展与未来展望，并重点分析了储备池架构、物理节点、输出层等构筑一个完整物理储备池系统的关键技术。

[7] Karniadakis, G. E., et al. "Physics-informed Machine Learning." Nature Reviews Physics, 2021.

<https://www.osti.gov/pages/servlets/purl/2282016>

推荐理由：这篇论文系统地调查并总结了物理信息机器学习领域，探讨了如何将物理先验知识（如PDEs/ODEs/SDEs、对称性约束和直观物理）嵌入到机器学习模型的不同部分（数据、架构、损失函数、优化方法和推理算法），并详细呈现了现有方法、挑战和未来方向。

[8] Zhang, Weihao, et al. "The development of general-purpose brain-inspired computing." Nature Electronics (2024): 1-12.

<https://www.nature.com/articles/s41928-024-01277-y>

推荐理由：这篇论文探讨了通用脑启发计算的发展，涉及神经形态电子系统、深度神经网络的大脑相似性、视觉传感器技术、大规模神经模拟以及相关硬件和软件工具的进步。

[9] Fan, Feng-Lei, et al. "On interpretability of artificial neural networks: A survey." IEEE Transactions on Radiation and Plasma Medical Sciences 5.6 (2021): 741-760.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9105427/>

推荐理由：这篇论文是一项关于人工神经网络可解释性的综述研究，提出了一个全面的可解释性分类体系，系统回顾了提高神经网络可解释性的最新研究，讨论了可解释性在医学等领域的应用，并探讨了未来研究方向，如与模糊逻辑和脑科学的关联。

[10] Cucchi, Matteo, et al. "Hands-on reservoir computing: a tutorial for practical implementation." Neuromorphic Computing and Engineering 2.3 (2022): 032002.

<https://iopscience.iop.org/article/10.1088/2634-4386/ac7db7/meta>

推荐理由：本文旨在为材料科学、化学和电子学等领域的读者提供实施水库计算（RC）实验的概览。RC为利用材料系统的非线性行为和短期记忆设计新型计算范式提供了框架，避免了传统前馈神经网络在硬件上实现时的问题。本文强调了RC在硬件神经网络中的潜力、优势及实施障碍，并希望降低科学家利用非线性系统进行机器学习和人工智能任务的门槛。

[11] Li, Shuaifeng, and Xiaoming Mao. "Training all-mechanical neural networks for task learning through in situ backpropagation." arXiv preprint arXiv:2404.15471 (2024).

<https://arxiv.org/abs/2404.15471>

推荐理由：本文提出了一种全新的机械神经网络训练方法——原位反向传播，通过局部信息高效计算梯度，实现行为学习和机器学习任务。实验证明，该方法在回归和分类任务中达到高精度，展现了机械学习硬件的潜力。

[12] Li, Zongyi, et al. "Fourier Neural Operator for Parametric Partial Differential Equations." International Conference on Learning Representations, 2021.

<https://arxiv.org/abs/2010.08895>

推荐理由：该论文提出的傅里叶神经算子开创了频域学习的新范式，在科学计算领域产生了重要影响。

[13] Pathak, Jaideep, et al. "FourCastNet: A Global Data-driven High-resolution Weather Model Using Adaptive Fourier Neural Operators." Nature, vol. 614, no. 7947, 2023, pp. 474-481.

<https://www.nature.com/articles/s41586-023-05847-6>

推荐理由：该论文将傅里叶神经算子应用于全球天气预报，实现了显著的计算加速。

[14] Wen, Guannan, et al. "Real-time High-resolution CO2 Geological Storage Prediction Using Nested Fourier Neural Operators." Energy & Environmental Science, vol. 16, no. 4, 2023, pp. 1732-1741.

<https://doi.org/10.1039/D2EE03266A>

推荐理由：该工作将嵌套傅里叶神经算子应用于地质储存预测，显著提升了预测效率。

[15] Azzizadenesheli, Kamyar, et al. "Neural Operators for Accelerating Scientific Simulations and Design." Nature Reviews Physics, 2024, pp. 1-9.

<https://www.nature.com/articles/s42254-024-00712-5>

推荐理由：该综述全面总结了神经算子在科学仿真与设计中的应用进展。

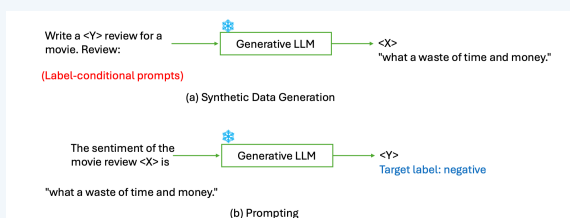
AIS 基础设施

2.9 合成数据和数据基础设施

2.9.1 背景介绍

在AI大模型时代，数据作为关键战略资源，既是模型训练的基础要素，也是制约性能提升的核心瓶颈，其质量与规模直接决定智能系统的能力边界。科学智能的发展同样面临着诸多数据挑战，其中最为关键的是数据稀缺问题[1]。一方面，在某些科学领域，往往难以获取大规模、高质量的标注数据[2]；另一方面，某些数据资源丰富的领域（如医疗）往往涉及隐私或法律限制[3]。此外，跨学科多源数据整合也面临巨大障碍，数据格式、标准和结构的差异，导致AI模型难以跨越多个领域进行通用学习，进一步加剧了数据稀缺问题。

合成数据（Synthetic Data）与**数据基础设施建设**是解决以上挑战的两个重要手段。在合成数据中，最基础也最关键的是**数据生成**（Data Generation）问题。与传统提示工程不同[13][14]，即通过提示对部署的LLMs输入文本数据X进行预测以生成标签Y，合成数据生成则要求LLMs根据条件化的标签Y提示生成文本数据X，以此适配多样化的下游任务的数据需求。



LLMs在生成特定标签的合成数据（a）与提示词预测生成标签（b）之间的比较 [13] | 图片来源: Xu Guo&Yiqiang Chen. (2024) "Generative AI for Synthetic Data Generation: Methods, Challenges and the Future"

然而，若直接将有限标签和任务信息嵌入到提示中，LLMs生成的数据可能与任务无关，缺乏多样性，规模也很有限。因此需要更先进的提示技术：如**属性控制提示**（Attribute-controlled prompt），通过指定一组属性、并在提示模板中进行属性混合，以此定义任务获取合成数据的混合软提示方法[15]，或从LLMs中直接提取特定属性的提示，并查询生成特定数据的AttrPrompt[16]；如**词汇化技术**（Verbalizer），通过将作为条件的属性类扩展为一组语义相似的提示来促进多样化数据生成，例如MetaPrompt[17]首先从LLMs获取扩展提示，然后利用丰富的提示进一步提示LLMs生成数据。

Method	Generator	Classifier	Benchmark
ZeroGen [9]	GPT2-XL [35]	LSTM [36] DistilBERT [40]	SST-2 [37], IMDb [38], QNLI [39] RTE [41], SQuAD [39] AdversarialQA [42]
ZeroGen+ [10]	GPT2-XL [35]	LSTM [36] DistilBERT [40]	IMDb [38], SST-2 [37], Amazon [43] Rotten Tomatoes [44], Yelp [45] Subj [46], AGNews [45], DBpedia [45]
SuperGen [8]	CTRL [47]	COCO-LM [48] RoBERTa [50] GPT-2 [35]	GLUE [49]
FewGen [7]	CTRL [47]	RoBERTa [50]	GLUE [49]
ReGen [12]	Condenser [51]	RoBERTa [50]	AGNews [45], DBpedia [45], MR [44] NYT [52], Yahoo [45], Amazon [43] Yelp [45], SST-2 [37], IMDb [38]
ProGen [11]	GPT2-XL [35]	LSTM [36] DistilBERT [40]	SST-2 [37], IMDb [38], Elec [43] Rotten Tomatoes [44], Yelp [45]
AttrPrompt [53]	ChatGPT [6]	BERT [3] DistilBERT [40]	NYT [52], Amazon [54] Reddit [55], StackExchange [55]
MixPrompt [13]	FLAN-T5 XXL [56]	GODEL [57]	NLU++ [58], TOPv2 [59] CrossNER [60]

LLMs生成特定属性任务训练数据的方法。同上[13]。

当然，以上合成数据生成方法，主要直接针对于文本数据等结构化的、序列化的数据，对于要保持全局一致性的复杂数据（如合成医学影像和3D图形），这些技术需要进行迁移和扩展。

除了数据生成，合成数据另外两个重要子领域是**科学数据表示**（Scientific Data Representation）和**模型自我改进方向**（Model Self-Improvement）的探索。前者重点研究如何有效地编码和组织合成数据，后者探索如何让模型通过合成数据来提升自身性

能。高效的数据表征方法通过提取关键特征提升生成数据的保真度，而模型的自迭代优化机制则能持续扩展数据覆盖范围，二者的协同作用实现了高质量与规模化数据生成的同步突破。

下文中将重点介绍关于合成数据的数据生成、科学数据表示、模型自我改进三个领域的最新进展和代表性工作，以及科学数据基础设施建设的 π -HuB项目。

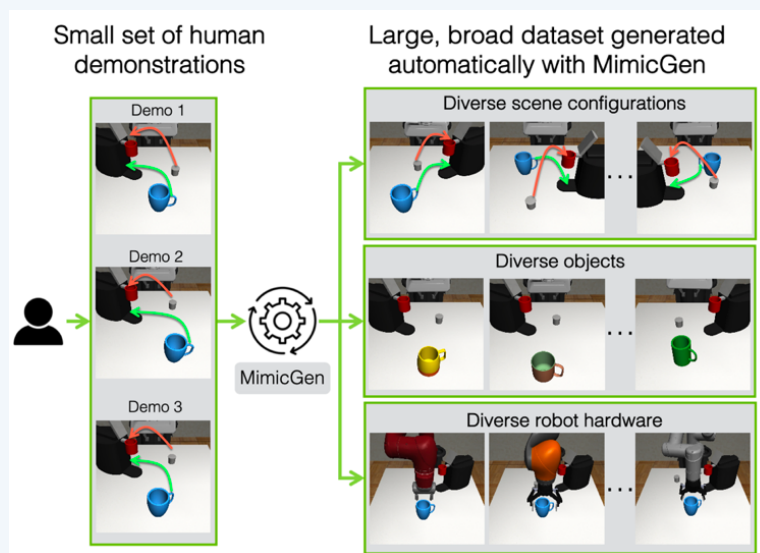
2.9.2 研究进展

2.9.2.1 合成数据生成：从序列数据到复杂数据

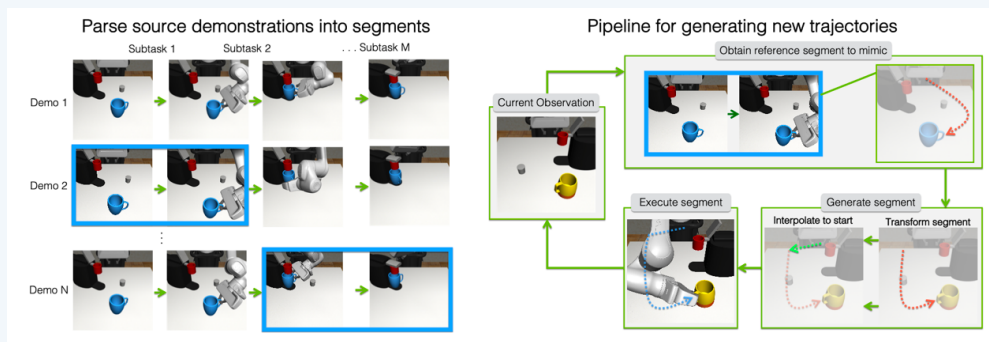
推荐理由：合成数据的核心优势不仅在于可大规模生成数据，而且可以根据特定需求进行定制。通过引入可控的变化确保不同类别数据的平衡表示，增强模型的鲁棒性。在这个领域中，从简单的序列数据到复杂的多模态数据，都出现了一些突破性的研究成果。

在众多创新工作中，英伟达于2023年开发的MimicGen系统展现出了特别的价值[4]。这个系统展示了如何从有限的人类示范中构建大规模训练数据集的有效方法。通过仅使用约200个人类演示样本，MimicGen成功生成了包含超过5万个示范的大规模数据集。这些数据涵盖了18种不同任务，并在多种场景配置、对象实例和机器人手臂操作中展现出良好的多样性。

MimicGen的工作流程中，系统首先对源数据集中的演示进行精细化解析，将每个演示分解为多个以对象为中心的子任务单元（图2左）。在生成新数据时，系统会选择合适的参考段，并将其智能地转换到新场景中，通过调整对象位姿来适应不同环境。最终，系统使用末端执行器控制器来实现这些转换后的目标位姿序列（图2右），确保生成数据的实用性和可靠性。



MimicGen合成数据生成示意图 | 图片来源: Mandlekar et al. (2023) "Mimicgen: A data generation system for scalable robot learning using human demonstrations."



MimicGen 系统流程 | 图片来源: Mandlekar et al.(2023) "Mimicgen: A data generation system for scalable robot learning using human demonstrations."

在实际应用中，这些技术可以进行跨模态整合，将文本领域的控制机制与视觉生成相结合，通过属性控制提示定义视觉目标，再利用词汇化技术扩展视觉表达，最终由视觉生成模型完成具体生成任务。这种技术迁移不仅扩展了原有技术的应用范围，也为复杂数据的生成提供了更精确的控制机制。

这种方法不仅显著提高了数据生成的效率，更重要的是保证了生成数据的质量和多样性。通过这种方式，MimicGen为解决人工智能领域中的数据瓶颈问题提供了一个可行的解决方案，同时也为未来合成数据生成技术的发展指明了方向。

2.9.2.2 科学数据表示：从图神经网络到多模态数据表征

推荐理由：科学数据表示是一个致力于开发和优化用于表示、处理和分析科学数据的方法和技术的研究领域。这一领域的核心目标是找到有效方式来捕捉和表达复杂科学数据中的本质特征和关系，使这些数据能够被机器学习模型更好地理解 and 利用。

科学数据表示早期的研究主要集中在传统的数据

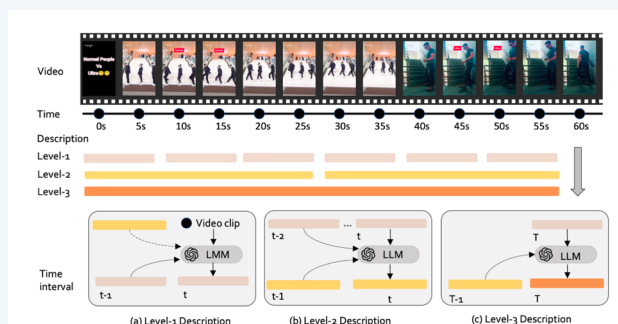
结构和统计方法上，随着深度学习和大语言模型兴起，这个领域开始向更复杂和强大的表示方法发展。图神经网络、图卷积网络等开创性地将深度学习与图结构数据处理结合，变分自编码器在生物学和物理学领域展现出强大的建模能力，自然语言处理技术的进步也为科学文献的表示和理解带来了新的可能性。特别是2023年提出的图文本联合表示方法，成功地将文本语义和引用关系结构统一起来，为科学知识的表示开辟了新途径。

此外，科学数据表示领域面临的主要挑战包括如何处理多模态科学数据、如何确保表示的可解释性，以及如何在保持数据完整性的同时实现高效的压缩表示等。在这方面，南洋理工大学、北京邮电大学和字节跳动合作训练的LLaVA-Video模型[5]，通过生成跨模态的合成数据，帮助模型学习不同模态间的关联关系，大幅提升了多模态表征能力。

LLaVA-Video研究团队开发了一个创新的视频理解系统，其核心是基于大规模跨模态合成数据集LLaVA-Video-178K[5]。这个数据集包含了178K个视频样本和1.3M个指令跟随样本，涵盖了视频描述、开放式问答和多项选择问答等多种任务形式。

研究团队采用了一种基于GPT-4的三层级递归生成架构来构建视频内容的层级表示（如图 3 所示）。这种架构通过时序嵌入（Temporal Embedding）和跨模态注意力机制（Cross-modal Attention）来实现视觉和文本信息的有效整合。具体来说，对当前层级的时间点 t ，以及最后的时间点 T 。系统的三个层级分别承担不同的表示任务：

（a）在第一级，为时间点 t 生成字幕时，参考了当前时间点的帧画面、前一个时间点的字幕，以及（如果适用）最近的第二级总结描述。（b）在第二级，为时间点 t 生成字幕时，基于前一个第二区间的字幕和最近三个第一区间的字幕。（c）在第三级，为最终时间点 T 生成整体字幕时，参考了最近的第二区间字幕和当前的第一区间字幕，生成对整个视频内容的全局语义表示。



LLaVA-Video-178K视频详细描述生成流程 | 图片来源: Zhang et al. (2024) "Video Instruction Tuning With Synthetic Data."

这种多层次表示方法的创新之处在于它不仅解决了视频内容的时序依赖问题，还实现了视觉和语言模态的深度融合。通过递进式的多层级表示来捕捉视频内容的不同粒度特征，从而实现从局部细节到全局语义的完整表示框架。这种方法不仅提高了模型的性能，也为未来的多模态内容理解研究提供了新的思路。

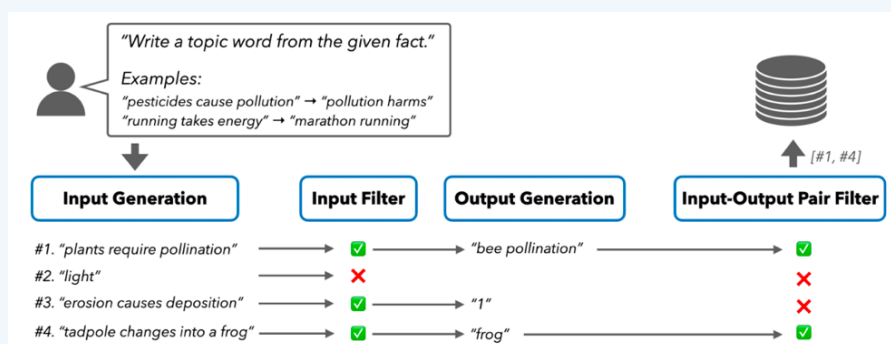
随着新技术的不断涌现和交叉学科的深度融合，科学数据表示领域有望在推动科学发现和促进学科发展方面发挥更加重要的作用。

2.9.2.3 模型自我改进：通过合成数据提升AI系统性能

推荐理由：在人工智能研究领域中，模型自我改进代表了一个极具前景的发展方向。这种方法让AI系统能够通过生成和利用合成数据来增强自身能力，无需过度依赖外部数据源或更强大的教师模型。这一技术不仅降低了对高质量训练数据的依赖，更开创了AI系统自主学习和进化的新范式。随着大语言模型的发展，这种自我改进机制展现出越来越重要的价值。

模型自我改进的研究聚焦于通过合成数据增强模型性能。这个领域的一个代表性工作是自训练（Self-training）方法，模型首先在有限的标注数据上训练，然后生成新的训练样本来改进自身。另一个创新性的研究是模型蒸馏（Model Distillation）与合成数据的结合，通过生成特定的训练样本来优化知识迁移过程。这些方法展示了如何通过合成数据来实现模型能力的持续提升。

在这一领域的最新突破中，卡内基梅隆大学和清华大学研究团队开发的SELF-GUIDE方法[6]展示了显著成果。这种创新方法通过设计高效的多阶段生成机制，使语言模型能够自主生成任务特定的训练数据。系统通过逐步生成“输入-输出”数据对，并经过严格的筛选过程，确保生成数据的质量和相关性。这些自生成的数据随后被用于模型的进一步微调，形成一个良性的自我提升循环。



SELF-GUIDE 针对生成任务的流程 | 图片来源: Zhao et al. (2024) "Self-guide: Better task-specific instruction following via self-synthetic finetuning."

研究结果令人振奋：SELF-GUIDE在多个任务上都取得了显著的性能提升。在分类任务中，模型性能获得了约15%的绝对提升；在生成任务中，提升幅度更是达到了约18%。这些数据充分证明了该方法在提升模型能力方面的有效性。

SELF-GUIDE的成功不仅证明了模型自我改进策略的可行性，也为未来AI系统的发展提供了新的思路。这种自主学习和改进的能力，可能会带来更智能、适应性更强的AI系统，推动整个领域向着更高水平发展。通过持续的自我改进，AI系统有望在更多复杂任务中展现出更强的性能和适应能力。

2.9.2.4 数据基础设施：人体蛋白质组计划

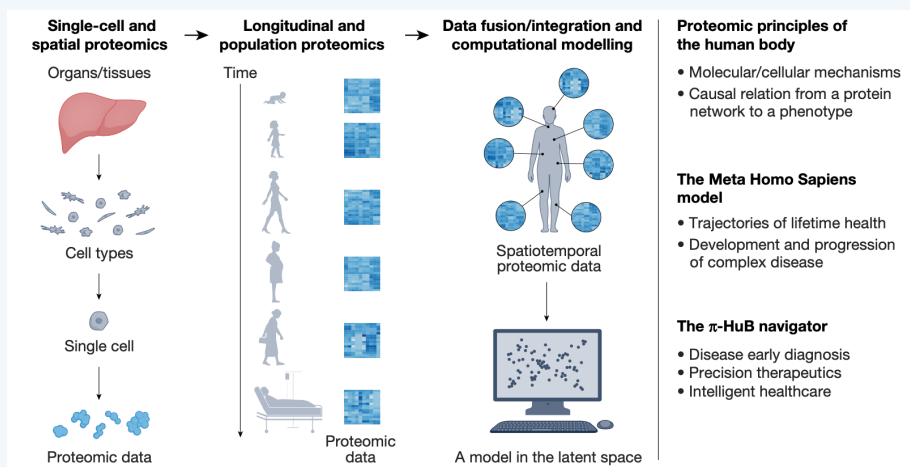
推荐理由：首次提出构建人体蛋白质组的精确“导航系统”[18]，不仅将彻底改变我们对人体生命活动的认识，随着单细胞蛋白质组学等技术[19]的快速发展，更有望推动医学范式从被动治疗向主动预防和精确医疗转变。

在生命科学研究中，基因组告诉我们生命的可能性，蛋白质组则展示了生命的现实状态。人体内约

37万亿个细胞虽然共享相同的基因组，却能展现出丰富多样的形态和功能。这种神奇的分化和调控过程，正是由蛋白质的精确表达和调控网络所主导。然而，目前我们对人体蛋白质组的认识仍然十分有限，这极大地制约了精准医疗的发展。

π -HuB项目提出了三个突破性的研究目标：首先是揭示人体的构建原理，通过最新的单细胞蛋白质组学技术，绘制不同类型细胞中蛋白质的精确组成及其调控网络；其次是建立“元人类”（Meta Homo Sapiens）计算模型，追踪记录人体蛋白质组在不同生命阶段的动态变化规律；最后是开发 π -HuB导航系统，将蛋白质组学数据转化为疾病预防、诊断和治疗的精确指导。

在技术路线上，该项目采用了多层次、多维度的研究策略。在细胞水平，运用最新的单细胞蛋白质组学技术，实现对人体主要器官细胞类型的精确分析；在群体水平，通过大规模队列研究[20]，揭示生活方式、环境因素等对蛋白质组的影响；在临床应用层面，通过建立标准化的生物样本库和数据分析平台，推动蛋白质组学在精准医疗中的转化应用。



π -HuB项目的总体目标 | 图片来源: He, F., Aebersold, R., Baker, M.S. et al. π -HuB: the proteomic navigator of the human body. Nature 636, 322–331 (2024)."

研究成果的应用前景令人振奋。在第一阶段（2024-2033年），项目将重点实现三个突破：完成人体主要器官的细胞类型蛋白质图谱；建立基于蛋白质组学的健康评估体系；开发新的疾病早期诊断标志物和治疗靶点。这些成果将为实现更具有实践智慧的精准医学奠定坚实基础，推动医学模式从被动治疗向主动预防转变。

π -HuB项目的创新性不仅体现在其科学目标上，也体现在其组织模式上。项目采用开放科学的理念，建立了国际化的研究团队网络，并承诺将研究数据和分析工具向全球科研界开放共享。这种协作模式将大大加速蛋白质组学研究的进展。 π -HuB项目的启动，标志着人类探索生命奥秘和追求健康的征程又迈出了重要一步。

2.9.3 挑战与展望

虽然OpenAI的CEO Sam Altman预见 AI 最终将能产生足够优质的合成数据来训练自身，但有研究表明[21]，劣质信息和不当的训练方法仍可能导致LLMs“模型崩溃”。因此合成数据虽然潜力巨大，但在实际应用中仍面临着诸多挑战。

数据质量与保真度问题至关重要。合成数据的质量直接影响模型性能。基于虚假、幻觉或有偏差的数据训练的模型不仅可能在现实场景中表现不佳，缺乏泛化能力[22]，如果设计和验证不当，还可能会放大已有偏差或引入新的偏差[23]。**去污染评估（decontamination evaluation）难度问题**仍需进一步探索。由于合成数据可能包含重述的基准数据版本，传统的词级别去污染方法可能会失效[24]，导致无法有效区分模型是真正理解和学习了新知识，还是仅仅在记忆和重复训练数据中的内容。**隐私与伦理问题**也依然严峻。尽管合成数据提供了一些不损害个人隐私的途径[25]，但在敏感领域中使用合成数据时

仍存在伦理问题。有研究[26]表明可以从训练数据集中提取特定信息，这意味着合成数据可能会在无意中暴露基础训练数据的某些敏感信息和隐私内容[27]。

当然，在面对挑战的同时，合成数据未来同样有下面几个具有前景的研究发展方向值得期待。

数据质量与多样性提升。合成数据未来研究应专注于开发基于GANs和扩散模型等的新技术，并结合领域特定知识，通过RAG等方法确保数据质量和多样性，拓展到包括医疗、金融和社会科学等应用领域。

数据质量监管机制研究。随着AI复杂性提升，传统数据评估方法已不足以应对挑战。需要建立更系统化的监管框架，实现合成数据的全面质量评估、自动化筛查和多场景验证。

探索合成数据的规模效应。鉴于一些经过精心训练的小型语言模型，能超出Chinchilla定律（模型性能与训练数据规模和模型参数量呈正比）的预测，未来研究需要探索合成数据规模的“质量-数量”权衡机制，以找到提升模型性能的最优数据策略。

自我改进能力的涌现研究。最新研究显示出积极进展，但仍需深入探索其理论基础、局限性和潜在风险，从而推动更具适应性和自主性的AI学习过程。

多源数据融合基础设施建设。需要解决数据标准化和语义对齐问题，特别是在处理跨组织、跨领域的敏感数据时，如何在促进数据共享的同时确保数据安全与隐私问题。近期，Anthropic开源的「模型上下文协议」MCP（Model Context Protocol）[28]已

经迈出了重要一步。

总之，虽然目前仍面临诸多挑战，但正如许多研究者所预见的，合成数据不仅是解决当前AI发展瓶颈的工具，更可能成为开启下一代人工智能革命的钥匙。特别是在自我改进能力方面的研究，能帮助我们实现从狭义AI到通用人工智能的跨越，推动人类和AI向更智能、更加美好的未来迈进。

2.9.4 参考文献

[1] Villalobos, Pablo, et al. "Will we run out of data? an analysis of the limits of scaling datasets in machine learning." arXiv preprint arXiv:2211.04325 (2022)

推荐理由：深入分析机器学习数据集扩展的极限问题，对当前大模型发展中的数据瓶颈提供了重要见解。

[2] Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. "ChatGPT outperforms crowd workers for text-annotation tasks." Proceedings of the National Academy of Sciences 120.30 (2023) : e2305016120.

推荐理由：首次系统地证明ChatGPT在文本标注任务上优于人工众包，为AI辅助数据标注提供了实证研究支持。

[3] Abay, Nazmiye Ceren, et al. "Privacy preserving synthetic data release using deep learning." Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18. Springer International Publishing, 2019.

推荐理由：提出了基于深度学习的隐私保护合成数据生成方法，在数据隐私和效用之间取得了良好平衡。

[4] Mandlekar, Ajay, et al. "Mimicgen: A data generation system for scalable robot learning using human demonstrations." arXiv preprint arXiv:2310.17596 (2023) .

推荐理由：创新性地提出了基于人类示范的机器人学习数据生成系统，为解决机器人学习中的数据瓶颈提供了新思路。

[5] Zhang, Yuanhan, et al. "Video Instruction Tuning With Synthetic Data." arXiv preprint arXiv:2410.02713 (2024) .

推荐理由：探索了视频指令微调的合成数据生成方法，为多模态大模型的训练提供了新的视角。

[6] Zhao, Chenyang, et al. "Self-guide: Better task-specific instruction following via self-synthetic finetuning." arXiv preprint arXiv:2407.12874 (2024) .

推荐理由：提出了自生成式指令微调方法，显著提升了模型对特定任务的理解能力。

[7] Chen, Zixiang, et al. "Self-play fine-tuning converts weak language models to strong language models." arXiv preprint arXiv:2401.01335 (2024) .

推荐理由：创新性地提出了自对弈式微调方法，证明了弱模型可以通过自我提升变强。

[8] Wu, Tianhao, et al. "Meta-rewarding language models: Self-improving alignment with llm-as-a-meta-judge." arXiv preprint arXiv:2407.19594 (2024) .

推荐理由：提出了创新的元奖励机制来改进语言模型对齐，为模型自我提升提供了新框架。

[9] Liang, Yiming, et al. "I-SHEEP: Self-Alignment of LLM from Scratch through an Iterative Self-Enhancement Paradigm." arXiv preprint arXiv:2408.08072 (2024) .

推荐理由：提出了从零开始的LLM自对齐迭代增强范式，为模型对齐提供了全新思路。

[10] Van Breugel, Boris, Zhaozhi Qian, and Mihaela Van Der Schaar. "Synthetic data, real errors: how (not) to publish and use synthetic data." International Conference on Machine Learning. PMLR, 2023.

推荐理由：深入分析了合成数据使用中的常见错误，提供了实用的合成数据发布和使用指南。

[11] Barbierato, Enrico, et al. "A methodology for controlling bias and fairness in synthetic data generation." Applied Sciences 12.9 (2022) : 4619.

推荐理由：提出了一种在合成数据生成中控制偏差和公平性的方法，旨在提高数据的公正性，该方法在减少算法偏差和促进公平性方面具有重要意义，特别是在敏感应用领域。

[12] Mattern, Justus, et al. "Membership inference attacks against language models via neighbourhood comparison." arXiv preprint arXiv:2305.18462 (2023) .

推荐理由：这项研究提出了基于邻域比较的创新成员推理攻击方法，有效揭示了语言模型在训练数据隐私保护方面的潜在漏洞。

[13] Xu Guo&Yiqiang Chen. (2024) "Generative AI for Synthetic Data Generation: Methods, Challenges and the Future"

推荐理由：全面梳理了生成式AI在合成数据生成领域的最新进展和技术挑战，为研究者提供了完整的领域发展全貌。

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014

推荐理由：这是GAN领域的开山之作，开创了生成对抗网络的研究方向并奠定了理论基础。

[15] D. Chen, C. Lee, Y. Lu, D. Rosati, and Z. Yu, "Mixture of soft prompts for controllable data generation," in Findings of the Association for Computational Linguistics: EMNLP 2023, H. Bouamor, J. Pino, and K. Bali, Eds.

Singapore: Association for Computational Linguistics, Dec. 2023, pp. 14 815–14 833.

[Online]. Available: <https://aclanthology.org/2023.findings-emnlp.988>

推荐理由：在EMNLP发表的研究创新性地提出了混合软提示方法来实现可控的数据生成，为提示工程提供了新思路。

[16] Y. Yu, Y. Zhuang, J. Zhang, Y. Meng, A. Ratner, R. Krishna, J. Shen, and C. Zhang, "Large language model as attributed training data generator: A tale of diversity and bias," in Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2023.

[Online]. Available: <https://openreview.net/forum?id=6hZLfAY9GD>

推荐理由：深入探讨了大语言模型在训练数据生成过程中的多样性和偏见问题，为提升AI系统的公平性提供了重要参考。

[17] L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," 2021.

推荐理由：突破性地探索了超越少样本范式的提示编程方法，为大语言模型的提示工程开辟了新方向。

[18] He, F., Aebersold, R., Baker, M.S. et al. π -HuB: the proteomic navigator of the human body. Nature 636, 322–331 (2024). <https://doi.org/10.1038/s41586-024-08280-5>

推荐理由：这篇Nature文章介绍了突破性的人体蛋白质组导航器 π -HuB平台，为理解人体生理机制和疾病发展提供了新视角。

[19] Kelly, R.T. Single-cell proteomics: progress and prospects. Mol. Cell. Proteomics 19, 1739–1748 (2020).

推荐理由：系统总结了单细胞蛋白质组学的技术进展和应用前景，为该领域研究者提供了重要参考。

[20] Eldjarn, G.H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* 622, 348–358 (2023).

推荐理由：提在Nature发表的大规模血浆蛋白质组学研究，通过与遗传学和疾病关联分析揭示了重要的生物学机制。

[21] Shumailov, I., Shumaylov, Z., Zhao, Y. et al. AI models collapse when trained on recursively generated data. *Nature* 631, 755–759 (2024).

<https://doi.org/10.1038/s41586-024-07566-y>

推荐理由：Nature上发表的重要研究，揭示了AI模型在递归生成数据训练时的崩溃现象，对当前生成式AI的局限性提供了关键洞察。

[22] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 19–27

推荐理由：开创性地探索了电影和图书内容的跨模态对齐问题，为视觉叙事理解提供了重要的研究基础，是多模态学习领域的经典文献。

[23] C. Peng, X. Yang, A. Chen, K. E. Smith, N. PourNejatian, A. B.

Costa, C. Martin, M. G. Flores, Y. Zhang, T. Magoc et al., "A study of generative large language model for medical research and healthcare," *arXiv preprint arXiv:2305.13523*, 2023.

推荐理由：系统研究了大语言模型在医疗研究和健康护理领域的应用前景，深入分析了其潜力和局限性。

[24] S. Moore, R. Tong, A. Singh, Z. Liu, X. Hu, Y. Lu, J. Liang, C. Cao, H. Khosravi, P. Denny et al., "Empowering education with llms-the next-gen interface and content generation," in *International Conference on Artificial Intelligence in Education*. Springer, 2023, pp. 32–37.

推荐理由：探讨了大语言模型在教育领域的创新应用，特别关注了接口设计和内容生成方面的突破。

[25] N. Rane, "Role and challenges of chatgpt and similar generative artificial intelligence in business management," Available at SSRN 4603227, 2023.

推荐理由：Rane (2023) 详细分析了ChatGPT等生成式AI在商业管理中的角色定位和面临的挑战，为企业应用提供了实用的参考框架。

[26] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, and L. Sun, "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt," *arXiv preprint arXiv:2303.04226*, 2023.

推荐理由：全面回顾了AI生成内容(AIGC)的发展历程，从GAN到ChatGPT的技术演进，是理解生成式AI发展的重要综述。

[27] A. Bauer, S. Trapp, M. Stenger, R. Leppich, S. Kounev, M. Leznik, K. Chard, and I. Foster, "Comprehensive exploration of synthetic data generation: A survey," *arXiv preprint arXiv:2401.02524*, 2024

推荐理由：全面综述了合成数据生成的各种方法和应用，为研究者提供了系统性的参考框架。

[28] <https://www.anthropic.com/news/model-context-protocol>

推荐理由：通过MCP协议为大语言模型在长文本处理能力上的突破性进展，为大模型架构优化提供了新思路。

2.10 新型智能计算

2.10.1 背景介绍

在当今数字化时代，以神经网络为代表的智能计算不断生成海量数据，同时也推动了对算力的更高需求。传统计算所面临的挑战也逐步展现：

- 大模型训练过程中的能耗和成本过高[1]；
- 当前训练大模型的计算需求增长速度远快于现代计算机硬件的改进速度[2]；
- AI算法可能被恶意复制和传播，对人类安全带来不可预测的风险[3]。

这一趋势正在深刻改变计算机体系架构的格局。为满足特定应用的需求，设计专门的领域处理器已成为一种重要趋势，如谷歌的TPU和华为的昇腾AI芯片会对神经网络在芯片上的运行进行优化。这种软硬件一体化的方案显著提升了应用程序的运行效率。除了在传统计算架构内的创新外，软硬件结合的新型计算模式也逐渐浮出水面。例如，凡人计算（Mortal Computation）[4, 5]、光学计算（Optical Computing）[6, 7]和量子计算（Quantum Computing）等新兴理念，展现出在更高效、安全和节能的框架下推进AI技术发展的潜力。

凡人计算通过将硬件与计算模型紧密结合，减少了冗余的数据传输和热量散发，从而显著降低了计算系统的能耗并提升了计算效率。此外，软硬件结合的模式在某种程度上限制了AI的无限复制能力，从本质上降低了未来超级智能失控的风险[4, 5]。

光学计算是一个新兴的学科领域，它利用光学原

理和技术来实现信息的处理、存储和传输。与传统的电子计算相比，光学计算在速度、功耗和并行性等方面具有独特的优势，在人工智能、大数据处理和科学计算等领域具有广阔的应用前景。光学神经网络是光学计算与人工智能领域的一个交叉方向，它利用光学技术模拟神经网络，克服传统计算平台在深度神经网络加速方面的瓶颈，并为未来的AI应用提供更高效的计算方案[6]。

量子计算是一种革命性的计算模式，基于量子力学的特性，如叠加态和量子纠缠，展现出比经典计算机更强大的信息存储和并行处理能力。因此，量子计算有望解决传统计算机难以应对的特定问题，包括量子系统模拟、量子化学、密码学、机器学习加速及组合优化等领域[8]。在量子计算与人工智能的交叉点上，两者的合作虽被寄予厚望，并被认为是孕育着巨大潜力，但科研实践中如何更高效地推进量子计算与人工智能的结合，仍然是一个开放的问题。

在“AI for Quantum”方面，AI技术在量子计算硬件、量子算法设计、量子纠错和量子误差缓解等领域正发挥着日益重要的作用[9]。首先，在量子噪声处理方面，谷歌团队最近推出了一种基于Transformer架构的递归神经网络解码器AlphaQubit[10]。该解码器在处理谷歌Sycamore量子处理器的真实数据时，表现优于其他先进解码器。同时，IBM也在探索将机器学习模型应用于量子错误缓解技术[11]，以在不牺牲计算精度的前提下降低错误缓解成本。除了量子噪声处理，人工智能还被引入到量子编译（Quantum Compilation）中。量子线路模型是广泛使用的量子计算模型之一，如何高效地将设计好的量子算法映射为量子线路中的量子比特和量子门操作的过程被称为量子编译。近期，研究人员利用生成人工智能中

的扩散模型（Diffusion Models）来生成量子编译所需的量子门[12]。在AI辅助的量子算法层面，AI通过预先优化参数来预测新实例的最佳参数，缓解了变分量子算法面临的训练中贫瘠高原（Barren Plateau）问题[9]。

另一方面，在“Quantum for AI”中，研究人员不断构建经典机器学习的量子版本，并在实验中追求更多量子比特以较低错误率执行更深层次的量子线路，力求通过实际案例展示量子优势[13]。在经典计算机上训练机器学习模型通常需要消耗大量的计算资源和时间，一些研究提出将机器学习的训练问题表述为二次无约束二进制优化（Quadratic Unconstrained Binary Optimization, QUBO）问题，并使用绝热量子计算机求解QUBO问题[14]。在“Quantum for AI”中具有代表性的算法包括量子神经网络（Quantum Neural Networks）和HHL算法（Harrow-Hassidim-Lloyd Algorithm）。HHL算法是一种用于对线性方程组进行数值求解的量子算法[15]，由于线性系统在科学和工程的绝大多数领域被普遍应用，因此线性方程组的量子算法具有广泛的应用潜力。对于特定类型的矩阵，HHL算法能够将线性方程求解的复杂度从经典最优算法的指数级降低至多项式级。意识到HHL算法在量子机器学习领域的重要性后，Scott Aaronson分析了可能限制该算法的因素[16]，指出量子态的制备成本和读出成本可能会消除指数加速。量子神经网络（QNN）是一种引入量子计算机制的神经网络模型，Abbas、Amira等人的研究通过有效维度这一尺度指出，量子神经网络相较于经典神经网络具有显著优势[17]。然而，量子神经网络同样面临训练中贫瘠高原（Barren Plateau）问题[18]。此外，《自然》杂志的一篇新闻特写文章指出[19]，“量子计算机是否能为机器学习提供优势，

应该通过实验来验证，而不是仅仅依赖数学证据来判断其优越性。”由于量子计算机的技术实现仍处于初级阶段，目前的量子神经网络模型大多属于理论构想，尚待在物理实验中进行全面验证。

因此，当前量子计算领域的重要任务是实现具有更高保真度并且规模足够大的量子计算机。在实现量子计算的物理载体中，超导体系已被广泛接受，其次是高保真度门操作的离子阱系统。然而，近年来，中性原子体系以黑马的姿态迅速崛起[20-23]。2018年，加州理工大学教授John Preskill提出了NISQ（Noisy Intermediate-Scale Quantum）概念[24]，强调当前量子计算硬件的局限性，包括受噪声影响的不完美量子操控和量子比特数量的限制（通常在10到1000之间）。NISQ概念提出后，量子计算领域在这两个维度上取得了最新进展：

1. 在量子比特数量方面，走在最前面的两家商业公司：量子计算公司Atom Computing利用中性原子构建的量子计算平台已实现1,180个量子比特[25]，而IBM的Condor超导量子处理器则包含1121个量子比特[26]。

2. 在处理噪声引起的错误方面，主要采用量子纠错（QEC）和量子错误缓解（QEM）两种方法。目前，量子计算领域的重要工作大多与量子纠错相关，例如在中性原子量子计算体系和超导量子体系上实现量子纠错码。

量子计算与传统计算之间并非替代关系，而是相辅相成。未来，量子计算有望与传统计算结合，形成混合计算架构，充分发挥各自的优势，推动计算技术的进一步发展。

2.10.2 研究进展

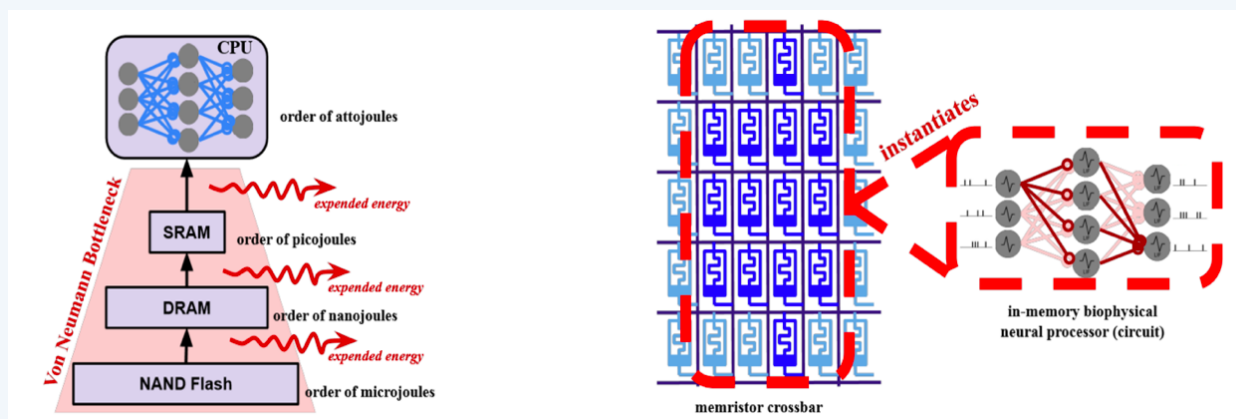
2.10.2.1 凡人计算

推荐理由：杰弗里·辛顿（Geoffrey Hinton）提出了基于硬件的“凡人计算”概念[4]。他指出，生物系统中的信息处理所依赖的计算过程是与其执行的物理基质不可分割的，即“软件”无法脱离“硬件”而存在，这与计算机科学中的计算概念形成了鲜明对比。Hinton 提出使用前向-前向传播的算法（Forward-Forward, FF）来实现凡人计算。

从“永生”软件向“凡人计算”范式的演进，为实现基于自然主义的机器智能提供了新的技术路径。研究[5]指出，不朽计算（图2左）和凡人计算（图2右）在热力学消耗上存在显著差异。在典型的冯·诺

依曼架构中，计算机体系结构中不同类型的存储之间存在“能量壁”，需要花费能量将信息最终载入CPU。而在神经形态芯片中，凡人计算程序的处理直接在存储之上或存储内进行，避开了冯·诺依曼架构中所需的大量能耗。

凡人计算也意味着一旦实现该程序的硬件媒介失效或“死亡”，其所拥有的知识、行为和特定功能，包括其特性也将随之消失。这类似于生物有机体在无法维持自身时，其所获得的知识 and 行为也随之终止，从而使得AI具备了自我限制和自我调节能力，降低了超级智能失控的可能。

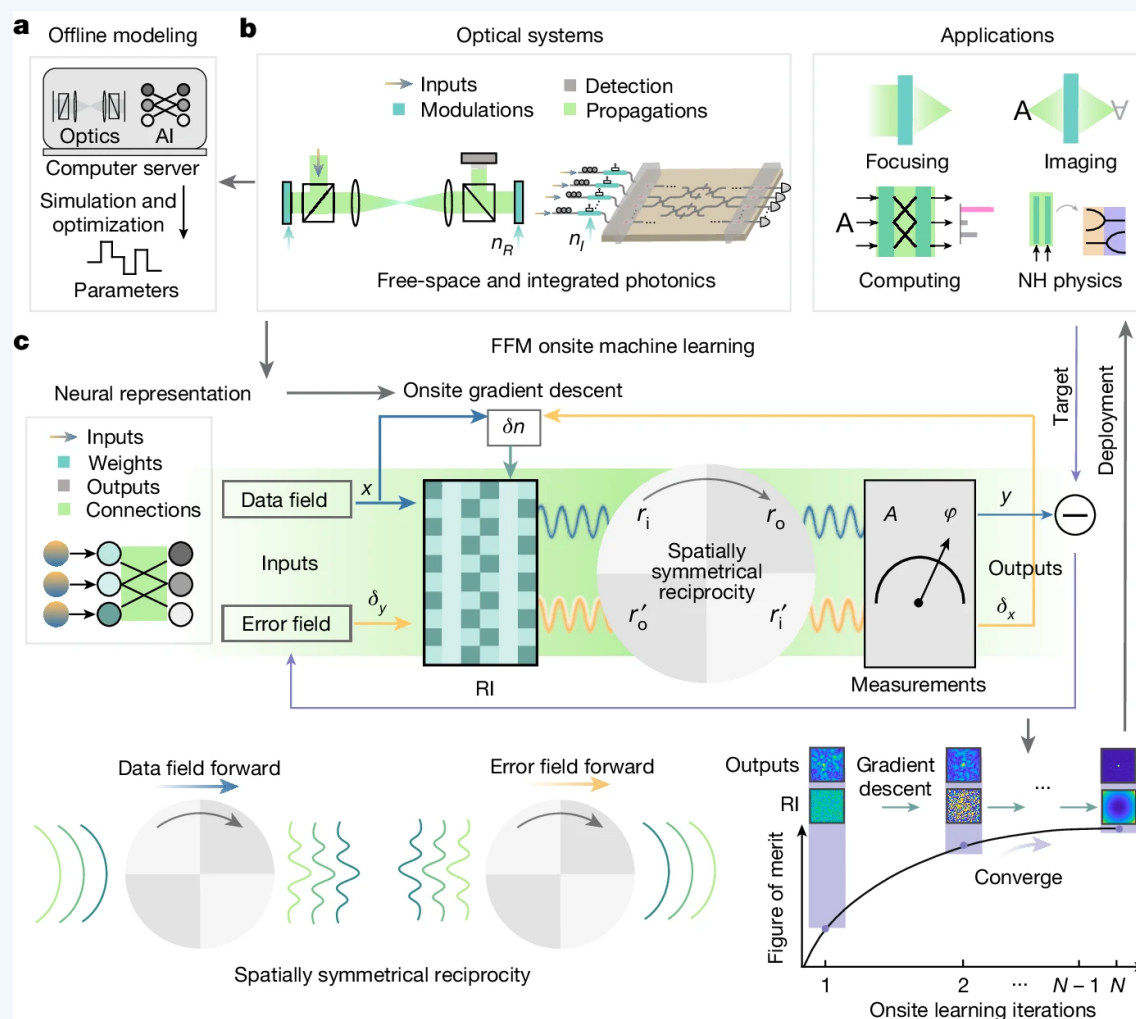


不朽计算（左）和凡人计算（右）在热力学消耗上的差异 | 图片来源：Ororbia and Karl (2023).
"Mortal computation: A foundation for biomimetic intelligence."

2.10.2.2 光学计算

光学计算作为一种新型计算范式，为深度神经网络工作负载提供加速，从而克服电子计算在该领域的局限性。清华大学研究团队创新性地提出了全前向模式学习（Fully Forward Mode Learning, FFM），通过构建基于光学系统对称性的可微分物理模型，在自由空间光子神经网络中实现了原位梯度计算与参数优化[7]。该方法利用洛伦兹互易性原理，将误差传

播转化为前向物理过程，突破性地解决了光学反向传播对精准系统建模的依赖问题。在8层百万参数量级自由空间光学神经网络实验中，该方法在Fashion-MNIST数据集上取得92.5%分类准确率，相较离线数值模拟理想模型（96%）展现出优异的物理系统适配能力，验证了全光域深度学习训练的可行性。



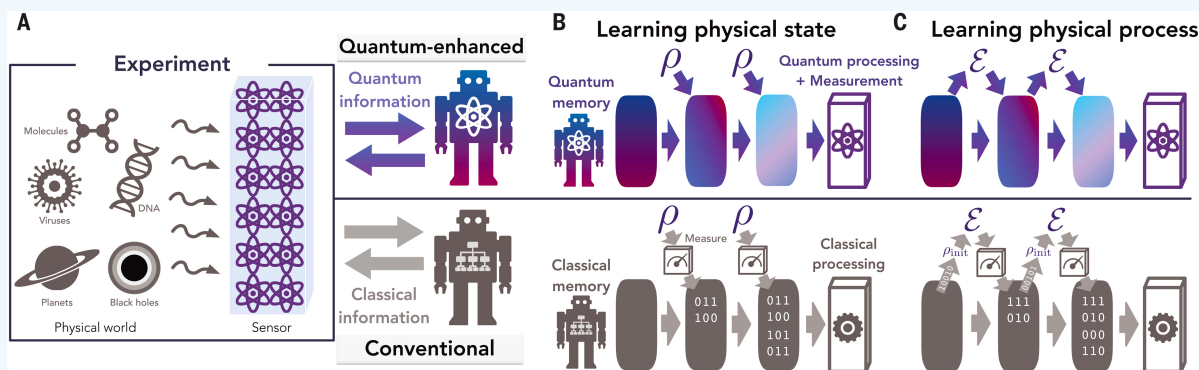
传统光学系统建模与基于FFM（Full-Forward Mode）的现场机器学习之间的区别与联系。| 图片来源：Xue, Z., Zhou, T., Xu, Z., Yu, S., Dai, Q., & Fang, L. (2024). Fully forward mode training for optical neural networks. *Nature*, 632(8024), 280-286.

2.10.2.3 量子计算

物理实验中的量子优势

推荐理由：人类通过实验认知自然，但迄今为止这种认知能力一直受限于通过经典视角观测量子世界。理查德·费曼（Richard Feynman）提出量子计算的动机之一是意识到经典计算机在模拟量子系统时的根本性局限：“自然不是经典的，如果你想模拟自然，最好用量子力学。”近期一项研究体现了费曼的思想，也许在量子实验相关的任务可以找到量子计算的优势场景[27]，相较于传统实验模式（即对物理系统进行测量后使用经典计算机处理结果），一种将物理系统数据转导至稳定量子存储器、并利用量子计算机处理数据的实验架构，可能具有显著优势。

这种指数优势体现在3个方面：（1）物理系统性质预测、（2）含噪量子态的主成分分析以及（3）物理动力学近似模型学习等任务中。举例来说，在物理系统性质预测的任务上，作者提到了一个案例，即预测Ising模型的磁性。传统方法需要测量大量的自旋状态，并进行复杂的计算才能预测其磁性。而量子方法可以通过制备Ising模型的量子态，并进行简单的测量，就能更有效地预测其磁性。作者对这三种学习任务进行了数学分析以证明具有指数级量子优势，并在Google Sycamore处理器上进行了实验，证明了即使使用目前的噪声量子设备，量子增强实验也能胜过传统实验。



量子增强实验与传统实验在学习物理系统性质方面差异的核心图示。上半部分展示了量子增强实验的流程。下半部分展示了传统实验的流程。信息存储方式和处理方式不同：量子增强实验使用量子存储器存储量子信息，并且使用量子设备处理量子信息。传统实验使用经典设备存储与处理经典信息。处理量子信息又细分为学习物理态与学习物理过程。| 图片来源：Huang, Hsin-Yuan, et al. "Quantum advantage in learning from experiments." Science 376.6598 (2022): 1182-1186.

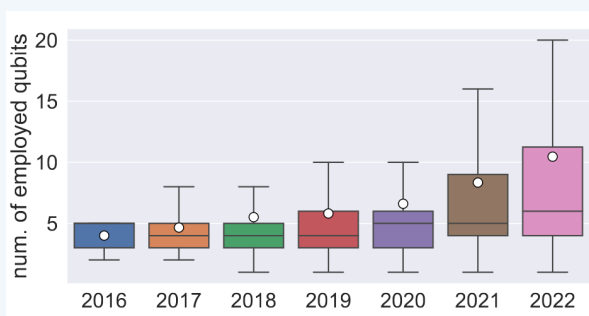
实现量子优势的瓶颈

推荐理由：迄今为止，仅在少数实验中证明了量子优势。2019年，Google AI Quantum 团队宣布了

一项被称为“量子计算优越性”的里程碑事件[28]。Google 团队利用超导量子技术构建了一台名为Sycamore 的可编程量子计算机，能在相邻量子比特之间执行纠缠的双量子比特门。目前已知的最佳模拟方法

表明，Sycamore几分钟内的计算结果，在现有最强大的经典超级计算机上至少需要几天时间来完成[29]。

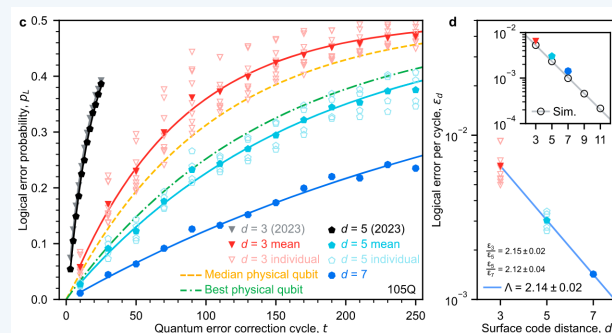
要充分发挥量子计算在解决实际问题上的加速优势，量子比特的数量和质量必须满足一定的要求。尽管量子处理器的量子比特数量持续增长，且通过云计算服务逐渐向学术界开放，但Nature Reviews Physics的一项研究[30]显示：2016年至2022年间quant-ph类别预印本中，研究所用的量子比特平均数量随量子体积（QV）增长升至10.5，中位数仅为6。该数值偏低源于NISQ设备抗噪能力有限，当量子比特数量增加时，计算结果可靠性显著下降。即使在具备更多量子比特的情况下，为了减少噪声对算法结果的干扰，研究者只能降低量子比特的使用数量。近来清华团队的研究从新的视角揭示了噪声对量子计算的巨大危害[31]。研究指出，在噪声强度某个阈值时，关联生成中量子优势突然消失。



各年使用的量子比特数量没有异常值的统计图。白色圆圈显示平均值。| 图片来源：Ichikawa, T., Hakoshima, H., Inui, K., Ito, K., Matsuda, R., Mitarai, K., ... & Fujii, K. (2024). Current numbers of qubits and their uses. Nature Reviews Physics, 1-3.

量子纠错

推荐理由：这些现象显示，要发挥量子计算的实质性优势，我们必须对量子噪声这一“绊脚石”发起挑战。量子计算中处理噪声引起的错误，主要源于量子纠错(Quantum Error Correction, QEC)这一方法。量子纠错最常见的方法是使用多个物理量子比特编码一个逻辑量子比特。尽管上个世纪90年代理论上就不断有量子纠错码被提出，但是实验上实现这些纠错码主要是近些年的事情。2024年，谷歌发表在Nature的论文显示[32]，其超导量子处理器Willow拥有105个物理比特，构建了码距为3、5、7的表面码。当表面码的码距每增加2时，逻辑错误率被 2.14 ± 0.02 倍抑制，这使得每个周期的错误为 $0.143\% \pm 0.003\%$ 。随着码距的持续扩展，逻辑量子比特的错误率将维持随码距呈指数下降的特性。需特别说明，当表面码码距为7时，构建单个逻辑量子比特需消耗97个物理比特（码距d与物理比特数量呈二次方关系），并额外配置4个物理比特用于泄漏抑制。



表面码的码距d=3提升到码距d=5，编码错误率降低 2.15 倍。表面码的码距d=5提升到码距d=7，编码错误率降低 2.12 倍。| 图片来源：Google Quantum AI and Collaborators. Quantum error correction below the surface code threshold. Nature (2024). <https://doi.org/10.1038/s41586-024-08449-y>

可以看出量子纠错的代价非常高。Gidney的研究指出[33]，如果使用Shor的算法破解RSA-2048需要2000万个含噪量子比特、运行8小时。在这2000万个量子比特中，只有0.064%的量子比特用于计算，剩下的99.936%都是用于纠错的冗余。

量子错误缓解

推荐理由：从长远来看，量子纠错技术有望成为量子计算的终极解决方案；然而在当前硬件条件受限的短期阶段，采用误差缓解而非完全消除错误的过渡性策略更为可行，这便是量子错误缓解（Quantum Error Mitigation, QEM）[34]被广泛研究的原因。变分量子算法常采用的量子错误缓解包含零噪声外推（Zero-noise Extrapolation, ZNE）、概率性错误消除（Probabilistic Error Cancellation, PEC）、测量错误缓解（Measurement error mitigation）。

中性原子量子计算

推荐理由：在量子计算的物理载体竞争中，超导体系和离子阱体系长期以来一直是研究者的重点。然而，近年来，中性原子量子计算迅速崛起，成为这一领域中进展最显著的体系。与其他量子计算体系相比，根据DiVincenzo判据，中性原子量子计算在多个维度上展现出明显优势。

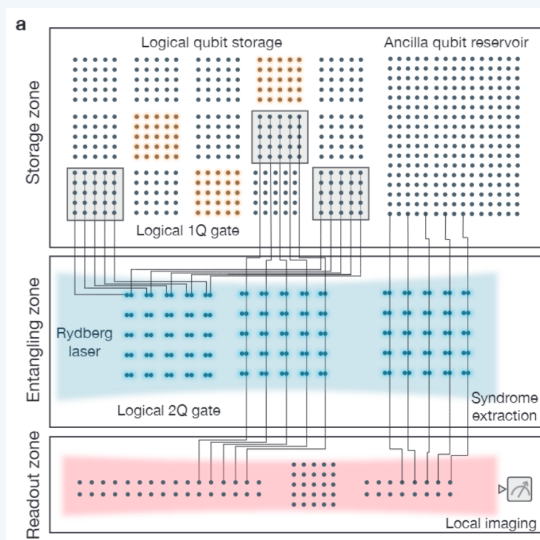
1 可拓展性：中性原子量子计算能够操控大量的原子，实验中已经实现了多达6100个原子的阵列[21]；

2 相干时间：中性原子的退相干时间可达到12.6秒[21]；

3 量子门构建：通过偶极相互作用，中性原子能

够构建高保真度的量子门；此外，研究者可以通过光镊（optical tweezers）移动原子阵列实现非近邻量子比特之间的门操作，这为量子电路的灵活设计提供了新的可能性[22]，实现了99.5%保真度的双比特纠缠门，超过了表面码的纠错阈值。

在中性原子量子计算领域，哈佛大学Lukin组及其合作者的研究成果尤为突出。他们成功地利用280个物理量子比特实现了多种类型的量子纠错编码：表面码（Surface Code）、颜色码（Color Code）和三维码（3D Code）[23]。其中物理量子比特通过受困在光镊中的铷原子（87Rb）的超精细能级进行物理编码。该实验展示了使用逻辑量子比特进行可扩展纠错和量子信息处理的关键技术。



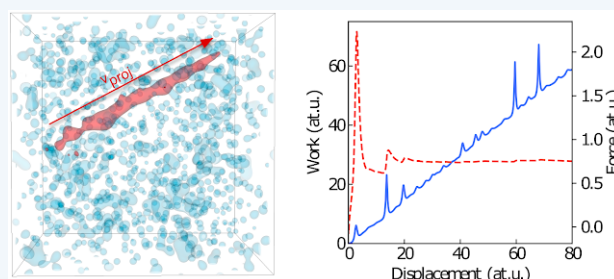
基于可重构原子阵列的可编程逻辑处理器。处理器分为三个区域：存储、纠缠和读出。通过光镊（optical tweezers）操控87Rb原子。| 图片来源：Bluvstein, D., Evered, S. J., Geim, A. A., Li, S. H., Zhou, H., Manovitz, T., ... & Lukin, M. D. (2024). Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997), 58-65.

应用

2023年在arXiv上发布的一篇长达300多页的文章，对量子算法在多个潜在应用领域进行了深入调查[8]。这项研究不仅展示了量子计算在各个行业中的应用潜力，还强调了其在解决复杂问题时的独特优势。随着不断扩展的应用范围，量子计算的实用价值将更加凸显。

2022年，Lukin组在Science上发表的高被引用论文[35]，提出了基于中性原子阵列求解最大独立集问题的创新方法。该研究利用了Rydberg原子的Rydberg Blockade特性，能够有效地编码和求解最大独立集问题。相较于经典退火算法，实验结果显示在深层线路中求解精确解问题时存在超线性加速现象。随后一篇综述文章[36]，概述了中性原子量子计算可以解决多种与图论相关的其他问题，例如最大团问题（Maximum Clique Problem）、最小顶点覆盖问题（Minimum Vertex Cover Problem）、图着色问题（Graph Coloring Problem）等。众多的现实问题可以建模成图相关的组合优化问题，这些解决方案在实际应用中具有广泛的潜力。

在核聚变领域，谷歌与美国Sandia国家实验室等机构合作研究并提出了一种量子算法，能够用于计算核聚变实验中温致密物质的“制动能力”[37]。制动能力是指高能 α 粒子由于与周围等离子体的库仑相互作用而减速的速率。理解系统的制动能力对优化反应堆效率至关重要。当 α 粒子在等离子体中因减速作用损失动能时，其能量转移将引发电离气体温度上升。这种自加热效应是维持聚变燃烧等离子体的关键机制，而建立精准的动力学模型将为未来聚变堆工程设计提供重要理论支撑。



一个抛射体（红色）以初始速度穿过介质（蓝色）（左）。研究监测抛射体与介质间的能量传递（蓝色实线），并确定其平均斜率（红色虚线）（右）。| 图片来源：Rubin et al. (2023)

"Quantum computation of stopping power for inertial fusion target design." □

2.10.3 挑战与展望

目前这些新型计算的实现还存在诸多挑战。凡人计算试图将生物系统的有限性和自我调节应用于AI，但在理论与实践的融合方面存在复杂挑战，尤其是如何准确捕捉“凡人性”特征并将其有效转化到计算系统中。光学计算需攻克纳米光子集成与大规模制造工艺。量子计算目前面临的问题包括：高精度纠错和噪声控制、有限的量子比特数量与相干时间、材料和制造工艺的局限性等。这些因素使得实现大规模、通用的量子计算机在短期内仍然充满挑战。

展望未来，通用人工智能（AGI）的发展将呈现多元化技术路径并行的格局。凡人计算促使我们更换思路，从“需求导向、工具中心”的AI观念转向植根于生命本质的智能观念。光学计算带来的高速运算能力可以应用于需要快速响应的场景，例如高级驾驶辅助系统。由于传统计算机在速度和功耗方面存在限制，光学计算方案迅速崛起，成为替代电子计算机、构建人工智能神经网络骨干的有力选择。量子机器学习是一个

充满潜力的应用领域，它融合了量子计算与机器学习技术，通过量子加速提升训练和推理的效率。目前，像TensorFlow和Qiskit等平台已开始支持量子模型，为研究人员探索QML应用提供了有力的工具。尽管在该领域已有一些进展，谷歌在2024年I/O大会上指出，尚未有人在处理经典数据的机器学习中展示出明确的量子优势。这一现状突显出该领域对创新方法的迫切需求，亟需证明量子计算在机器学习中的实际效能。在未来的研究和应用中，除了单独使用某种计算方式，还有望能够通过组合计算，结合各技术的优点，更高效地解决各种复杂的计算问题。

2.10.4 参考文献

- [1] Rillig, Matthias C., et al. "Risks and benefits of large language models for the environment." *Environmental Science & Technology* 57.9 (2023): 3464-3466.
<https://pubs.acs.org/doi/10.1021/acs.est.3c01106>
概述：大模型对环境的直接和间接影响
- [2] Conklin, Alexander A., and Suhas Kumar. "Solving the big computing problems in the twenty-first century." *Nature Electronics* 6.7 (2023): 464-466.
<https://www.nature.com/articles/s41928-023-00985-1>
概述：21世纪重大计算问题的概述
- [3] Silver, David, et al. "Mastering the game of go without human knowledge." *nature* 550.7676 (2017): 354-359.
<https://www.nature.com/articles/nature24270>
概述：关于AlphaGo Zero的论文，使用一种完全独立的增强学习算法，没有人类数据和指导。
- [4] Hinton, Geoffrey. "The forward-forward algorithm: Some preliminary investigations." *arXiv preprint arXiv:2212.13345* (2022).
<https://arxiv.org/abs/2212.13345>
推荐理由：Hinton的论文，介绍一种新的神经网络学习过程，前向-前向算法（Forward-Forward Algorithm）
- [5] Ororbia, Alexander, and Karl Friston. "Mortal computation: A foundation for biomimetic intelligence." *arXiv preprint arXiv:2311.09589* (2023).
<https://arxiv.org/abs/2311.09589>
推荐理由：关于凡人计算的综述
- [6] Pasricha, Sudeep. "Optical Computing for Deep Neural Network Acceleration: Foundations, Recent Developments, and Emerging Directions." *arXiv preprint arXiv:2407.21184* (2024).
<https://arxiv.org/abs/2407.21184>
推荐理由：综述论文，主要关注光学计算作为一种新兴技术，如何克服传统计算平台在深度神经网络加速方面的瓶颈，并为未来的 AI 应用提供更高效率的计算方案。
- [7] Xue, Zhiwei, et al. "Fully forward mode training for optical neural networks." *Nature* 632.8024 (2024): 280-286.
<https://www.nature.com/articles/s41586-024-07687-4>
推荐理由：提出全前向模式（FFM）学习，首次在光学物理系统中直接实现深度学习训练，突破传统依赖数字模拟的离线建模限制。
- [8] Dalzell, Alexander M., et al. "Quantum algorithms: A survey of applications and end-to-end complexities." *arXiv preprint arXiv:2310.03011* (2023).
<https://arxiv.org/abs/2311.09589>
推荐理由：应用方向重要的综述该论文长达300多页，很全面的覆盖了量子计算在各领域的应用。第一部分从应用领域的角度出发，第二部分是从算法的角度出发。
- [9] Alexeev, Yuri, et al. "Artificial Intelligence for Quantum Computing." *arXiv preprint arXiv:2411.09131* (2024).
<https://arxiv.org/abs/2411.09131>
推荐理由：人工智能辅助量子计算的综述人工智能在量子计算各个流程中的辅助，包含量子纠错，测量，量子线路设计，控制和优化。
- [10] Bausch, Johannes, et al. "Learning high-accuracy error decoding for quantum processors." *Nature* (2024): 1-7.
<https://www.nature.com/articles/s41586-024-08148-8>
推荐理由：针对量子纠错码的解码过程，谷歌团队提出 AlphaQubit，基于机器学习的方法处理来自Sycamore量子芯片的真实数据。
- [11] Liao, Haoran, et al. "Machine learning for practical quantum error mitigation." *Nature Machine Intelligence* (2024): 1-9.
<https://www.nature.com/articles/s42256-024-00927-2>
概述：利用机器学习的方法进行量子错误缓解
- [12] Furrutter, F., Muñoz-Gil, G., & Briegel, H. J. (2024). Quantum circuit synthesis with diffusion models. *Nature Machine Intelligence*, 1-10.

<https://www.nature.com/articles/s42256-024-00831-9>

概述：利用人工智能中的diffusion模型，生成量子编译所需的量子门

[13] Wang, Y., & Liu, J. (2024). A comprehensive review of quantum machine learning: from NISQ to fault tolerance. Reports on Progress in Physics.

<https://iopscience.iop.org/article/10.1088/1361-6633/ad7f69>

推荐理由：量子机器学习方面的综述与量子机器学习相关的概念、算法，并且考虑了硬件限制。

[14] Date, Prasanna, Davis Arthur, and Lauren Pusey-Nazzaro. "QUBO formulations for training machine learning models." Scientific reports 11.1 (2021): 10029.

<https://www.nature.com/articles/s41598-021-89461-4>

概述：将线性回归、支持向量机 (SVM) 和平衡k均值聚类这三种机器学习模型的训练问题表述为QUBO问题，在绝热量子计算机上解决。

[15] Harrow, Aram W., Avinatan Hassidim, and Seth Lloyd. "Quantum algorithm for linear systems of equations." Physical review letters 103.15 (2009): 150502.

<https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.103.150502>

推荐理由：求解线性方程组的量子算法的起始论文

[16] Aaronson, Scott. "Read the fine print." Nature Physics 11.4 (2015): 291-293.

<https://www.nature.com/articles/nphys3272>

概述：Scott Aaronson分析了可能限制HHL算法实际量子优势的注意事项和因素。

[17] Abbas, Amira, et al. "The power of quantum neural networks." Nature Computational Science 1.6 (2021): 403-409.

<https://www.nature.com/articles/s43588-021-00084-1>

概述：论证量子神经网络优势的论文，通过有效维度这一尺度。

[18] McClean, Jarrod R., et al. "Barren plateaus in quantum neural network training landscapes." Nature communications 9.1 (2018): 4812.

<https://www.nature.com/articles/s41467-018-07090-4>

概述：量子神经网络的训练过程中遇到的困难Barren plateaus

[19] Castelvechi, Davide. "The AI-quantum computing mash-up: will it revolutionize science?." Nature (2024).

<https://www.nature.com/articles/d41586-023-04007-0>

概述：非学术论文，论述各方对量子与人工智能结合是否会带来优势的看法。

[20] Evered, S.J., Bluvstein, D., Kalinowski, M. et al. High-fidelity parallel entangling gates on a neutral-atom quantum computer. Nature 622, 268–272 (2023).

<https://www.nature.com/articles/s41586-023-06481-y>

概述：中性原子量子计算在量子门保真度方面取得的重要进展。

[21] Manetsch, H. J., Nomura, G., Bataille, E., Leung, K. H., Lv, X., & Endres, M. (2024). A tweezer array with 6100 highly coherent atomic qubits. arXiv preprint arXiv:2403.12021.

<https://arxiv.org/abs/2403.12021>

概述：中性原子量子计算在物理量子比特的进展，光镊最多抓取原子数量突破六千。

[22] Bluvstein, D., Levine, H., Semeghini, G., Wang, T. T., Ebadi, S., Kalinowski, M., ... & Lukin, M. D. (2022). A quantum processor based on coherent transport of entangled atom arrays. Nature, 604(7906), 451-456.

<https://www.nature.com/articles/s41586-022-04592-6>

概述：中性原子量子计算在量子门操作取得的进展，通过移动光镊实现任意量子比特之间的门操作。

[23] Bluvstein, D., Evered, S. J., Geim, A. A., Li, S. H., Zhou, H., Manovitz, T., ... & Lukin, M. D. (2024). Logical quantum processor based on reconfigurable atom arrays. Nature, 626(7997), 58-65.

<https://www.nature.com/articles/s41586-023-06927-3>

推荐理由：不论是量子计算物理体系，还是量子纠错，都是一篇重要的一篇进展。展示了中性原子量子计算多方面的优势。

[24] Preskill J. Quantum Computing in the NISQ era and beyond[J]. Quantum, 2018, 2: 79.

<https://quantum-journal.org/papers/q-2018-08-06-79/>

概述：Preskill提出NISQ概念的论文。

[25] Quantum startup Atom Computing first to exceed 1, 000 qubits

<https://atom-computing.com/quantum-startup-atom-computing-first-to-exceed-1000-qubits/>

概述：量子计算公司Atom Computing在物理量子比特数量上突破一千个。

[26] Keumars Afifi-Sabet (2023-12-11). "Scientists just built a massive 1, 000-qubit quantum chip, but why are they more excited about one 10 times smaller?". livescience.com. Retrieved 2023-12-27

<https://www.livescience.com/technology/computing/ibm-scientists-built-massive-condor-1000-qubit-quantum-computer-chip-133-qubit-heron-system-two>

概述：IBM的量子芯片Condor量子比特数量突破一千个

[27] Huang, Hsin-Yuan, et al. "Quantum advantage in learning from experiments." *Science* 376.6598 (2022): 1182-1186.

<https://www.science.org/doi/full/10.1126/science.abn7293>

概述：量子计算技术在下面任务中的优势（1）物理系统性质预测、（2）含噪量子态的主成分分析以及（3）物理动力学近似模型学习。

[28] Arute, Frank, et al. "Quantum supremacy using a programmable superconducting processor." *Nature* 574.7779 (2019): 505-510.

<https://www.nature.com/articles/s41586-019-1666-5>

概述：谷歌的量子芯片Sycamore运行随机电路采样（Random Circuit Sampling, RCS）实验实现“量子霸权”

[29] Huang, Cupjin, et al. "Classical simulation of quantum supremacy circuits." *arXiv preprint arXiv:2005.06787* (2020).

<https://arxiv.org/abs/2005.06787>

概述：指出谷歌实现“量子霸权”的论文中，超级计算机 Summit 大约10,000年完成的任务其实在他们的方法和模拟器下可以在不到20天完成。

[30] Ichikawa, T., Hakoshima, H., Inui, K., Ito, K., Matsuda, R., Mitarai, K., ... & Fujii, K. (2024). Current numbers of qubits and their uses. *Nature Reviews Physics*, 1-3.

<https://www.nature.com/articles/s42254-024-00725-0>

推荐理由：非综述，但是很有趣的一项针对量子计算领域的科学的工作，反应出在量子比特数量上去之后为什么还大量使用小数量比特的现象。

[31] Sun, W., Wei, F., Shao, Y., & Wei, Z. (2024). Sudden death of quantum advantage in correlation generations. *Science Advances*, 10(47), eadr5002.

<https://www.science.org/doi/10.1126/sciadv.adr5002>

推荐理由：非综述，定量描述了某个量子优势与噪声的关系。

[32] Google Quantum AI and Collaborators. Quantum error correction below the surface code threshold. *Nature* (2024).

<https://doi.org/10.1038/s41586-024-08449-y>

推荐理由：谷歌在量子处理器Willow上实现纠错码，展示逻辑量子比特的错误率随着码距增大被指数压制。

[33] Gidney, C., & Ekerå, M. (2021). How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. *Quantum*, 5, 433.

<https://quantum-journal.org/papers/q-2021-04-15-433/>

概述：分析了2048 bit RSA分解需要的时间和量子比特数量。

[34] Cai, Z., Babbush, R., Benjamin, S. C., Endo, S., Huggins, W. J., Li, Y., ... & O'Brien, T. E. (2023). Quantum error mitigation. *Reviews of Modern Physics*, 95(4), 045005.

<https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.95.045005>

推荐理由：量子错误缓解方面较新并且影响大的综述，讲述了QEM中各种方法之间的共性和局限性，如何根据存在的主要噪声类型选择缓解方法。

[35] Ebadi, Sepehr, et al. "Quantum optimization of maximum independent set using Rydberg atom arrays." *Science* 376.6598 (2022): 1209-1215.

<https://www.science.org/doi/10.1126/science.abo6587>

推荐理由：利用Rydberg原子的性质求解最大独立集问题

[36] Wurtz, Jonathan, et al. "Industry applications of neutral-atom quantum computing solving independent set problems." *arXiv preprint arXiv:2205.08500* (2022).

<https://arxiv.org/abs/2205.08500>

推荐理由：中性原子的求解图相关的组合优化问题

[37] Rubin, Nicholas C., et al. "Quantum computation of stopping power for inertial fusion target design." *Proceedings of the National Academy of Sciences* 121.23 (2024): e2317772121.

<https://www.pnas.org/doi/10.1073/pnas.2317772121>

概述：应用于核聚变实验中的量子算法

PART 3

展望

AI × Science

展望

3.1 特殊挑战

通观全局，AI×Science尚处起步阶段，但其发展速度却非常迅速。同时，这一领域虽然发展迅猛，但也存在着严重的挑战并蕴含着巨大的机遇。与一般的AI研究领域不同，我们认为，AI×Science在未来可能需要面对两大特殊的挑战：

3.1.1 Science for AI: AI如何与人类共享知识

这既包括如何将专家的知识融入AI系统，也包括如何提高AI系统的可解释性。

目前已知的大量AI系统架构上面的里程碑式进展，往往都伴随着人类将一大类先验知识融入到了AI体系架构的设计之中。例如卷积神经网络（CNN）的架构设计就是融入了空间平移对称性的考虑；基于扩散的生成式模型，则是由扩散物理过程转变而来的一种全新架构。知识图谱、向量数据库等技术则将人类知识转变为一种外部模块，从而被AI系统所应用。然而，这些进展只在一定程度上解决了共享人类专家知识的问题，如何更好地将专家知识转变为AI系统可用的模块、架构或外部知识仍然是未来的重要挑战之一。

其次，随着AI模型的规模越来越大，这些大模型的工作机理将会变得越来越不透明，特别是大模型涌现现象的存在，使得模型能力的提升变得扑朔迷离。因此，如何将AI大模型学习到的知识更好地同步给人类，甚至于找到AI大模型运作的基本原理，则是

未来有待解决的另一重大挑战。

我们认为这两个方面的挑战将成为未来Science for AI研究的重要方向。

3.1.2 AI for Science: AI如何具备实验思维

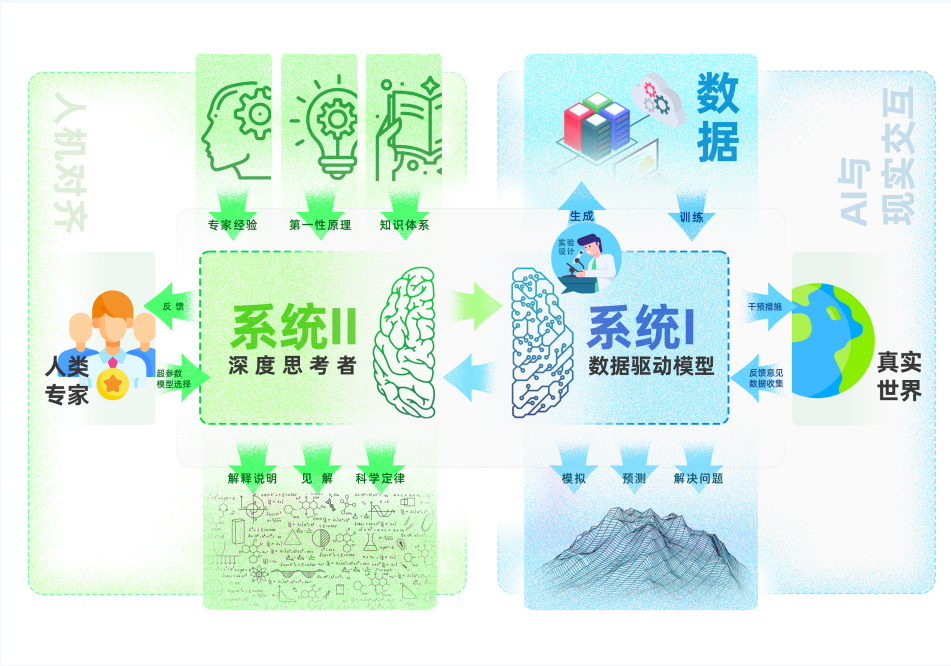
纵观人类数百年的科学活动，做实验是科学研究活动中不可或缺的重要组成部分。如何设计巧妙的实验，从而用最小成本获取最关键信息，是一种不同于被动学习、演绎和推理的全新的思维方式——即实验思维。

然而，目前的科学智能研究绝大部分都集中在如何让AI系统具备人类一样的逻辑思维和基于被动数据的归纳思维模式，却相对缺乏基于干预数据的反事实推理能力，以及与真实世界互动并主动获得数据的能力。之所以出现这一问题，部分原因在于目前的AI系统的犯错误成本过高，因此尚难以具备与真实世界互动的能力。特别是，由于AI系统的不可解释、不透明性，人类难以信任AI系统自行与真实世界互动。这一能力的缺乏才是导致数据稀缺的本质，也将大大限制AI科学家达到人类科学家的水平。

因此，我们认为，未来AI for Science研究的一大挑战将是如何巧妙地设计实验，自主从世界获取新的数据，以促进AI系统的学习，从而更有效的完成学习任务。

3.2 科学世界模型框架

为了应对这两大特殊挑战，我们尝试提出一个科学世界模型框架：



科学世界模型框架

该框架包含了两大智能主体和两个交互界面。两大智能主体包括：数据驱动为主的基础模型（Foundation Model）和人机交互为主的深思考者（Deep Thinker）。两个交互界面主要包括：人机对齐界面（Human-AI Alignment）和AI-现实互动（AI-Reality Interaction）界面。下面分别展开论述：

3.2.1 数据驱动模型

数据驱动模型是一个单一领域，或跨领域的AI模型，可以通过机器学习的方式从多源、多模态数据

之中提取有用的模式和规律。但该基础模型不仅包括学习和建模，理想化的终极形态应该是一个代理模型（Surrogated Model），该模型将可以替代真实的研究对象，从而完成预测、模拟、问题求解等一系列下游任务。

然而，面向科学智能的基础模型还应具备一种全新的能力，即“做实验”——如何为了学到更好的替代模型而直接与真实世界互动，从而获取新的干预数据。这部分干预世界、获取反馈的功能则将专门由实

验设计模块 (Experiment Designer) 完成。考虑到 AI 与世界互动的成本过高的问题, 如何在尽可能少而轻地与真实世界互动的前提下, 获取更多有用信息乃至补全数据, 将会是实验设计模块的重要任务。

从一定意义上说, 基础模型模块所完成的类似于人类大脑中的系统 I 的功能。但是, 与人脑中的系统 I 不同的是, 该基础模型还包括了一定程度上的问题求解能力和实验设计能力。

3.2.2 深思者

目前的 AI 系统无法替代人类科学家来单独探索世界。我们认为, 在今后较长的一段时间内, 科学研究仍将通过人类科学家和机器的协作来完成。在数据驱动模型之外, 融合第一性原理、先验知识和领域机理的深度推理模型必不可少, 类似 o1、o3 和 R1 这些推理模型, 以及神经符号体系中的 Lean, 这就是“深思者”。深思者在一定程度上类似于人脑中的系统 II, 数据驱动模型需要和深思者深度融合, 将深度学习的感知和符号推理的严谨、可解释相结合。

一方面, 深思者会充分获取人类专家的知识 and 经验, 这包括从大量科学文献中获取知识、完成推理; 也包括将人类专家的知识 (以知识图谱、概率图模型等形式呈现) 存储在单独的处理模块中, 并与基础模型深度结合起来; 还包括将人类专家的经验性知识、领域的第一性原理以归纳偏置 (Inductive Bias, 例如对称性的考量等) 的方式, 转变为基础模型模块融入神经网络架构设计之中, 这一功能可通过构建某种通用的神经网络架构设计语言来完成。

另一方面, 深思者还需要将基础模型学习到的隐性知识以对人类更友好的方式展示出来, 从而与人类科学家对齐。这包括但不限于生成具有启发性和洞察力的科学报告、友好的可视化界面, 以及对数据驱动模型思考过程的解释等等, 甚至于深思者自发总结的新科学规律。

3.2.3 AI 与现实的交互

为了更好地探索真实世界, 基础模型主体应具备重要的“做实验”的能力, 从而通过主动地与世界互动, 以获得更多反馈数据。这一交互过程将主要由 AI-现实交互界面完成。

一方面, 基础模型应通过“实验设计器”规划出一个如何干预世界以更好地获得信息的方案, 再根据这一规划来与真实世界互动。但是, 由于现实中的互动往往是成本高昂的 (需要考虑到实验成本和犯错误成本), 因此这一互动过程并非类似于具身智能那样频繁而直接, 它更应该以一种弱交互的方式来干预世界。

另一方面, 基础模型应能直接从真实世界获取反馈, 这既包括对智能体的评价反馈, 也包括从真实世界收集到的全新数据。这些新的数据往往可以反映出对现实世界进行干预之后的反馈, 因此也可以被理解作为一种干预数据。这些干预数据可以进一步更新到训练数据集之中, 从而更好地训练基础模型。这种以主动完成数据收集和机器学习的方式, 往往可以获得更好的效果。

3.2.4 人类专家与AI系统的对齐

人类-AI对齐界面主要负责人类与AI系统的信息和知识的共享，从而达到人机能够更好地协作这一重要目的。这包括两方面：

首先，对于任何学科，目前绝大部分知识是掌握在人类专家手中的，因此，如何将人类专家的知识、经验以及一些灵感性的启发式思考，融入基础模型这样的以机器学习为主的系统之中，就是一个亟待解决的难题。尽管目前已普遍应用的知识图谱、向量数据库等技术可以部分解决这个问题，但是，更新的技术仍有巨大的应用潜力。例如，可以根据人类的直觉经验而自动构造神经网络架构、选择合适超参的工具。

其次，可解释性问题仍然是一个桎梏AI发展的重要卡点。如何设计更友好的人机交互界面，从而将AI的思考、推理过程展现给人类是未来待解决的重要问题。

