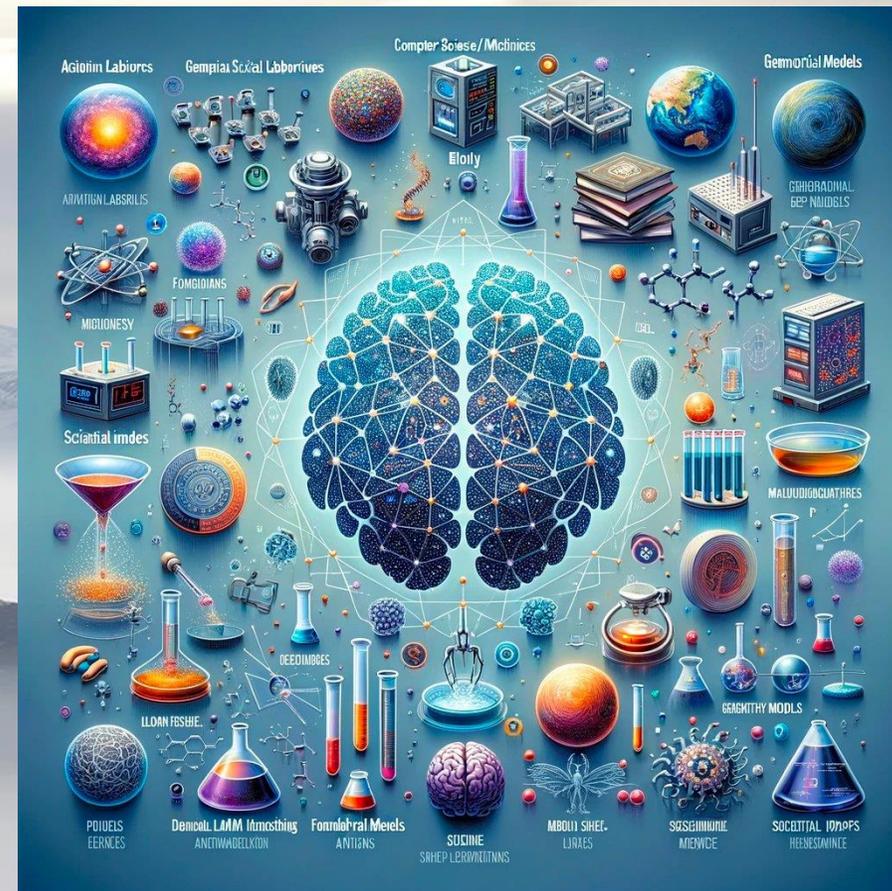


AI4Science, DeepSeek和 大科学创新

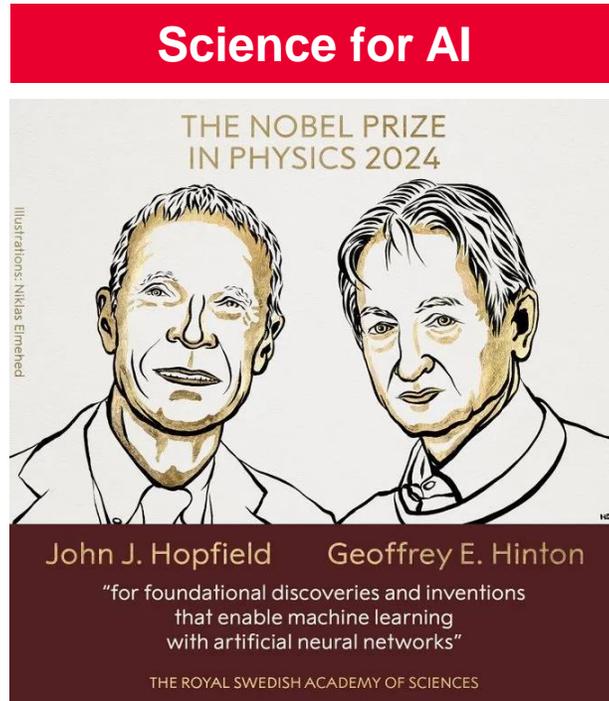
乔楠博士，华为AI4Science首席科学家
未来网络研究室



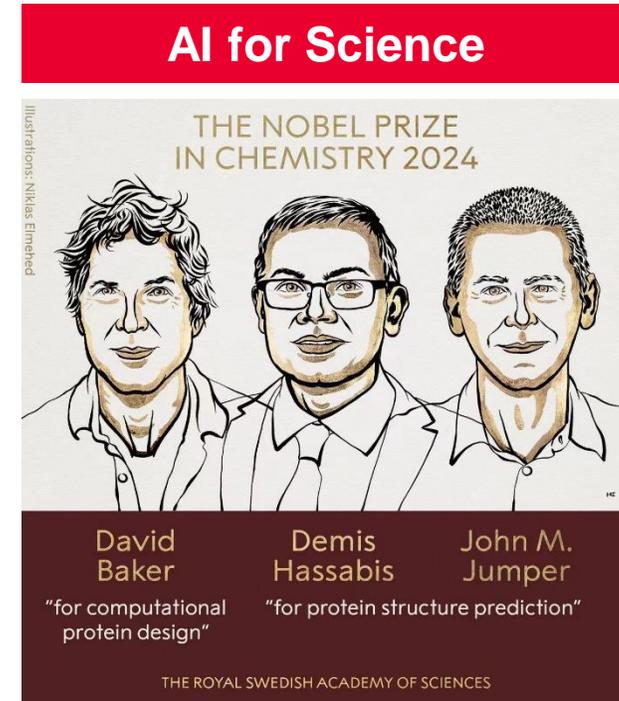
01

科学新范式

Nobel Prize for AI



2024年10月8日，瑞典皇家科学院宣布，将2024年诺贝尔物理学奖授予John J. Hopfield和Geoffrey E. Hinton，以表彰他们“为利用人工神经网络进行机器学习做出的基础性发现和发明”；

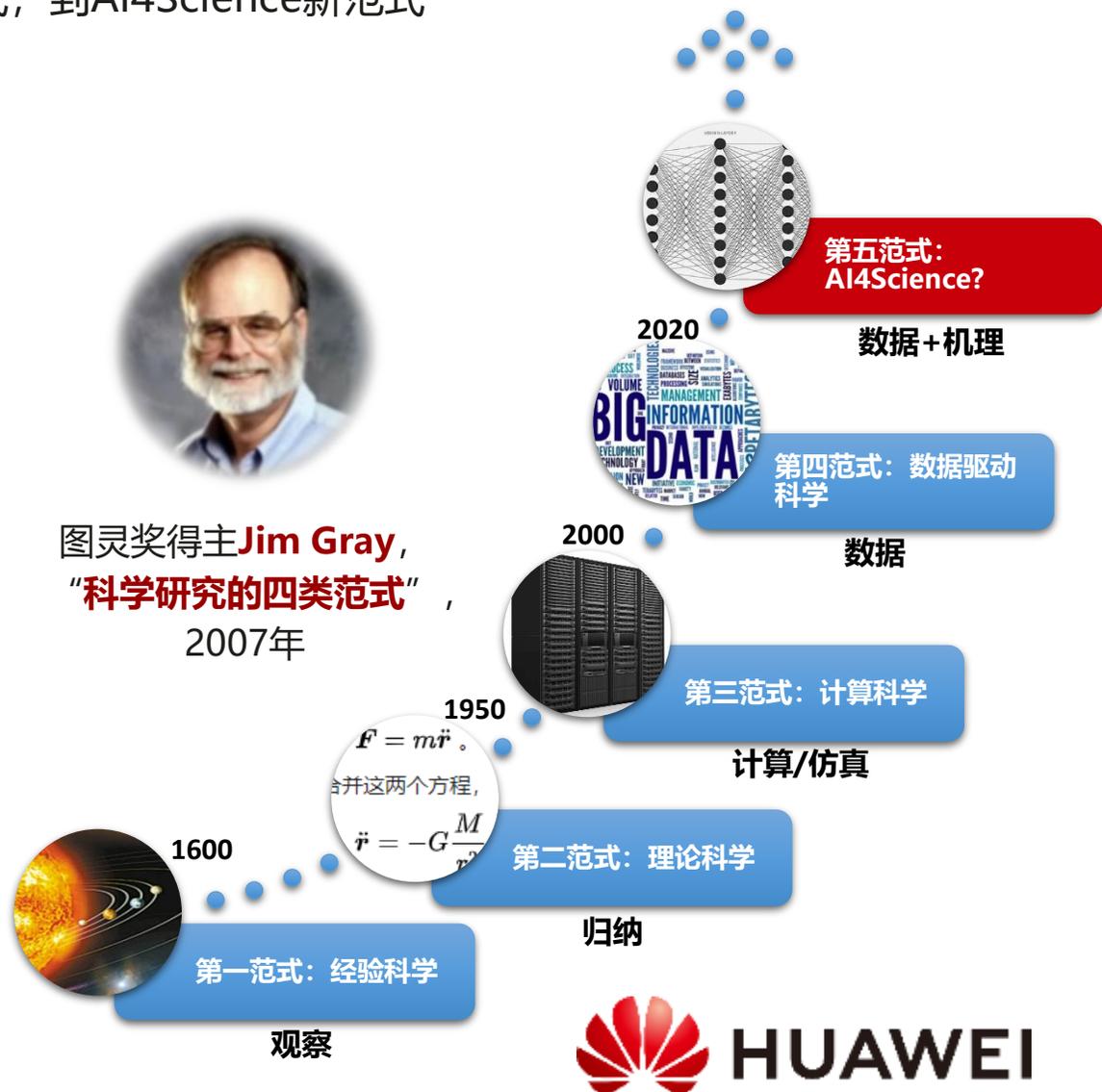


2024年10月9日，瑞典皇家科学院宣布，将2024年诺贝尔化学奖授予David Baker，以表彰其在计算蛋白质设计方面的贡献，另一半则共同授予Demis Hassabis和John M. Jumper，以表彰其在蛋白质结构预测方面的贡献。

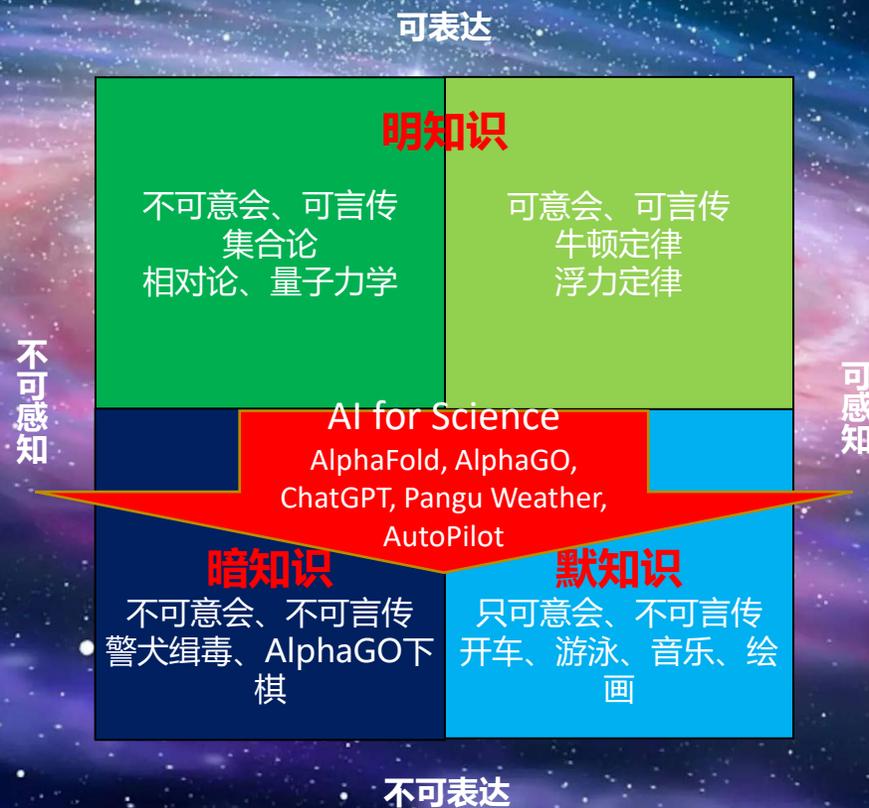
科学研究发展之路

从经验范式，数据驱动范式，到AI4Science新范式

- **第一范式：经验范式**，基于经验的观察，是科学家对万物万象的总结。比如著名的天文学家开普勒，他通过观察总结出天体运行的规律。
- **第二范式：理论范式**，指数学家对经验进行数学抽象和推演，比如用于描述经典力学的牛顿运动方程，用来描述电场磁场关系的麦克斯韦方程等。
- **第三范式：计算范式**，随着计算机的发明，人们开始有能力求解复杂的物理方程。比如，通过有限元或者有限差分方式求解流体方程，从而有助于人类对于天气预报进行精准预测。
- **第四范式：数据驱动范式**，这个过程中ML（机器学习）扮演着非常重要的角色，人们使用ML方法来分析数据，寻找规律，并进行预测。
- **第五范式：AI4Science范式？**，是前四种范式的有机结合，发挥了经验和理论各自的特长，又把AI和计算科学融合在一起。AI4Science是对科学发现更全面的认知。



AI for Science 扩展知识的边界



默（隐形）知识 Tacit Knowledges, Polanyi, 1958年

暗知识 Dark Knowledges, 王维嘉, 2019年

AI for Science 扩展知识的边界

AI通过对人类积累的数据学习对暗默知识进行建模

明知识

AlphaFold

AlphaGO

AlphaProof

PanguDrug

PanguWeather

ChatGPT

SORA

蛋白质
数据库

人类
棋谱数据

数学物理
定理

小分子药
物数据

70年气象
数据

人类
语言数据

视频
数据



预测未知
蛋白结构

超出人类
棋力

发现
新定理

设计
新药物

气象
预报

设计出创新
软件

设计出创新
视频

Nature of Science
understanding

Nature language/vision
understanding

暗知识
不可意会，不可言传

默知识
只可意会，不可言传



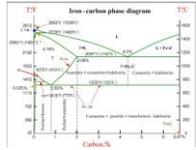
02

AI4Science学科场景

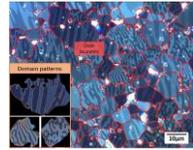
AI for Science

领域、学科、场景及方法

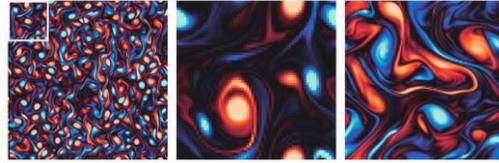
时间维度 ↑



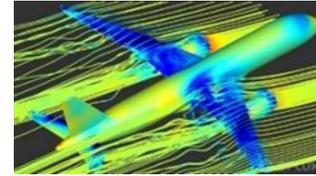
Phase diagram



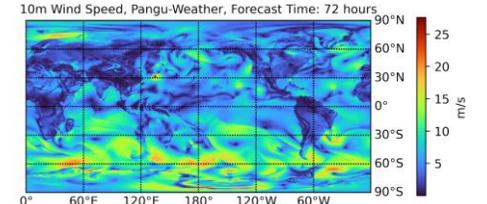
Material texture



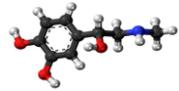
Larger domain Decaying More turbulent



Aerofoil simulation



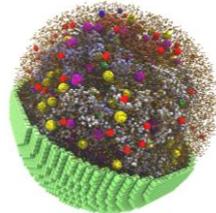
Weather forecasting



Small molecule



Proteins



Cell



Brain

宏观尺度 (Macroscale)

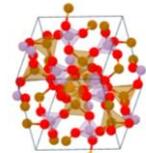
- 学科: 经典物理学、工程学、材料科学、地球科学等
- 问题: 结构力学分析, 流体动力学模拟, 热传导和热交换分析, 材料行为的大规模模拟, 地球科学模型
- 方法: 有限元分析 (FEA), 计算流体动力学 (CFD), 多体动力学 (MBD), 热力学和传热模型, 经典力学模型, 有限差分法和有限体积法, 地球科学和环境工程统计模型和经验公式

介观尺度 (Mesoscale)

- 学科: 材料科学、纳米科技、软物质物理、生物物理等
- 问题: 多粒子系统的动力学, 相界面与相分离现象, 自组装与纳米结构形成, 复杂流体的流变学性质, 材料的微观结构与宏观性质之间的关系
- 方法: 连续介质模型, 分子动力学 (MD) 模拟, 蒙特卡洛模拟, 多尺度建模, 非平衡动力学理论



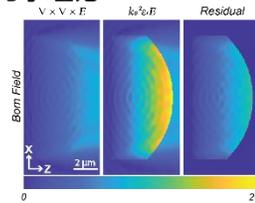
Atomic orbitals



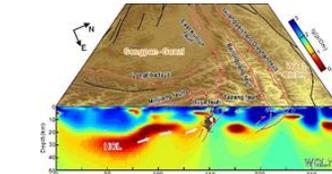
Crystal structure

微观尺度 (Microscale)

- 学科: 物理化学、分子物理、量子化学、分子生物学等
- 问题: 电子结构计算, 分子动力学模拟, 化学反应机理和路径分析, 分子间相互作用的研究, 材料的性质预测
- 方法: 量子力学计算 (如密度泛函理论DFT)、分子动力学模拟、分子力场、化学反应模拟



Electromagnetic simulations



Geophysical Electromagnetic Inversion



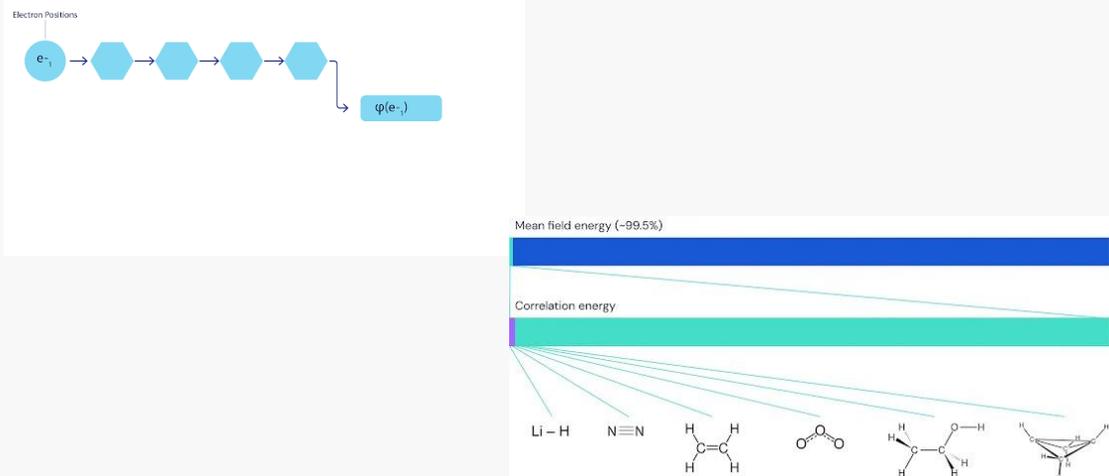
空间维度 →



AI for Quantum Mechanics

FermiNet (DeepMind 2024)

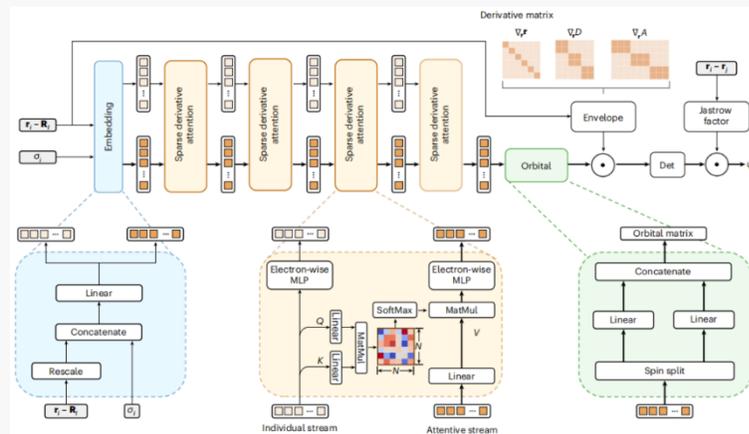
FermiNet 是由 DeepMind 提出的神经网络架构，具有参数化的波函数表示方法，可以高效处理多电子量子系统，提升了电子结构计算的精度和效率。FermiNet 计算原子和分子的能量，在双环丁烷 30 个电子组成的分子系统上仍然捕获了 97% 的相关能量。



David Pfau, et al. Accurate computation of quantum excited states with neural networks. Science, 2024

LapNet (北大&字节 2024)

LapNet 是由字节跳动与北京大学开发的基于神经网络的量子变分变分蒙特卡罗 NN-VMC 的深度学习架构，团队设计了一个 Forward Laplacian 的计算框架，通过有效的前向传播过程来计算与神经网络相关的 Laplacian (这是 NN-VMC 的瓶颈)。LapNet 实现了显著的加速率，从而极大地将 NN-VMC 的适用性扩展到更大的系统。



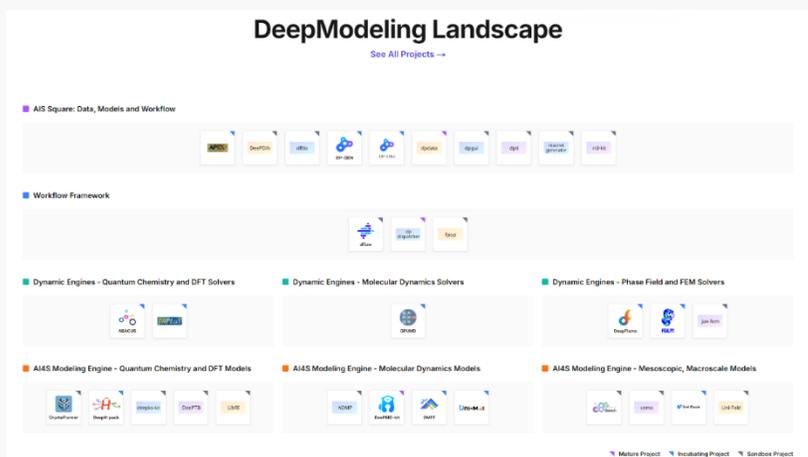
Ruichen Li, et al. A computational framework for neural network-based variational Monte Carlo with Forward Laplacian. Nature M. I., 2024

AI能够高效处理多体问题，加速波函数求解，为量子材料设计和量子信息处理提供了新的思路

AI for Molecular Dynamics

DeePMD (深势科技 2023)

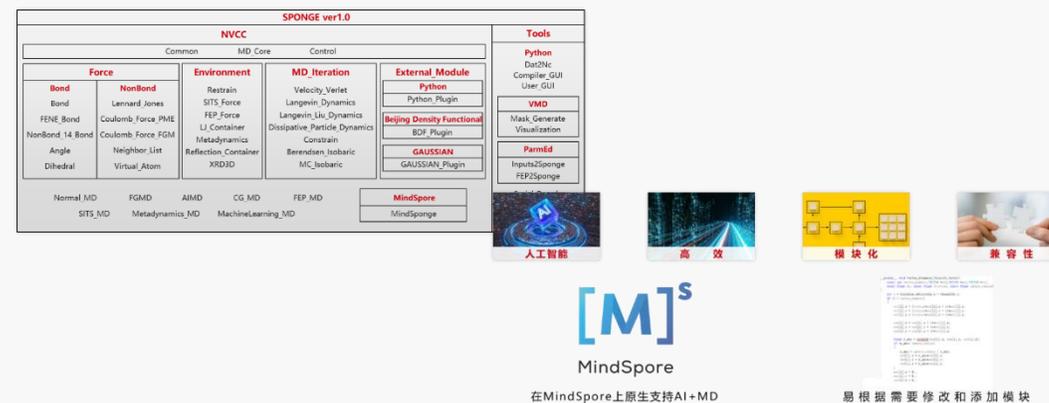
DeePMD-kit 是深势科技开发的一个具有很大潜力的深度学习包，通过将第一性原理建模、以机器学习为代表的驱动方法和高性能计算紧密结合，能够以第一性原理的精度处理超大规模的分子动力学问题。DeePMD-kit 能够在提高模拟精度的同时，显著提升计算效率，解决了传统模拟方法中的精度与计算量之间的矛盾。



Weinan E, et. al. DeePMD-kit v2: A software package for deep potential models. 2023

SPONGE (北京大学 2023)

SPONGE是北大和深圳湾实验室开发的国内首个开源的通用分子模拟软件框架，旨在解决现有分子模拟软件的可靠性和效率问题，并融合人工智能技术推动分子模拟领域的变革。SPONGE与华为MindSpore合作，推出了可与AI算法无缝融合的分子模拟库MindSpore。SPONGE的开发得到了学术界和产业界的支持，未来希望成为下一代分子模拟平台的领航者。



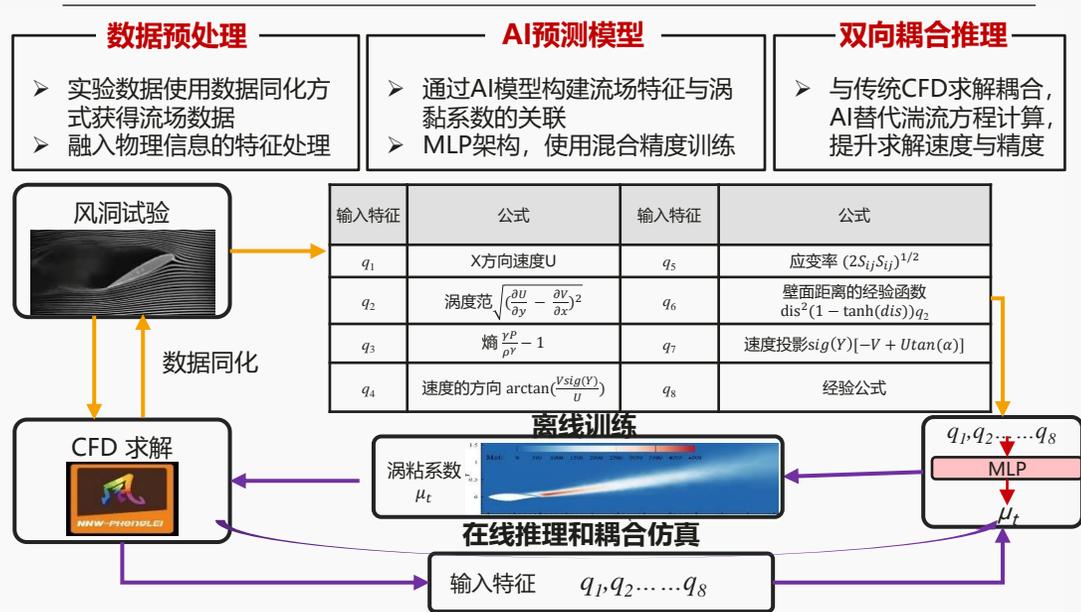
YQ Gao, et. al. SPONGE: A GPU-Accelerated Molecular Dynamics Package with Enhanced Sampling and AI-Driven Algorithms.

AI 加速原子级别的模拟过程，使得复杂分子系统的长时间尺度动态行为和多种物理化学过程的探索成为可能

AI for Fluid Mechanics

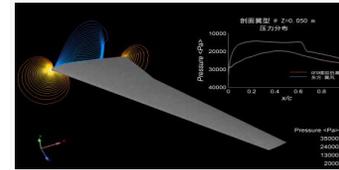
秦岭·翱翔 (西工大&华为, 2023)

西北工业大学团队联合华为, 推出面向飞行器的AI湍流大模型“秦岭·翱翔”, 突破复杂流动下流场准确预测, 使能双向耦合多机多卡并行, AI与CFD求解器耦合, 实现了湍流流场的高精度数值仿真。



东方·御风 (商飞&华为, 2023)

中国商飞联合华为打造东方·御风, 实现了千倍级仿真效率提升, 荣获23年WAIC最高奖SAIL奖



3D超临界翼型流场仿真

千万网格流场秒级推理

典型超临界机翼流场预测

降低仿真时间1000倍

效率突破:

AI模型替换传统Navier-Stokes方程求解, 提升典型场景仿真效率



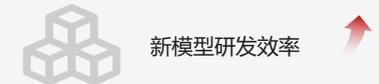
精度突破:

对流动剧烈变化区域特征进行精细捕捉, 提升整体AI流体仿真的预测精度



模型突破:

建立大数据样本下模型组件化与分布式并行能力, 大幅提升新模型研发效率



场景突破:

建立流体到AI的数据统一映射, 适用多种仿真场景



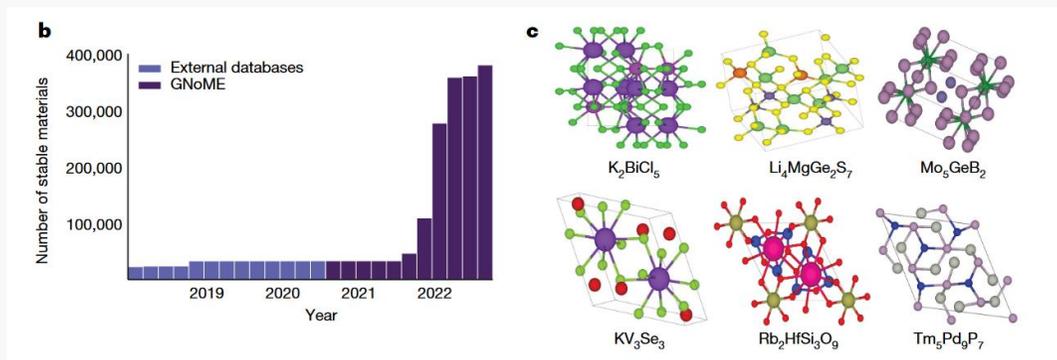
AI 通过数据驱动的模式和优化算法, 显著提升了复杂流动现象的预测能力、求解效率和设计创新

AI for Materials Science

GNoME (DeepMind, Nature 2023)

DeepMind发布的GNoME工具发现了220万种新晶体，其中38万个稳定结构有望合成，可能推动下一代电池、超导体等技术变革。

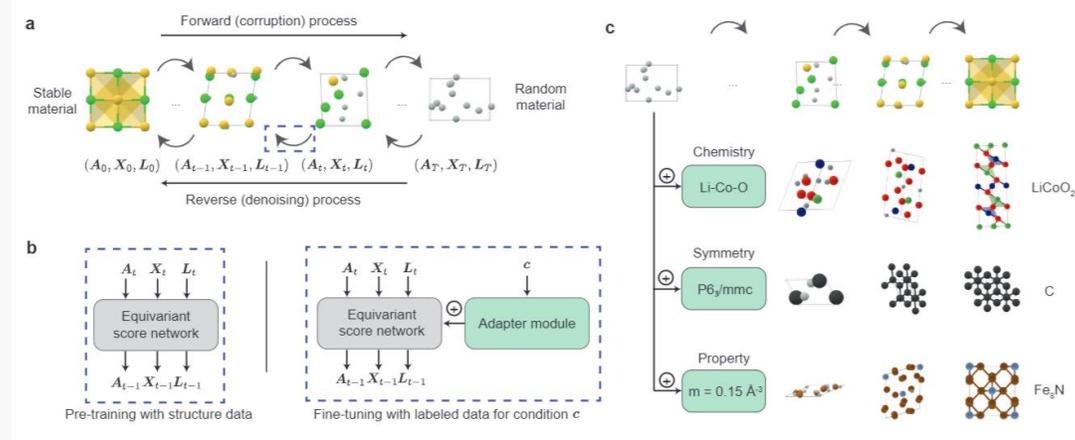
GNoME通过图神经网络（GNN）模型，可以高效地预测材料的晶体结构和稳定性，大幅加速新材料的发现过程。与传统的实验方法相比，AI模型能在短时间内预测数百万种材料，从而缩短了研究周期，节省了大量实验成本。



Amil Merchant et al. *Scaling deep learning for materials discovery*, Nature (2023)

MatterGen (Microsoft, 2024)

MatterGen 是微软研究院开发的一种新型生成模型，用于设计具有所需性质的无机材料。MatterGen能够生成稳定且多样的晶体结构，并通过新型的基于扩散的生成过程细化原子类型、坐标和周期性晶格。该模型可以根据特定属性要求进行微调，生成满足多种属性约束的新材料，如化学组成、对称性以及机械、电子和磁性等性能，进行定向材料设计。



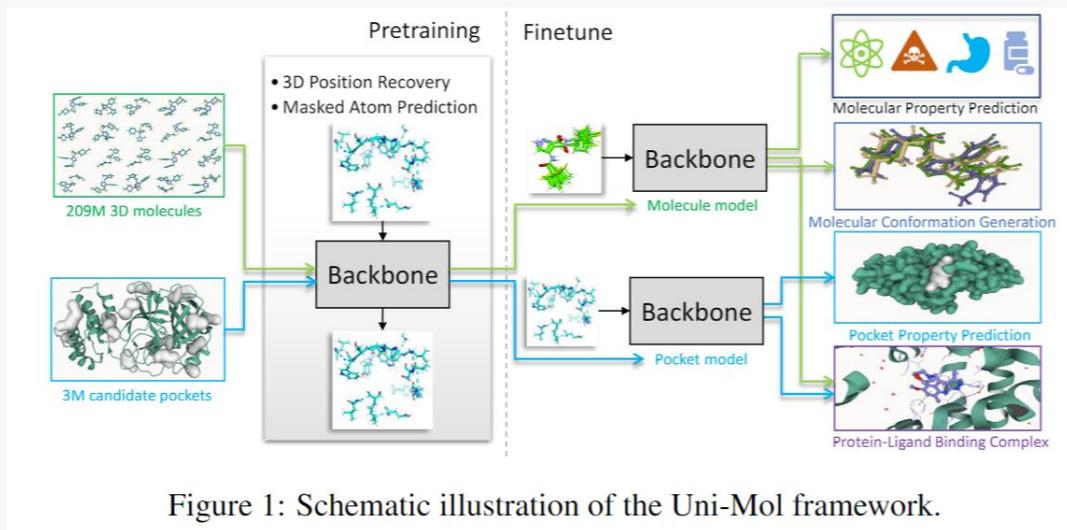
Tian Xie et al. *MatterGen: a generative model for inorganic materials design*. (2024)

AI 快速筛选和预测新材料的性能，促进了材料设计的创新、加速了实验进程

AI for Small Molecules

Uni-Mol (深势科技, 2023 LCLR)

Uni-Mol是深势科技于2022年5月发布的一款基于分子三维结构的通用分子表征学习框架。Uni-Mol将分子三维结构作为模型输入，并使用约2亿个小分子构象和300万个蛋白表面空腔结构，使用原子类型还原和原子坐标还原两种自监督任务对模型进行预训练。



Weinan E et al. Uni-Mol: A Universal 3D Molecular Representation Learning Framework, LCLR (2023)

Pangu Drug (华为&中科院, 2021 SCLS)

华为和中科院团队提出了盘古药物分子大模型，盘古采用不对称图到序列 (graph2seq) 结构，基于17亿类药化学分子进行预训练，能够预测分子性质、生成和优化分子。该模型在20个药物发现任务中取得领先成绩。盘古的应用有效提升了药物研发效率，推动了AI在制药领域的应用。



Nan, Qiao et al. PanGu Drug Model: learn a molecule like a human. SCLS (2021)

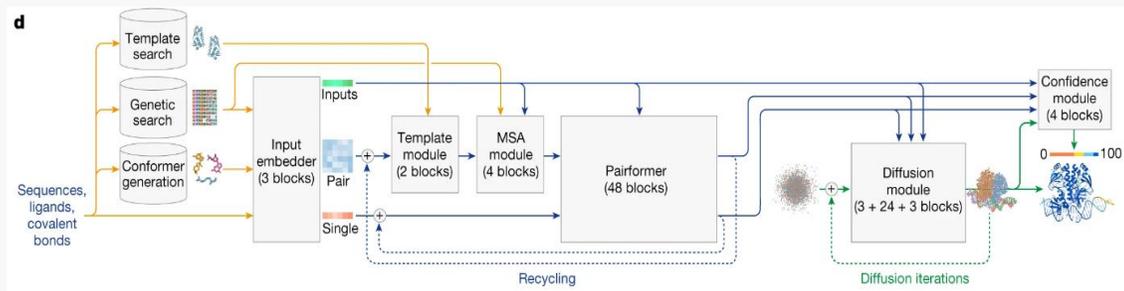
AI加速了分子设计、性质预测和药物发现的过程，推动了新药研发的突破

AI for Protein Science

AlphaFold3 (DeepMind 2024)

AlphaFold3能够以原子级别的精度预测多种生物分子的结构和相互作用，包括蛋白质、DNA/RNA、小分子和离子等。在蛋白质-配体相互作用、蛋白质-核酸相互作用等任务中，AlphaFold3的预测准确性显著高于现有工具。

- 输入/输出：蛋白质、DNA、RNA、配体等生物分子/生物分子的结构和相互作用
- 模型架构：Pairformer，MSA进行特征提取；生成模块使用Diffusion Module；激活函数使用SwiGLU

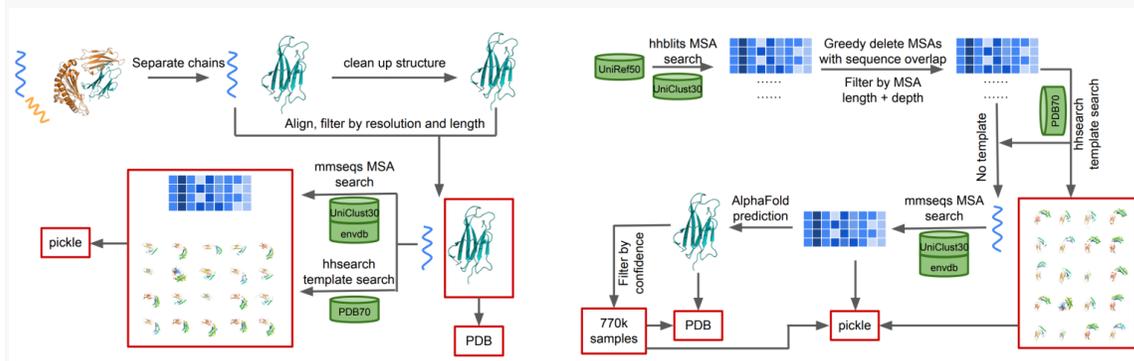


Abramson, J et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature (2024)

MEGA-Protein (北京大学 2023)

北大与华为团队的共同研发了MEGA-Protein工具，包括蛋白质结构预测功能MEGA-Fold，MSA生成修正工具MEGA-EvoGen，蛋白质结构评分工具MEGA-Assessment三部分组成，在CAMEO-3D蛋白质结构预测赛道多次获得月榜第一。

- 采用MMSeqs2的序列检索能力，相比于AlphaFold2速度提升2-3倍
- 突破“孤儿序列”结构预测准确度的问题
- 内存复用大幅提升内存利用效率，同硬件条件下支持更长序列推理



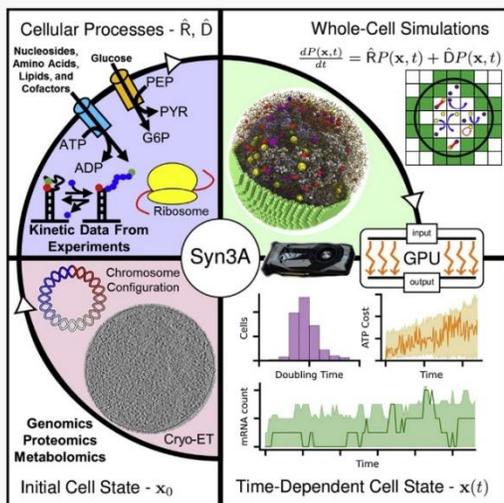
YQ Gao et al. PSP: Million-level Protein Sequence Dataset for Protein Structure Prediction. (2023)

AI快速预测蛋白质的结构、折叠过程及其相互作用，帮助加速药物设计、疫苗研发和蛋白质功能的理解

AI for Cell

Minimal Cell (University of Illinois & NIH, Cell 2022)

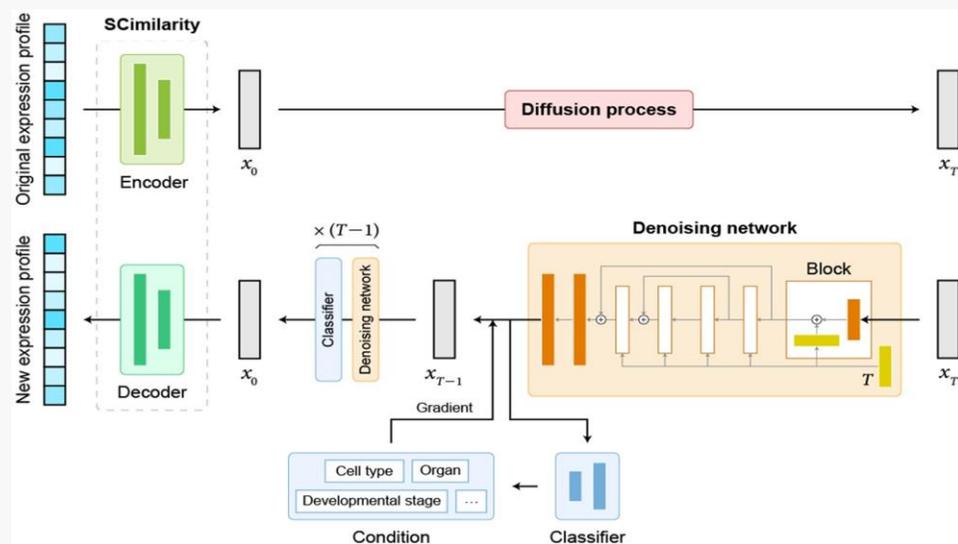
来自美国伊利诺伊大学和NIH的研究团队构建出活的“最小细胞”，其基因组只剩下最基本的要素，与此同时，他们开发了一种三维的、完全动态的活的最小细胞的动力学模型，模拟实际细胞中发生的事情。该模型为理解细胞如何平衡代谢、遗传过程和生长提供了深入洞察，有助于探索生命基本原理，并推动生命科学领域的研究进展。



Zaida Luthey-Schulten et al. Fundamental behaviors emerge from simulations of a living minimal cell. Cell (2022)

scDiffusion (清华大学, Bioinformatics 2023)

清华大学提出了一种基于扩散模型的单细胞大模型scDiffusion，能够在大规模单细胞数据集上稳定训练，学习复杂的数据分布，通过训练，scDiffusion可以在任何自定义的单细胞水平条件下生成基因表达谱。



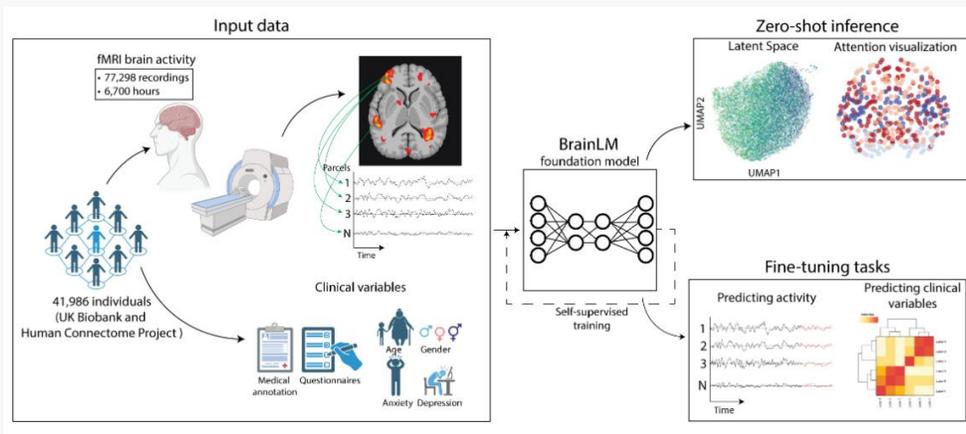
Xuegong Zhang et al. scDiffusion: conditional generation of high-quality single-cell data using diffusion model, Bioinformatics, 2024

AI 建模揭示细胞行为和生物过程的复杂模式，推动疾病诊断、个性化治疗以及生物学研究的深度探索

AI for Neuroscience

BrainLM (Baylor & Yale, 2024)

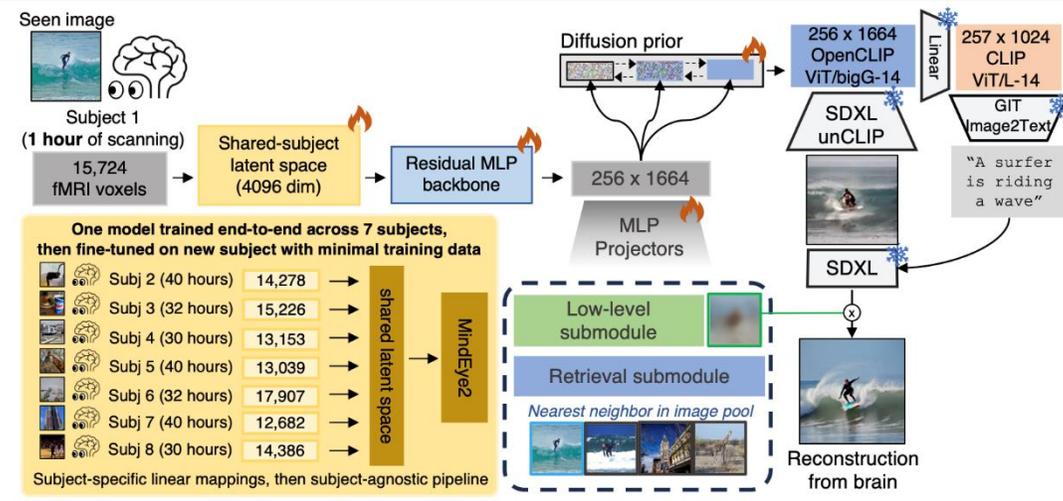
贝勒医学院和耶鲁大学的研究团队利用生成式人工智能创建了一个基础性大脑活动模型，称为大脑语言模型（BrainLM）。通过分析来自40,000名受试者的80,000个fMRI扫描数据，训练了BrainLM模型。BrainLM能够准确预测临床变量如年龄、焦虑和创伤后应激障碍，并预测未来的大脑状态。更重要的是，该模型对完全未见过的外部队列表现良好。



Josue Ortega Caro et al. BrainLM: A foundation model for brain activity recordings. ICLR (2024)

MindEye2 (Stability AI, 2024)

Stability AI的研究人员开发出MindEye2技术，仅用1小时的功能磁共振成像数据就能重建大脑中的图像。除了能够重建出某人看到的具体图像，它还能在有多个图像选项的情况下，准确识别出哪一个是那个人实际看到的图像。这一功能使得MindEye2在解读大脑活动方面具有更高的准确性。



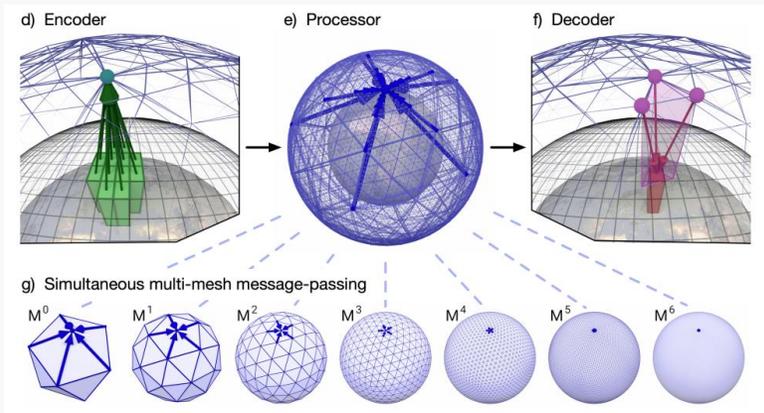
Paul S. Scotti et al. MindEye2: Shared-Subject Models Enable fMRI-To-Image With 1 Hour of Data. (2024)

AI 助力神经网络建模、脑疾病诊断以及脑机接口的开发，推动对大脑的理解和应用。

AI for Earth Science

GraphCast (DeepMind 2023, Science)

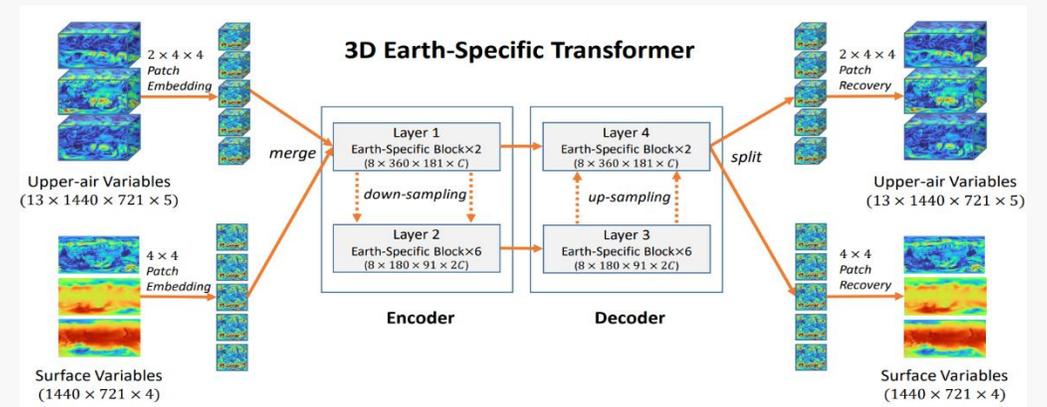
DeepMind发表的基于图神经网络的天气预测模型GraphCast，能够在1分钟内预测未来10天的数百个天气变量，且在准确度上超越传统气象预报系统。GraphCast在预测极端天气事件，如飓风路径和极端温度方面表现突出。相比传统方法，GraphCast能以更低能耗提供更快速、精准的预测，标志着天气预报技术的重大进步，并为未来气候研究开辟新方向。



Remi Lam et al. Learning skillful medium-range global weather forecasting, Science (2023)

Pangu Weather (HUAWEI 2023, Nature)

华为云发布的盘古气象大模型突破了AI天气预测的精度和速度瓶颈，成为首个精度超越传统数值预报的AI模型。该模型能在秒级完成全球气象预测，速度比传统方法提升10000倍。使用高分辨率数据，盘古气象大模型在短期至中期天气预报中表现出色，并能精准预测极端天气事件。与现有AI模型相比，其准确性和效率显著提高，且能提供更为细致的天气预测细节，标志着气象预报技术的重大进步。



Qi Tian et al. Accurate medium-range global weather forecasting with 3D neural networks. Nature (2023)

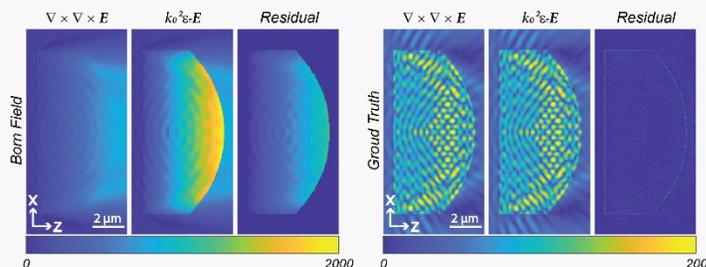
AI快速处理气候变化、地震活动、天气模式等复杂现象，为环境监测、自然灾害预测提供精准决策支持。

AI for Electromagnetic

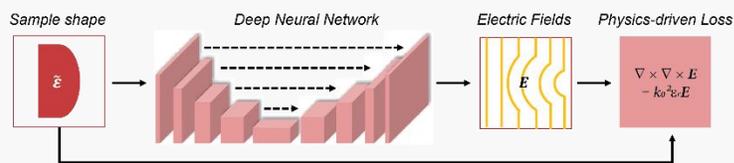
MaxwellNet (EPFL 2021)

传统的电磁场模拟依赖于计算密集型的电磁求解器，如有限元法 (FEM) 和有限差分法 (FDTD)，这些方法不仅耗时长，而且难以高效进行逆向设计。MaxwellNet通过将麦克斯韦方程的残差作为物理驱动的损失函数，成功训练出能够快速且准确地求解电磁场的神经网络，避免了对传统求解器的依赖。

(a) Physics-driven loss



(b) Main Idea

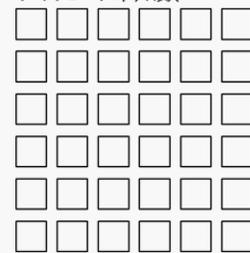


MaxwellNet: Physics-driven deep neural network training based on Maxwell's equations. Lim, Joowon and Psaltis, Demetri (2021)

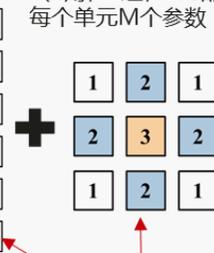
金陵.电磁脑 (东南大学&华为 2022)

大规模阵列天线在国防 (相控阵雷达天线) 和民用 (5G基站天线, 自动驾驶) 广泛应用, 但阵列天线的阵列规模大、单元构成复杂, 亟需快速精确仿真。金陵.电磁脑在大规模阵列天线等场景中, AI电磁仿真相比传统方法效率提升10+倍, 且随着阵列单元规模增大, 效率进一步提升。

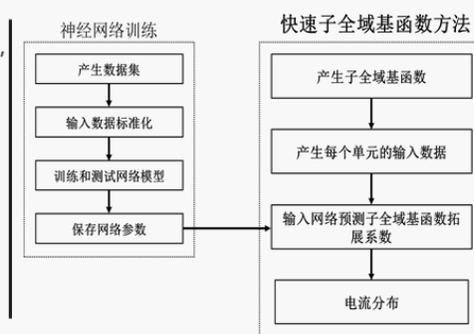
计算一: N 个单元, 每个单元1个未知数



计算二: 9个基础单元 (1角, 2边, 3内部), 每个单元 M 个参数



SED基函数法计算复杂度 $O(N + 9M)$: 单元个数 $N = N_x \times N_y$ (每个单元1个未知数), 9个基础单元 (每个单元 M 个未知数, 表征单元的细节)



AI融合方法: 单元个数 N 可达成千万, AI代替计算量最高的这部分, 计算复杂度从 $O(N + 9M) \rightarrow O(1 + 9M)$

AI革新电磁模拟和设计流程, 极大提升效率与精度, 推动电磁技术进步。

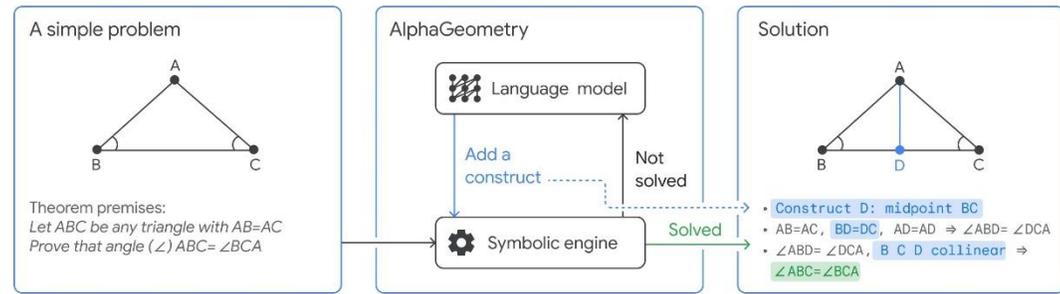
AI for Mathematics

AlphaGeometry (DeepMind 2024)

AlphaGeometry 它是一个神经-符号混合系统 (Neuro-Symbolic AI) , 其中的语言模型基于 Gemini, 并在比其前身多一个数量级的合成数据上从头开始训练。这有助于该模型解决更具挑战性的几何问题, 包括有关物体运动以及角度、比例或距离方程的问题。

AlphaGeometry 2 可以解决过去 25 年中所有 IMO 几何历史问题的 83%, 而其前身的解决率仅为 53%。在 IMO 2024 中,

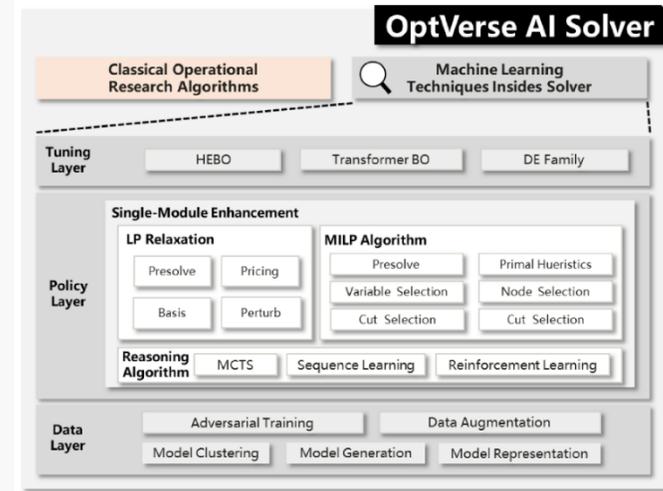
AlphaGeometry 2 在收到问题4后19秒内就解决了它。



Trieu H. Trinh et al. Solving olympiad geometry without human demonstrations. Nature (2024)

OptVerse AI Solver (上交大&华为 2024)

OptVerse AI Solver 是上交大和华为开发的AI优化求解器。它通过利用生成模型生成复杂的 SAT 和 MILP 实例, 模拟现实问题的多样结构, 缓解了数学规划实例稀缺的问题。在近日更新的国际权威数学优化求解器榜单Hans Mittelmann benchmark上再次实现突破, 5项重量级榜单荣登榜首。



Feng Mao et al. Machine Learning Insides OptVerse AI Solver: Design Principles and Applications.aRxiv (2024)

AI通过自动化推理, 推动数学理论的创新和应用

03

AI4Science发展趋势

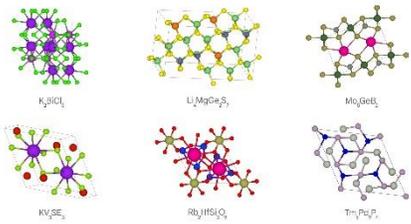
AI for Science

科学是从微观粒子到宏观系统的连续体系

Materials

421,000 new stable crystals

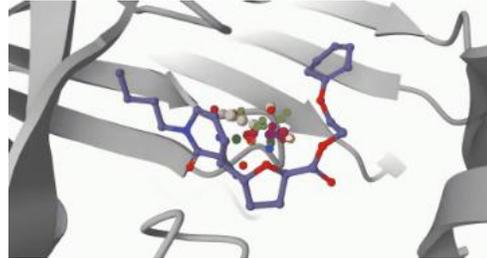
Computation crystals increase **↑10X**



Small Molecule

100 million new molecules

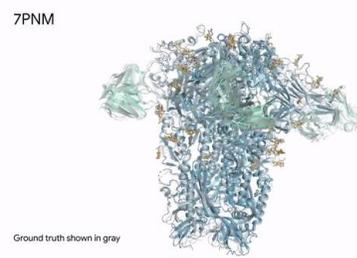
Drug design success rate **↑40%**



Protein

200 million protein structures

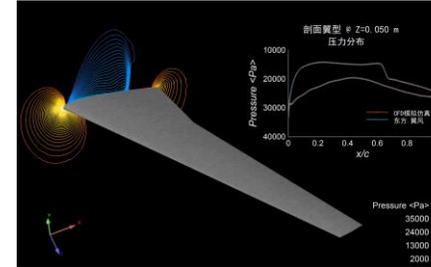
Accuracy increase **↑65%**



Aerofoil

Millisecond-level inference

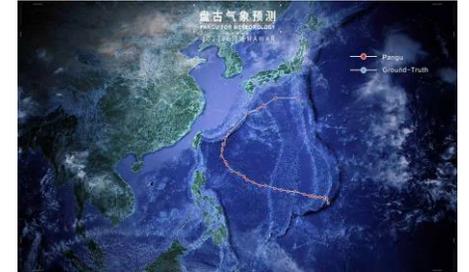
Simulation speed increase **↑1000X**



Weather

10% increase in forecast accuracy

Speed forecasting **↑10,000X**

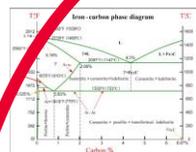


| Application Field | Mathematics | Physics | Chemistry | Biology | Fluid mechanics | Meteorology | |
|--------------------|---|---|--|--|--|---|-------|
| Objective | Geometry solver, theorem proving, algorithms discovery, operations optimization | Quantum chemistry, interatomic potential energy, force field, materials discovery | Molecule modeling, retrosynthesis, chemical synthesis yield | Protein structure prediction, pathogenic missense variants, Cell digital twin | Navier-Stokes equation, multiphase flow, turbulence, airfoil flow field, | Weather forecasting, extreme weather, atmospheric processes, climate modeling | |
| Peer Application | AlphaGeometry (DeepMind) AlphaProof (DeepMind) AlphaDev (DeepMind) | FermiNet (DeepMind) GNoME (DeepMind) DeepMD (Deep Potential) | Graphormer (Microsoft), MegaMolBART (NVIDIA), RetroSub (Microsoft) | AlphaFold (DeepMind), AlphaMissense (DeepMind), Minimal Cell (University of Illinois) | U-FNO (Stanford University) BNNRANS (Nanjing Institute of Technology) | GraphCast (DeepMind), Aurora (Microsoft), ClimaX (Microsoft) | |
| Huawei Application | OptVerse AI Solver (Chinese Academy of Sciences, Shanghai Jiaotong University) | SPONGE (Beijing University), PDEformer (Beijing University) | Pangu Drug Molecule Model (Chinese Academy of Sciences), Pangu Yeild (GuangZhou Lab) | Mega-Protein (Beijing University), AutoGenome (Chinese Academy of Sciences) | Dongfang · Yifeng (Commac), Qingling · Aoxiang (Northwestern Polytechnical University) | Pangu Weather Model (Shenzhen Meteorological Bureau, Hong Kong Observatory) | |

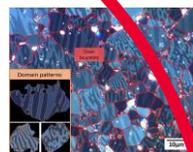
AI for Science

未来AlphaFold级别的大科学创新需要多学科和多尺度数据的融合

时间维度



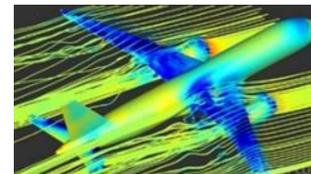
Phase diagram



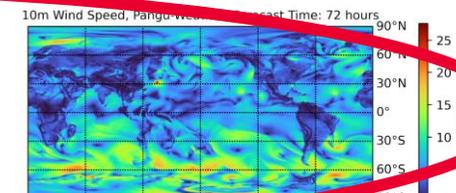
Material texture



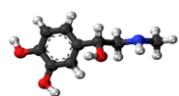
Larger domain Decaying More turbulent



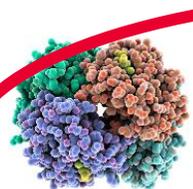
Aerofoil simulation



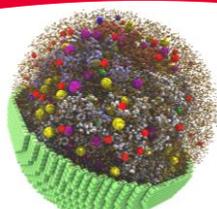
weather forecasting



Small molecule



Proteins



Cell



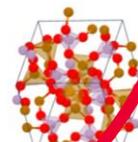
Brain

宏观尺度 (Macroscale)

- 学科: 经典物理学、工程学、材料科学、地球科学等
- 问题: 结构力学分析, 流体动力学模拟, 热传导和热交换分析, 材料行为的大规模模拟, 地球科学模型
- 方法: 有限元分析 (FEA), 计算流体动力学 (CFD), 多体动力学 (MBD), 热力学和传热模型, 经典力学模型, 有限差分法和有限体积法, 地球科学和环境工程统计模型和经验公式



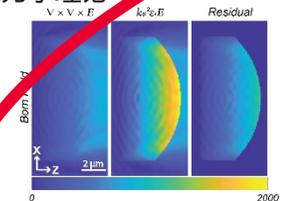
Atomic orbitals



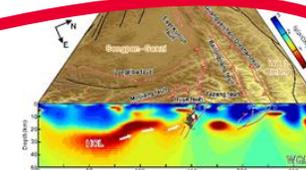
Crystal structure

介观尺度 (Mesoscale)

- 学科: 材料科学、纳米科技、软物质物理、生物物理等
- 问题: 多粒子系统的动力学, 相界面与相分离现象, 自组装与纳米结构形成, 复杂流体的流变学性质, 材料的微观结构与宏观性质之间的关系
- 方法: 连续介质模型, 分子动力学 (MD) 模拟, 蒙特卡洛模拟, 多尺度建模, 非平衡动力学理论



Electromagnetic simulations



Geophysical Electromagnetic Inversion



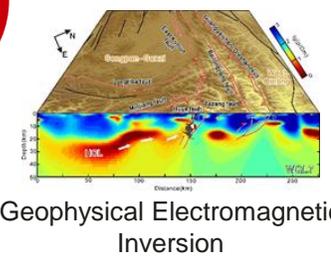
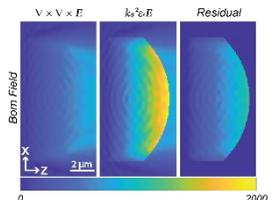
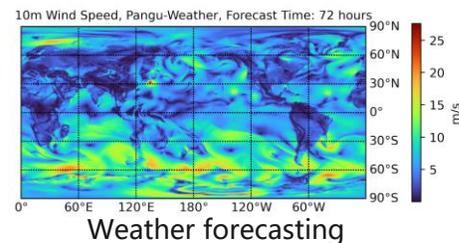
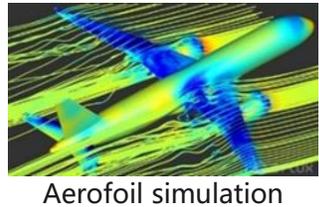
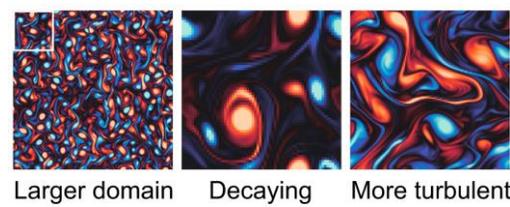
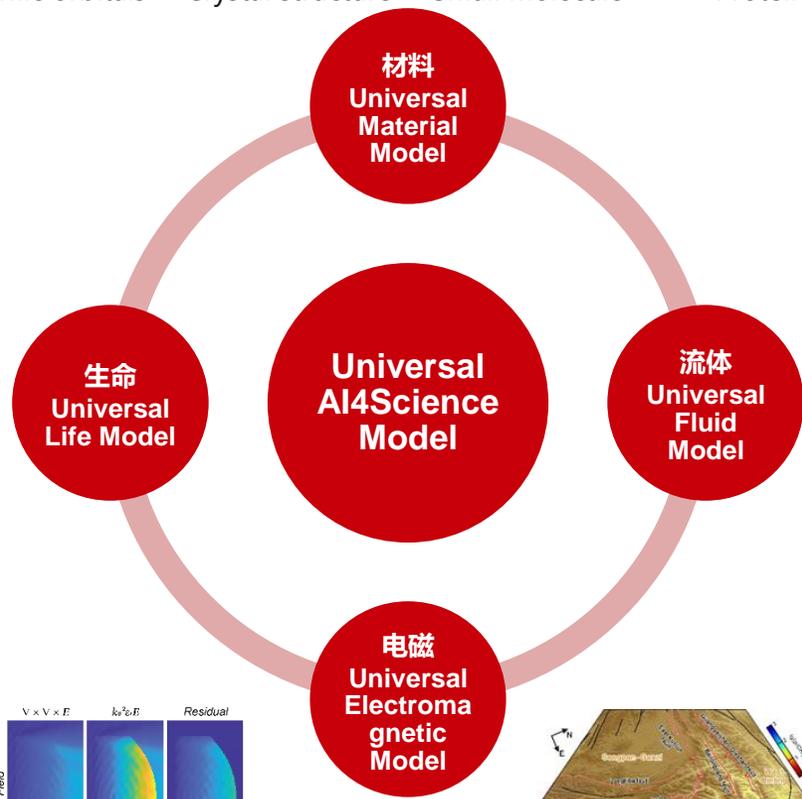
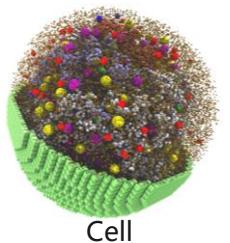
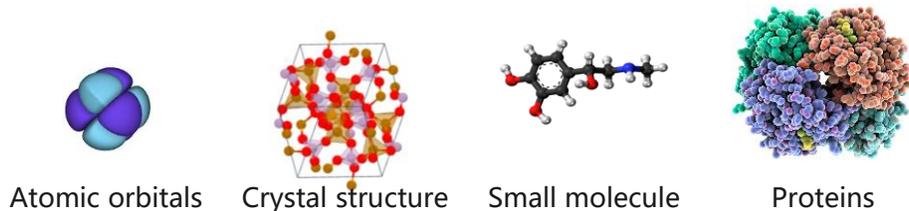
空间维度

微观尺度 (Microscale)

- 学科: 物理化学、分子物理、量子化学、分子生物学等
- 问题: 电子结构计算, 分子动力学模拟, 化学反应机理和路径分析, 分子间相互作用的研究, 材料的性质预测
- 方法: 量子力学计算 (如密度泛函理论DFT)、分子动力学模拟、分子力场、化学反应模拟

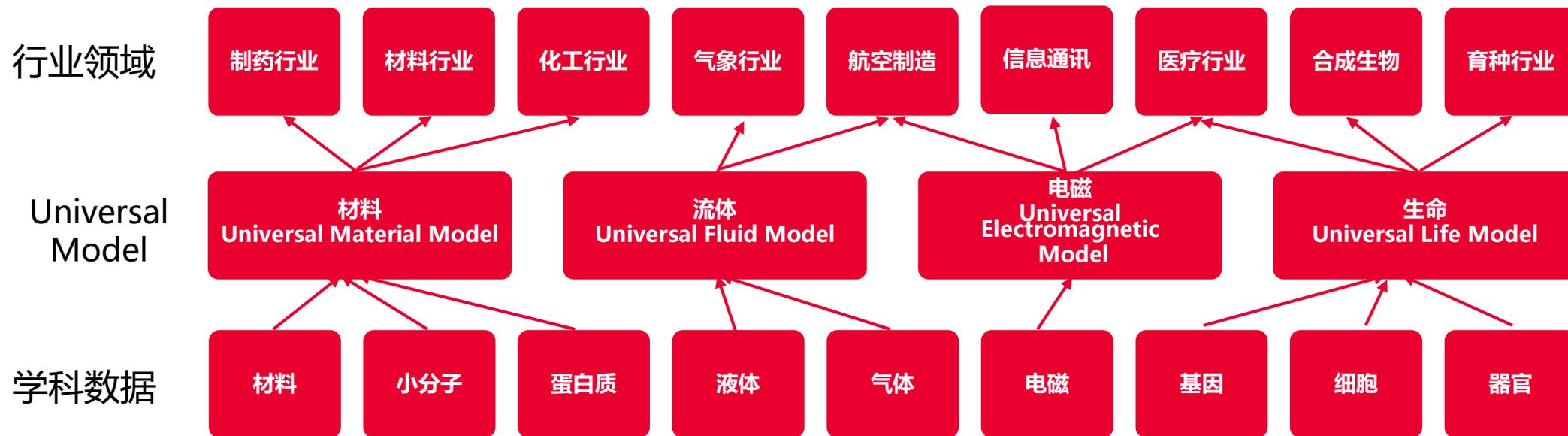
Scaling Law in AI for Science

融合跨学科跨尺度数据训练生命、物质、流体、电磁四大基础科学模型



Emergence in AI for Science

打通行业场景海量数据实现智能涌现



AI for Science 2035

AI4Science根技术的突破释放巨大的行业创新驱动力

技术趋势

科学大模型到行业场景端到端打通，海量行业场景数据推动科学大模型实现涌现。

跨学科多尺度数据融合，聚焦到物质、流体、场、生命四大科学大模型的构建。

AI在更多学科如数学，量子力学，材料，流体，细胞，脑科学，地球科学等实现突破

行业趋势

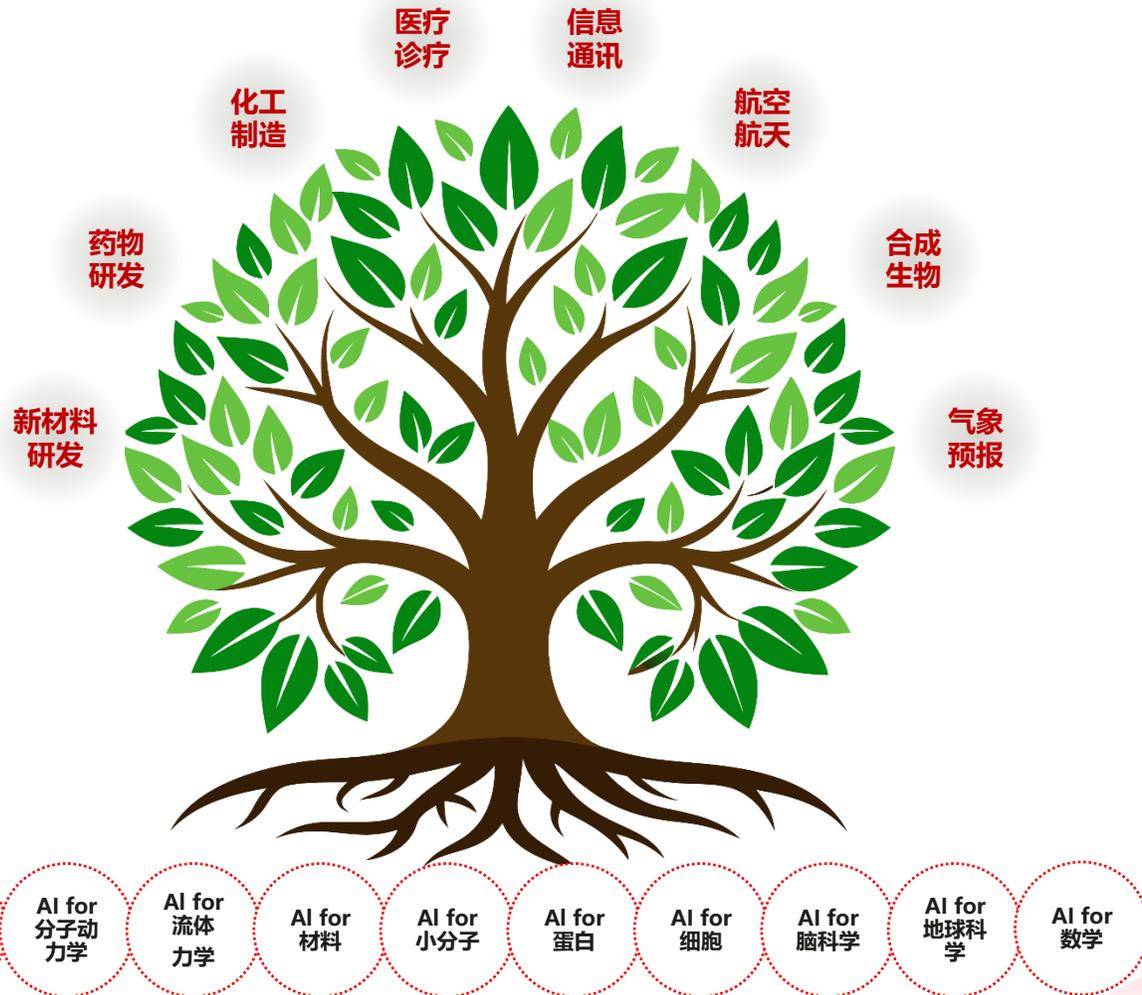
- 2035
1. AI细胞在临床试验，个性化医疗中广泛应用
 2. AI育种广泛应用
 3. AI准确预测极端天气
 4. AI广泛用于工业设计，化工优化，合成生物

2030

1. AI设计材料在超导体、半导体，电池，特种合金等领域实现突破。
2. AI设计细胞得到验证
3. AI飞行器设计得到验证

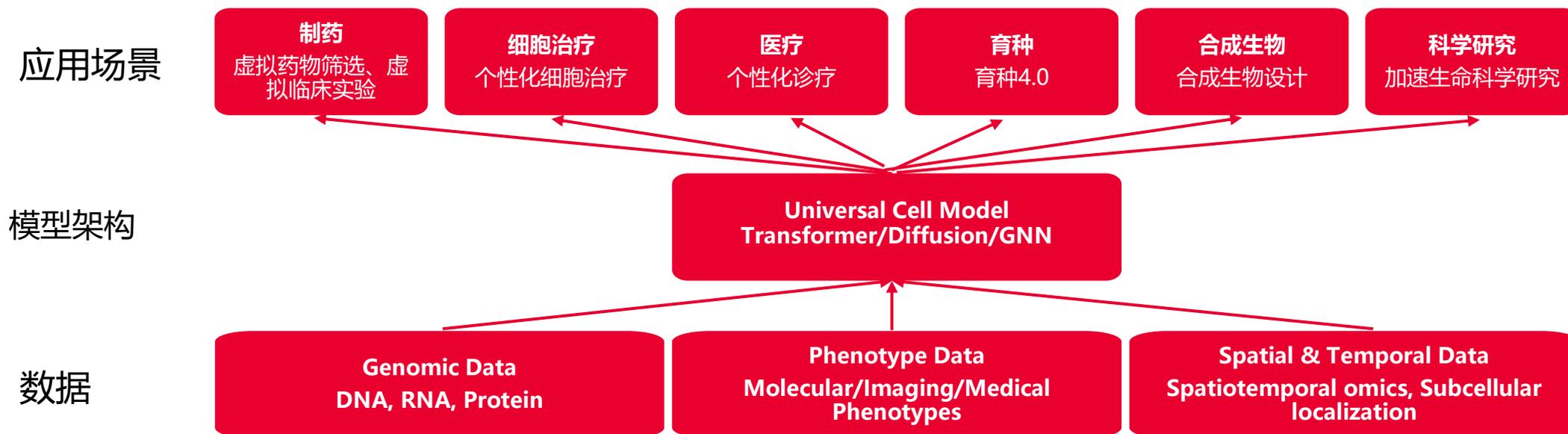
2025

1. AI药物完成临床上市
2. 基于卫星的气象大模型实现分钟级全球预报
3. AI材料设计得到验证，相图预测实现突破



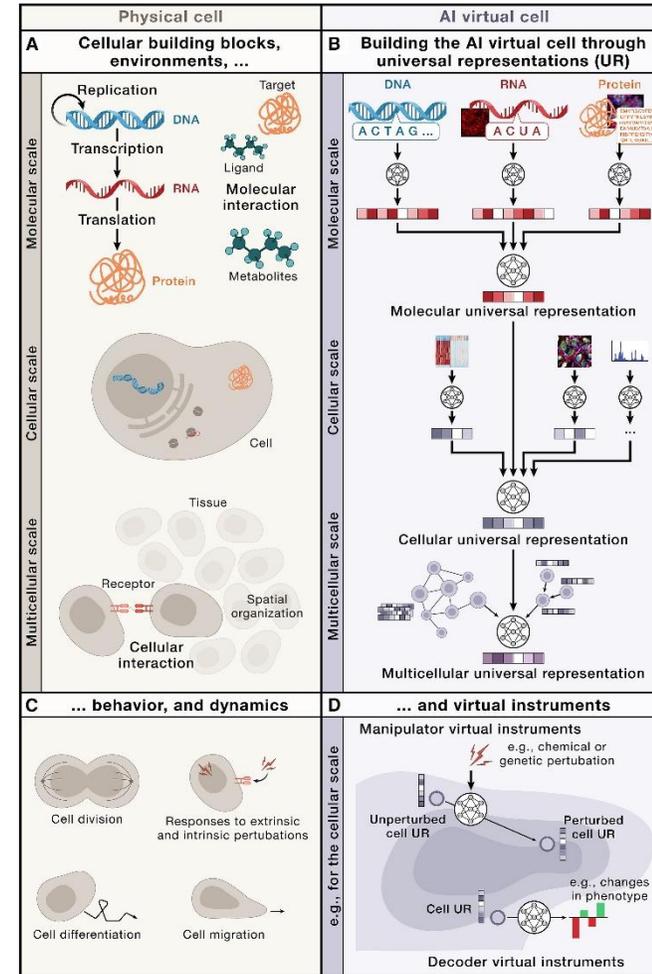
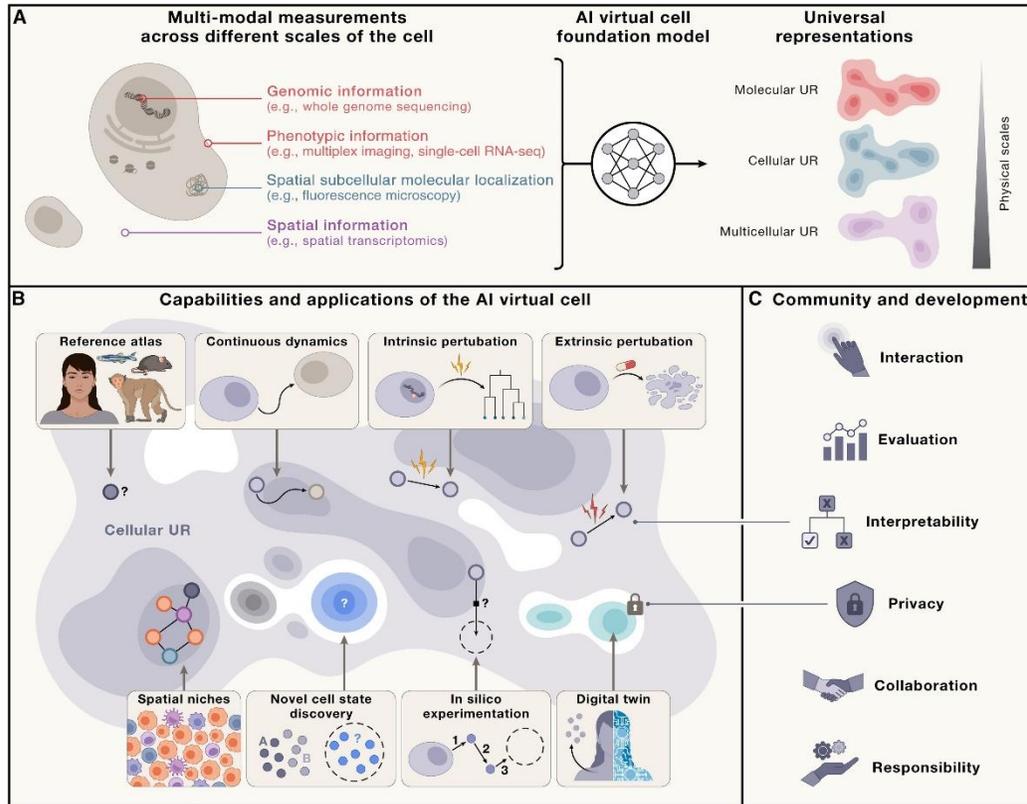
Universal Cell Model

通过统一分子、细胞、器官三个尺度的数据构建统一细胞模型，实现对生命建模模拟



Universal Cell Model

美国斯坦福大学、基因泰克制药公司和陈-扎克伯格基金会的科学家团队Cell发文：AI虚拟细胞时机已至

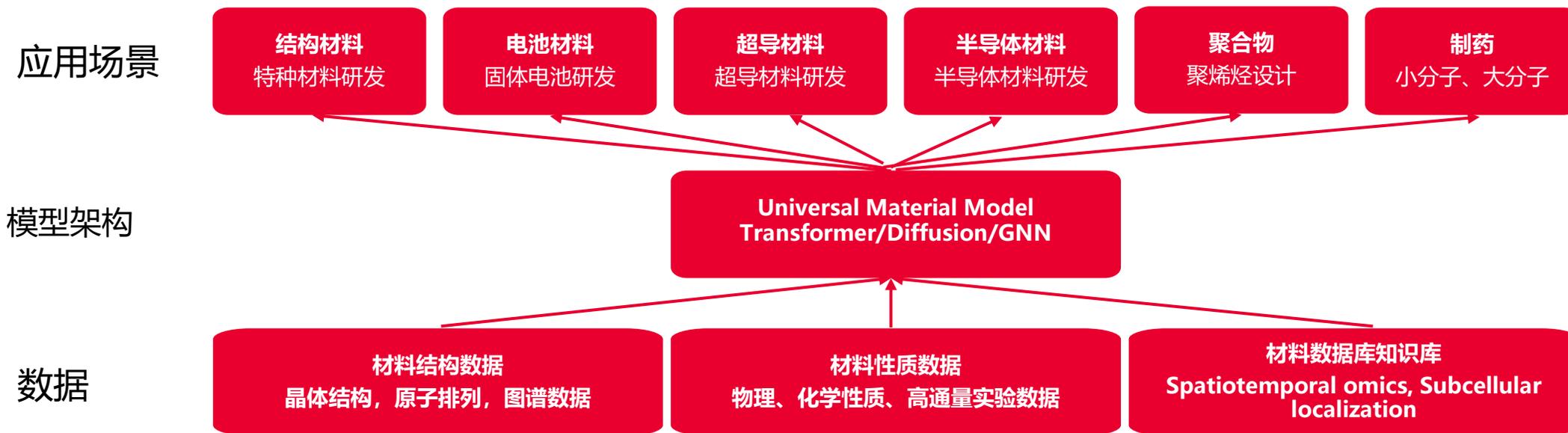


How to build the virtual cell with artificial intelligence: Priorities and opportunities, Cell 2024

Learning the language of life with AI, Science 2025

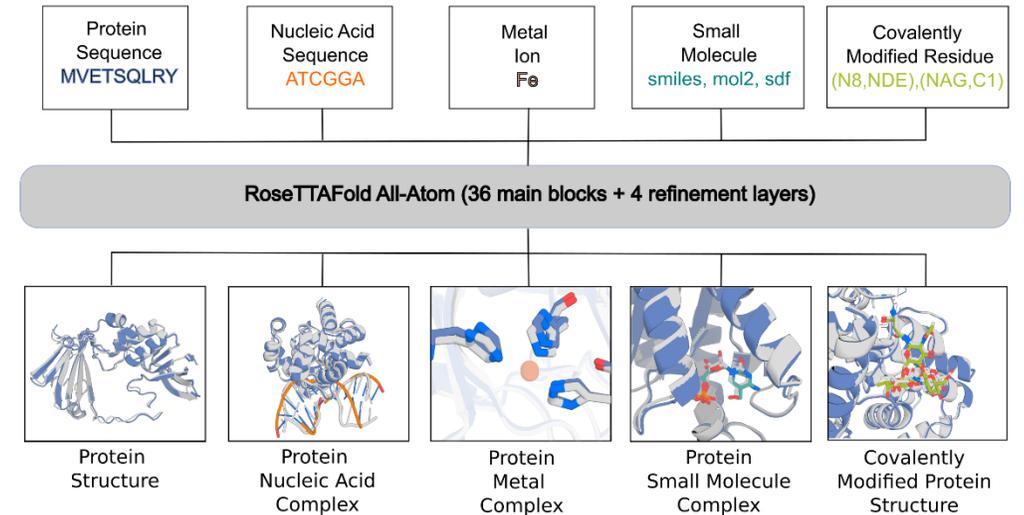
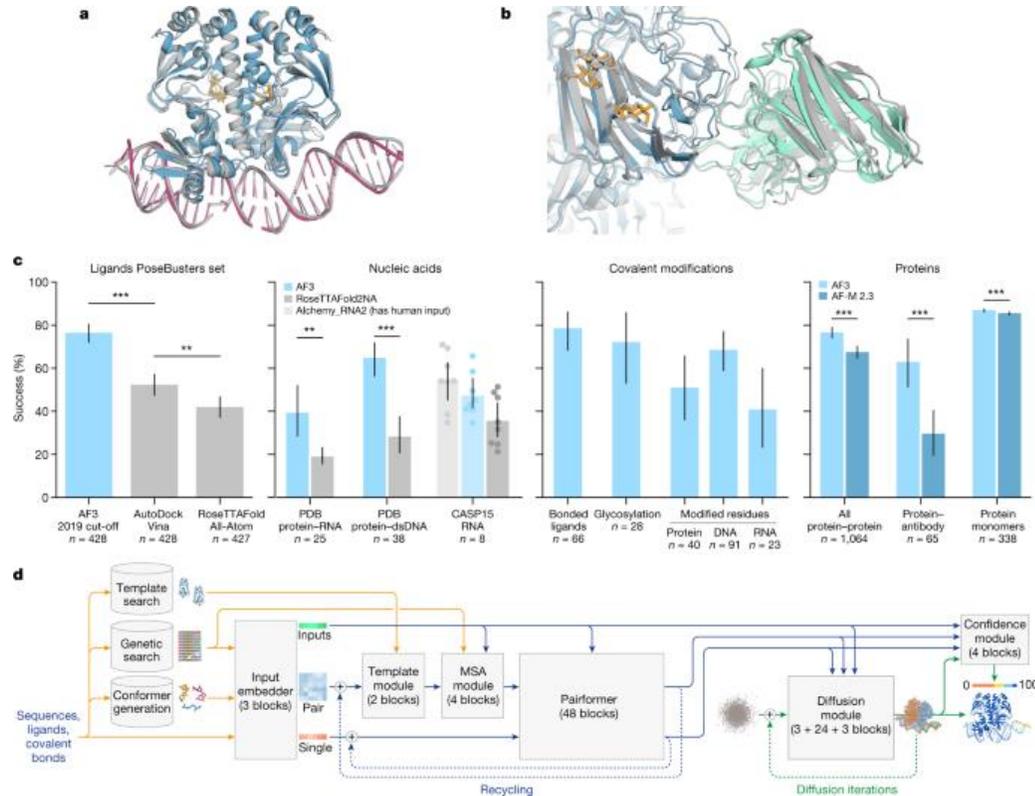
Universal Material Model

融合不同类别的材料数据构建统一材料大模型



Universal Material Model

AlphaFold & RoseTTAFold All-Atom for Biomolecular



Universal Material Model

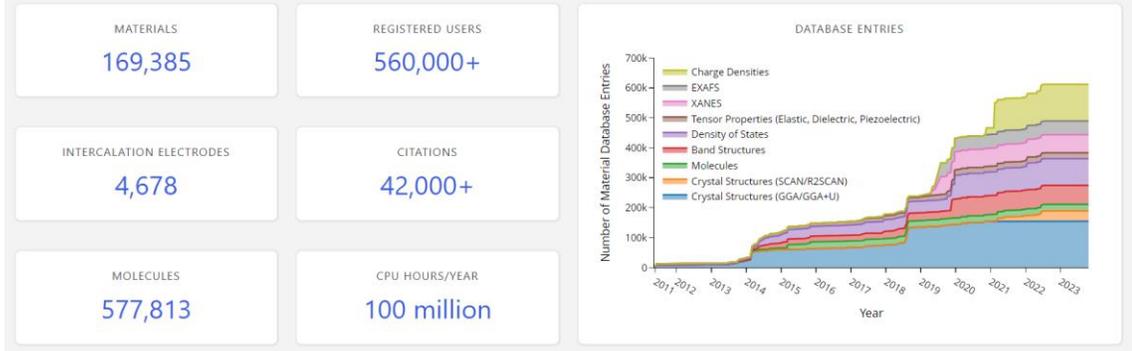
Material Project & MACE-MP-0 for inorganic materials

The Materials Project

Harnessing the power of supercomputing and state-of-the-art methods, the Materials Project provides open web-based access to computed information on known and predicted materials as well as powerful analysis tools to inspire and design novel materials.

[Login or Register](#) [See a Random Material](#) [Browse Apps](#)

The Materials Project by the numbers



Materials Project 由劳伦斯实验室及麻省理工学院在2011年发起，旨在整合各方资源，利用理论计算，让原子性质完整公开，以缩短材料相关科技产业的研究周期，在无机化合物、晶体结构、电池材料等领域提供了丰富的数据支持。

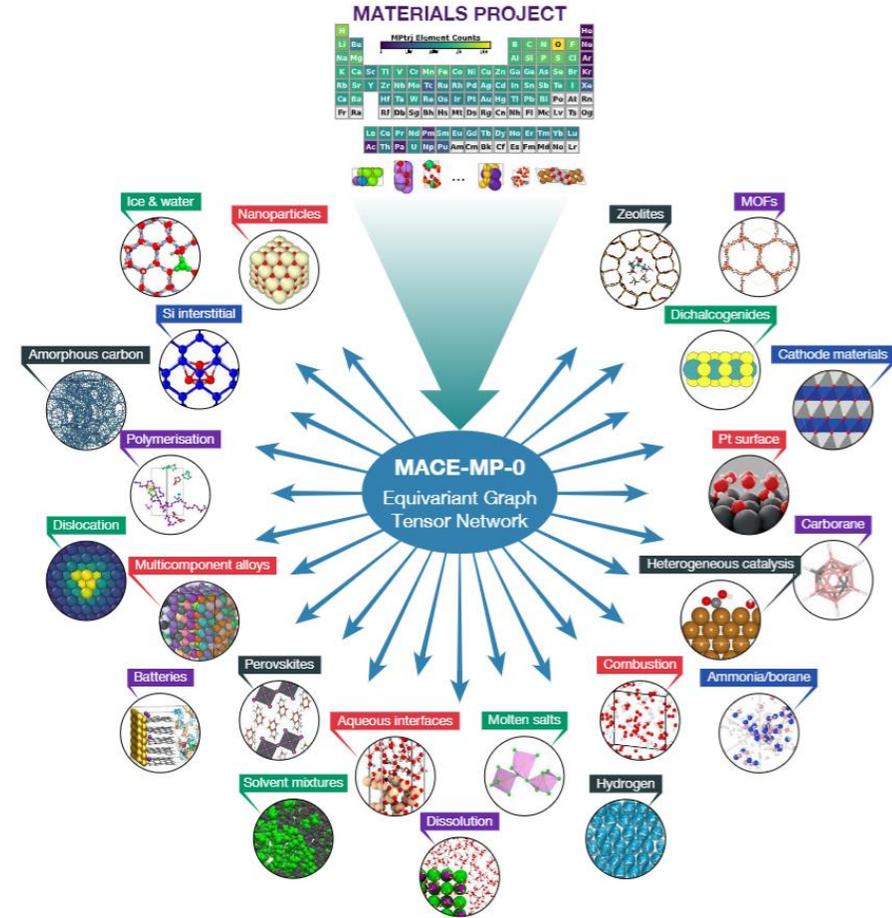
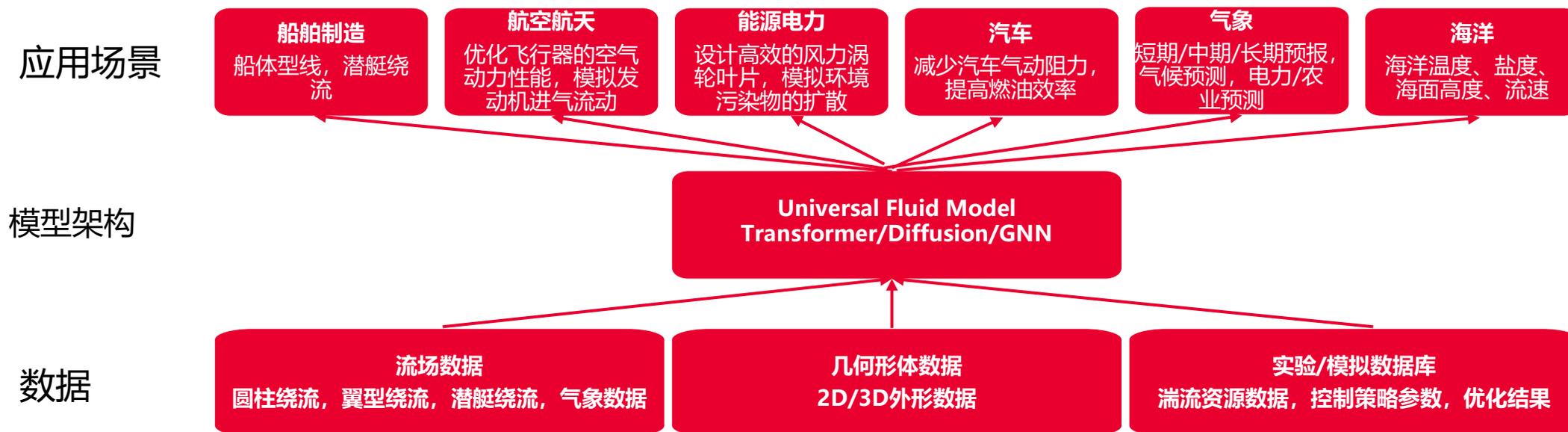


Figure 1: A foundation model for materials modelling. Trained only on Materials Project data [\(19\)](#)

Universal Fluid Model

打造通用的数据驱动流体大模型，对液体、气体及大气等不同类型的流体进行建模和预测。



联合中国商飞打造东方·御风，仿真效率千倍提升，荣获23年WAIC最高奖SAIL奖

现有气动设计手段耗时长、成本高



地面设备模拟飞行

风洞试验
~ 数万次



求解Navier-Stokes方程

数值仿真
~ 近亿CPU 小时



模型试飞和真机飞行

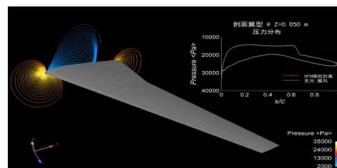
飞行试验
~ 数千飞行小时

MindSpore Flow 流体仿真套件

东方·翼风：千倍级仿真效率提升

全新模型

SAIL奖



3D超临界翼型流场仿真

千万网格流场秒级推理

典型超临界机翼流场预测

降低仿真时间1000倍

效率突破：

AI模型替换传统Navier-Stokes方程求解，提升典型场景仿真效率



精度突破：

对流动剧烈变化区域特征进行精细捕捉，提升整体AI流体仿真的预测精度



模型突破：

建立大数据样本下模型组件化与分布式并行能力，大幅提升新模型研发效率



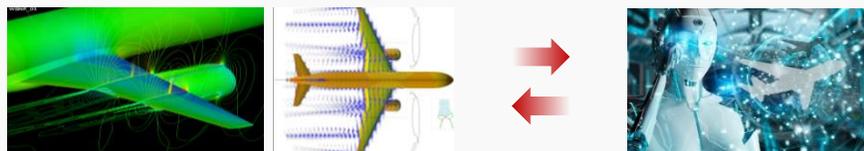
场景突破：

建立流体到AI的数据统一映射，适用多种仿真场景



AI加速和优化科学计算，助力大飞机高效设计

千万网格流体特征提取 · AI分布式并行训练&秒级推理 · 流体物理性质嵌入AI



“飞机的空气动力学模拟仿真运算量非常大，需要借助超算，而有了“东方·御风”大模型的帮助，在精度一样的情况下，仿真时间只需要原来的1/25”^{注1}

——中国工程院 吴光辉院士

注1：东方·御风(2022)案例基于二维场景，仿真时间缩短为原来的1/25；东方·翼风(2023)案例基于三维场景，仿真时间缩短为原来的1/1000

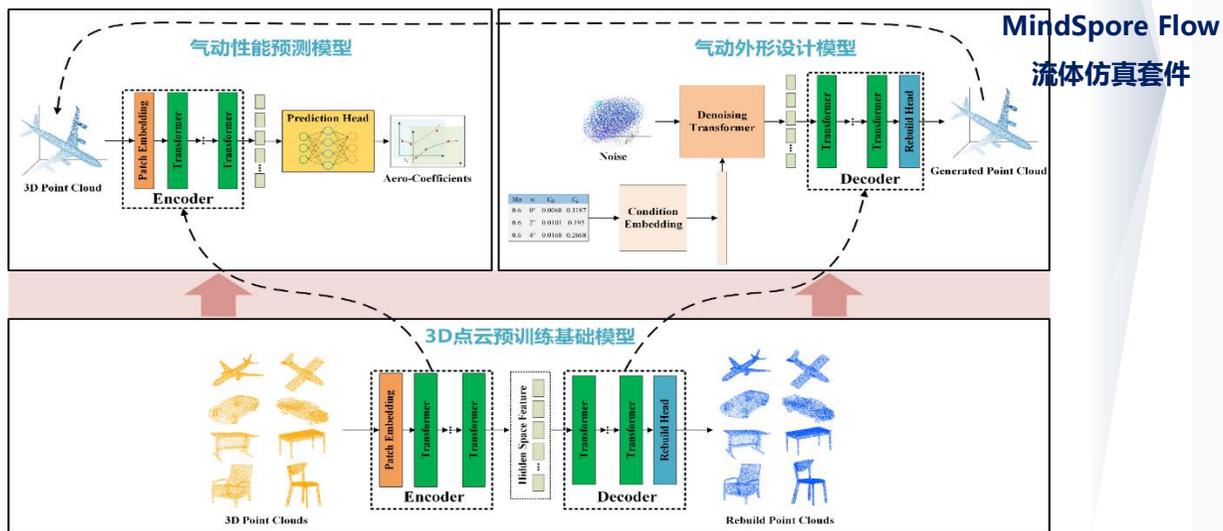
联合中国空气动力研究与发展中心，打造“风雷”生成式气动设计大模型平台，满足概念设计要求

气动外形设计依赖专家经验且通用性差

- 传统方法强烈依靠专家经验
- 仅能针对特性构形，进行迭代优化设计
- 存在初始方案提出难、参数化设计局限多、气动设计耗时长等问题



MindSpore提供AI赋能的一站式气动设计大模型平台



生成式气动设计结果满足概念设计要求



复杂三维气动外形表征

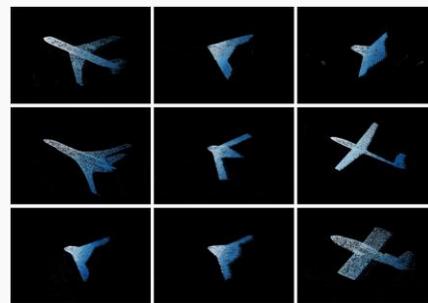
- 采用点云气动外形，突破飞行器拓扑结构的限制
- Encoder+Decoder实现非结构点云数据的低维表征

气动性能端到端预测

- 预训练Encoder+回归网络，气动性能数据微调
- 实现点云外形在给定工况下的气动性能秒级预测

三维气动外形端到端生成

- 以设计指标为条件生成气动外形的特征空间表达
- 对接预训练的Decoder，点云外形分钟级生成



模型可生成出多个不同拓扑的气动外形

具有明显的多样性特征

经预测模型筛选后，条件生成结果和指标偏差 < 5%

满足概念设计要求

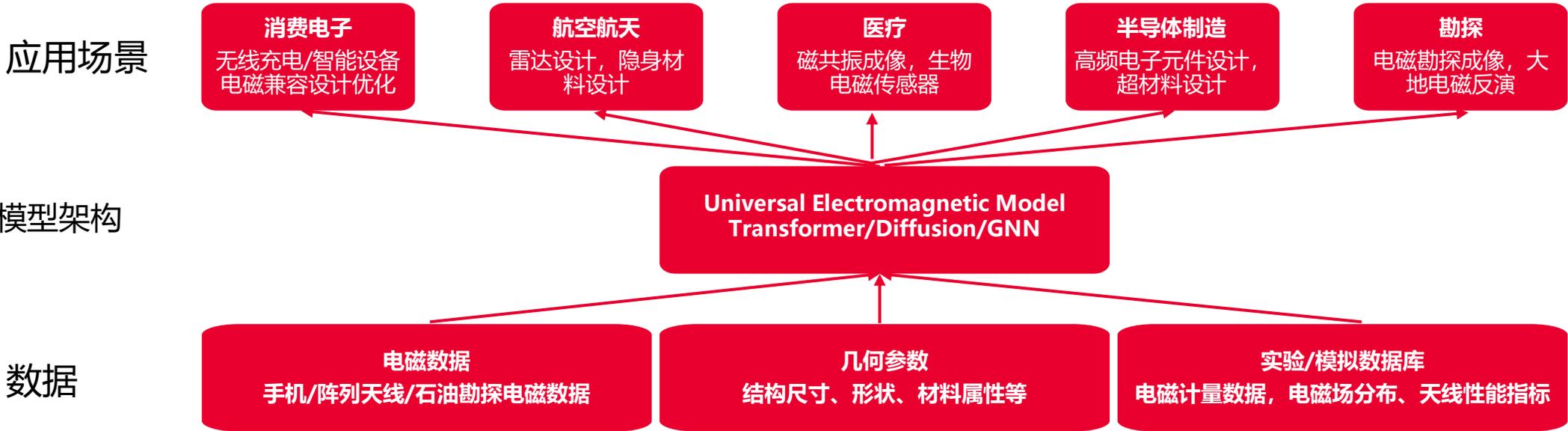


风雷生成式气动设计大模型平台打破传统设计范式，设计时长由月级缩短到分钟级，满足概念设计要求。未来该平台将扩展到航空航天/船舶/高铁/能源/汽车等多个行业领域，引领工业装备的设计和制造能力跨越式发展。

——中国工程院 唐志共院士

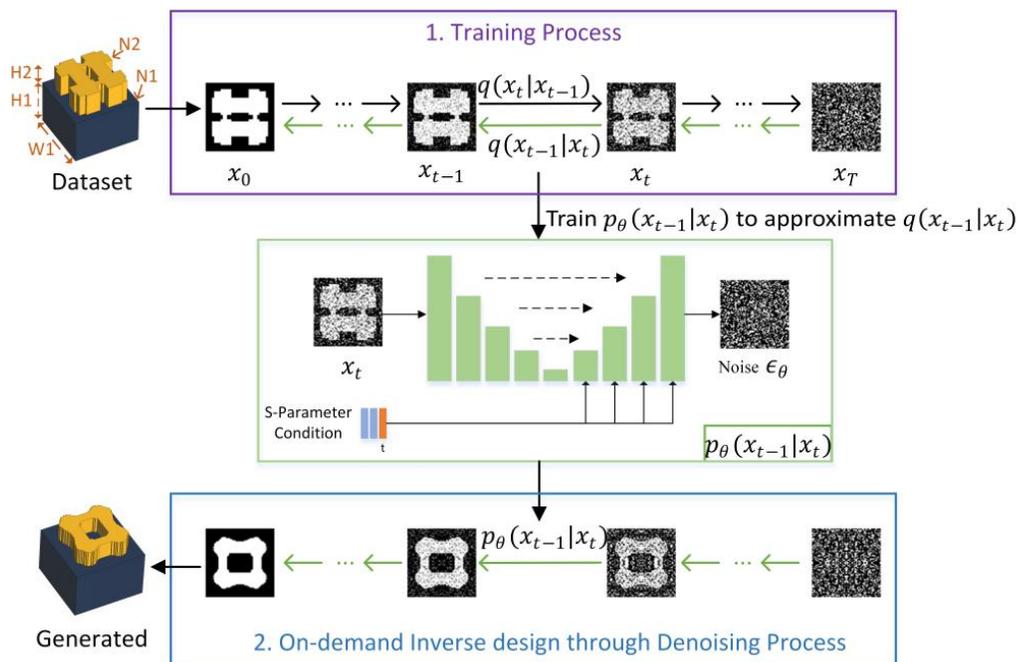
Universal Electromagnetic Model

构建能够处理电磁问题的通用场模型，支持不同频段、不同几何结构和材料属性的电磁预测



基于扩散模型的超材料设计

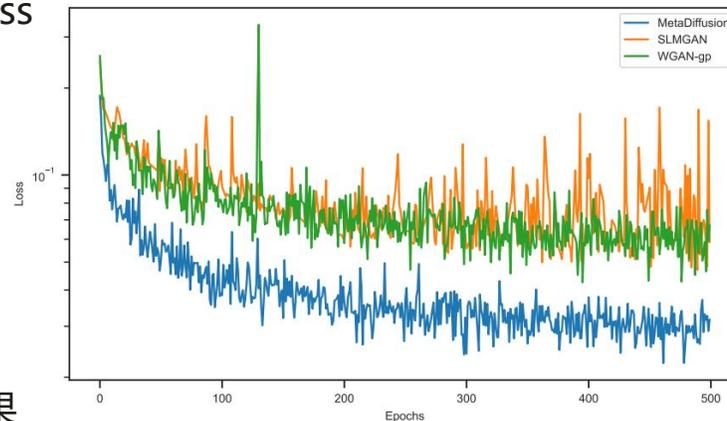
场景介绍：超材料具有超长电磁特性的人工复合材料（负折射、电磁隐身、异常透射），被Science评为**十大科技突破**。传统超材料设计严重依赖先验知识和全波模拟的试错，导致流程长耗时。黑盒优化方法如遗传算法、拓扑优化等，很难做到多目标优化。深度学习方法（GAN/Diffusion等）更有助于超材料设计



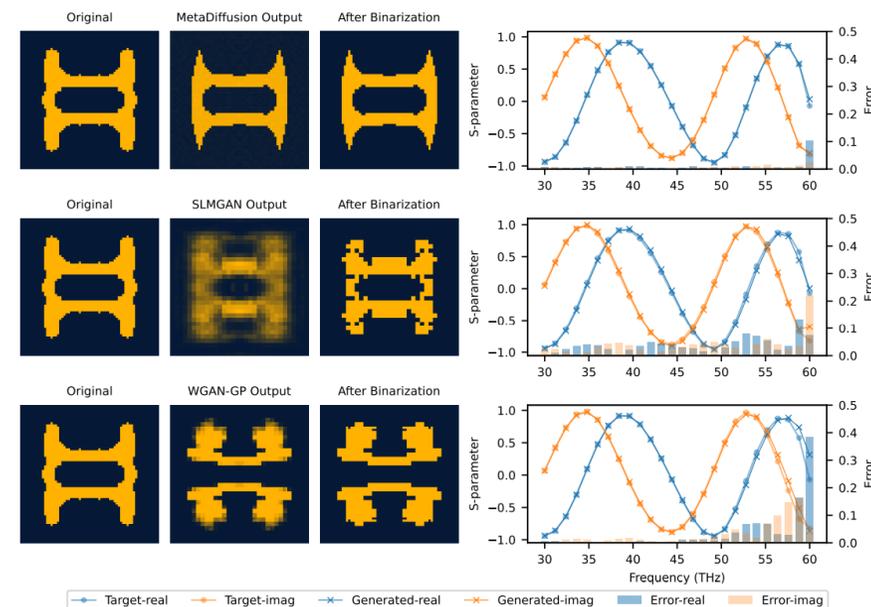
Zezhou Zhang Chuanchuan Yang* Yifeng Qin* Hao Feng Jiqiang Feng Hongbin Li, Diffusion Probabilistic Model Based Accurate and High-Degree-of-Freedom Metasurface Inverse Design

• **实验结果：**扩散模型收敛速度和精度都优于GAN网络

• 训练过程的loss



• 逆设计的结果

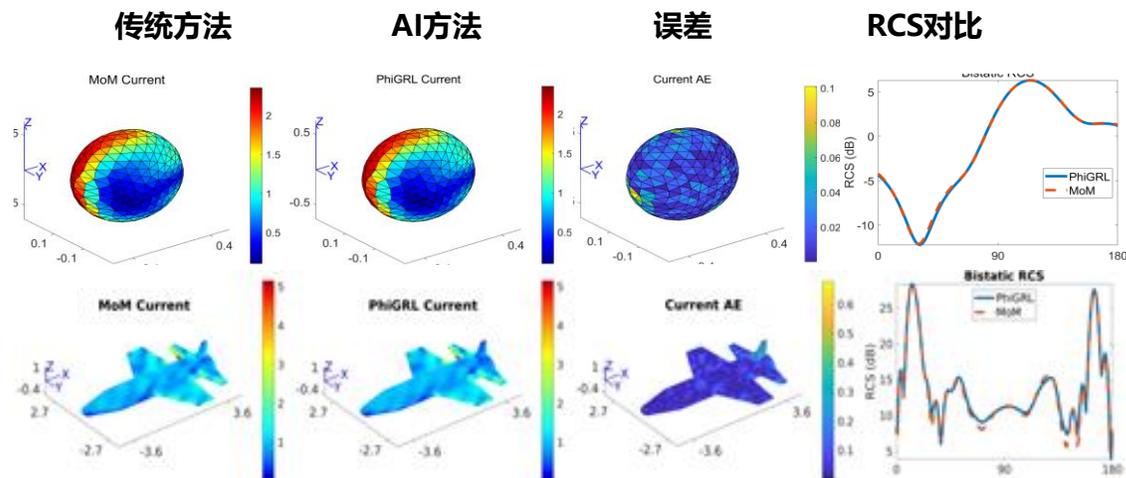


基于图神经网络的飞行器电磁仿真

- **场景介绍:** 电磁仿真有助于飞行器（飞机/无人机等）安全飞行，传统方法计算量随问题规模呈指数增长，AI方法有助于提升性能

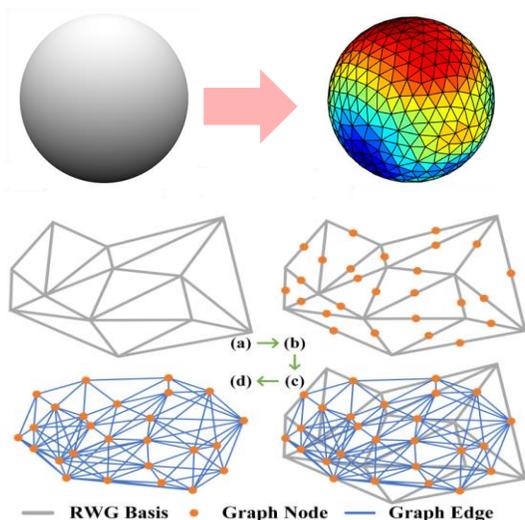


- **实验结果:** AI方法与传统MoM法精度保持一致，性能有数倍至百倍提升



- **方案设计:** 图神经网络拟合电流密度，加速电磁仿真

(a) 构图:



(b) 模型:

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \Delta_k$$

$$\Delta_k = \Psi(\mathbf{R}_k, \Theta_k)$$

$$\mathbf{R}_k = \mathbf{b} - \mathbb{Z}\mathbf{u}_k$$

无监督:

$$res = \frac{1}{2} \left[\frac{\|\mathbf{b}^r - [\mathbb{Z} \cdot \mathbf{u}_g]^r\|_2^2}{N_{\mathbf{u}_g}} + \frac{\|\mathbf{b}^i - [\mathbb{Z} \cdot \mathbf{u}_g]^i\|_2^2}{N_{\mathbf{u}_g}} \right]$$

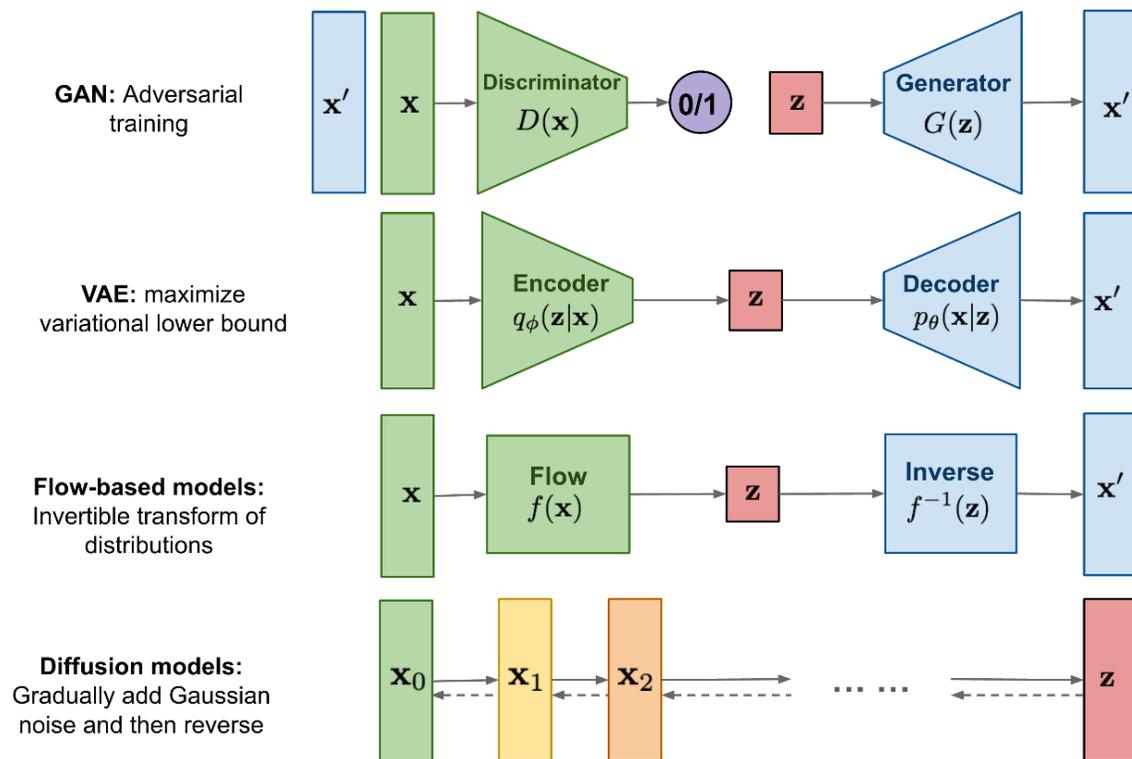
有监督:

$$MSE = \frac{1}{2} \left[\frac{\|\mathbf{u}_m^r - \mathbf{u}_g^r\|_2^2}{N_{\mathbf{u}_g}} + \frac{\|\mathbf{u}_m^i - \mathbf{u}_g^i\|_2^2}{N_{\mathbf{u}_g}} \right]$$

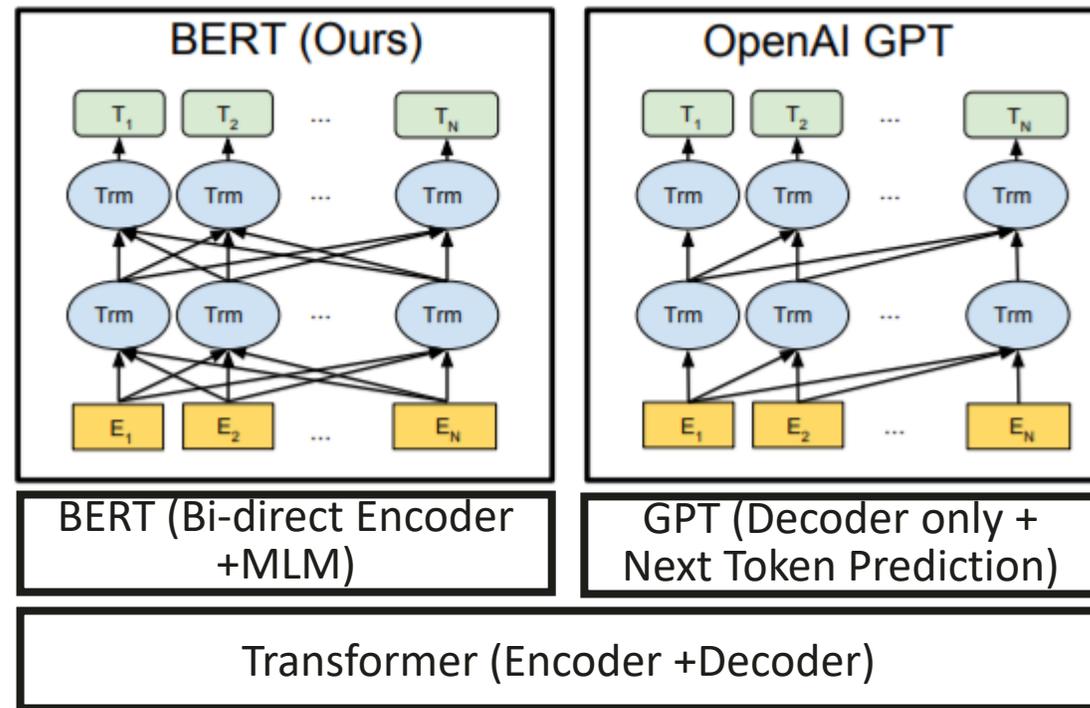
| 三维目标 | 方法 | 性能 | 未知量 |
|------|-----|---------------|------|
| 椭球 | 传统 | 0.14s | 930 |
| | GNN | 0.01s | |
| 飞行器 | 传统 | 33.2s | 6396 |
| | GNN | 0.087s | |

科学大模型架构讨论

NLP和Vision的成功架构可以借鉴到AI4Science大模型的设计上



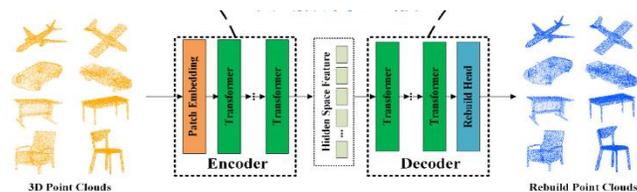
Vision生成模型架构，可抽象到具备欧式空间特征的数据上



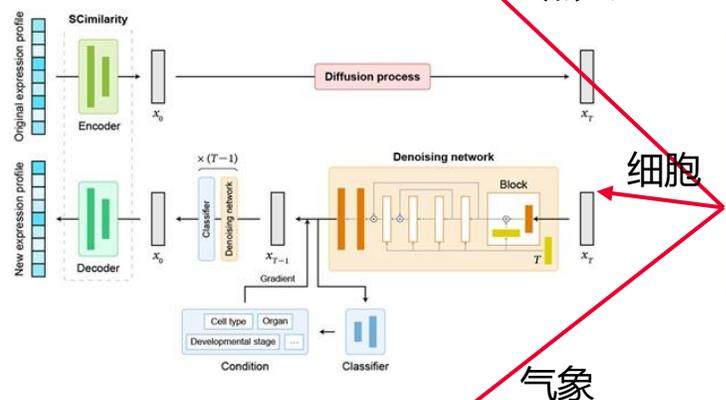
NLP 预训练大模型架构，可抽象到具备序列特征的数据上

科学大模型架构讨论

扩散模型的架构具有普适性，通过结合各种神经网络结构（CNN/Transformer），能够适应不同领域的需求

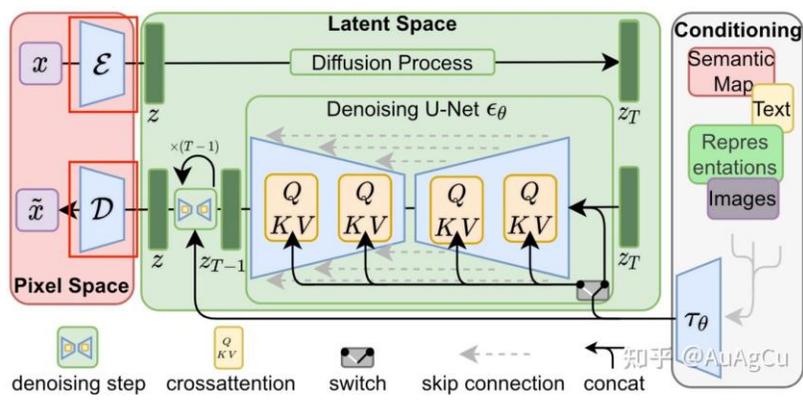


3D点云



细胞

气象



材料

小分子

大分子

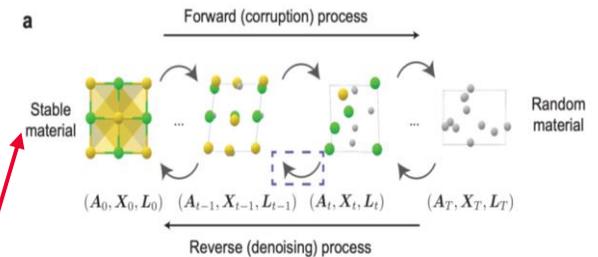
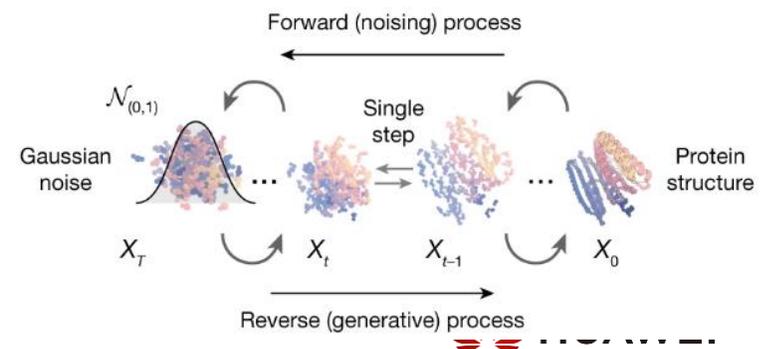
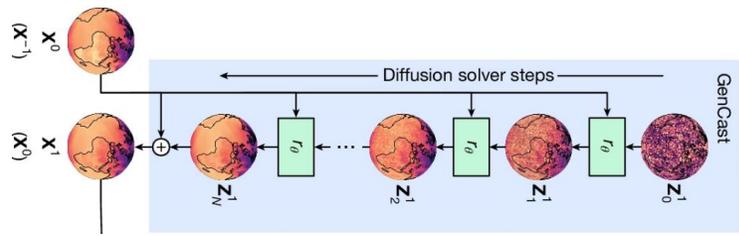


Figure 2. Overview of the Equivariant Diffusion Model. To generate molecules, coordinates x and features h are generated by denoising variables z_t starting from standard normal noise z_T . This is achieved by sampling from the distributions $p(z_{t-1}|z_t)$ iteratively. To train the model, noise is added to a datapoint x, h using $q(z_t|x, h)$ for the step t of interest, which the network then learns to denoise.

Diffusion model



扩散模型可以适用于多种类型的数据（序列，2D，3D，Graph，点云）



04

解码DeepSeek的创新之道 三代模型演进路线图



解码DeepSeek创新之道：三代模型演进路线图

DeepSeek模型历经三代演进，从DeepSeek-LLM (V1) 的全栈框架完善，到V2的MLA和DeepSeekMoE架构创新及GRPO后训练，再到V3通过全栈协同训练实现更大规模MoE模型。DeepSeek R1作为AI领域里程碑创新，为下一代智能发展指明了方向。



DeepSeek-LLM (v1)

<https://arxiv.org/html/2401.02954v1>

1. 高质量数据的构建：构建了一个包含约2万亿 tokens 的多语言数据集，主要涵盖中文和英文。通过数据去重、过滤和重混 (remixing) 三个阶段，确保数据多样性和高质量。
2. 模型架构借鉴llama：模型在总体上借鉴了 LLaMA 的设计，但在细节上进行了调整。例如，7B 模型采用30层网络，而67B模型采用95层网络，同时在部分模块（如注意力机制中采用了 Grouped-Query Attention 以降低推理成本）上有所不同。
3. 超参数设置：讨论了批次大小和学习率与模型规模之间的关系，并提出了基于大规模实验的多步学习率衰减策略。与传统的余弦调度器相比，这种策略不仅能保证性能，还方便模型的持续训练。
4. Scaling Laws研究：重新审视了模型规模和数据规模的扩展定律，提出了基于“非嵌入 FLOPs/token”的新模型规模表示方法，并通过 IsoFLOP 分析得出最优的模型与数据扩展分配策略。发现不同数据集之间的质量差异会对扩展定律产生明显影响，高质量数据更有助于训练更大规模的模型。
5. 自研完善基础infrastructure：基于 HAI-LLM 的轻量级训练框架，集成了数据并行、张量并行、序列并行和流水线并行等多种技术，并利用 Flash Attention 和 ZeRO-1 技术进一步提升硬件利用率和训练效率。
6. 完善了pretraining 和 alignment (SFT+DPO) 策略，证明SFT和DPO提升模型性能都是有效的。
7. 完善evaluation数据和框架，构建并强化了 safty evaluation整体策略。

整体来讲，DeepSeek v1搭建了大模型的全栈技术框架，主要以解决开源工作为主，模型架构创新很少，infrastructure层自研创新很多，完善了大模型架构设计策略，超参数策略，alignment及evaluation。

DeepSeekMath

<https://arxiv.org/pdf/2402.03300>

1. 数学领域大规模预训练语料的构建策略：提出了一套迭代式数据采集流程，用于从 Common Crawl 中提取数学相关内容。利用 OpenWebMath 作为种子数据，首先使用 fastText 分类器进行初步筛选，再通过域划分和人工标注扩充种子数据，从而不断优化分类器，最终构建出包含约 120B 数学 tokens 的 DeepSeekMath Corpus。与现有数学预训练语料（如 MathPile、OpenWebMath、Proof-Pile-2）相比，该语料库规模更大、质量更高，并且具有多语言（主要是英语和中文）覆盖能力。
2. 预训练模型策略：基于选用 DeepSeek-Coder-Base-v1.5 7B 作为预训练起点，证明先进行代码预训练可以显著提升模型的数学推理能力。实验结果显示，在数学预训练后，即使模型参数量较小（7B），其数学问题解决能力已接近甚至超越某些大规模闭源模型（如 Minerva 540B）。
3. 提出创新的强化学习算法GRPO（重要原创创新）：Group Relative Policy Optimization (GRPO) 作为 Proximal Policy Optimization (PPO) 的变体，取消了对 critic 模型的依赖，而是通过“组内评分”来估计基线，从而大幅降低了训练时的内存消耗和计算资源。该方法在数学指令调优阶段得到了显著提升，在同一数据条件下，通过 GRPO 进一步提高了模型在各类数学基准（如 GSM8K、MATH 以及其他外域数学任务）的表现。
4. 同时讨论了 Rejection Sampling Fine-Tuning (RFT)、Direct Preference Optimization (DPO) 和 PPO 等方法归纳为直接或简化的强化学习技术，系统探讨了在线与离线训练、结果与过程监督以及单回合与迭代强化学习等关键因素，这为后续V3和R1的训练奠定了很强的理论和数据基础。

总结，DeepSeekMath的探索为R1奠定了很强的理论和数据基础，关键原创创新有两点：

1. 大规模高质量数学语料构建：通过精心设计的筛选流水线和多次迭代，构建出远超现有资源的 120B tokens 数学语料，为数学预训练奠定了坚实基础。
2. GRPO 强化学习算法：提出了取消 critic 模型的 GRPO，通过组内评分来优化训练，既提高了数学指令调优阶段的效率，又降低了内存和计算资源的消耗，从而在数学基准上取得了领先表现。

解码DeepSeek创新之道：三代模型演进路线图

DeepSeek模型历经三代演进，从DeepSeek-LLM (V1) 的全栈框架完善，到V2的MLA和DeepSeekMoE架构创新及GRPO后训练，再到V3通过全栈协同训练实现更大规模MoE模型。DeepSeek R1作为AI领域里程碑创新，为下一代智能发展指明了方向。



DeepSeek V2

<https://arxiv.org/html/2405.04434v5>
<https://arxiv.org/html/2401.06066v1>

1. 高质量数据：DeepSeek V2继续做了大量工作以构建高质量的训练数据8.1T Token。
2. 提出了核心架构创新MLA和DeepSeekMoE
 - **Multi-head Latent Attention (MLA)** 的发明核心是为了解决KV缓存瓶颈：MLA 通过对 Key 与 Value 进行低秩联合压缩，将高维表示映射到一个较低维的潜在空间中，再在推理阶段通过上投影恢复所需维度。这样既能大幅减少存储 KV 缓存所需的内存，又能保留足够的信息用于准确的注意力计算。采用 MLA 后，模型在保持甚至超过传统 MHA 性能的同时，其 KV 缓存量显著降低，从而大幅提升了推理效率，允许更大批量和更长序列的处理。
 - **DeepSeekMoE** 架构的发明核心在探索MoE的极限：现有的 MoE 架构（如 GShard）在专家划分上往往较为粗粒度，容易导致知识冗余；同时，在专家选择和负载均衡方面存在分布不均、通信开销过高等问题，这些问题都会增加训练成本和不稳定性。
 - 细粒度专家划分：DeepSeekMoE 对专家进行更细致的分割，使每个专家在特定领域或任务上更专精，从而提高知识获取的准确性和效率。
 - 共享专家隔离：同时设计了一部分共享专家，用于捕捉通用知识，避免各个路由专家之间的重复计算（知识冗余），实现更合理的专家职责分配。
 - 专家并行与负载均衡机制：在训练时采用专家并行策略，并引入辅助损失（包括专家级、设备级和通信级平衡损失）以及 token-dropping 策略，确保每个专家的负载相对均衡，控制跨设备通信开销，从而降低训练成本并提升整体训练效率。
3. 引入DeepSeekMath的GRPO策略通过RL训练 DeepSeek V2 Chat策略并证明有效提升。

总结，DeepSeek-V2 通过在 Transformer 架构中引入 MLA 和 DeepSeekMoE 两项关键创新，有效解决了传统模型在大规模预训练和长序列推理过程中面临的内存和计算瓶颈问题，同时在保证高性能的前提下实现了更经济、更高效的训练与推理。这两个创新为后续V3和R1沿用。

DeepSeek V3 & R1

<https://arxiv.org/html/2412.19437v1>
<https://arxiv.org/html/2501.12948v1>

V3创新点：

1. 数据：持续构建高质量的训练数据，提升到14.8T Token。
2. DeepSeek V3架构沿用V2的MLA和DeepSeekMoE，又提出了两个小的创新点用于提升模型性能：
 - auxiliary-loss-free strategy for load balancing：对于每个专家，在路由时为其加上一个偏置项（仅用于路由选择），在训练过程中根据当前专家是否过载进行动态调整（增加或减少偏置），这一策略保持专家负载均衡而不引入额外损失，从而不会对模型性能造成明显干扰，同时节点限制路由确保在专家并行时跨设备通信最小化，使得大规模 MoE 模型训练更稳定、高效。
 - Multi-Token Prediction (MTP)：引入了 MTP 模块和相应训练目标，既能提升模型预测精度，也可用于推理时的 speculative decoding，从而加速生成过程。
3. 提出一系列核心解决怎么在受限硬件上训练更大规模的MoE模型
 - 首创FP8 Mixed Precision Training Framework：首次在极大规模模型上引入并验证了 FP8 混合精度训练框架。通过支持 FP8 运算和存储实现了训练加速和减少内存占用。
 - 训练框架优化：DualPipe 算法实现流水线并行，减少流水线空泡实现了计算与通信重叠；开发了专门的跨节点 All-to-All 通信内核，以充分利用 InfiniBand (IB) 和 NVLink 的带宽，从而确保各节点之间的数据交换高效且低延迟。
4. 两阶段训练的上下文扩展策略：pre-training后通过两阶段的训练策略将上下文扩展到128K。
5. 后训练蒸馏DeepSeek-R1提升V3推理能力，经过 DeepSeek-R1 蒸馏的数据能显著提升模型在 LiveCodeBench 和 MATH-500 两个基准测试中的 Pass@1 指标，知识蒸馏虽然能够提升模型性能，但有时也会大幅增加平均响应长度。
6. **DeepSeek-V3可以作为生成式奖励模型**：
 - DeepSeek-V3 在 RewardBench 上的表现与最优版本的 GPT-4o-0806 和 Claude-3.5-Sonnet-1022 相当，甚至超过了其他版本的表现，这表明其判断能力非常出色。
 - DeepSeek-V3探讨了self-rewarding策略，通过模型自身的投票评估结果来生成奖励信号，实现自我优化。这种方法不仅提高了模型的对齐效果，还为未来在更广泛任务中的奖励机制设计提供了新的思路。



解码DeepSeek创新之道：三代模型演进路线图

DeepSeek模型历经三代演进，从DeepSeek-LLM (V1) 的全栈框架完善，到V2的MLA和DeepSeekMoE架构创新及GRPO后训练，再到V3通过全栈协同训练实现更大规模MoE模型。DeepSeek R1作为AI领域里程碑创新，为下一代智能发展指明了方向。



R1创新点:

1. R1的目标是完全通过无标注的数据进行推理能力的训练，最终希望实现模型的自我进化。
2. Reasoning领域里程碑R1-Zero: R1-Zero直接将强化学习应用于基础模型，使得模型能够使用思维链(CoT)来解决复杂问题，模型也展现了自我验证、反思和生成长思维链等能力。这是业界首个公开研究证明了无需使用SFT，仅仅通过RL就可以显著增强LLM的推理能力，是Reasoning领域的里程碑性工作。
 - 强化学习使用DeepSeekMath中提出的GRPO策略。
 - Reward Modeling: 1). Accuracy Reward: usually rule based model, 2). Format reward: 强制模型生成cot过程。
 - Aha Moment of R1-Zero: 在RL训练过程中会出现Aha moment，模型突然“恍然大悟”，并分配更多的思考时间重新考虑初始条件策略，这说明了通过提供正确的激励，模型可以自主发展出先进的解决问题的策略。“Aha Moment”是通过强化学习，解锁下一层LLM智能的可能路径。
3. DeepSeek-R1: 为了解决R1-Zero生成结果可读性差和语言混合的问题，通过在冷启动阶段加入推理数据做SFT的方法训练了R1，解决了R1-Zero的问题。

DeepSeek未来发展方向:

1. 持续优化模型架构: 继续研究和改进模型架构，进一步提升训练和推理效率。未来目标之一是实现对无限上下文长度的高效支持，同时突破 Transformer 现有的架构局限，推动模型表达能力的边界。
2. 提升训练数据的数量和质量: 持续迭代和扩充训练数据，不仅要增加数据量，还要探索引入更多样化的训练信号，从多个维度推动数据规模的提升，从而进一步提高模型性能和泛化能力。
3. 增强深度思考能力: 针对模型的深度推理和问题解决能力，将进一步研究如何扩展模型的推理长度和深度，以增强模型在复杂认知任务中的智能表现和问题求解能力。
4. 开发多维度综合评测方法: 探索更加全面、多维度的模型评估方法，避免仅针对固定基准进行优化，防止对模型能力产生误导性的评估。通过多角度考量模型在各类任务和实际应用中的表现，确保对模型能力有更为准确和客观的基础性评估。

V3和R1的相互促进:

1. R1是基于V3-Base模型开发的。
2. V3在post training环节使用了R1产生的高质量推理数据，显著提升了V3模型的推理能力。
3. V3在post training环节同样使用了和R1一样的RL策略，提升推理能力并对齐人类偏好。
4. V3/R1的post training环节都使用了V3作为Reward model对非数学编程场景提供反馈。
5. Distilling R1 for V3: Distilling R1可以提升V3的推理能力，但会影响到处理一般问题的能力，增加反应长度，考虑到模型准确性和计算效率，V3主要蒸馏了R1的数学和编程能力。这个方向未来可探索空间很大，基模型和推理模型之间的边界很淡，未来是否会融合，迈入下一层智能，值得期待。



05

DeepSeek R1 for 科学研究

从“工具”向“协作者”的跃迁



DeepSeek R1 for 科学研究： 从“工具”向“协作者”的跃迁

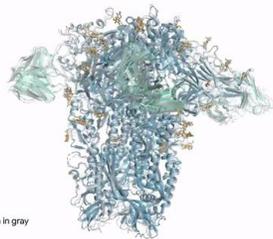
DeepSeek R1可以在科学研究全流程中作为“协作者”，帮助确立研究方向、设计实验、分析数据优化结果、促进跨学科协作以及撰写报告。

科学研究和AI

昨天在《洞悉DeepSeek三代模型演进：不断探索迭代的系统创新》中从深度学习的角度总结了DeepSeek的系统创新，今天想面向科研人员，从科学研究的角度来讨论一下DeepSeek R1可能为科学研究领域带来的改变。

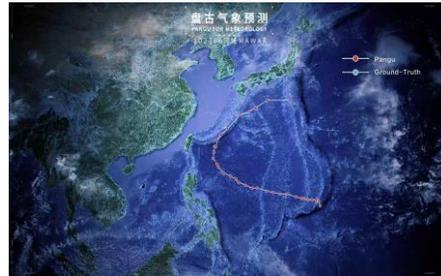
科学研究通过系统的观察、实验、分析和推理，探索自然界和社会现象的规律与本质，不断获得新的知识和理论。在探索未知的过程中，研究人员往往需要面对复杂的问题和繁琐的实验，不断的试探、调整和修正，在这个过程中，AI作为工具对科学数据和问题进行分析和建模，在科学的很多领域取得了显著的成果，如解决了蛋白质结构预测问题的AlphaFold，对气象要素进行准确预测的PanguWeather。

7PNM



Ground truth shown in gray

AlphaFold3



Pangu Weather

DeepSeek R1 的深度思考

DeepSeek R1直接将强化学习应用于基础大语言模型，使得模型能够增强思考过程，利用思维链（CoT）来解决复杂问题，R1模型已经展现了很强的自我验证、反思和生成思维链等能力。深度思考的能力会对于科研人员带来很大的帮助。

在DeepSeek R1的训练过程中，随着训练迭代次数的增加，模型的思考深度也逐渐自发的增加，这也意味着模型解决复杂问题的能力在不断提升。科研的过程中也同样的需要研究人员不断思考，引入更多领域的信息或数据，对科研问题进行反复论证。

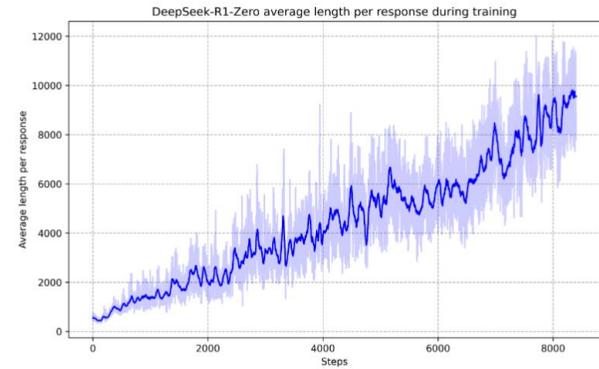


Figure 3 | The average response length of DeepSeek-R1-Zero on the training set during the RL process. DeepSeek-R1-Zero naturally learns to solve reasoning tasks with more thinking time.



DeepSeek R1 for 科学研究： 从“工具”向“协作者”的跃迁

DeepSeek R1可以在科学研究全流程中作为“协作者”，帮助确立研究方向、设计实验、分析数据优化结果、促进跨学科协作以及撰写报告。



在DeepSeek R1训练过程中的另一个有趣发现是Aha moment，模型突然“恍然大悟”，重新考虑初始条件策略，并分配更多的思考时间。科研人员对于Aha moment也不会陌生，DNA双螺旋结构的发现，可能也是沃森的Aha moment思考所得。

Question: If $a > 1$, then the sum of the real solutions of $\sqrt{a - \sqrt{a+x}} = x$ is equal to

Response: <think>
 To solve the equation $\sqrt{a - \sqrt{a+x}} = x$, let's start by squaring both ...
 $(\sqrt{a - \sqrt{a+x}})^2 = x^2 \implies a - \sqrt{a+x} = x^2$.
 Rearrange to isolate the inner square root term:
 $(a - x^2)^2 = a+x \implies a^2 - 2ax^2 + (x^2)^2 = a+x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$
 ...
Wait, wait. Wait. That's an aha moment I can flag here.
 Let's reevaluate this step-by-step to identify if the correct sum can be ...
 We started with the equation:
 $\sqrt{a - \sqrt{a+x}} = x$
 First, let's square both sides:
 $a - \sqrt{a+x} = x^2 \implies \sqrt{a+x} = a - x^2$
 Next, I could square both sides again, treating the equation: ...
 ...

Table 3 | An interesting “aha moment” of an intermediate version of DeepSeek-R1-Zero. The model learns to rethink using an anthropomorphic tone. This is also an aha moment for us, allowing us to witness the power and beauty of reinforcement learning.

DeepSeek R1的性能测试中，有三块成绩对于科研是重要的，数学能力（AIME 2024, MATH-500），编程能力（Codeforces, SWE-bench），科学研究能力（GPQA Diamond），R1都取得了很好的成绩。值得一提的是，GPQA数据集是一个非常具有挑战性且高质量的数据集，专为评估生物学、物理学和化学领域的多项选择问题而设计，它的难度即便是对于领域专家都是巨大的，DeepSeek R1的这一成绩可以减轻大部分科研人员对于模型专业性上的顾虑。

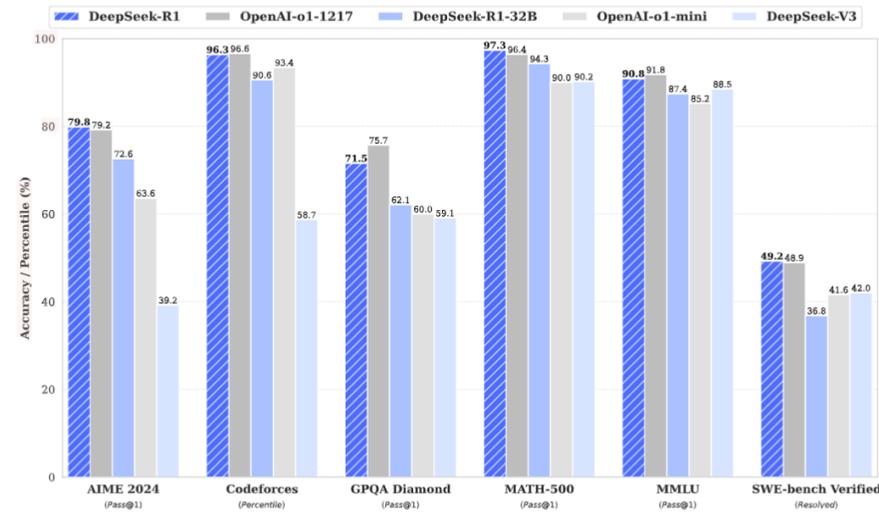


Figure 1 | Benchmark performance of DeepSeek-R1.





DeepSeek R1 for 科学研究： 从“工具”向“协作者”的跃迁

DeepSeek R1可以在科学研究全流程中作为“协作者”，帮助**确立研究方向、设计实验、分析数据优化结果、促进跨学科协作以及撰写报告。**

研究方向确立

DeepSeek R1可以做为Copilot，帮助科研人员发现潜在的知识空白方向，构思并生成初步的研究计划或项目提案。这些提案将包括问题背景、研究目标、研究方法以及预期成果等，帮助科研人员清晰阐述研究思路，推动项目立项过程。

如果期望做出更加专业的领域洞察报告或项目立项报告，需要和R1多次交流，不断迭代明确目标和方向，或者借助增强搜索（RAG）和Agent能力，使得模型考虑到更多专业的资料，系统的进行报告的生成。

NextChat
github.com/ChatGPTNextWeb/ChatGPT-Next-Web

模型: deepseek-r1-distill-qwen-32b
消息: 2
主题: New Conversation
时间: 2025/2/8 11:03:01

我是研究阿尔兹海默症的研究人员，请帮我深入分析和洞察最近该领域的研究情况，发现可能的知识空白方向，并分析该方向是否可以作为我的课题下一步研究方向，并给出研究建议和计划。

<think> 嗯，用户是研究阿尔兹海默症的，现在想让我帮他深入分析最近的研究情况，找出可能的知识空白，并判断是否可以作为下一步的研究方向，最后给出建议和计划。

首先，我需要回顾最近阿尔兹海默症的研究进展。这包括病因、病理机制、诊断方法和治疗方法等方面。最近的研究可能集中在生物标志物、基因研究、神经影像学、早期诊断、精准治疗以及生活方式的影响等方面。

接下来，我需要识别这些领域的知识空白。例如，尽管已经发现了一些基因突变与AD相关，但环境因素和基因的相互作用仍不明确。此外，现有的生物标志物可能不足以全面评估疾病进展，尤其是针对不同亚型的患者。

然后，分析这些空白是否可以作为研究方向。例如，基因与环境的交互作用是一个有潜力的领域，因为它可能解释为什么有些携带风险基因的人没有患病，而其他人则患病。这可能为预防策略提供新的思路。

接下来，我需要为用户提供具体的研究建议和计划。这可能包括确定研究目标，选择合适的方法，设计实验，以及规划时间表。

在与用户沟通时，我需要确保我的建议符合他们的兴趣和资源。例如，如果用户对基因研究感兴趣，我可以重点放在基因与环境的交互作用上；如果他们更关注诊断技术，可以建议开发新的生物标志物。

实验设计

在实验设计阶段，DeepSeek R1能够帮助科研人员设计合理的实验方案，预测可能的实验结果，并提供数据分析建议。并且能够根据历史数据和科学规律推测出不同实验条件下可能的结果，帮助研究人员避免实验设计中的常见陷阱。

NextChat
github.com/ChatGPTNextWeb/ChatGPT-Next-Web

模型: deepseek-r1-distill-qwen-32b
消息: 2
主题: New Conversation
时间: 2025/2/8 10:27:27

请考虑到下文的研究背景，帮助我设计实验计划，详细列出实验步骤

The advent of epidermal growth factor receptor tyrosine kinase inhibitor (EGFR-TKI) has revolutionized the treatment of non-small cell lung cancer (NSCLC) to a great extent (1). EGFR genotype detection is the most common method to identify patients sensitive to EGFR-TKI in clinical practice and clinical trials of novel EGFR-TKIs. Previous studies have demonstrated that third-generation EGFR-TKI can effectively overcome the acquired TKI resistance [nd by secondary T790M mutation, and there are a total of 24 third-generation EGFR-TKIs being developed globally (2,3). While EGFR-TKIs (1,2,3), based on EGFR genotype detection, can identify sensitive patients, tumor heterogeneity and resistance could stratify NSCLC patients into different groups.

<think> 嗯，我现在需要设计一个实验计划，基于用户提供的研究背景。让我先仔细阅读一下背景内容，理解研究的目标和问题。

研究背景主要讨论了EGFR-TKI在治疗非小细胞肺癌（NSCLC）中的应用，特别是第三代EGFR-TKI克服T790M突变后的耐药性。然而，仍有约30%的患者对第三代药物无反应，这提示仅通过EGFR基因型检测可能不足以预测治疗效果。此外，肿瘤异质性和晚期患者获取组织样本的困难，使得需要寻找非侵入性的生物标志物来辅助治疗决策。

研究还提到了CT影像和AI结合在预测EGFR-TKI反应中的潜力，以及利用电子病历（EMR）数据和实验室检查结果开发风险模型的可能性。现有的模型如mPLCom2012和GBDT在诊断和预测方面表现良好，但需要进一步验证和优化。

基于这些信息，实验计划需要综合使用AI、EMR数据、实验室检查和影像学数据，来构建预测模型，评估患者对EGFR-TKI的反应，优化治疗决策。

首先，实验目标应该是构建一个基于AI的预测模型，整合EMR、实验室和影像数据，预测EGFR-TKI的疗效和毒性。同时，评估模型的性能，并比较不同数据源的贡献。

实验设计需要分为几个阶段：数据收集与预处理、模型开发与验证、结果分析与解释，以及临床应用。

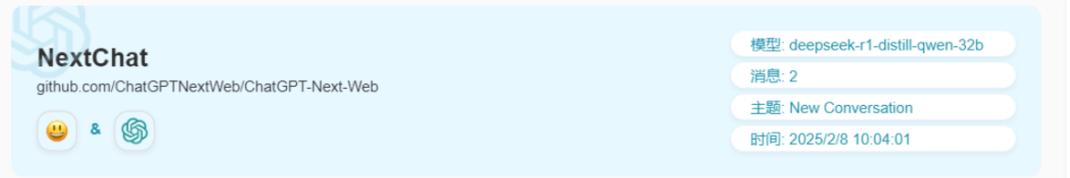


DeepSeek R1 for 科学研究： 从“工具”向“协作者”的跃迁

DeepSeek R1可以在科学研究全流程中作为“协作者”，帮助确立研究方向、设计实验、分析数据优化结果、促进跨学科协作以及撰写报告。

实验数据分析

DeepSeek R1具备强大的编程能力，可以做为Copilot方便的支持研究人员针对实验数据生成分析代码，对实验数据进行建模分析，并生成相应的图表。推荐使用VSCode+Countinue插件调用DeepSeek R1进行本地编程。



下面的内容是我的研究实验的设计和实验数据情况，请帮我生成数据分析的完整代码，包括数据预处理，特征提取，使用CoxMoE模型进行深度学习建模，模型的验证和评估。

Study design and participants
The workflow of this study is graphically summarized in Figure 1. Initially, we assembled 177 patients in abivertinib phase I clinical trial into a training cohort (n=177) and patients from phase II clinical trial of Abivertinib were used as validation cohort.

<think> 嗯，我需要帮助用户生成一个完整的数据分析代码，基于他们提供的研究背景、实验设计和数据情况。用户的研究涉及非小细胞肺癌患者对EGFR-TKI治疗的反应，使用CoxMoE模型进行建模，并进行验证和评估。首先，我需要理解用户的具体需求。他们的数据包括训练集和两个验证集，涉及多个特征，如人口统计学、实验室指标、药代动力学指标等。他们希望进行数据预处理、特征提取，然后使用CoxMoE模型进行建模，并验证模型的性能。

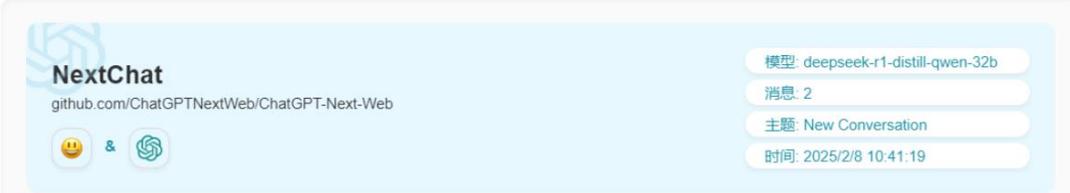
接下来，我需要考虑数据预处理的步骤。数据可能包含缺失值、异常值，以及需要标准化的数值特征。此外，分类变量可能需要进行编码处理。这些步骤对于模型的性能至关重要。

特征提取方面，用户提到使用CoxNet和CoxSVM来选择和缩减特征。这可能意味着他们需要一些特征选择的方法，如Lasso回归或逐步回归，来筛选出最重要的特征。这可能需要使用到特定的Python库，如scikit-survival。

然后是模型的建立，用户提到了CoxMoE模型。我需要查找是否有现成的Python库支持这种模型，如果没有，可能需要自己实现。考虑到时间和资源限制，可能需要寻找替代方案，比如使用Cox比例风险模型或集成学习方法。

实验结果的分析和优化

在科研过程中，DeepSeek R1可以和研究人员一起跟踪实验的进展，并根据新的数据和反馈优化研究方法。它能够根据科研人员的需求和实验结果不断调整推理和分析策略，提供定制化的建议和反馈，促进科研工作持续改进。



这是我当前的数据情况和初步试验结果，请帮我分析实验结果，给出实验下一步优化的建议：

Data sources and characteristics

Three hundred patients were included in the training and validation cohorts. The training cohort (n=177) was assembled from the phase I clinical trial, and the validation cohort (n=59) was assembled from the phase II clinical trial. The training cohort was stratified by age (median 68 years, range 59-80 years) and sex (50% male, 50% female). The validation cohort was stratified by age (median 68 years, range 59-80 years) and sex (50% male, 50% female). The training cohort was stratified by age (median 68 years, range 59-80 years) and sex (50% male, 50% female). The validation cohort was stratified by age (median 68 years, range 59-80 years) and sex (50% male, 50% female).

<think> 好的，我现在需要帮助用户分析他们的实验结果，并给出下一步优化的建议。用户提供了详细的数据情况和初步试验结果，包括数据来源、特征、模型性能等。我需要仔细阅读这些信息，理解实验的设计和结果，然后找出可能的改进点。

首先，用户的数据包括326名接受第三代EGFR-TKI治疗的晚期NSCLC患者，分为训练集、验证集1和验证集2。初步结果显示，CoxMoE模型在预测治疗反应和PFS方面表现良好，但验证集2的AUC略低于训练集和验证集1。

接下来，我需要分析实验结果，找出可能的问题，并提出优化建议。以下是可能的思考步骤：

- 数据分布和平衡：**检查患者在不同医院和治疗反应类别中的分布是否平衡。如果存在不平衡，可能会影响模型的泛化能力。
- 模型性能：**虽然CoxMoE在训练集和验证集1中表现良好，但在验证集2中AUC略低。这可能表明模型在不同数据集上的泛化能力存在差异，需要进一步验证。
- 特征选择：**用户已经使用了一些特征，但可能还有其他潜在的重要特征未被考虑，如基因表达数据或更详细的影像特征。



DeepSeek R1 for 科学研究： 从“工具”向“协作者”的跃迁

DeepSeek R1可以在科学研究全流程中作为“协作者”，帮助确立研究方向、设计实验、分析数据优化结果、促进跨学科协作以及撰写报告。

跨学科的协作与创新

DeepSeek R1能够处理多种学科的数据和知识，促进跨学科的合作与创新。例如，R1能够将生物学、化学、物理学、计算机等领域的知识相结合，提供多学科交叉的解决方案。这种跨学科的协作有助于科研人员在不同领域的知识中找到联系，从而产生创新性思维，解决复杂的科学难题。

NextChat
github.com/ChatGPTNextWeb/ChatGPT-Next-Web

模型: deepseek-r1-distill-qwen-32b
消息: 2
主题: New Conversation
时间: 2025/2/8 10:55:36

医疗影像数据的加入是否能够提升模型对EGFR-TKI治疗预测的准确性，我需要获取哪些医疗影像数据，加入医疗影像数据后模型需要进行怎么调整？

<think> 嗯，用户现在想知道医疗影像数据是否能提升模型对EGFR-TKI治疗预测的准确性。他们需要知道需要获取哪些影像数据，以及加入这些数据后模型需要进行哪些调整。

首先，我需要考虑医疗影像数据在肺癌治疗预测中的作用。影像数据通常包括CT、MRI等，这些数据可以提供肿瘤的形态、大小、位置等信息，对预测治疗反应可能有帮助。

接下来，用户需要获取哪些影像数据呢？可能包括胸部CT扫描，特别是薄层CT，因为它们能提供详细的肺部结构信息。此外，增强CT可能显示肿瘤的血管供应情况，这对预测治疗反应可能有用。另外，如果有PET-CT数据，那也能提供代谢信息，可能对模型有帮助。

然后，关于模型的调整，用户需要考虑如何将影像数据与现有的临床数据（如基因型、实验室结果）结合起来。这可能涉及到多模态学习，其中模型同时处理结构化数据和非结构化影像数据。

在模型构建方面，可能需要使用卷积神经网络（CNN）来处理影像数据，提取特征，然后与其他数据（如基因型）结合使用。这可能需要一个深度学习架构，例如将CNN与循环神经网络（RNN）或全连接层结合，形成一个多任务学习模型。

另外，数据预处理也是一个关键点。影像数据通常需要标准化，可能需要归一化处理，调整图像大小，进行数据增强（如旋转、翻转、缩放）以增加数据的多样性，防止过拟合。

结果解读与报告撰写

在科研工作完成后，DeepSeek R1可以帮助科研人员分析实验结果，并生成高质量的报告或论文草稿。它们能够从实验数据中提炼出关键信息，生成直观的图表和解释，帮助科研人员快速理解结果。

NextChat
github.com/ChatGPTNextWeb/ChatGPT-Next-Web

模型: deepseek-r1-distill-qwen-32b
消息: 2
主题: New Conversation
时间: 2025/2/8 11:11:29

请详细解读实验结果并生成高质量的报告：

Data sources and characteristics
Three hundred twenty-six advanced NSCLC patients with EGFR mutations receiving third-generation EGFR-TKI therapy were included. One hundred seventy-seven patients from the phase I abivertinib trial were used for model training validation.

<think> 好，我现在需要详细解读用户提供的实验结果，并生成一份高质量的报告。首先，我需要理解实验设计和数据来源。

实验分为训练集和两个验证集。训练集来自Abivertinib I期试验，包含177名患者，分为训练集和内部验证集，比例为8:2。验证集1来自Abivertinib II期试验，有106名患者。验证集2来自BPI-7711试验，随机选取了43名患者。这些患者的年龄、性别分布相似，但验证集2女性比例较高。

接下来，我需要分析患者的PFS（无进展生存期）数据。R患者（治疗反应较好）和NR患者（治疗反应较差）的中位PFS在不同队列中有所不同，这可能反映了患者群体的异质性。

然后，评估模型性能。CoxMoE在训练集和验证集中的AUC值分别为0.832、0.728和0.732，显示出良好的预测能力。此外，CoxMoE在C-index（一致性指数）上优于其他模型，如CoxNet、CoxSVM和DeepSurv，说明其在生存分析中的优势。

风险评分的分组显示，高风险组和低风险组在PFS上有显著差异，这表明模型在临床应用中的潜在价值。特别是在Abivertinib和BPI-7711试验中，应用CoxMoE筛选低风险患者，显著提高了中位PFS，这可能为临床决策提供支持。



DeepSeek R1 for 科学研究： 从“工具”向“协作者”的跃迁

DeepSeek R1可以在科学研究全流程中作为“协作者”，帮助确立研究方向、设计实验、分析数据优化结果、促进跨学科协作以及撰写报告。

另外，通过结合现有的科学知识和理论，R1可以提出对实验结果的可能解释，增强对于结果的反思和讨论，并在此基础上帮助科研人员撰写具有创新性和深度的科研报告或论文。

DeepSeek R1 标志着AI从“工具”向“协作者”的跃迁

科学研究的人机协作：

未来科研将越来越多地依赖于 AI 与人类专家的合作。DeepSeek R1 不仅是一个工具，更可能成为“协作者”，参与论文撰写、数据分析和模型构建，推动开放、协同、迭代的科研模式。

加速科学发现与创新：

DeepSeek R1拥有强大的链式推理能力，能够在面对复杂问题时“连环”式地展开推导，帮助科研人员捕捉前沿动态和跨学科知识，从大量零散的实验数据和文献中发现潜在联系，通过整合不同领域的理论和数据，模型有望提出新的研究假说或预测实验结果，激发颠覆性的创新。

跨学科整合与多领域协同：

科研现今正趋向跨学科融合，而DeepSeek R1具备广泛知识覆盖和深度推理能力，能够打破学科壁垒，将物理、生物、化学、计算机等领域的知识关联起来，构建跨学科的理论模型；同时促进协同研究，为多领域科研人员提供共同的平台，通过统一的推理框架开展合作研究，缩短从理论到实验的转化周期。

智能化科研决策与实验设计：

借助深度学习和推理链技术，科研人员可以利用DeepSeek R1优化实验设计，模型能够模拟实验条件，进行参数优化与敏感性分析，从而为实验方案提供建议；辅助科研决策，在面对复杂变量与多种可能性的情形下，模型可对不同研究方向进行评估，帮助科研团队聚焦最有潜力的课题。

DeepSeek R1 现在仍不完美：

1. 受限于科学研究的思维链数据集少，当前R1的能力上限还是比不上领域专家。
2. 不同于常见的问题，更长的思维链对科学研究的帮助会更大，R1未来值得往深度上继续探索极限。
3. 图像数据在科研中非常常见，虽然DeepSeek-VL2中已经提出了基于DeepSeekMoE支持图像数据完整方案，当前R1还不支持图像。
4. 当前思维链长度为8K，在实际使用时经常还是不够的。



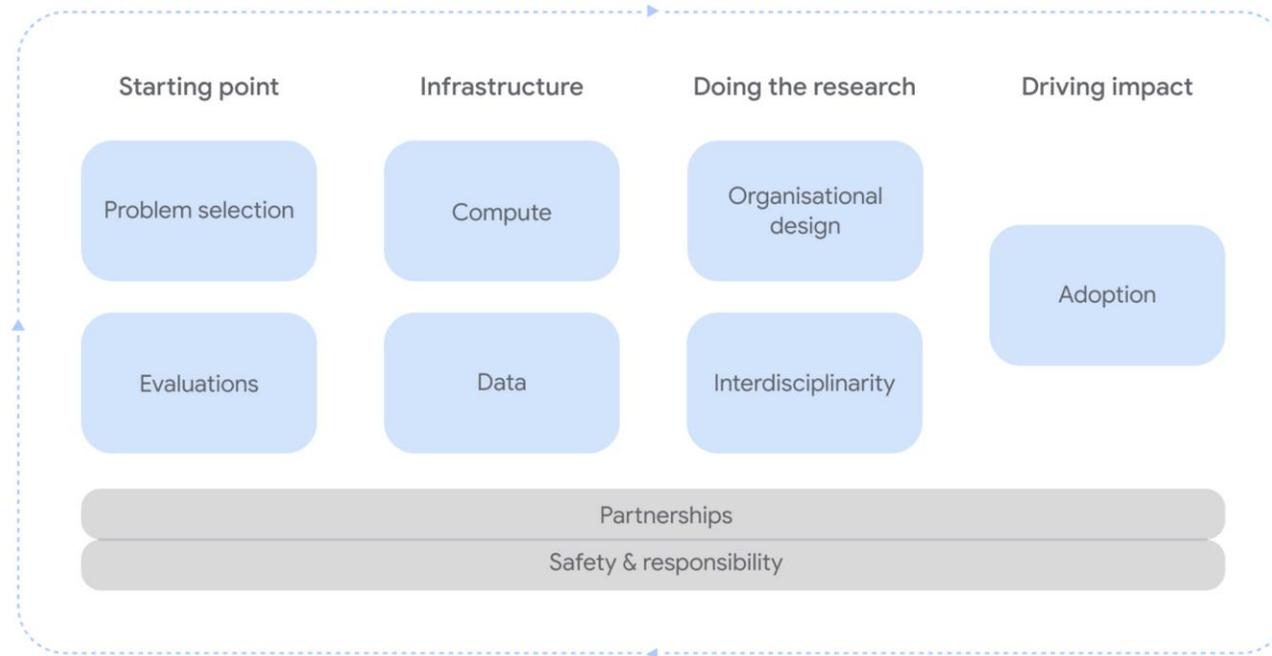
06

DeepSeek的创新机制 借鉴到AI4Science领域 “大科学创新”

大科学创新

AI for Science研究成功的九大要素

The AI for science production function

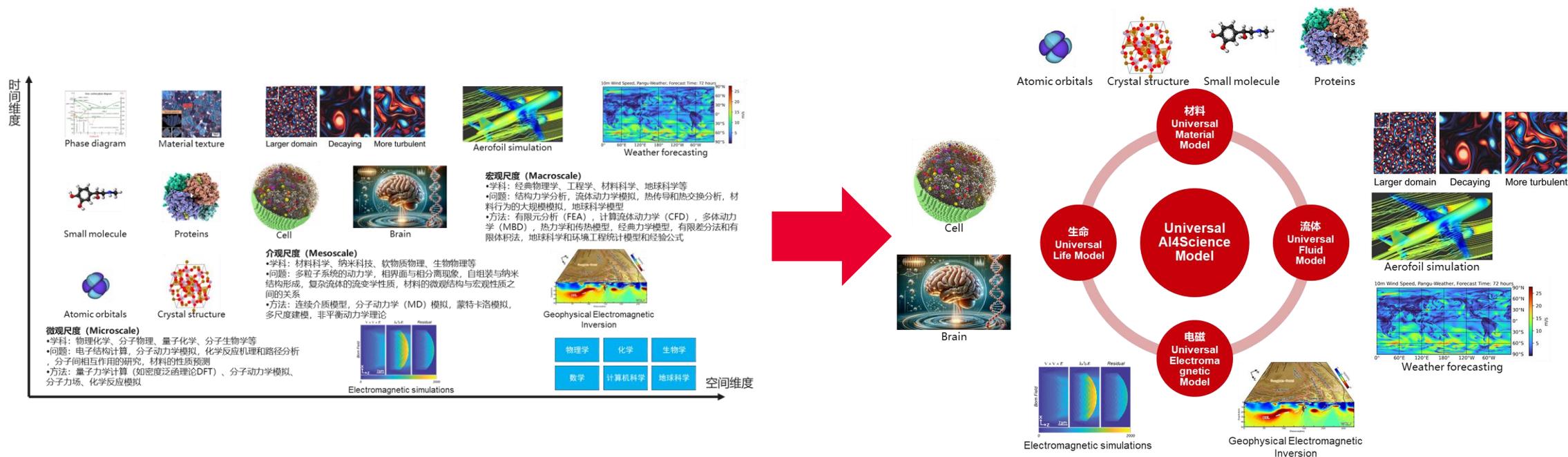


Perspectives, A. P. A new golden age of discovery. <https://www.aipolicyperspectives.com/p/a-new-golden-age-of-discovery> (2025).

问题选择

从点状创新发展到更加立体的大科学创新

从跨学科、多尺度的数据融合趋势，正在促进科研人员思考和设计更大的科学基础模型，从原来的点状创新，慢慢发展到更加立体的大科学创新。



评估方法

更高质量的评估方法也是大模型往更高的智能目标进行挑战的关键

在AI for Science领域，当我们明确了问题的选择，接下来围绕重大科学问题，系统的构建高质量的、公开公正的评估数据、评估方法，并围绕问题推动相关高质量的竞赛或研究方向，是我们迈向问题解决第一步。

| Benchmark (Metric) | # Shots | DeepSeek-V2 | Qwen2.5 | LLaMA-3.1 | DeepSeek-V3 |
|------------------------------------|---------|-------------|-------------|--------------|-------------|
| | | Base | 72B Base | 405B Base | Base |
| Architecture | - | MoE | Dense | Dense | MoE |
| # Activated Params | - | 21B | 72B | 405B | 37B |
| # Total Params | - | 236B | 72B | 405B | 671B |
| Pile-test (99%) | - | 0.606 | 0.638 | 0.542 | 0.548 |
| BBH (9M) | 3-shot | 78.8 | 79.8 | 82.9 | 87.5 |
| MMLU (9M) | 5-shot | 78.4 | 85.0 | 84.4 | 87.1 |
| MMLU-Redux (9M) | 5-shot | 75.6 | 83.2 | 81.3 | 86.2 |
| MMLU-PRO (9M) | 5-shot | 51.4 | 58.3 | 52.8 | 64.4 |
| DROP (9) | 3-shot | 80.4 | 80.6 | 86.0 | 89.0 |
| ARC-Easy (9M) | 25-shot | 97.6 | 98.4 | 98.4 | 98.9 |
| ARC-Challenge (9M) | 25-shot | 92.2 | 94.5 | 95.3 | 95.3 |
| HellaSwag (9M) | 10-shot | 87.1 | 84.8 | 89.2 | 88.9 |
| PIQA (9M) | 0-shot | 83.9 | 82.6 | 85.9 | 84.7 |
| WinoGrande (9M) | 5-shot | 86.3 | 82.3 | 85.2 | 84.9 |
| RACE-Middle (9M) | 5-shot | 73.1 | 68.1 | 74.2 | 67.1 |
| RACE-High (9M) | 5-shot | 52.6 | 50.3 | 56.8 | 51.3 |
| TriviaQA (9M) | 5-shot | 80.0 | 71.9 | 82.7 | 82.9 |
| NaturalQuestions (9M) | 5-shot | 38.6 | 33.2 | 41.5 | 40.0 |
| AGIEval (9M) | 0-shot | 57.5 | 75.8 | 60.6 | 79.6 |
| HumanEval (9M/99%) | 0-shot | 43.3 | 53.0 | 54.9 | 65.2 |
| MBPP (9M/99%) | 3-shot | 65.0 | 72.6 | 68.4 | 75.4 |
| LiveCodeBench-Base (9M/99%) | 3-shot | 11.6 | 12.9 | 15.5 | 19.4 |
| CRUXEval-I (9M) | 2-shot | 52.5 | 59.1 | 58.5 | 67.3 |
| CRUXEval-O (9M) | 2-shot | 49.8 | 59.9 | 59.9 | 69.8 |
| GSM8K (9M) | 8-shot | 81.6 | 88.3 | 83.5 | 89.3 |
| MATH (9M) | 4-shot | 43.4 | 54.4 | 49.0 | 61.6 |
| MGSM (9M) | 8-shot | 63.6 | 76.2 | 69.9 | 79.8 |
| CMATH (9M) | 3-shot | 78.7 | 84.5 | 77.3 | 90.7 |
| CLUEWSC (9M) | 5-shot | 82.0 | 82.5 | 83.0 | 82.7 |
| C-Eval (9M) | 5-shot | 81.4 | 89.2 | 72.5 | 90.1 |
| CMMLU (9M) | 5-shot | 84.0 | 89.5 | 73.7 | 88.8 |
| CMRC (9M) | 1-shot | 77.4 | 75.8 | 76.0 | 76.3 |
| C3 (9M) | 0-shot | 77.4 | 76.7 | 79.7 | 78.6 |
| CCPM (9M) | 0-shot | 93.0 | 88.5 | 78.6 | 92.0 |
| Multilingual MMLU-non-English (9M) | 5-shot | 64.0 | 74.8 | 73.8 | 79.4 |

Table 3: Comparison among DeepSeek-V3-Base and other representative open-source base models. All models are evaluated in our internal framework and share the same evaluation setting. Scores with a gap not exceeding 0.3 are considered to be at the same level. DeepSeek-V3-Base achieves the best performance on most benchmarks, especially on math and code tasks.

Humanity's Last Exam

On [Humanity's Last Exam](#), a recently released evaluation that tests AI across a broad range of subjects on expert-level questions, the model powering deep research scores a new high at 26.6% accuracy. This test consists of over 3,000 multiple choice and short answer questions across more than 100 subjects from linguistics to rocket science, classics to ecology. Compared to OpenAI o1, the largest gains appeared in chemistry, humanities and social sciences, and mathematics. The model powering deep research showcased a human-like approach by effectively seeking out specialized information when necessary.

| Model | Accuracy (%) |
|-------------------------------|--------------|
| GPT-4o | 3.3 |
| Grok-2 | 3.8 |
| Claude 3.5 Sonnet | 4.3 |
| Gemini Thinking | 6.2 |
| OpenAI o1 | 9.1 |
| DeepSeek-R1* | 9.4 |
| OpenAI o3-mini (medium)* | 10.5 |
| OpenAI o3-mini (high)* | 13.0 |
| OpenAI deep research** | 26.6 |

* Model is not multi-modal, evaluated on text-only subset.

**with browsing + python tools

计算基础设施

发展适用于科学领域的计算基础设施，软件计算框架和开放平台

AI for Science 开放平台

AI for 生命

AI for 材料

AI for 气象

AI for 物理等

AI4Science
计算框架

高度灵活与可定制
支持动态图和静态图，可自定义层、损失函数、优化器及其他运算模块

高效自动微分机制
支持高阶求导、隐式函数求导，支持正向传播和反向传播，支持非标准的计算图结构

互操作性与集成能力
对接NumPy、SciPy、MATLAB等科学计算工具，支持支持多种数据格式(如HDF5、NetCDF等)

高性能与硬件加速
硬件加速支持(如GPU、NPU、Ascend等)，支持FP8、FP16、BF16等多种计算精度模式

开放生态与社区支持
并提供丰富的示例、教程和扩展模块，易于扩展与定制，允许用户开发新的模块或插件。

AI4Science
计算基础设施

分布式计算

芯片互联、节点互联、实现高效的数据并行和模型并行

异构计算

CPU/ARM/Kunpeng、GPU/NPU/Ascend、Anton3

混合计算

量子计算、光计算、生物计算

数据基础设施

利用AI for Science数据基础设施持续生产AI Ready的数据

AI for Science 开放平台

AI for 生命

AI for 材料

AI for 气象

AI for 物理等

AI Ready
数据集

级别：0 (Not AI-Ready)

数据一致性角度：未进行内部一致性的检查
数据访问角度：仅通过请求或订单系统对公众开放使用
元数据角度：没有数据字典可用，或数据字典不是机器可读格式（例如PDF）

级别：1 (Minimal)

•数据一致性角度：手动一致性检查
•数据访问角度：只有一个非程序化的访问选项，例如文件下载
•元数据角度：数据字典是机器可读格式（例如CSV、XML、JSON）

级别：2 (Intermediate)

•数据一致性角度：一致性检查部分自动化，部分结果有文档记录
•数据访问角度：多种访问方式，且至少包括一个可编程访问方法（API）
•元数据角度：数据字典使用机器可读的元数据标准

级别：3 (Optimal)

数据一致性角度：完全自动化的内部一致性检查和报告；且考虑了与社区数据集的外部一致性
数据访问角度：除多种交付选项外，还有基于云和高性能计算提供的数据即服务（DaaS）
元数据角度：机器可读的元数据标准；参与行业标准或其他机构数据集标准协调一致

AI4Science
数据基础设施

现有学科数据

现有跨学科、跨尺度的高质量科学基础知识库、数据库

科学大设施

基于科学大设施，产生新的、大规模、高质量、全场景科学数据。

计算模拟数据

设计高质量的模拟数据产生的流程，优化和加速模拟仿真速度，设计准确高效的数据验证策略。

AI Generated Data

当AI模型的能力足够强，AI生成的数据、或者AI标注的数据，可以作为新的高质量数据。

组织设计、跨学科和合作

发展适用于大科学前沿无人区领域探索的创新组织架构

大科学创新组织架构

目标：动态目标管理

无人区的创新具有高度的不确定性，组织需要采用动态目标管理，允许团队在探索中调整方向，同时保持对关键里程碑的把控。

结果：敏捷迭代与里程碑

无人区的创新需要同时应对快速出现的新技术和新趋势，组织架构需要在敏捷迭代和长期投入之间找到平衡点，既支持短期快速实验，又保障长期技术积累。

创新文化：自下而上

创新的核心是激发员工的主动性与创造力。组织需要营造一种开放、包容的文化氛围，鼓励员工提出新想法，并为创新提供试验和试错的空间。

部门：跨职能协同

大模型的创新涉及算法、算力、数据、应用场景等多个领域，需要不同部门和技术团队的深度协同。组织架构需要打破传统的部门壁垒，建立跨职能团队，确保信息流畅和资源高效整合

团队：跨学科协同

跨学科合作（**学科专家、数学专家、AI科学家、工程师**）是解决复杂科学问题的核心驱动力。跨学科团队的成功不仅依赖于各个领域的专业知识，还需要团队成员之间的紧密协作、跨领域的沟通以及对问题的深入理解。

产业：科研与企业合作

AI for Science的最终目标是解决实际问题并推动产业进步，**实现技术落地与产业升级**。科研单位与传统行业企业的合作，是实现这一目标的关键。

资源配置：自上而下

大科学的创新往往需要大量资源投入，包括算力、数据和人才。建立灵活的资源分配机制，根据项目进展和优先级动态调整资源投入，确保里程碑目标的同时鼓励敏捷的创新。

Adoption 采纳和推广

取之开源、用之开源，打造AI for Science的DeepSeek时刻

AI for Science 开放平台

在AI for Science领域，借鉴Hugging Face的成功经验，设计出更加高效的开放平台，整合资源，促进合作与创新，进而推动科学与技术的发展

建立开放的协作平台：建立一个类似Hugging Face的平台，允许科学家和开发者上传、下载和共享AI模型、数据集和工具。促进资源的广泛共享和合作，支持多领域模型的开发和应用。

建立模型评估和比较体系：为每个模型提供详细的性能评估报告，包括准确率、收敛速度等关键指标。开发工具帮助用户比较不同模型在特定任务上的表现，选择最适合的解决方案。

确保透明度和可重复性：提供模型的详细配置文件和训练数据，确保研究结果的可重复性。记录和分享模型在不同研究中的实际应用案例，为其他研究者提供参考。

激励模型开发和贡献：为贡献优秀模型、工具或案例的研究者提供奖励，激发社区的创新能力。支持开源项目，为开发者提供资金和技术支持，鼓励更多人参与开发。

促进社区互动与协作：鼓励科学家和开发者在平台上分享经验、讨论问题和展示成果。组织线上和线下的学术活动，如workshops、hackathons和学术会议，促进科学界与AI社区的互动与协作。

致谢

未来网络研究室：党文栓

战略研究院：周红

终端：田奇

加研所：Tong Wen

计算产品线：朱照生

诺亚：姚俊，陈祥

MindSpore：于璠，刘红升

在本材料的编写过程中，我们与中国科学院及多个研究所的教授进行了深入交流，获得了宝贵的建议和意见，在此一并致谢。

