



浙江大學
ZHEJIANG UNIVERSITY

人工智能重塑科学与工程研究

朱霖潮

浙江大学计算机科学与技术学院人工智能研究所
zhulinchao@zju.edu.cn

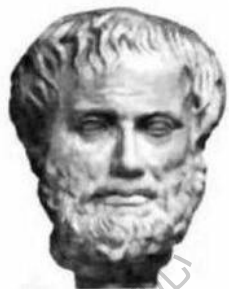
Recognition, LEarning, Reasoning

- 科学研究方法
- 以蛋白质结构预测为例
- AI驱动科学研究的全过程
- 小结

- **第一范式：经验驱动**——基于观察和归纳的实验研究
 - 以**经验主义**和**人的思考**为主导的科学研究范式
 - **实验**是开展研究的主要手段
 - 起源可追溯至**古希腊**和**中国**的早期实验方法
 - 数千年文明史中，人类绝大多数技术发展源于对自然现象的观察和实验总结



甲骨文反映了早期古人通过观察自然现象(如天气、农作物生长、天象等)，并进行记录的做法。

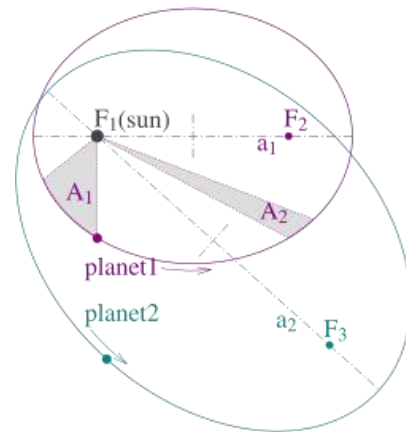


亚里士多德
(Aristototele)
(公元前384-前322)

亚里士多德认为自由落体运动中，物体越重，下落越快；物体越轻，下落越慢。



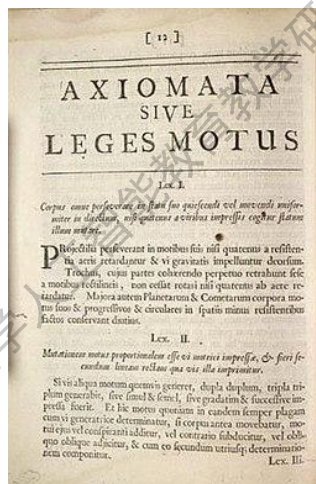
伽利略的落体实验：通过实验验证物体落下速度与质量无关



开普勒发现行星的移动遵守着三条定律

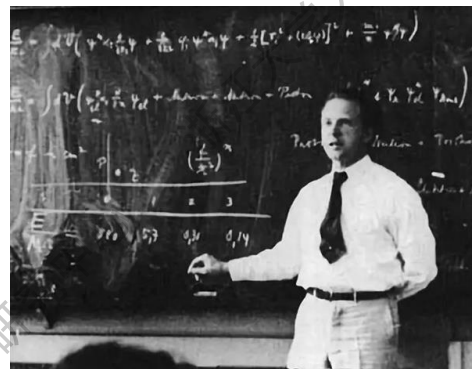
● 第二范式：理论驱动——基于科学假设和逻辑演绎的理论研究

- 当实验条件不具备时，第一范式难以为继
- 不再局限于描述经验事实，使用数学工具研究更精确的自然现象
- 从“知其然”到“知其所以然”，对自然界规律做出背后原理性的解释
- 探索**第一性原理**（First Principle），提供精确描述自然规律的语言，形成可分析的理论模型



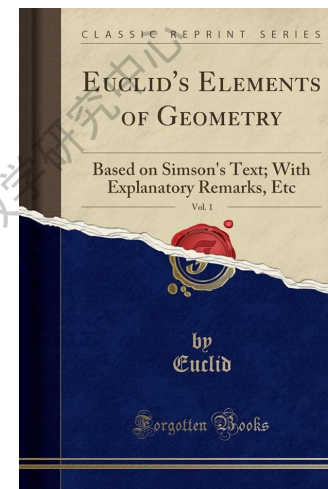
牛顿定律

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}$$



薛定谔方程

$$\hat{H}\Psi = i\hbar\frac{\partial}{\partial t}\Psi$$



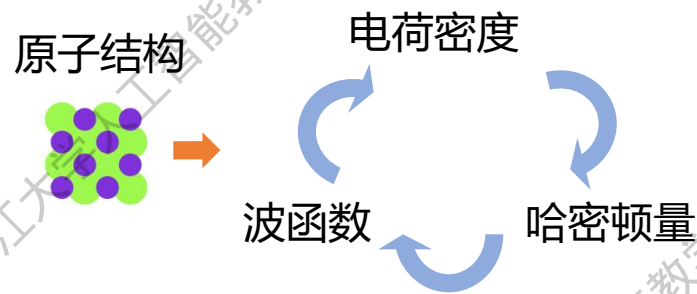
欧几里德《几何原本》

- 23个定义明确了点、线、面等基本几何概念
- 5条公设确立了几何作图的基本操作
- 5条公理奠定了几何推理的基础规则

- **第三范式：计算驱动**——以计算和仿真为主导的科学研究

- 肯尼斯·威尔逊博士（计算物理学家、诺贝尔奖获得者）指出，**计算**是与**理论**和**实验**并列的科学方法之一
- **高性能计算机**和**大规模并行计算**兴起
- 利用计算机精确、大规模求解方程组
- 探索无法通过实验和理论推导解决的复杂问题

- 将量子力学方程转换为计算机高效求解的形式
- 处理复杂多电子体系问题



密度泛函理论 (DFT)



有限元仿真的汽车碰撞测试



超级计算机系统

- 第四范式：数据驱动——数据密集型科学研究方法
 - 图灵奖获得者吉姆·格雷提出**数据密集型**科学研究
 - 利用海量数据采集取代传统观察，以机器学习、统计学等技术替代人类归纳
 - 大数据：例如，美国国家海洋和大气管理局每天收集数十T的环境数据，包含卫星、雷达、船舶、气象
 - 传统模式（先提出假设再验证） vs. 让数据本身“说话”

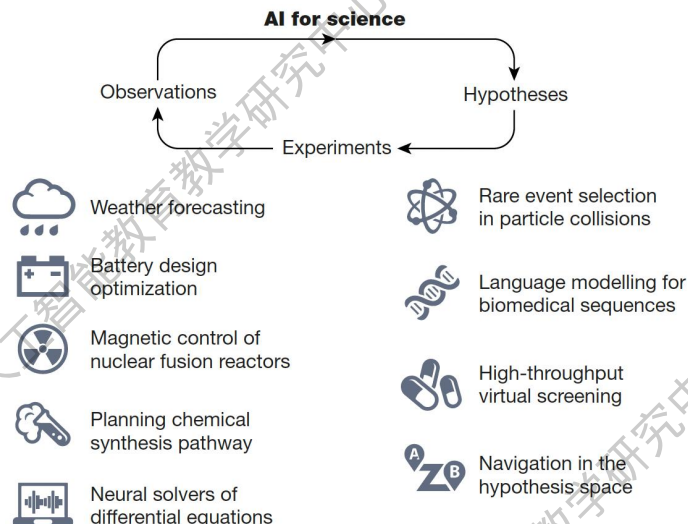
项目名称	英文全称	数据量
DPOSS	帕洛马数字天空巡天 (The Palomar Digital Sky Survey)	3 TB
2MASS	两微米全天巡天 (The Two Micron All-Sky Survey)	10 TB
GALEX	星系演化探测器 (The Galaxy Evolution Explorer)	30 TB
SDSS	斯隆数字天空巡天 (The Sloan Digital Sky Survey)	40 TB
SkyMapper	南天巡天 (Southern Sky Survey)	500 TB
GBT	绿岸望远镜 (Green Bank Telescope)	20 PB
PanSTARRS	全景巡天望远镜和快速反应系统 (The Panoramic Survey Telescope and Rapid Response System)	预计约40 PB
LSST	大型综合巡天望远镜 (The Large Synoptic Survey Telescope)	预计约200 PB
FAST	五百米口径球面射电望远镜	19年内1 EB
SKA	平方公里阵列 (The Square Kilometer Array)	预计约4.6 EB, 每周1 PB

天文大数据

时期/特性	数据统计	备注
年度数据产生量	约70拍字节(PB)/年	主要来自ATLAS和CMS实验
累计数据总量	>1,000拍字节(1艾字节)	截至2024年初的所有LHC实验
数据处理能力	>10亿次计算任务/天	通过WLCG网格计算系统
全球计算中心	约170个	分布在全球各地的WLCG网络
数据转换比例	原始数据约1%被保存	大部分原始碰撞数据经过筛选
Run 2数据量(2015-2018)	约300拍字节	13-14 TeV能量下的碰撞
Run 3预计数据量(2022-2025)	>500拍字节	能量提升至13.6 TeV
HL-LHC预计年数据量(2029后)	>600拍字节/年	高亮度LHC升级后

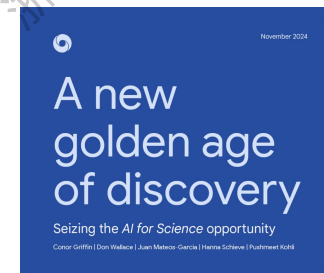
大型强子对撞机(LHC)实验数据统计

- **第五范式：智能驱动（AI for Science）** ——人工智能驱动的科学研究的科学研究
 - 计算驱动范式准但不够快，擅长处理小规模的科学问题
 - 数据驱动范式中，AI主要作为数据分析工具；然而仅靠统计分析，难以在复杂系统的研究上实现突破
 - 第五范式中，AI驱动科学发现的**全过程**，包括提出假设、设计实验等，应对计算复杂性高的组合爆炸问题，实现实验、理论、计算和数据科研范式的融合



求是

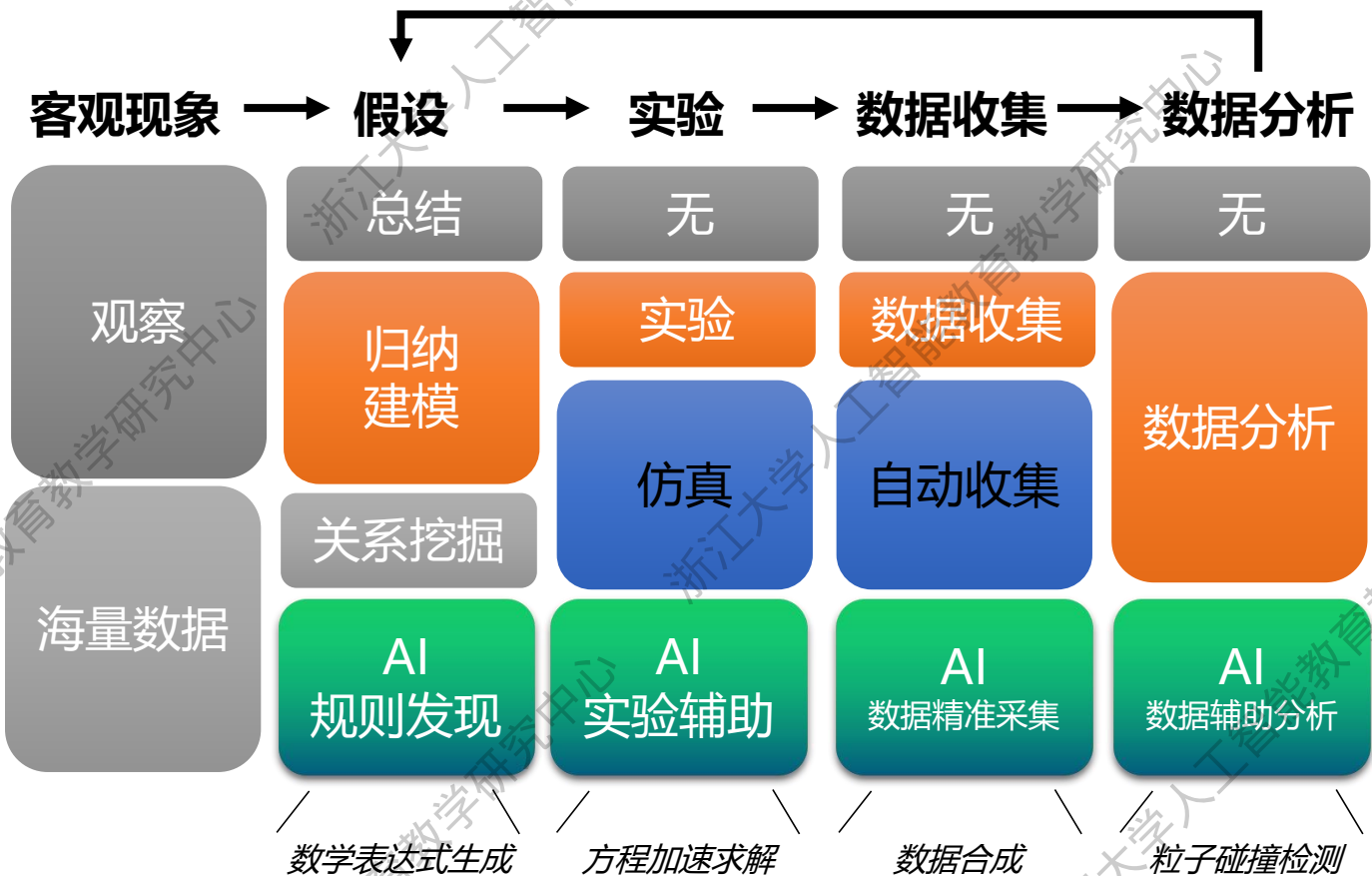
科学发现是一个多方面的过程，包含几个相互关联的阶段，包括假设形成、实验设计、数据收集和分析。人工智能有望通过增强和加速研究过程的各个阶段来重塑科学发现¹。



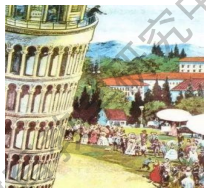
知识：改变科学家获取和传递知识的方式
数据：生成、提取和标注大规模科学数据集
实验：模拟、加速并指导复杂实验
模型：建模复杂系统及其组件的相互作用
解决方案：为大规模搜索问题提供方案

1. Scientific discovery in the age of artificial intelligence, Nature 2024.
2. A New Golden Age of Discovery | Seizing the AI for Science opportunity, Google DeepMind

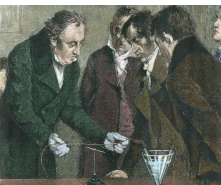
科学方法



科学成果



比萨斜塔



麦克斯韦尔方程



易染-感染-免疫
传播模型模拟



数据驱动机器学习

应用：方程求解

斯托克斯方程

$$\rho \frac{D\mathbf{v}}{Dt} = \nabla \cdot \mathbb{P} + \rho \mathbf{f}$$

气压梯度力的向量式

$$\mathbf{G} = -\frac{1}{\rho} \nabla p$$

$$\nabla = -\frac{1}{\rho} \left(\frac{\partial p}{\partial x} \vec{i} + \frac{\partial p}{\partial y} \vec{j} + \frac{\partial p}{\partial z} \vec{k} \right)$$

湍流系统

给定时间内在垂直于流动的方向上的热通量和动量传递（由剪切应力 τ 表示）为

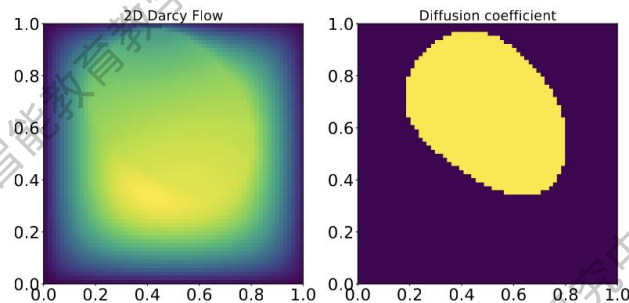
$$q = \underbrace{v'_y \rho c_p T'}_{\text{experimental value}} = -k_{\text{turb}} \frac{\partial \bar{T}}{\partial y};$$

$$\tau = \underbrace{-\rho \overline{v'_y v'_x}}_{\text{experimental value}} = \mu_{\text{turb}} \frac{\partial \bar{v}_x}{\partial y};$$

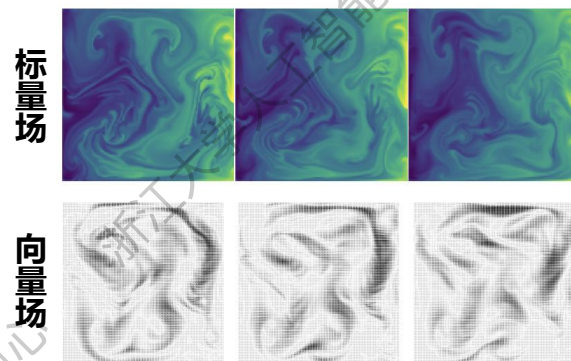
地转方程

从左到右分别是地转相对涡度、行星涡度和伸展涡度。

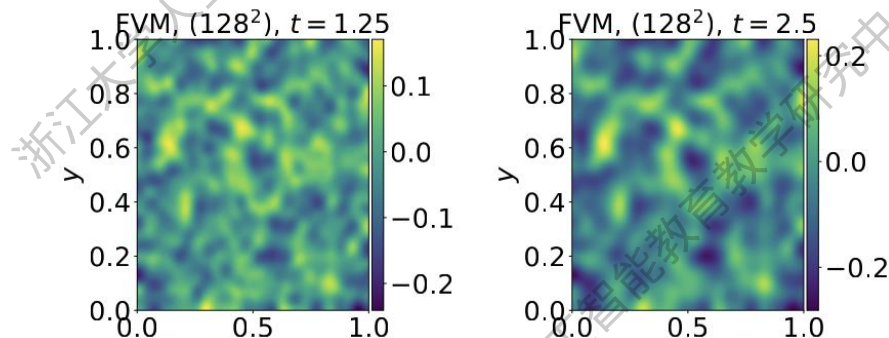
$$q = \frac{1}{f_0} \nabla^2 \Phi + f + \frac{\partial}{\partial p} \left(\frac{f_0}{\sigma} \frac{\partial \Phi}{\partial p} \right)$$



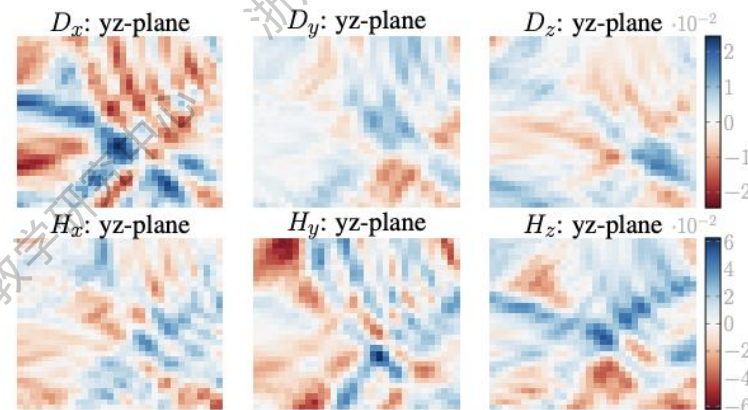
二维Darcy Flow方程



二维Navier-Stokes方程

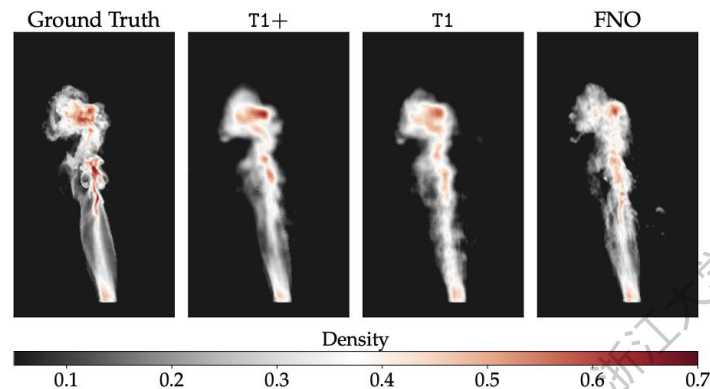


二维Diffusion-Reaction方程

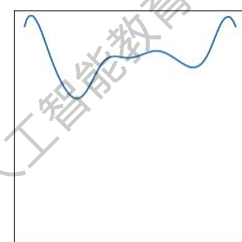


三维Maxwell方程

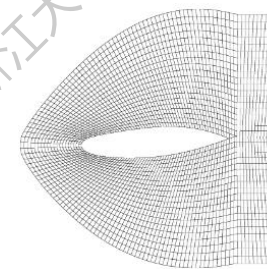
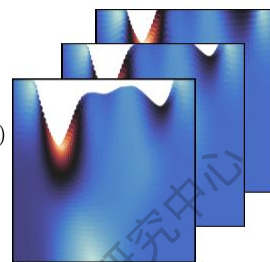
应用：工程仿真



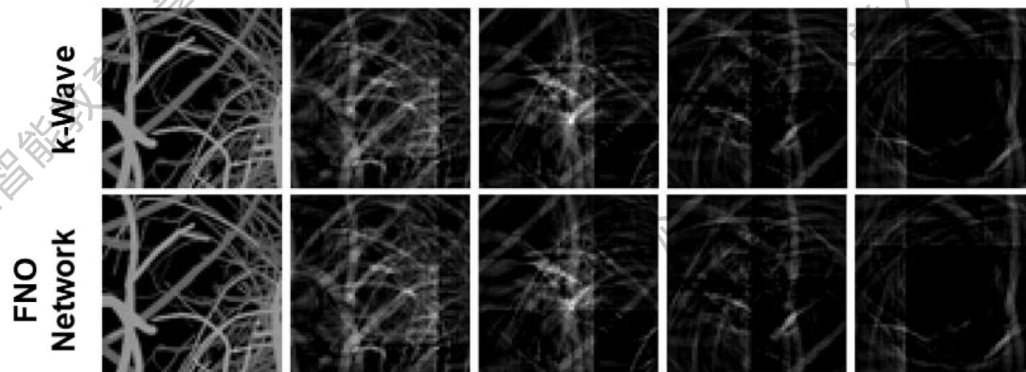
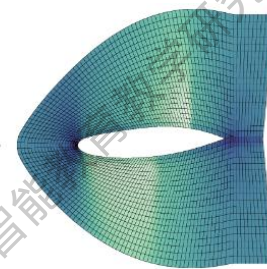
湍流烟雾仿真



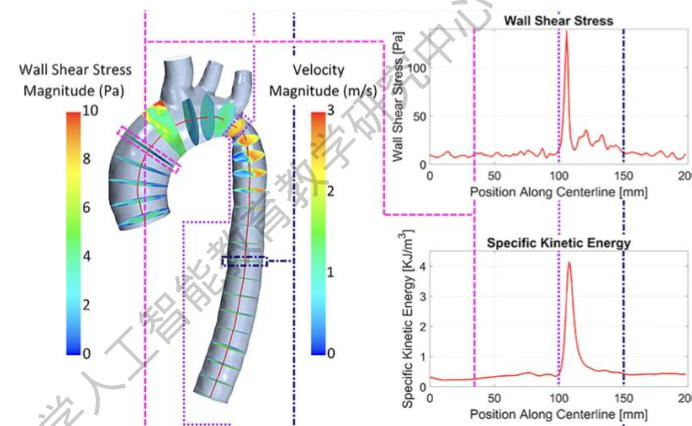
塑形锻造仿真



机翼空气流动仿真

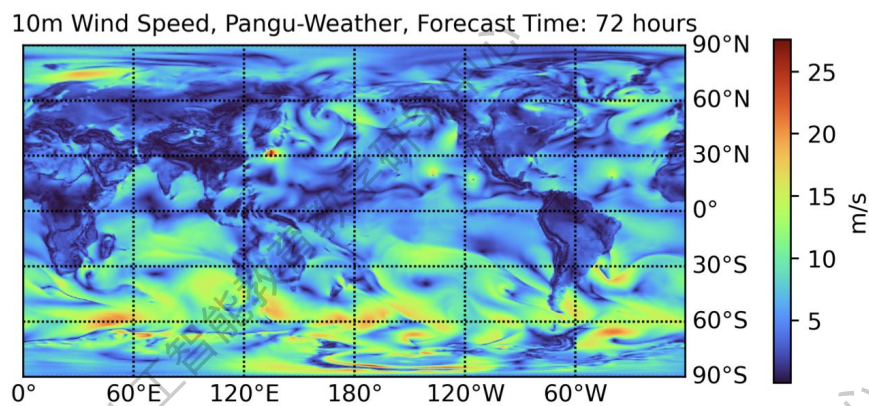
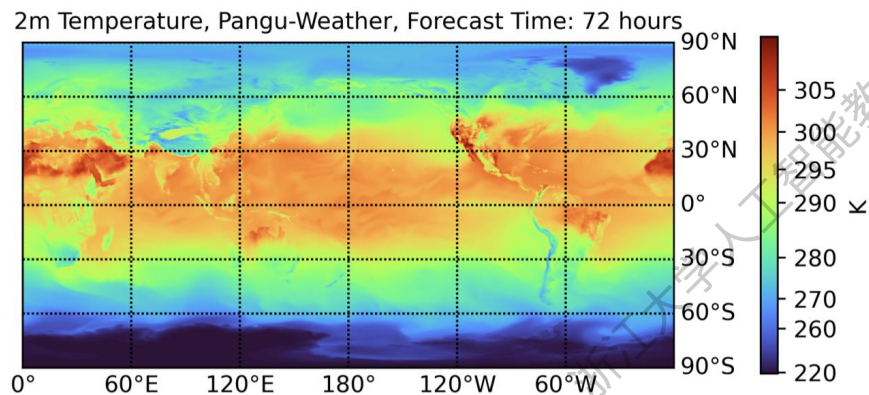


光声成像仿真

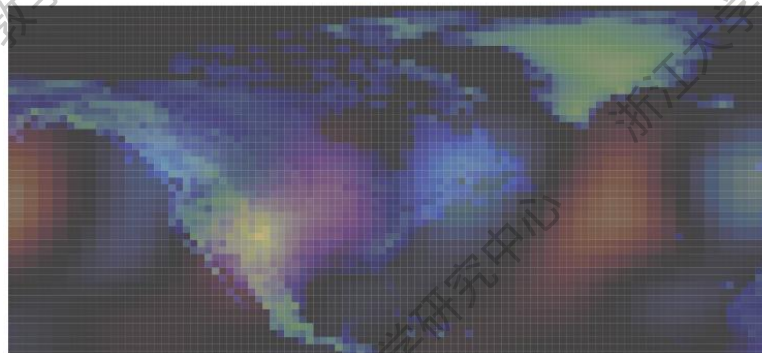


血管液体流动仿真

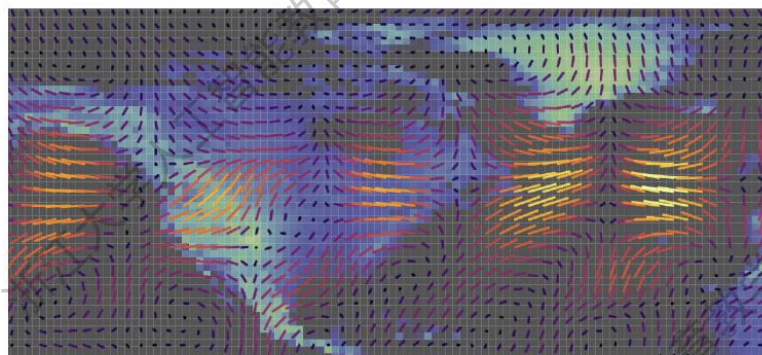
应用：地球科学



地球气候状态预测

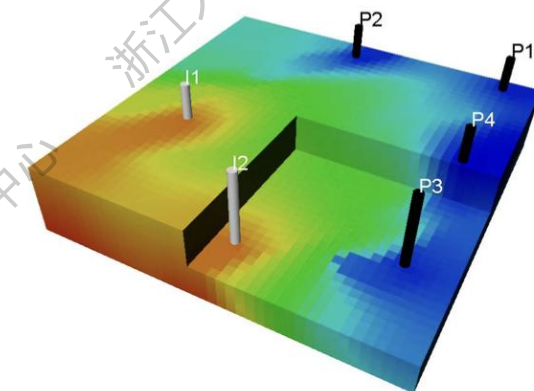
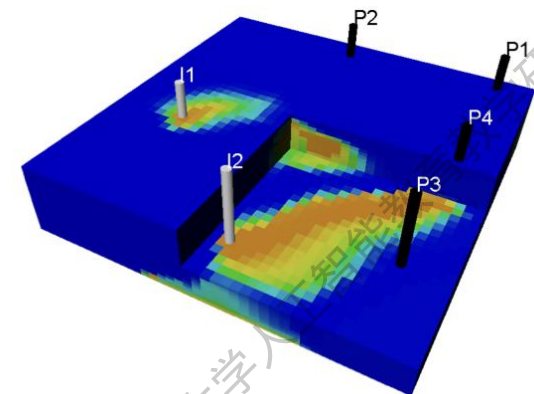


(a) Scalar pressure field



(b) Vector wind velocity field

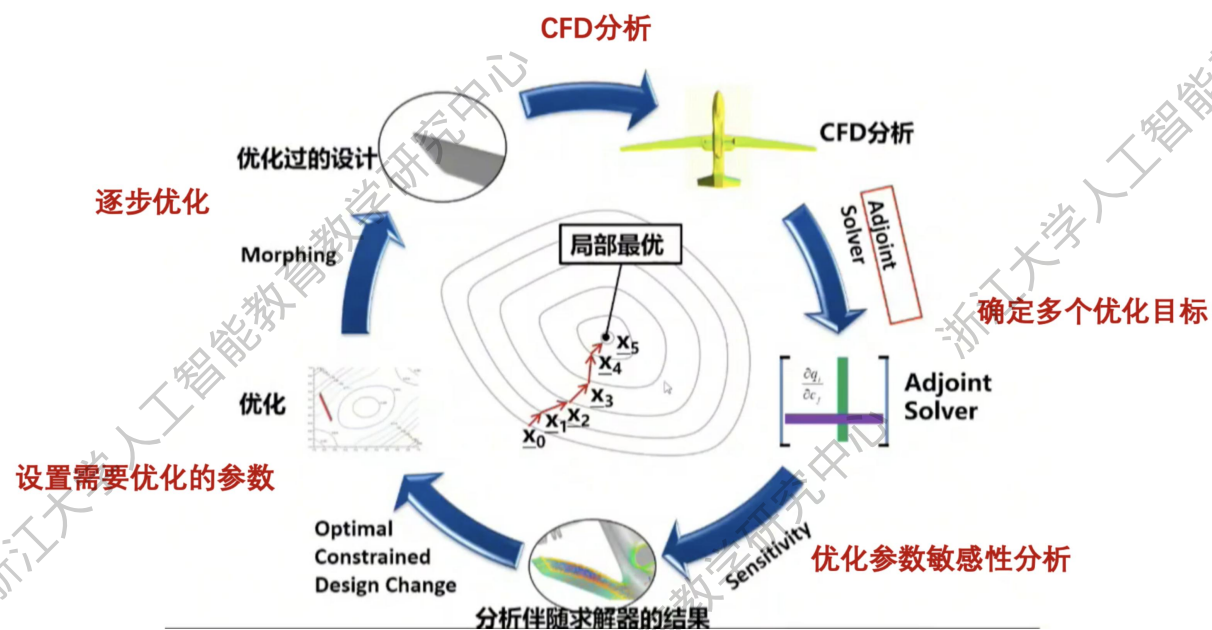
地球浅水状态预测



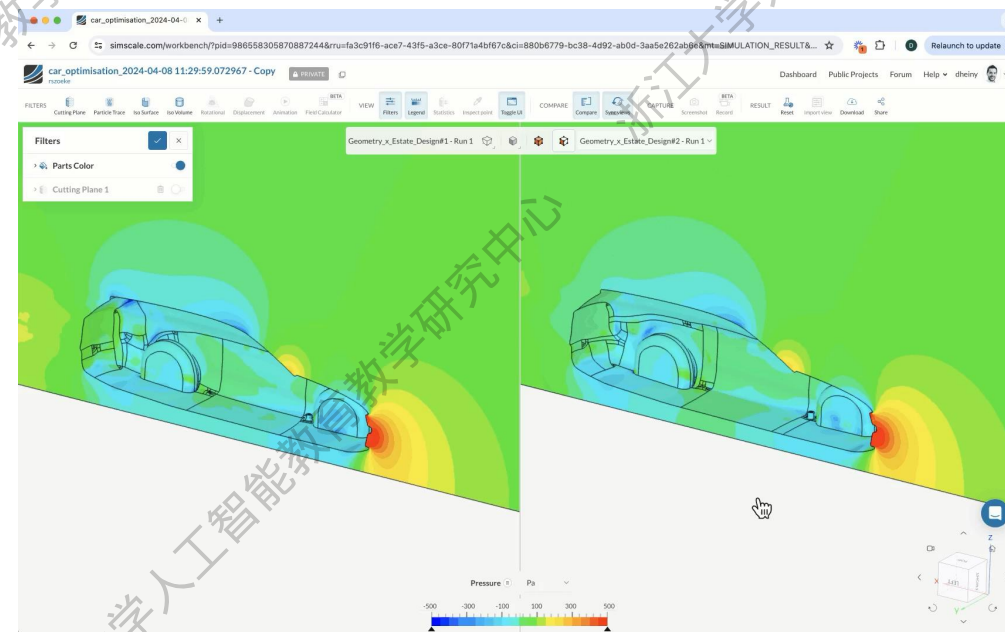
地下水流状态预测

应用：工程设计

- 飞机机型设计的优化流程
 - 首先确定需要优化的设计参数，然后进行CFD分析评估气动性能，确定多个优化目标，进行参数敏感性分析
 - 根据结果实施最优约束设计变更，通过形态变换调整机身外形，最终得到空气动力学性能更优的机型设计

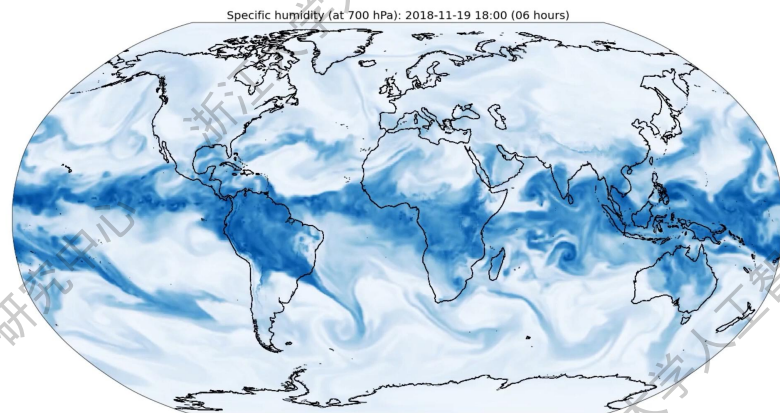
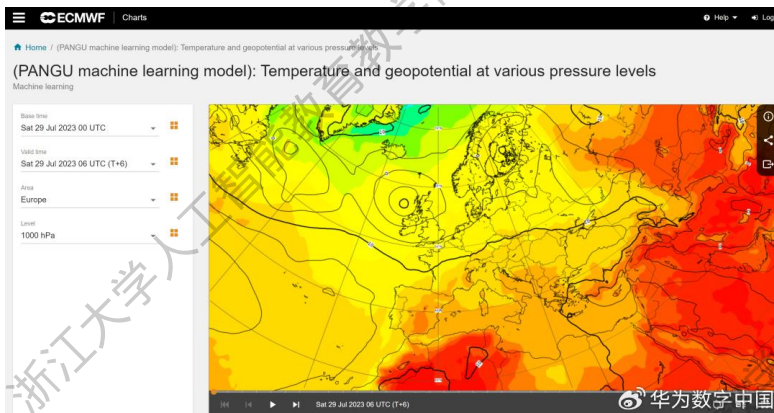


- Ansys发布Ansys SimAI™，通过结合AI，将设计流程加速10-100倍
- 雷诺集团利用 Ansys SimAI，加速了汽车零部件的设计和测试过程，数分钟内完成测试，减少了产品开发周期



应用：气象预报

- 华为盘古气象模型是首个精度超过传统数值预报方法的AI模型，速度相比传统数值预报提速10000倍以上
- 气象模型能够提供全球气象秒级预报，其气象预测结果包括位势、湿度、风速、温度、海平面气压等
- Google DeepMind的GraphCast：利用图神经网络，不到1分钟内生成10天全球预报，在90%的1380个验证目标上优于传统系统



中华人民共和国中央人民政府

www.gov.cn

我国2035年要基本建成气象强国

2021-01-19 21:32 来源：新华社

字号：默认 大 超大 | 打印

新华社北京1月19日电（记者 高敬）中国气象局局长庄国泰19日表示，到2035年，我国要基本建成监测精密、预报精准、服务精细的气象业务体系，气象综合实力达到世界先进水平，气象深度融入民生保障和行业发展，气象强国基本建成，为全面建成社会主义现代化强国作出更大贡献。

庄国泰在当天举行的全国气象局长会议上介绍，“十三五”时期，我国气象灾害造成的死亡失踪人数由“十二五”年均约1300人下降到900人以下，经济损失占GDP的比例由0.6%下降到0.3%。

“十三五”时期，气象业务现代化整体实力再上台阶。综合气象观测能力达到世界先进水平，建成了乡镇全覆盖的近7万个地面站、216部雷达、7颗风云气象卫星组成的地空天一体化综合观测体系。全国卫星遥感应用体系基本建成。

同时，气象预报能力稳步提升，强对流天气预警时间提前至30分钟，台风路径预报24小时误差缩小到70公里。气象信息化水平显著提升，物联网、大数据、人工智能等新技术广泛应用。

庄国泰表示，“十四五”时期，要加快构建更高水平的气象科技创新体系、观测体系、预报体系、服务体系和治理体系。

他介绍，2021年，我国要加快推进气象关键核心技术攻关，持续强化气象业务现代化建设。在观测方面，将启动下一代天气雷达和智能地面气象观测站研发。筹备发射风云四号B星和风云三号E星，推进风云四号微波星立项研制。增补建设重点流域及青藏高原边缘地区观测站网，开展西南区域和海洋观测站网优化设计。

在预报方面，今年将加强全球智能预报能力建设，升级改进全球和区域数值预报模式。此外，要构建“智能预报+气象服务”业务体系，升级改进暴雨诱发中小河流洪水、山洪和地质灾害气象风险预警系统，发展台风精细化风险预警业务，探索开展基于网络机器人的在线定制气象服务。

扫描二维码，在手机端打开网页

相关链接

· 中国气象局与航天科工集团签署战略合作协议 推进航天高科技融入气象强国建设

链接： 全国人大 | 全国政协 | 国家监察委员会 | 最高人民法院 | 最高人民检察院

国务院办公厅网站 | 地方政府网站 | 驻港澳机构网站 | 驻外机构

中国政府网 | 关于本网 | 网站声明 | 联系我们 | 网站地图

主办单位：国务院办公厅 运行维护单位：中国政府网运行中心

版权所有：中国政府网 中文域名：中国政府网.政务

网站标识码bm01000001 京ICP备05070218号 京公网安备1101020200001号

国务院客户端 国务院客户端小程序 微博

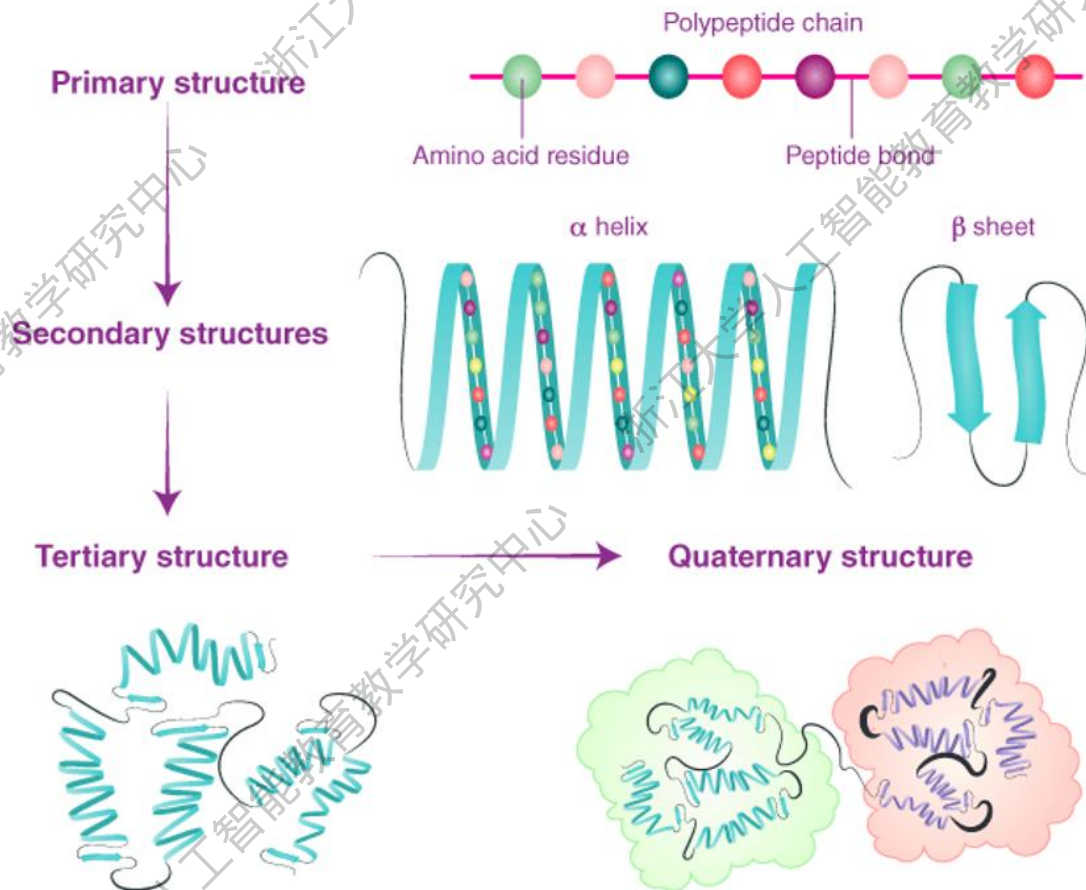


人工智能重塑科学与工程研究

——以蛋白质结构预测为例

蛋白质结构预测：为什么蛋白质结构如此重要？

- 生命的基本构件和功能执行者
- 结构决定功能
 - 主要结构 (Primary structure)：由氨基酸残基 (Amino acid residue) 通过肽键 (Peptide bond) 连接形成的多肽链 (Polypeptide chain)
 - 二级结构 (Secondary structures)：多肽链局部折叠形成的规则结构，包括 α 螺旋 (α helix) 和 β 折叠 (β sheet)
 - 三级结构 (Tertiary structure)：整个多肽链在三维空间中的折叠结构
 - 四级结构 (Quaternary structure)：由多个蛋白质亚基组合在一起形成的复合蛋白质结构
- 疾病与蛋白质结构异常
- 药物开发与蛋白质靶点



蛋白质结构预测：蛋白质折叠问题

● 蛋白质折叠

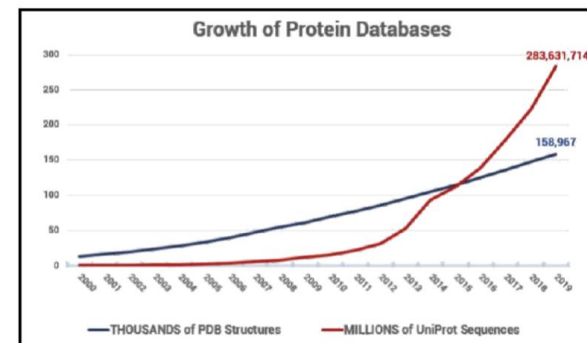
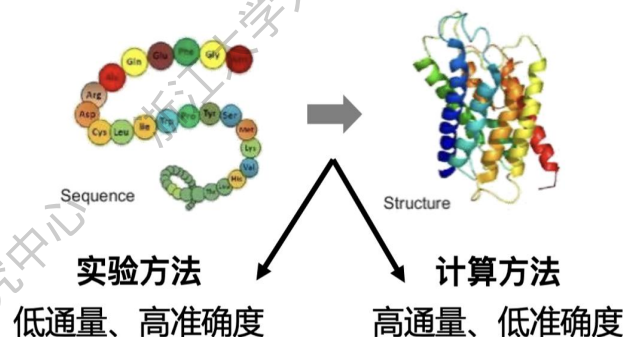
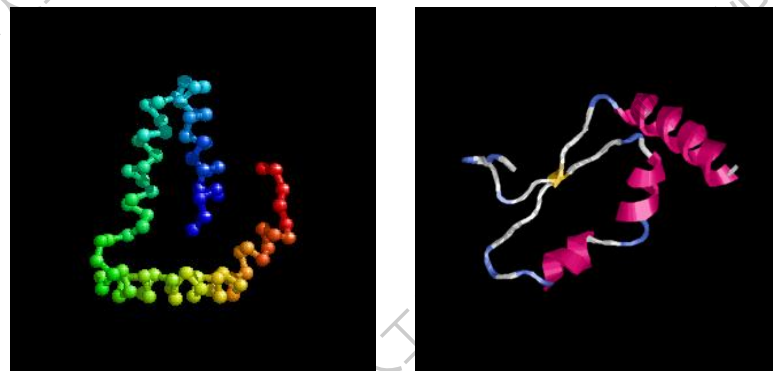
- **序列决定结构**：蛋白质的氨基酸序列包含了所有必要的信息，决定了其最终三维结构。
- 一条线性氨基酸链如何在水溶液中自发折叠成特定的三维结构？
- **折叠机制的复杂性**：蛋白质折叠涉及多种分子力的精确平衡：氢键形成、疏水相互作用（疏水核心的形成）、范德华力、静电相互作用、熵效应、溶剂效应

● 实验挑战

- 直接观察蛋白质折叠过程极其困难
- 建立完整的折叠理论困难

● 计算挑战

- 从计算角度看，蛋白质折叠问题搜索空间巨大，需要模拟复杂的物理和化学过程
- 头计算折叠过程十分缓慢



海量的序列信息，少量的结构信息
需要高精度高通量发现蛋白质结构的方法

蛋白质结构预测：CASP竞赛

● 早期挑战与背景

- 结构预测的可靠性问题：1980年代末到1990年代初，蛋白质结构预测领域处于混乱状态。

● John Moult和Jan Pedersen的愿景

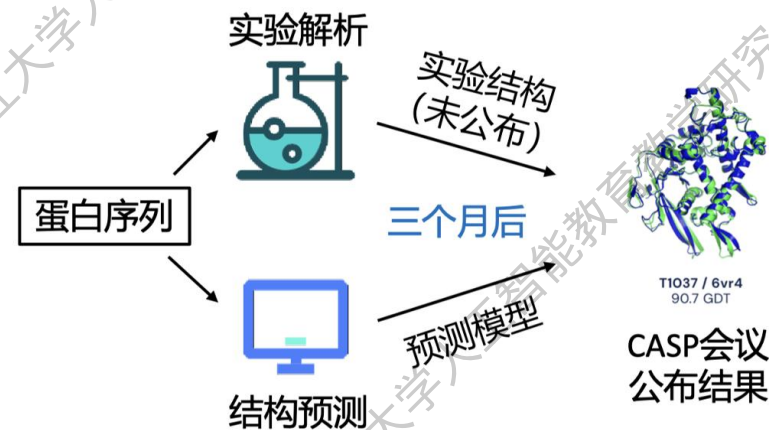
- 完全透明、公正的“盲测”竞赛

● CASP的建立与运作机制

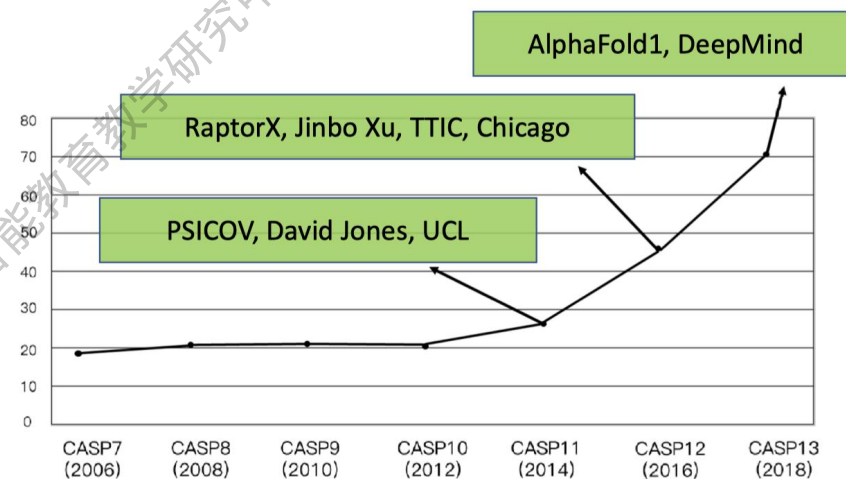
- 首届CASP竞赛（1994年）
- 独特的盲测机制

● CASP的演变与影响

- CASP1到CASP9（1994-2010）：稳步进展
- CASP10-**CASP13**（2012-2018）：深度学习兴起
- **CASP14**（2020）：AlphaFold2



评估计算机预测蛋白质结构方法的准确度



蛋白质结构预测：AlphaFold（2018年）

● CASP13

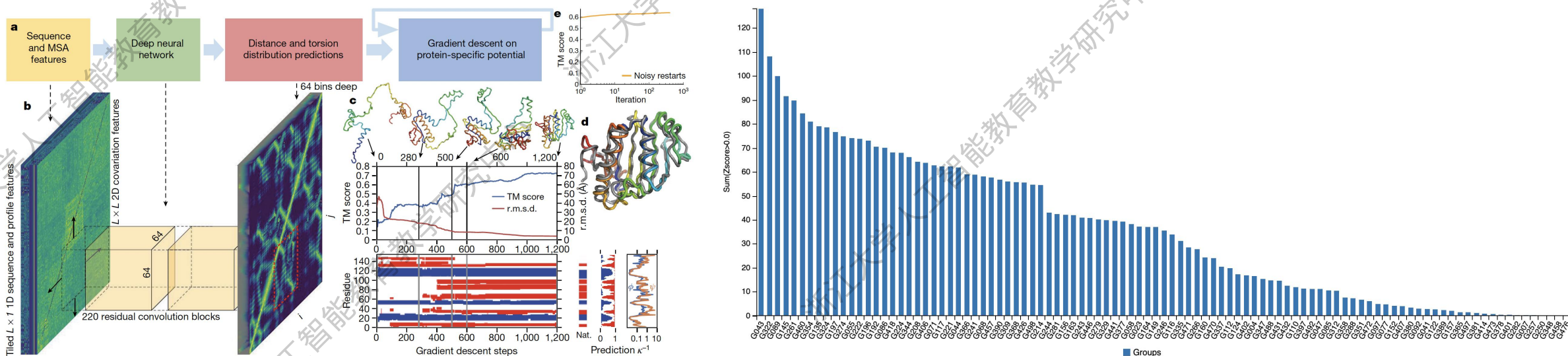
- 2018年12月，AlphaFold在CASP13获全球第一，平均得分明显高于其他参赛者。

● 技术方案

- 使用多序列比对(MSA)收集同源蛋白质序列数据，通过分析序列变异模式找出蛋白质中的残基对接触可能性
- 将这些信息输入深度残差神经网络，预测蛋白质内各残基之间的距离和角度约束
- 使用梯度下降优化算法，将预测的距离和角度约束转化为完整的三维结构

● 引起轰动但尚未完全突破

- 尽管AlphaFold表现出色，仍未达到实验精度水平

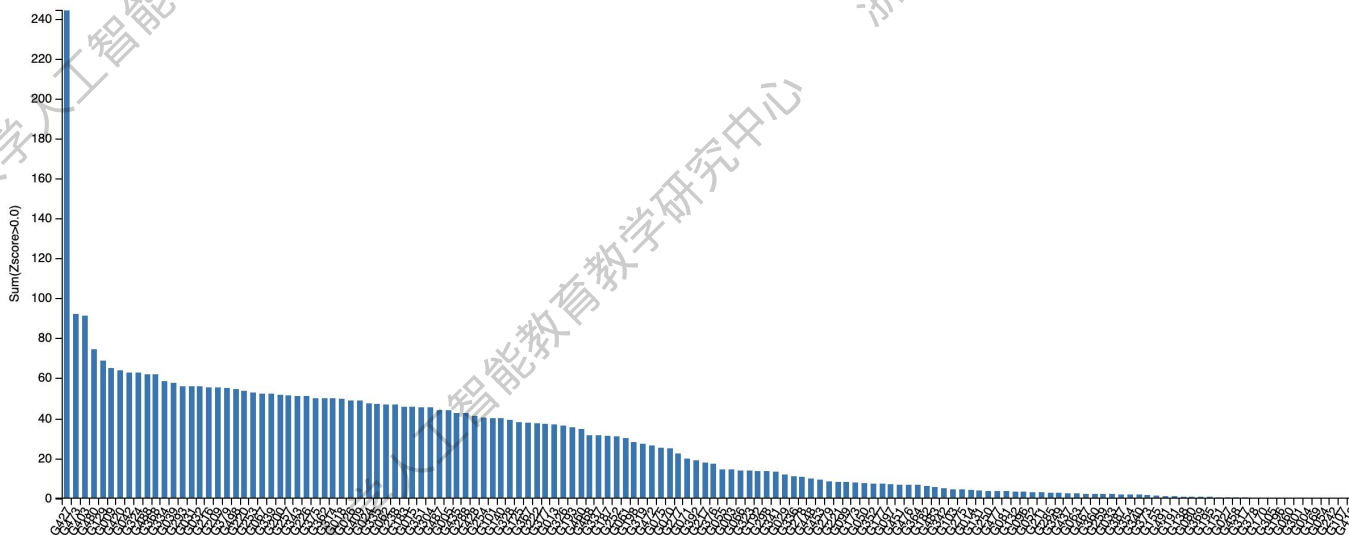
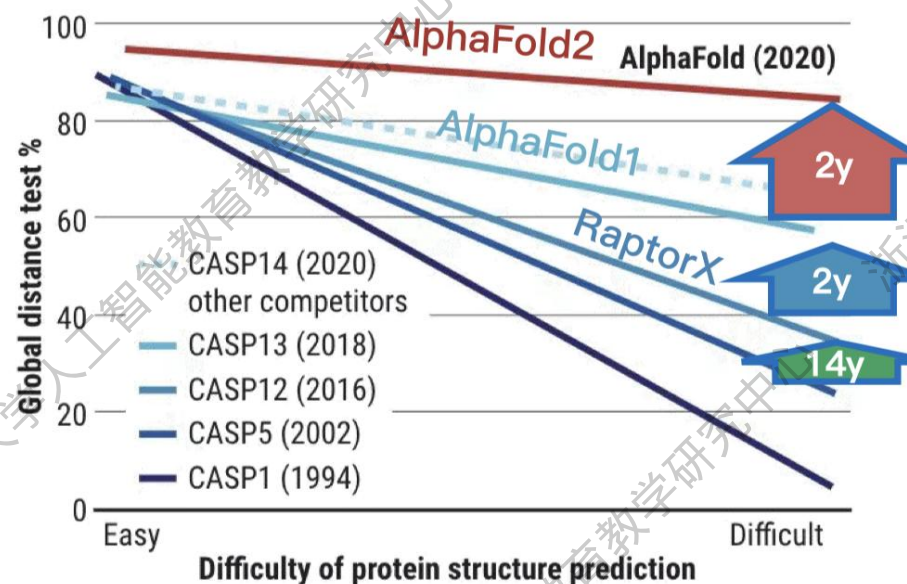


蛋白质结构预测：AlphaFold 2（2020年）

● “AlphaFold时刻”：CASP14上的惊人表现

- 2020年11月30日，在CASP14的线上会议上，DeepMind团队公布了AlphaFold2的结果，平均GDT_TS得分达到92.4分（满分100）
- 在100个测试蛋白质中，有三分之二的预测达到了与实验结构几乎无法区分的精度水平（GDT_TS>90）

● 《科学》杂志评为2021年度科学突破



蛋白质结构预测：AlphaFold 2 (2020年)

● 技术方案：采用端到端架构

1. 多序列比对(MSA)处理

- AlphaFold2首先对输入蛋白质序列进行大规模数据库搜索，找到同源蛋白质序列，并构建MSA

2. Evoformer模块

- **Transformer变体**：包含多个交替的行(序列)和列(残基位置)注意力机制

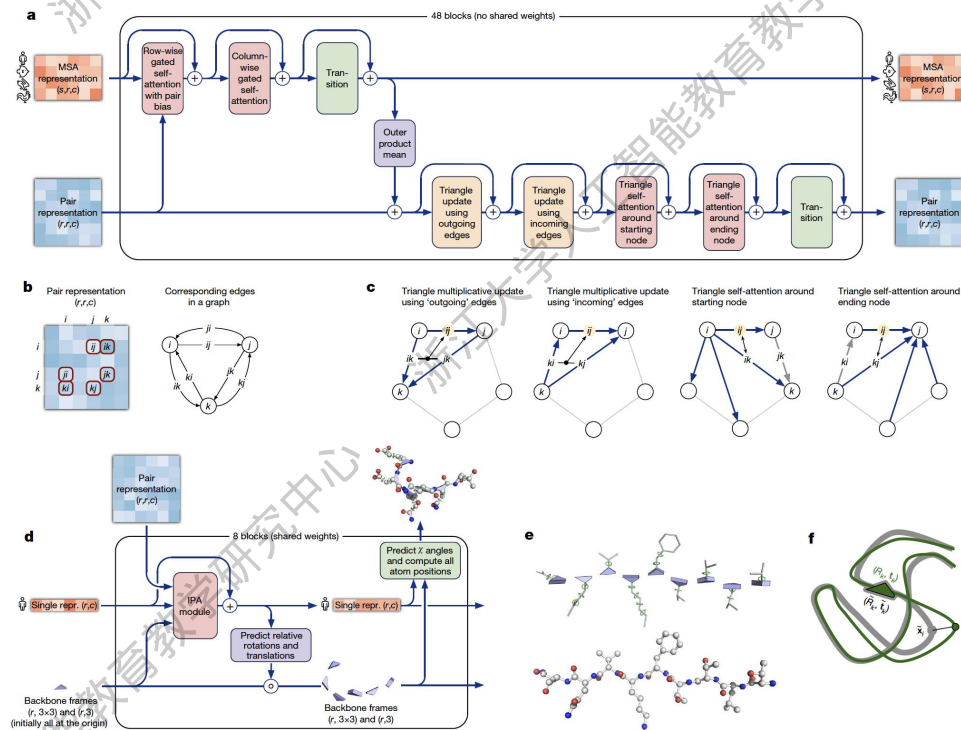
- 三角乘法更新操作维护残基间距离和角度的几何一致性

3. 结构模块：将Evoformer处理的信息转换为三维坐标

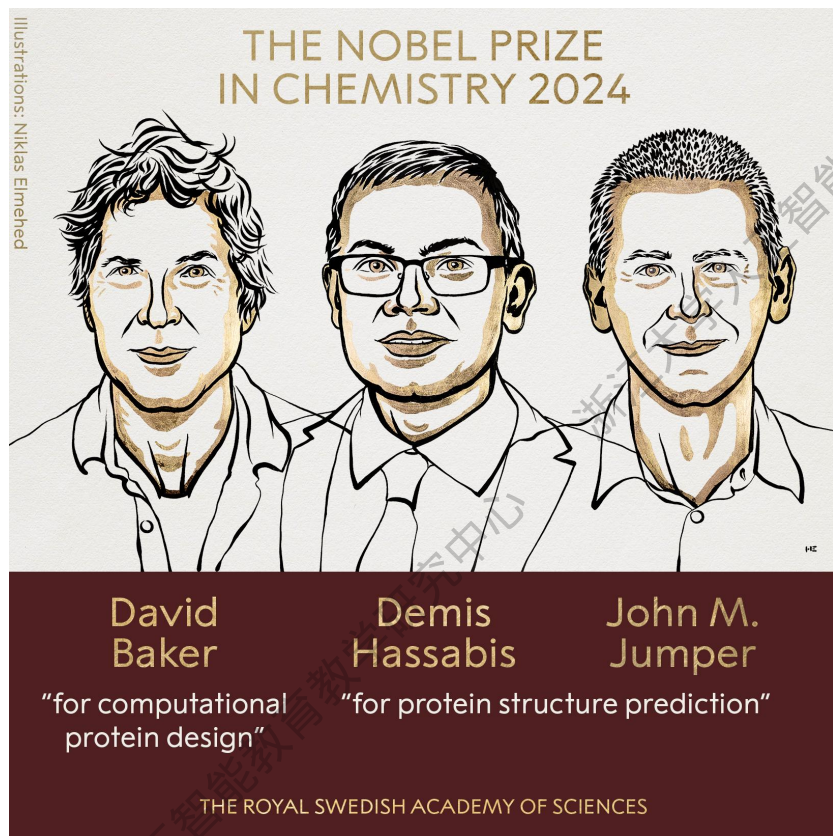
- 预测每个氨基酸残基的主链角度(φ , ψ , ω)和侧链角度(χ)
- 使用前馈网络预测3D坐标框架
- 采用等变神经网络保持空间变换不变性

4. 损失函数设计

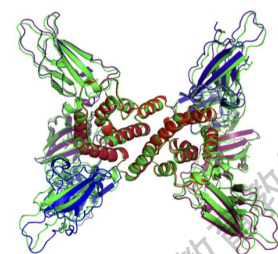
- LDDT(局部距离差异测试)损失：评估局部结构准确性
- TM-score(模板建模评分)损失：评估全局结构相似性
- 几何约束损失：确保蛋白质物理合理性



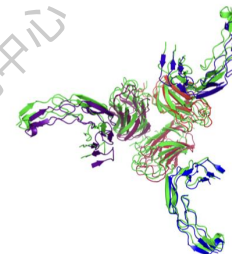
2024年诺贝尔化学奖 (2024年10月9日)



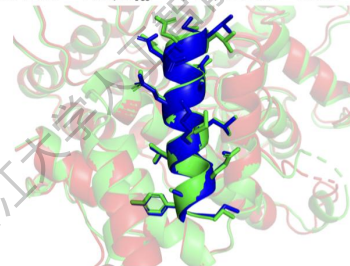
- 2024年诺贝尔化学奖聚焦于人工智能在蛋白质结构预测领域的突破性进展
 - 大卫·贝克 (David Baker) 表彰其在计算蛋白质设计领域的工作
 - 戴密斯·哈萨比斯 (Demis Hassabis) 和约翰·朱姆珀 (John M. Jumper) 表彰其在使用人工智能进行蛋白质结构预测方面取得的进展。
- 人工智能应用于基础科学研究首次获得诺贝尔奖，标志着计算方法在化学、生物学领域的重要性得到认可



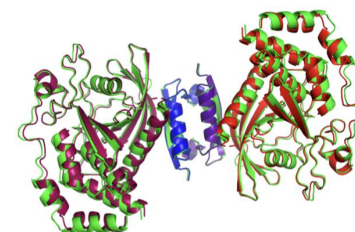
(a) A2B2C2 heteromer
TM-score = 97.4, $N_{\text{res}} = 1,246$, PDB ID = 6E3K



(b) A3B3 heteromer
TM-score = 85.4, $N_{\text{res}} = 795$, PDB ID = 7KHD

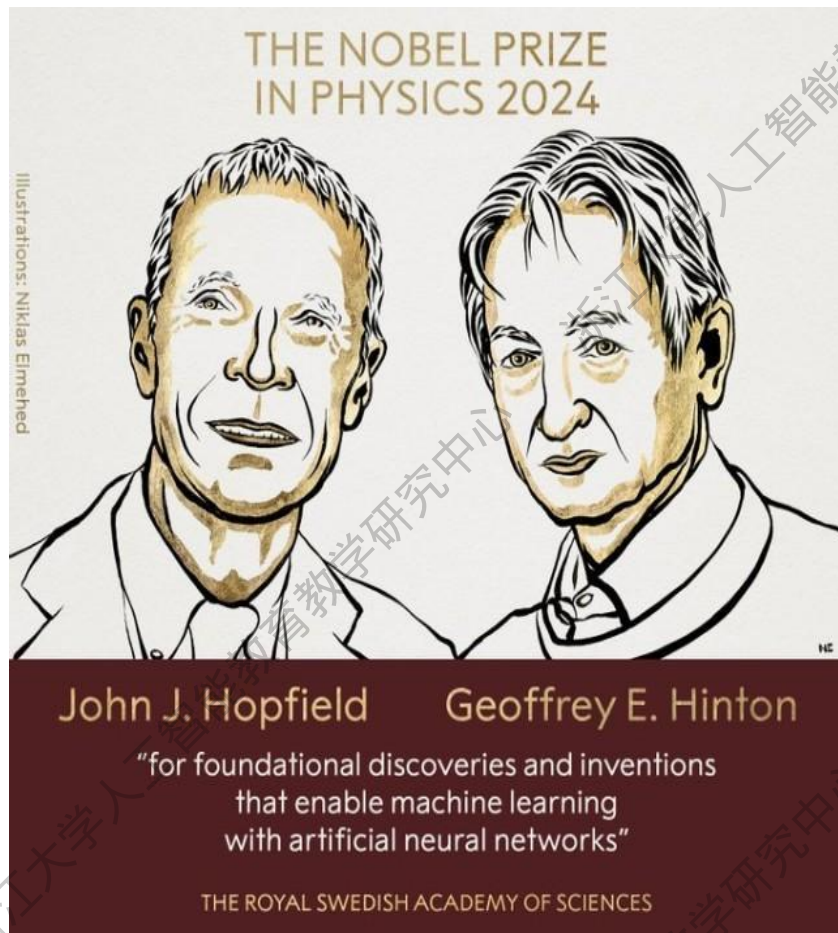


(c) Protein-peptide complex
TM-score = 96.6, DockQ = 0.954,



(d) A2B2 heteromer
TM-score = 98.5, $N_{\text{res}} = 716$, PDB ID = 6IWD

2024年诺贝尔物理学奖 (2024年10月8日)



- John Hopfield, Neural networks and physical systems with emergent collective computational abilities (《具有涌现集体计算能力的神经网络和物理系统》), 1982, PNAS

1949年赫布(Hebbian theory) 提出: 神经元之间持续重复经验刺激可导致突触传递效能增加(**Neurons that fire together, wire together**), 即学习和记忆的本质是大脑结构在后期经验刺激下改变神经元之间强弱联结的权重的结果。Hopfield以**能量最小**这一物理学视角优化神经网络而得到局部最优解。

- Ackley David H, Hinton Geoffrey E, Sejnowski Terrence J, A learning algorithm for Boltzmann machines, Cognitive science, Elsevier, 9 (1): 147-169, 1985

Boltzmann机能够达到**全局优化**原因在于基于统计物理所设计的Monte Carlo方法, 以一定的概率重新设置神经元的值, 选择的概率保证网络收敛于由**“能量”完全确定的热平衡态**。

从神经网络的生物学机制阐释到物理学可计算模拟

小结

- 蛋白质结构预测：存在近50年的科学难题
 - AlphaFold2 (DeepMind) 和RoseTTAFold (Baker实验室) 两个系统的开发，将蛋白质结构预测精度从之前的40-50%提高到90%以上
 - AI不只是大模型：专用AI系统
- AlphaFold是科学吗？
 - 尽管AlphaFold2可以精确预测结构，但它并不能解释蛋白质如何折叠或为何采取特定结构
 - 传统研究试图通过理解基本物理过程来解决蛋白质折叠问题，DeepMind通过关注结果而非过程，采取了不同路径。
 - 如果你能治愈癌症，你关心它是如何真正工作的吗？
- 蛋白质折叠只是“冰山一角”
 - 模拟整个细胞
 - 蛋白质复合物
 - 蛋白质设计
 - 蛋白质动力学
 - 蛋白质构象变化
 - 突变解释
 - ...

Not an end, but a Beginning:
The door is open to?:

- Protein complexes
- Accuracy estimation
- Protein design
- Protein dynamics
- Protein conformational change
- Preferred conformations of disordered proteins
- Mutation interpretation
- Ligand docking

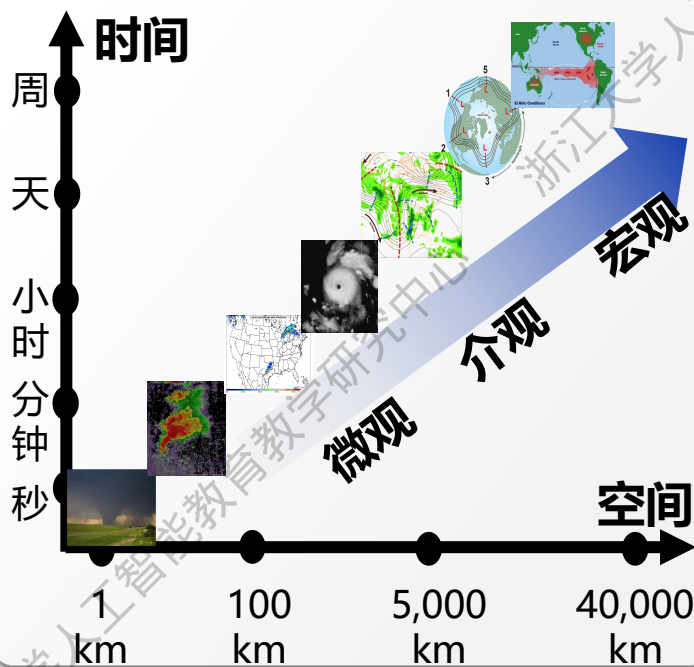




AI驱动科学研究的全过程

科学研究的主要挑战

多尺度



- 非线性耦合：不同尺度之间的相互作用并不是简单的线性叠加，而往往是复杂的、非线性耦合
- 某一尺度上的微小变化，可能会通过耦合机制对其他尺度产生巨大影响。

大数据/小数据

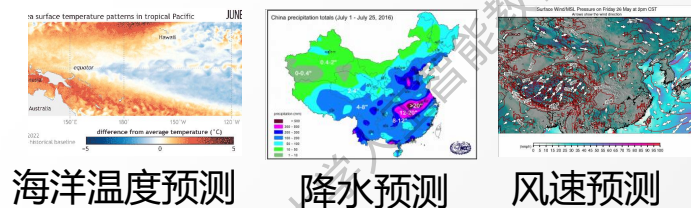
数据类型	数据量规模	主要来源/项目	特点
基因组测序数据	>20,000拍字节	人类基因组计划、千人基因组计划	单个人类基因组约200GB
蛋白质组学数据	>5拍字节	人类蛋白质组计划(HPP)	高复杂度、多维数据结构
医学影像数据	>2,000拍字节	医院、研究机构	单次MRI扫描可达1GB
电子健康记录	>500拍字节	医疗系统、保险公司	结构化与非结构化混合
微生物组数据	>100拍字节	人类微生物组计划	环境样本数量庞大
药物发现数据	>50拍字节	制药公司、研究机构	高通量筛选结果
单细胞测序	>200拍字节	人类细胞图谱计划	快速增长的新兴领域
生物医学文献	>10拍字节	PubMed、科研数据库	每年增加>100万篇文章
临床试验数据	>30拍字节	ClinicalTrials.gov	复杂的多中心、多时间点数据

PDB（蛋白质数据库）中的结构数量

- 1982年有100个结构
- 1993年有1,000个结构
- 1999年有10,000个
- 2014年有100,000个
- ...

- 数据量大、类型多，包括蛋白质序列、学术出版物、电子医疗记录、基因测序和影像等领域。每种类别都具有独特的属性
- 某些科学领域数据稀缺

任务多



- 环境科学领域
- 航空航天领域
- 医疗健康领域
- 材料科学领域
- 地质学与资源勘探领域
- ...

- 科学领域非常广泛，知识体系复杂
- 各学科发展出各自专业术语、研究方法和理论框架，导致跨学科交流困难
- 学科壁垒和专业分化使得研究人员难以全面掌握相关领域知识，解决综合性问题困难

从聊天出发: Artificial General Intelligence (AGI)



浙江大学
ZHEJIANG UNIVERSITY

Stage Level 5: Organizations,
AI that can do the work of an organization.

Stage Level 4: Innovators,
AI that can aid in invention.

Stage Level 3: Agents,
systems that can take actions.

Stage Level 2: Reasoners,
human-level problem solving.

Stage Level 1: Chatbots,
AI with conversational language.

1-5 级说明

L5: 可以完成整个组织工作的人工智能。

L4: 可以创造新事物的人工智能。

L3: 可以代表用户采取行动的人工智能代理。

L2: 像博士一样解决问题，无需工具。

L1: 当前AI，类似ChatGPT，可以与人类对话的人工智能。





AI驱动科学研究的全过程

知识：改变科学家获取和传递知识的方式

数据：生成、提取和标注大规模科学数据集

实验：模拟、加速并指导复杂实验

知识：改变科学家获取和传递知识的方式

● 挑战

- 科学知识呈指数级增长，专业化程度不断提高
- 跨学科合作需求增加，但知识壁垒阻碍学习与交流

● 自动文献管理与分析

- Semantic Scholar有超过2.14亿篇论文

● 图表理解与信息提取

- 结合图像、表格、公式和文本，分析复杂科学文献

● 代码辅助生成与优化

- 使用大模型将难维护的代码进行转换与注释

● 推理->科研助手 (Science Copilot)

- 陶哲轩：AI 在数学和编程领域变得越来越有用，但人类的洞察力和创造力仍是关键

- <https://www.bilibili.com/video/BV1c946eVEPu/>

量子力学中哈密顿量与电子密度是什么关系？



在量子力学中，哈密顿量 (Hamiltonian) 和电子密度 (electron density) 之间的关系可以通过密度泛函理论 (Density Functional Theory, DFT) 的核心思想来理解。以下是两者的关键联系：

1. 传统量子力学中的哈密顿量与电子密度

- 哈密顿量 ($\hat{H}$) 是描述量子系统总能量的算符，通常包括动能项、电子与核的吸引势 (外势)，以及电子间的排斥势：

$$\hat{H} = -\frac{\hbar^2}{2m} \sum_i \nabla_i^2 + \sum_i V_{\text{ext}}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|},$$

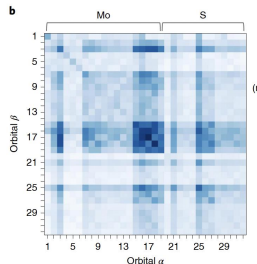
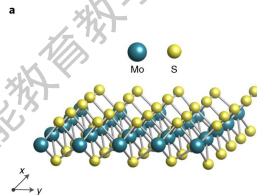
其中 V_{ext} 是外势 (如原子核产生的库仑势)。

- 电子密度 $n(\mathbf{r})$ 是描述电子在空间分布的标量场，定义为：

$$n(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_2 \dots d\mathbf{r}_N,$$

其中 Ψ 是系统的多体波函数。

在传统量子力学中，哈密顿量决定了系统的波函数，进而通过波函数计算电子密度。然而，对于多电子体系，直接求解多体波函数极其复杂。



这个图什么含义？



上半部分 (a)：晶体结构

内容描述

- 图中显示了一个二维晶体结构，由两种原子组成：
 - 蓝色球代表钼 (Mo) 原子。
 - 黄色球代表硫 (S) 原子。
- 原子通过化学键连接，形成了一个六边形晶格结构。这种结构类似于过渡金属二硫化物。
- 晶体结构在平面内延伸，具有周期性。
- 图中标注了坐标轴和晶格常数。

含义

- 上半部分 (a) 展示了二维材料 (如 MoS_2) 的晶体结构，强调了其六边形晶格和原子排列。

这是一个典型

六边形晶格是

应用场景

这种结构在电

- 下半部分 (b) 展示了轨道相互作用矩阵，反映了钼和硫原子轨道之间的耦合强度，这对于理解材料的电子结构和能带特性至关重要。
- 这类图常见于材料科学和凝聚态物理学的研究中，特别是在研究二维材料 (如 TMDs) 的电子性质、能带结构和器件应用时。
- 轨道相互作用矩阵可以帮助研究人员预测材料的导电性、光学响应和量子效应等。

总结

这张图的整体含义是：

1. 上半部分 (a)：展示了二维材料 (如 MoS_2) 的晶体结构，体现了其原子排列和周期性。
2. 下半部分 (b)：展示了轨道相互作用矩阵，反映了钼和硫原子轨道之间的耦合强度，对于理解材料的电子结构和能带特性具有重要意义。

下半部分 (b)：矩阵图

内容描述

- 这是一个矩阵图，横轴表示 "Orbital (alpha)" (轨道 (alpha))，纵轴表示 "Orbital (beta)" (轨道 (beta))。
- 矩阵中的每个单元格的数值表示耦合强度，右侧的颜色条说明了颜色与数值的对应关系：
 - 蓝色越深，数值越大 (单位为 meV，毫电子伏特)。
 - 白色表示数值接近零。
- 矩阵被分为两部分：
 - 左侧标注为 "Mo"，右侧标注为 "S"。
 - 这表示矩阵的行和列分别对应钼原子和硫原子的轨道。

能量 (单位为 meV)。

展示了不同原子类型之间或同一原子类型内部的轨道相

域 (接近零) 表明相互作用较弱或不存。

挑战一：如何实现复杂科学任务的高效推理？

专业领域知识迭代频繁， 模型难以及时学习新推理链条

介绍今年GaN充电器发展



深度思考 (用时 23 秒)

嗯，用户让我介绍今年GaN充电器的发展情况。首先，我得回忆一下GaN是什么，全称是氮化镓，对吧？这种材料在充电器中的应用应该和它的特性有关，比如高效率和小体积。不过，用户可能更关注的是今年的进展，所以需要聚焦在2023年的新技术和市场变化。

首先，我应该考虑用户的背景。他们可能是普通消费者，想了解是否值得升级到GaN充电器，或者是科技爱好者，寻求更详细的技术细节。也有可能是一些业内人士，需要市场趋势分析。不过问题比较简短，没有特别说明，所以需要兼顾不同层次的信息。

接下来，我需要收集今年GaN充电器的主要发展点。比如，功率提升方面，可能有多口高功率产品，比如200W以上的充电器，支持多个设备同时快充。体积和散热技术的改进，可能涉及到新的结构设计或散热材料，使得大功率下仍保持小巧。快充协议的支持，比如PD 3.1，可能让充电器兼容更多设备，尤其是笔记本电脑。此外，价格下降可能让GaN充电器更普及，市场渗透率提高。

训练数据更新慢：基于2023年12月前数据的DeepSeek R1无法获取最新知识，导致在介绍“GaN充电器发展”时，无法结合更新的行业知识。

更新知识成本高：模型需要额外的训练学习更新的内容，需要额外的算力资源和时间成本。

挑战1.1：
数据更新快

行业知识更新快

求是创新

推理结论多样且非唯一最优， 难以构建精准的奖励函数评估推理结论质量

问题：患者，男，45岁，出现胸痛、气短和左臂麻木，有高血压和糖尿病史。

思维链：

1. 首先分析症状组合：胸痛、气短和左臂麻木是心脏问题的典型信号，特别是左臂麻木常见于冠状动脉疾病
2. 评估风险因素：患者有高血压和糖尿病史，这两种慢性病都是冠心病的高风险因素
3. 鉴别诊断考虑：
4. 需要立即的检查：心电图、心肌酶谱、胸部影像
5. 初步诊断：急性冠脉综合征可能性最大，需要紧急治疗方案

答案：最可能是急性冠脉综合征，建议立即进行心电图检查和心肌酶谱检测，同时启动胸痛中心流程。

医疗诊断示例

医疗诊断的正确性难以给出基于规则的评估标准，无法直接由程序自动化评估。

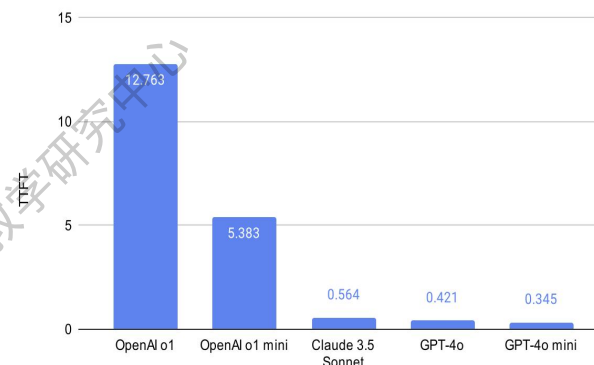
挑战1.2：
奖励难定义

推理结论评估困难，
难以定义奖励标准

推理模型消耗大量思考时间， 计算资源需求大

推理模型在输出答案之前需要花费大量时间思考，导致实时性差。并且难度越大的问题思考时间越长，因此消耗更多的推理资源。

Time to first token (time in seconds)



挑战1.3：
推理成本高

复杂推理资源
消耗大

挑战二：如何实现复杂科研流程的智能执行

OpenAI Deep Research （2025年2月2日，目前仅供 Pro 订阅用户使用，每月 200 美元）

Deep research以OpenAI o3推理模型作为支撑，能够独立完成**联网资料查询**，海量信息处理，综合报告生成。依托其强大的搜索分析能力，在金融、政策、工程等领域都能提供可靠的帮助。



帮我查找过去10年，GDP前十的发达国家的iOS和Android市场占有率，将这些信息整理成表格。

GDP排名前10的发达经济体手机OS市场份额

国家	iOS市场占有率	Android市场占有率
美国	58%	42%
日本	69%	31%
...	...	...



Deep research在Humanity' s Last Exam数据集上和其他模型的对比

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

*模型不是多模态的，仅在文本子集上进行评估

**联网检索信息+python工具

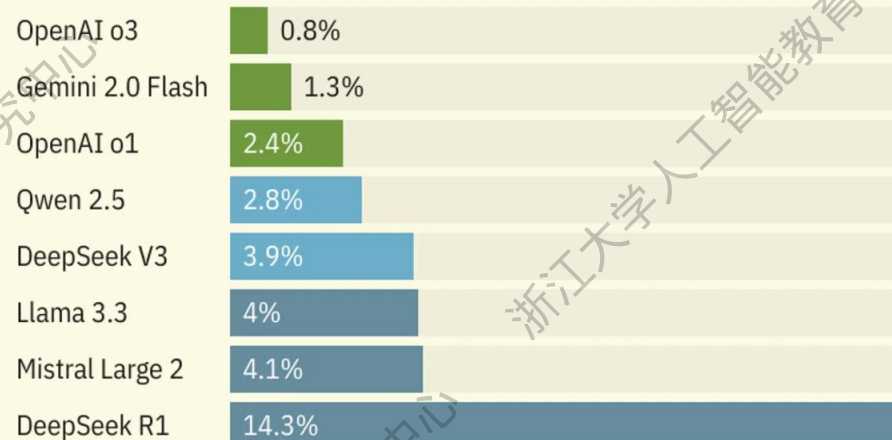
- 推理模型OpenAI o1和DeepSeek-R1相比GPT-4o等，具有更高的性能
- Deep research以OpenAI o3作为基础，通过合理的工作流在复杂任务上取得了当前最好性能

挑战三：如何抑制推理大模型的幻觉现象

当前推理大模型知识边界不确定，难以把握知识覆盖范围和可靠程度，推理幻觉严重

Hallucination rate of AI models:

Reasoning-enhanced Open-sourced Both



DeepSeek R1的
幻觉问题相对严重

问：8,8,3,4算24点

已深度思考（用时 312 秒） 嗯，今天老师布置了一个算24点的题目，四个数字是8、8、3、4。我得好好想想怎么用这四个数字通过加减乘除来得到24。首先，我应该回忆一下算24点的基本方法，通常是先尝试用乘法或者除法来得到较大的数，或者组合出能够相互抵消的情况，比如 $3 \times 8 = 24$ 这样的组合。不过这里有两个8，一个3和一个4，可能需要结合其他运算。

首先，我先把这四个数字列出来：8、8、3、4。我需要用每个数字各一次，通过加减乘除以及可能的括号来组合它们得到24。让我尝试不同的组合方式。

首先，我注意到如果有两个8的话，或许可以先用这两个8进行某种运算。比如 $8 + 8 = 16$ ，不过这样的话剩下的3和4要怎么办呢？16加上3和4的话， $16 + 3 + 4 = 23$ ，不够。或者16乘以 $(4 - 3) = 16 \times 1 = 16$ ，还是不够。或者16乘以3再除以4的话， $16 \times 3 = 48$ ， $48 \div 4 = 12$ ，还是不够。或者 $16 \times (4 + 3) / \text{某个数}$ ，但这样可能也不太行。

.....

或者，用减法来调整，比如 $8 \times 4 = 32$ ，然后用剩下的8和3来减，比如 $32 - 8 = 24$ ，但这样用了三个数字，剩下的3没用上。或者 $32 - (8 - 3) = 32 - 5 = 27$ ，也不对。

再试试其他的组合方式，比如用8、8、4这三个数先组合，比如 $8 + 8 + 4 = 20$ ，然后 $20 + 3 = 23$ ，不够。或者 $8 \times 8 = 64$ ， $64 \div 4 = 16$ ，然后 $16 + 3 = 19$ ，还是不够。

或者考虑用除法来结合，例如， $(8 \times (4 - 3)) \times 8$ ，这样就是 $8 \times 1 \times 8 = 64$ ，不对。或者 $8 \times (4 - (8/3))$ ，不过这样会涉及到分数，可能比较复杂，而且结果可能不是整数。比如 $8 \times (4 - 8/3) = 8 \times (4/1 - 8/3) = 8 \times (12/3 - 8/3) = 8 \times (4/3) = 32/3 \approx 10.666$ ，不是24。

有没有可能用加减乘除混合运算？比如， $8 \times 3 = 24$ ，然后剩下的8和4如何组合成0，这样 $24 + 0 = 24$ 。但如何用8和4得到0呢？比如 $8 - 8 = 0$ ，但这里已经用了两个8，剩下的还有3和4。不对，因为题目中每个数字只能用一次。所以，这样的话，可能不行。

再想想，可能用减法来调整。例如， $8 \times 4 = 32$ ，然后32减去 $(8 - 3) = 5$ ，得到 $32 - 5 = 27$ ，还是不对。或者 $8 \times 3 = 24$ ，然后用剩下的8和4组成0，但同样的问题，无法用剩下的两个数得到0。

....

每个数字均被使用一次，符合规则。最终答案为：

$$8 * 3 * (8 / 4 - (4 - 3)) = 24$$

思考长度达6672，
最终答案仍有错误



AI驱动科学研究的全过程

知识：改变科学家获取和传递知识的方式

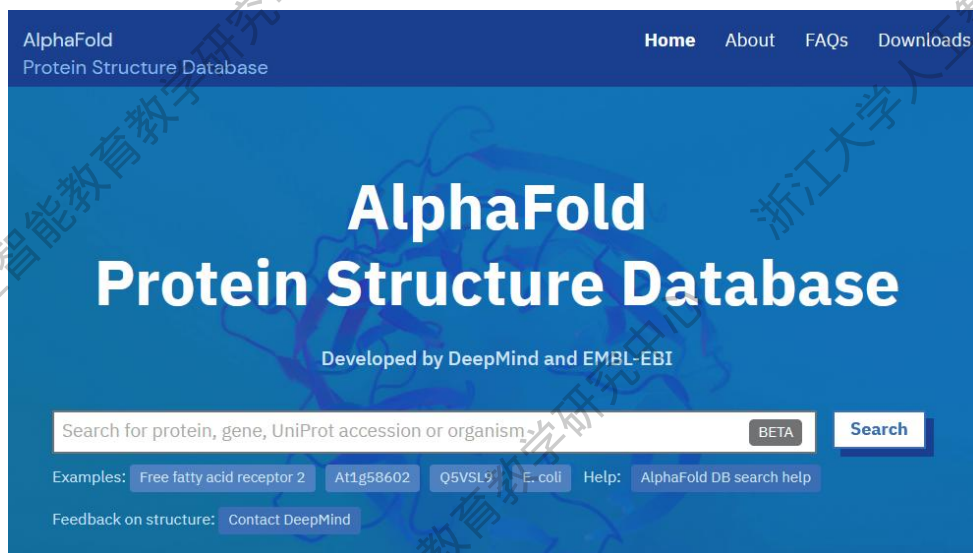
数据：生成、提取和标注大规模科学数据集

实验：模拟、加速并指导复杂实验

数据：生成、提取和标注大规模科学数据集

● AlphaFold DB

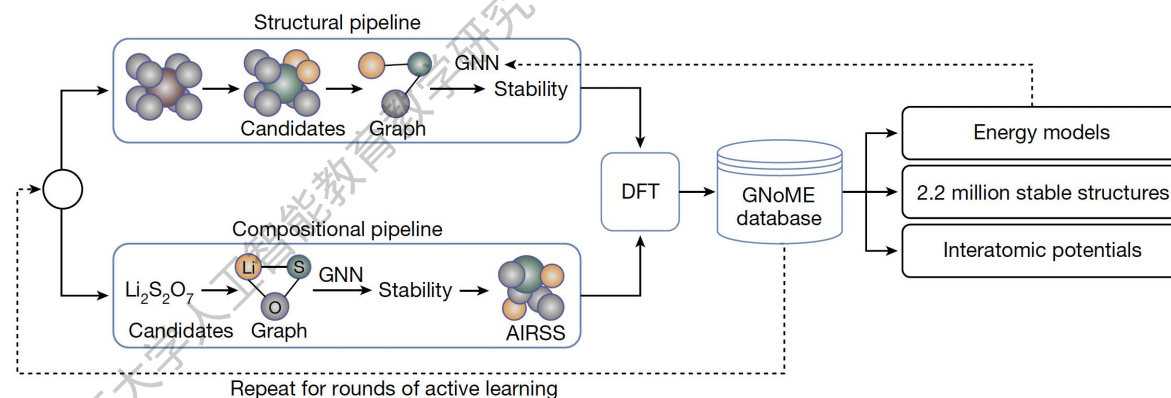
- 由DeepMind与欧洲生物信息学研究所合作开发的大型蛋白质结构数据库
- 超过200万用户来自190个国家使用
- 高准确度：提供的结构预测在多数情况下接近实验确定的结构
- 包含超过2.14亿个蛋白质结构



● 基于主动学习的数据清洗与合成¹

1. 基于已知的稳定材料生成候选结构
2. 对候选结构进行筛选
3. 基于密度泛函理论验证稳定性
4. 将新材料放到GNoME数据集，重复迭代

● 提供了超过220万种无机材料

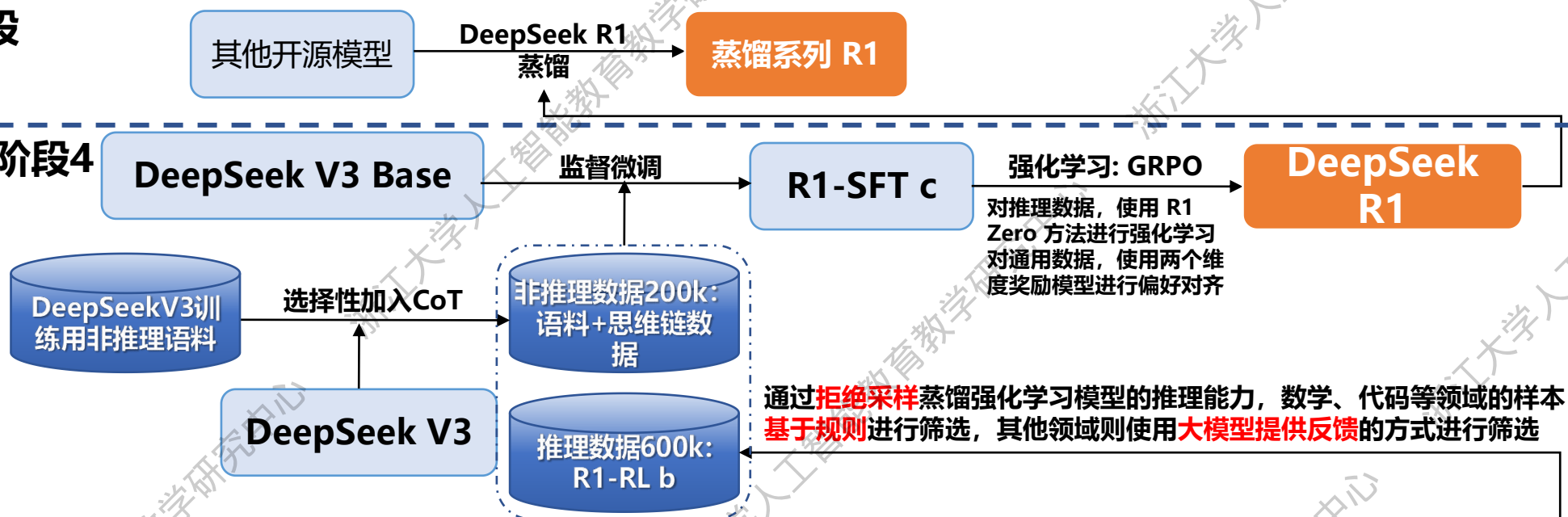


[1] Scaling deep learning for materials discovery, Nature 2023

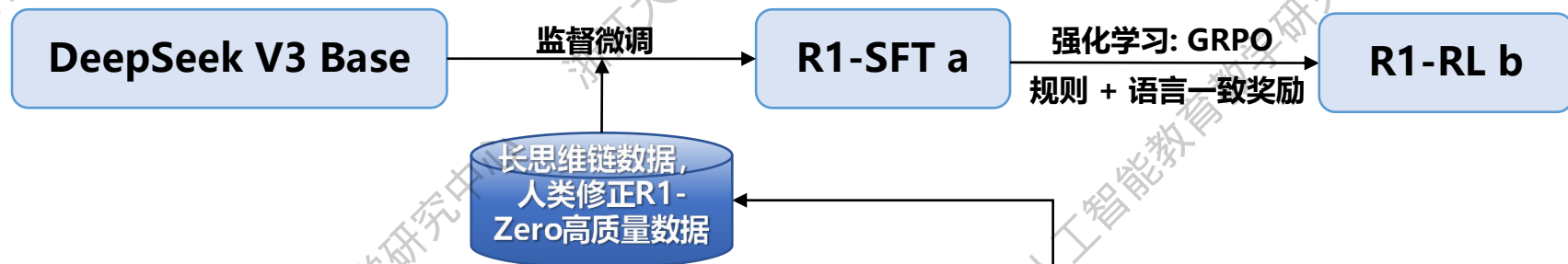
数据：生成、提取和标注大规模科学数据集

蒸馏阶段

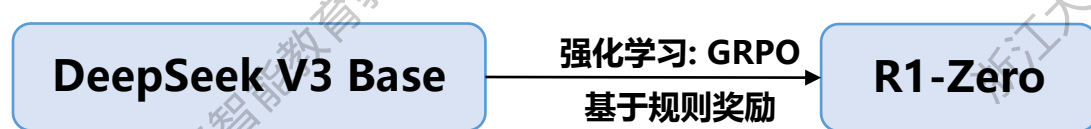
阶段3+阶段4



阶段1+阶段2



阶段0



R1

V3



AI驱动科学研究的全过程

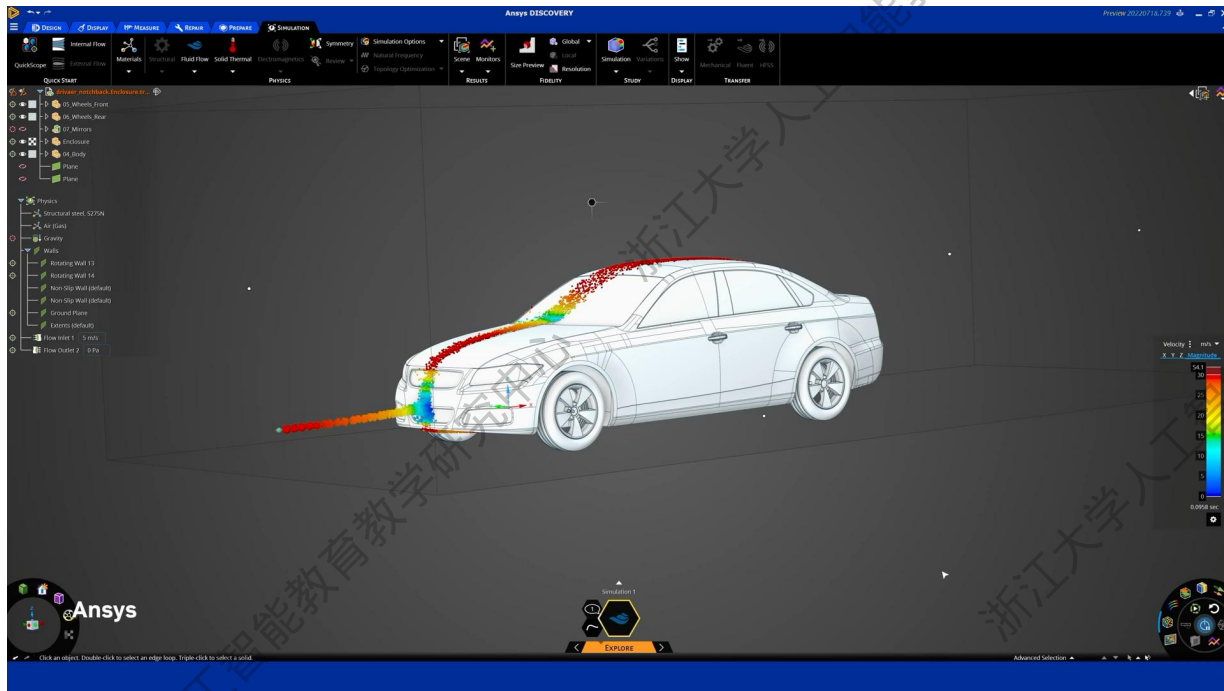
知识：改变科学家获取和传递知识的方式

数据：生成、提取和标注大规模科学数据集

实验：模拟、加速并指导复杂实验

实验：模拟、加速并指导复杂实验

AI通过仿真和自动化加速并指导实验过程，显著提高工程设计和科学发现的效率



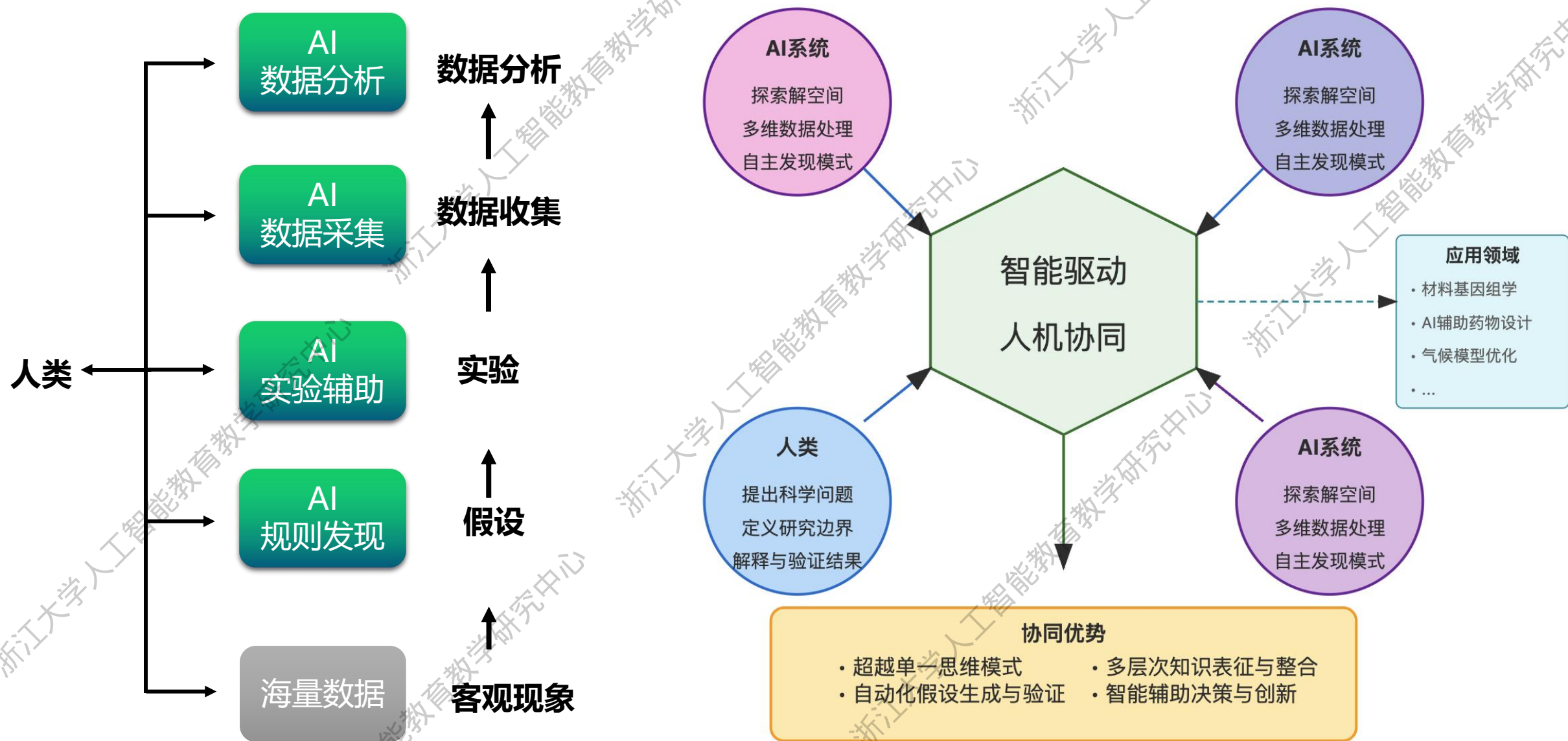
2024年11月，英伟达发布实时仿真平台，实现了实时**虚拟风洞**，通过简化CFD仿真的创建过程，以推进汽车、飞机、船舶等产品的设计优化。



2020 年登上《Nature》封面的 **AI 化学家**，利用实验室自动化和人工智能的系统，8 天完成 688 次实验，一周时间研究了 1000 种催化配方，并发现了一种新的催化剂。

A mobile robotic chemist, Nature 2020

智能驱动，人机协同



- **定位“基本”问题**
- **科学研究范式的演进**
 - 第一范式：经验驱动 — 基于观察和归纳的实验研究
 - 第二范式：理论驱动 — 基于科学假设和逻辑演绎的理论研究，探索第一性原理
 - 第三范式：计算驱动 — 以计算和仿真为主导，利用高性能计算机求解复杂问题
 - 第四范式：数据驱动 — 数据密集型研究方法，利用海量数据和机器学习技术
 - 第五范式：智能驱动 — AI参与科学研究全过程，整合前四种范式的优势
- **蛋白质结构预测：AI for Science的成功案例**
- **AI驱动科学研究的全过程**
 - 知识获取与传递 — 自动文献管理、图表理解、跨学科知识整合
 - 数据生成与处理 — 创建大规模科学数据集、智能数据清洗与合成
 - 实验模拟与指导 — 实时仿真加速设计、自动化实验系统
 - 模型构建与预测 — 处理多尺度、多物理场复杂系统建模
- **人机协同的未来科学研究模式**
 - AI协同处理大规模数据和复杂计算，人类提出问题并解释验证结果
 - 超越单一思维模式，实现多层次知识整合
 - AI不仅是工具，而是科学发现全过程的参与者和加速器



谢谢！ 敬请批评指正！

朱霖潮

浙江大学计算机科学与技术学院人工智能研究所
zhulinchao@zju.edu.cn