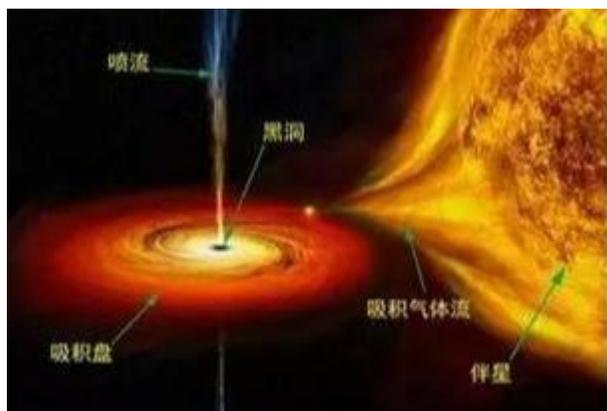


# 人工智能在天文学中的应用

张彦霞

中国科学院国家天文台

2023.6.13



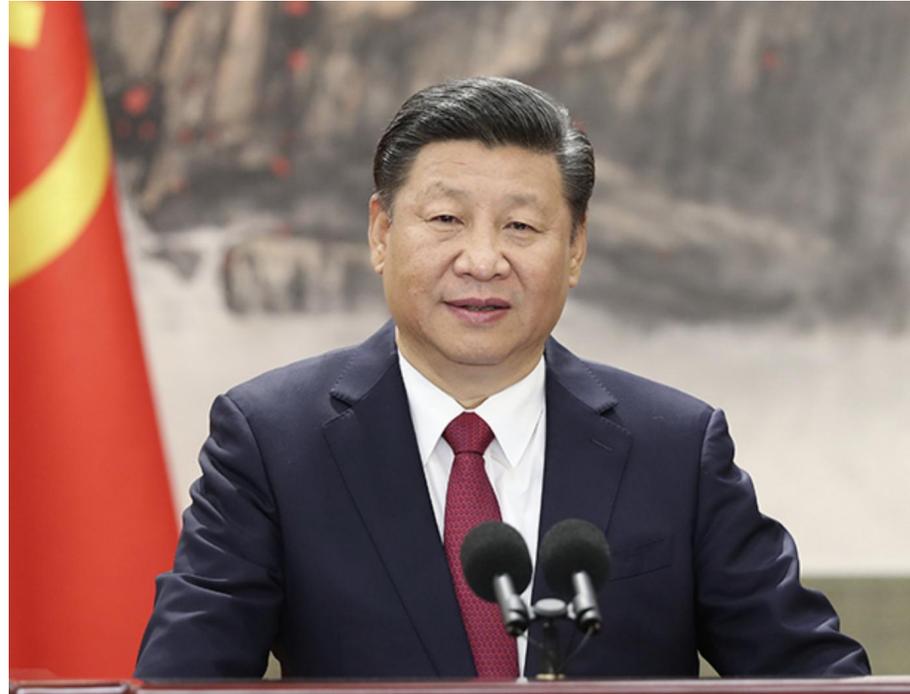
# 报告提纲

---

- AI背景
- 天文大数据
- 机器学习
  - 概念
  - 分类
  - 任务
  - 应用
  - 挑战
- 展望

2022年习近平总书记指出，“把新一代**人工智能**作为推动科技跨越发展、产业优化升级、生产力整体跃升的驱动力量，努力实现高质量发展”。党的十八大以来，以习近平同志为核心的党中央高度重视智能经济发展，促进人工智能和实体经济深度融合，为高质量发展注入强劲动力。

2018年9月17日世界人工智能大会在上海开幕，习近平致信祝贺。“新一代**人工智能**正在全球范围内蓬勃兴起，为经济社会发展注入了新动能，正在深刻改变人们的生产生活方式。”习近平在贺信中强调，中国正致力于实现高质量发展，人工智能发展应用将有力提高经济社会发展智能化水平，有效增强公共服务和城市管理能力。



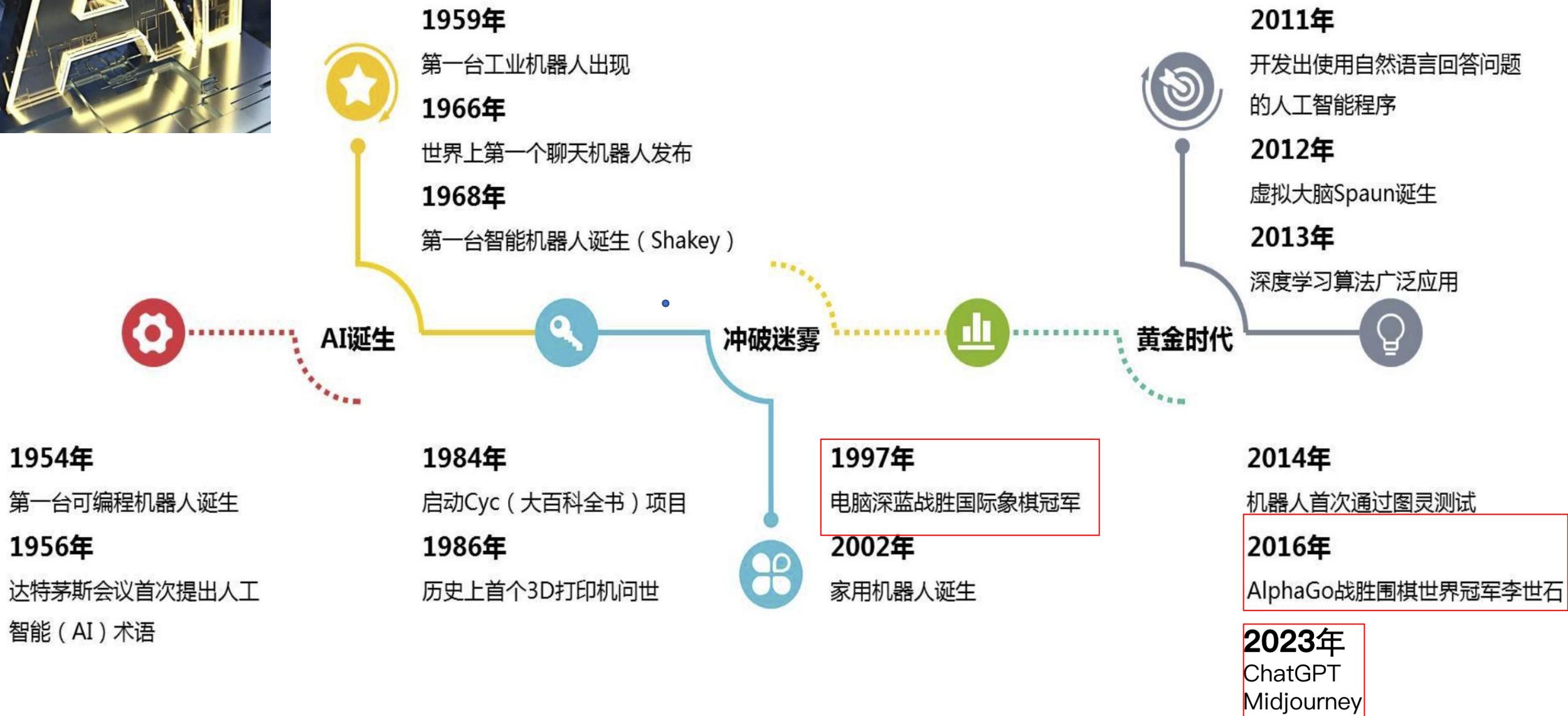
在危机中育新机，  
变局中开新局

习近平总书记强调，**人工智能**是新一轮科技革命和产业变革的重要驱动力量，加快发展新一代人工智能是事关我国能否抓住新一轮科技革命和产业变革机遇的战略问题。要深刻认识加快发展新一代人工智能的重大意义，加强领导，做好规划，明确任务，夯实基础，促进其同经济社会发展深度融合，推动我国新一代人工智能健康发展。（2023年3月15日人民日报）



# 人工智能的前世今生

来源：通联数据整理



## 第一次人工智能浪潮



**20世纪  
60~70年代**

- 60年代：因“推理和探索”时取得重大进展而繁荣。
- 70年代：因“推理和探索”对现实问题束手无策而衰落。

## 第二次人工智能浪潮



**20世纪  
80~90年代**

- 因导入知识使机器变得更聪明而繁盛。
- 因知识描述和管理的能力低下而缺陷暴露。

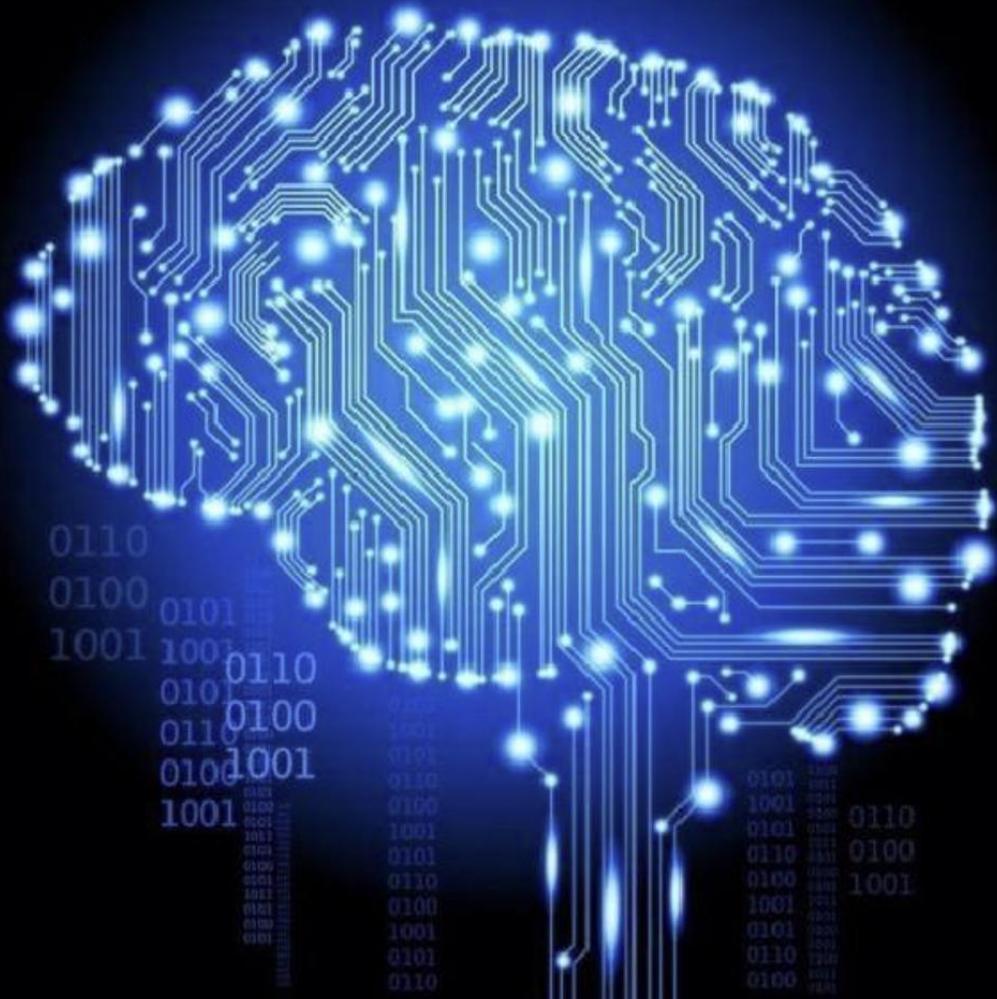
## 第三次人工智能浪潮



**21世纪  
至今**

- 三大引擎：深度学习、大数据、超强运算能力。

# ★ · 人工智能的定义



人工智能（Artificial Intelligence），英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来，理论和技术日益成熟，应用领域也不断扩大，可以设想，未来人工智能带来的科技产品，将会是人类智慧的“容器”。人工智能是对人的意识、思维的信息过程的模拟。人工智能不是人的智能，但能像人那样思考、也可能超过人的智能。



Artificial Intelligence

*能为我们做什么？*

# 人工智能

## 让机器像人一样思考

- 机器学习
- 自动推理
- 人工意识
- 知识展示
- .....

## 让机器像人一样听懂

- 语音识别



## 让机器像人一样看懂

- 视觉识别

## 让机器像人一样运动

- 运动控制

人工智能是一门前沿的综合学科

计算机科学

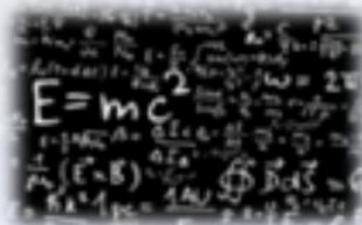


识别



人工智能可以代替人类实现多种功能

统计学



认知



脑神经学



分析



社会科学



决策



# 人工智能产业链

来源：通联数据整理



颠覆性  
涌现性  
工程化  
通用性  
密集型

# ChatGPT 能做的49件事

问&答	关键字提取	机器学习机器人	产品取名
内容概况	广告设计	文本情绪分析	程序语言转换
程序命令生成	句子简化	生成SQL语句	程序文档生成
Stripe国际API生成	颜色生成	修复代码Bug	代码压缩
结构化生成	段落创作	语言聊天机器人	人称转换
语法纠正	故事创作	清单制作	头脑风暴
生成OpenAi的代码	摘要说明	航空代码抽取	ESRB文本分类
语言翻译	好友聊天	抽取联系信息	点评生成
SQL语句生成	美食制作	文字转表情符号	面试
信息分类	摆烂聊天	程序代码翻译	知识学习
Python代码解释	提纲生成	代码解释	分解步骤
时间复杂度计算	AI聊天	问题解答	高级情绪评分
高级情绪评分	表格填充数据		

本质差距：  
创新机制  
创新生态  
创新文化

有志者，事竟成，破釜沉舟，百二秦关终属楚；  
苦心人，天不负，卧薪尝胆，三千越甲可吞吴。

——清代·蒲松龄·斋联

人工智能将引领下一波计算浪潮。与之前的重大转型类似，人工智能将构建出更加美好的时间

# 人工智能时代 已经来临

人工智能的潜力将大大激发，为企业和社会创造更积极的影响

# 人工智能的分类及相关介绍

## 弱人工智能

弱人工智能也称限制领域人工智能或者应用型人工智能。是指专注于且只能解决特定领域问题的人工智能。目前我们看到的所有人工智能算法和应用都属于弱人工智能的范畴，AlphaGo就是弱人工智能的一个最好实例。

## 强人工智能

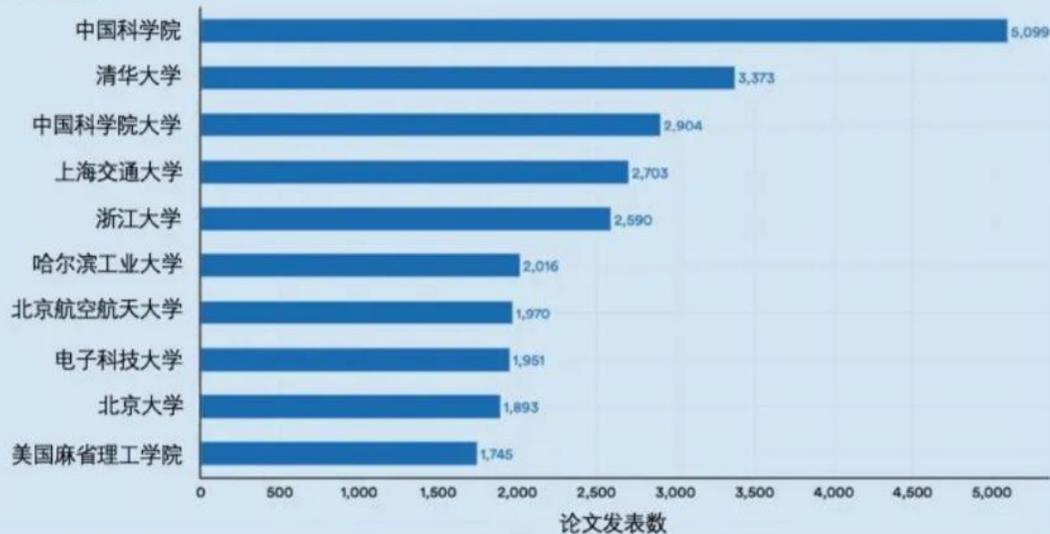
强人工智能又称通用人工智能或完全人工智能，指的是可以胜任人类所有工作的人工智能。其一般认为是有知觉的，有自我意识的，可以独立思考问题并制定解决问题的最优方案，有自己的价值观和世界观体系。

## 超人工智能

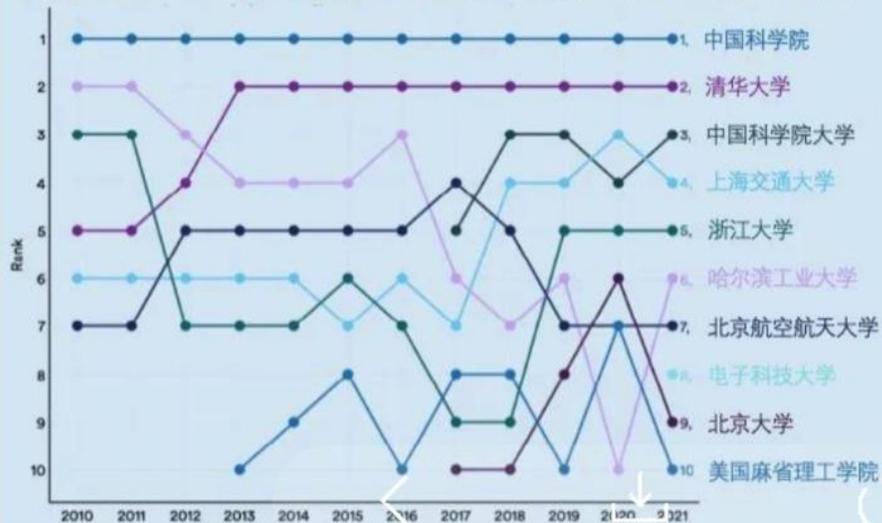
假设计算机程序不断发展，可以比世界上最聪明的人类还聪明，那么由此产生的人工智能系统可以被称为超人工智能。目前，以我们的科技水平，还远远达不到这种程度，因此，对其的定义也十分模糊。

## 中国：AI 论文大户

2021 年，AI 论文发表量全球 TOP 10 机构中，**中国机构占据 9 席**，美国麻省理工学院排名第十。



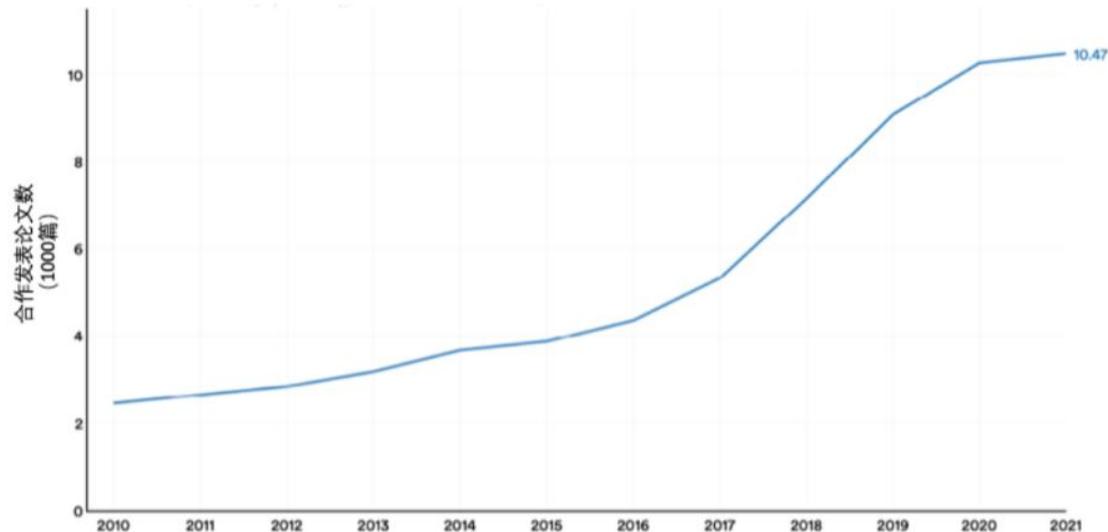
从 2010 年起，**中国科学院**就一直占据 AI 论文发表量世界第一的宝座。



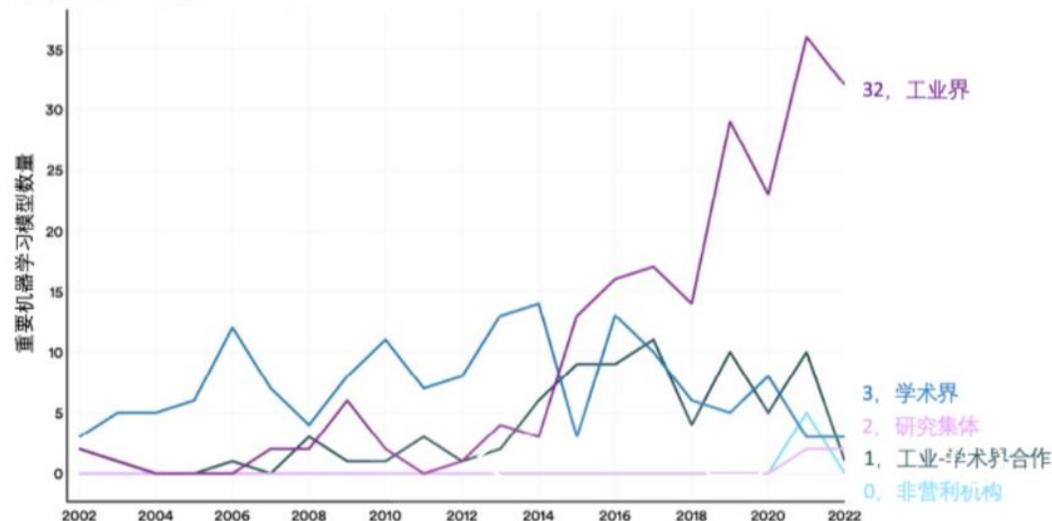
来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

## 中美合作放缓 学界与业界差距加大

中美 AI 合作研究数量持续上升，但**增速放缓**。

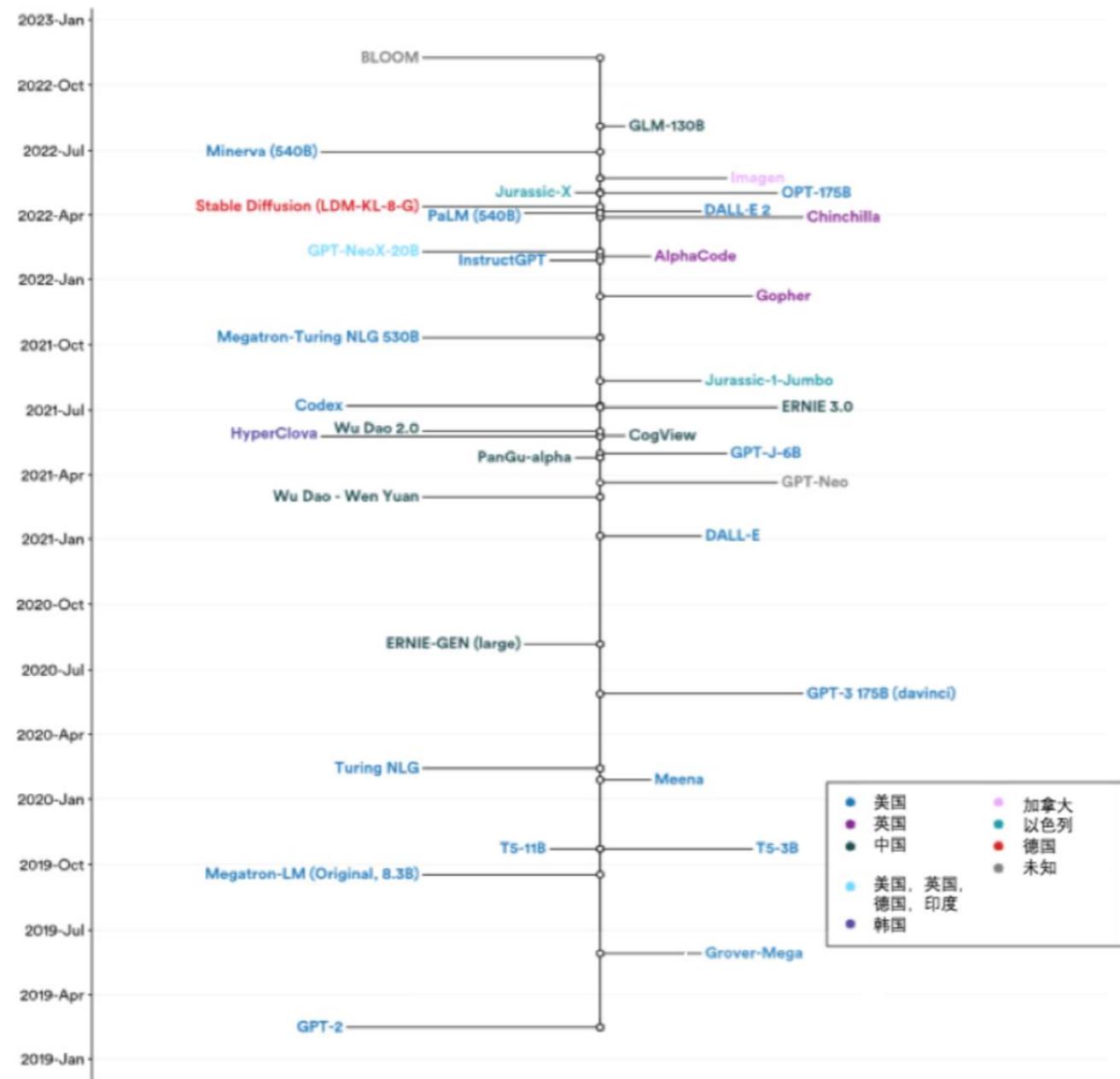


工业界领先于学术界：2022 年，**工业界贡献了 32 个重要机器学习模型**，而学术界只创造了 3 个。



来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

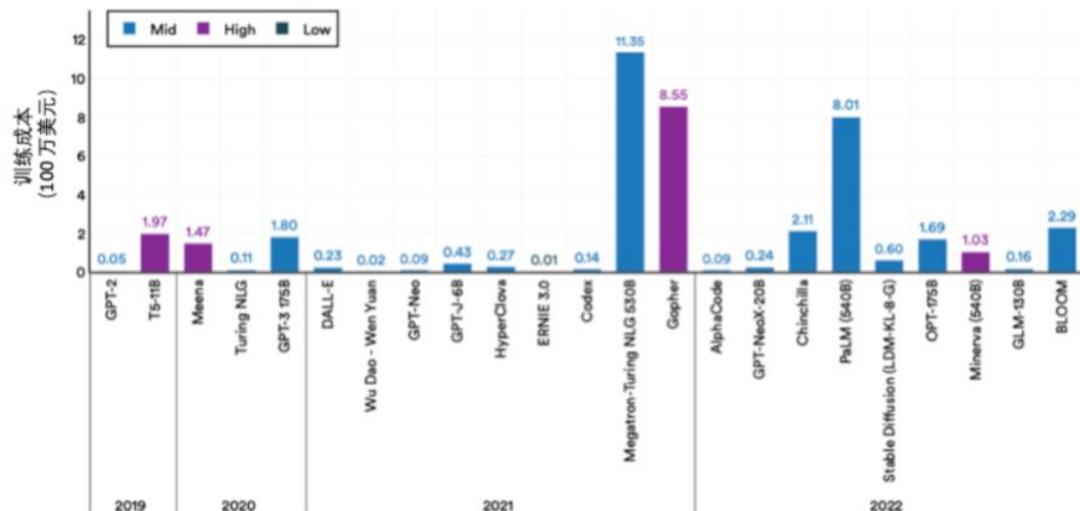
2021 年起，大型语言和多模态模型的“玩家”越来越多。



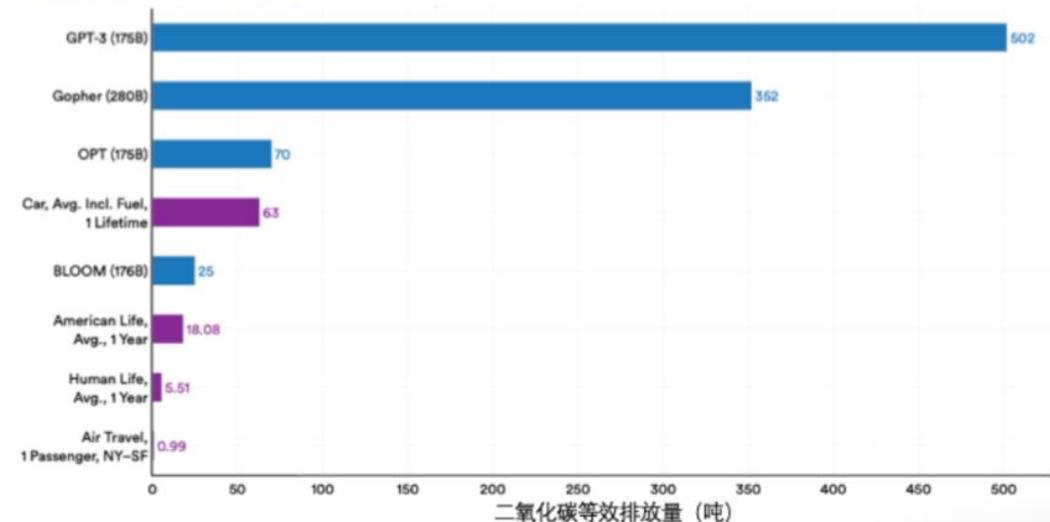
来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

## 越来越大，越来越烧钱

2022 年发布的 PaLM 有 5400 亿参数，成本估计为 800 万美元——参数比 3 年前发布的 GPT-2 大了 360 倍，预估成本高了 160 倍。



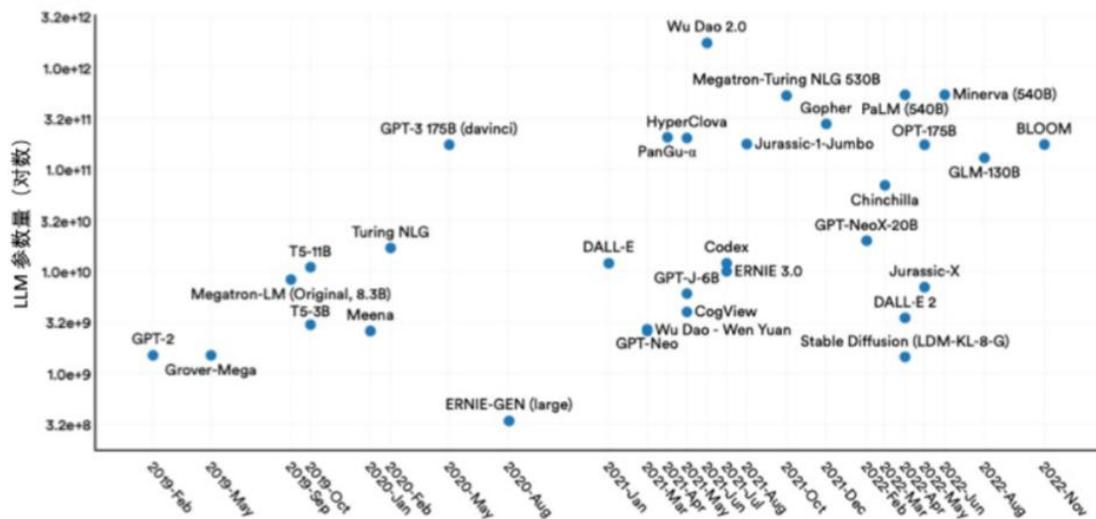
GPT-3 一年的二氧化碳排放量达到了 502 吨，相当于一名旅客从纽约飞往旧金山产生排放量的 500 多倍。



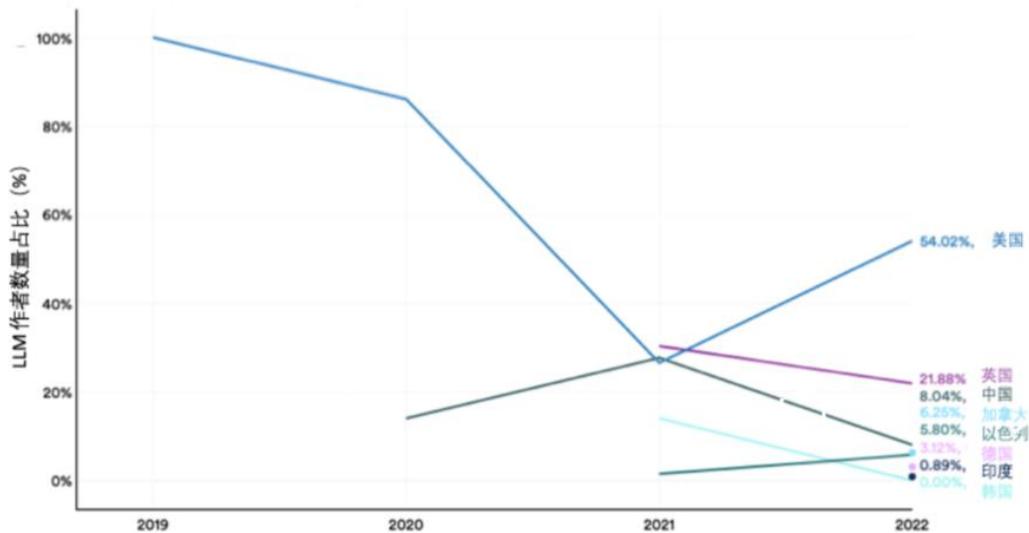
来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

## 谁在领先？

大型语言模型的体量竞争：中国的“悟道 2.0”（Wu Dao 2.0）是目前参数量最大的 LLM。



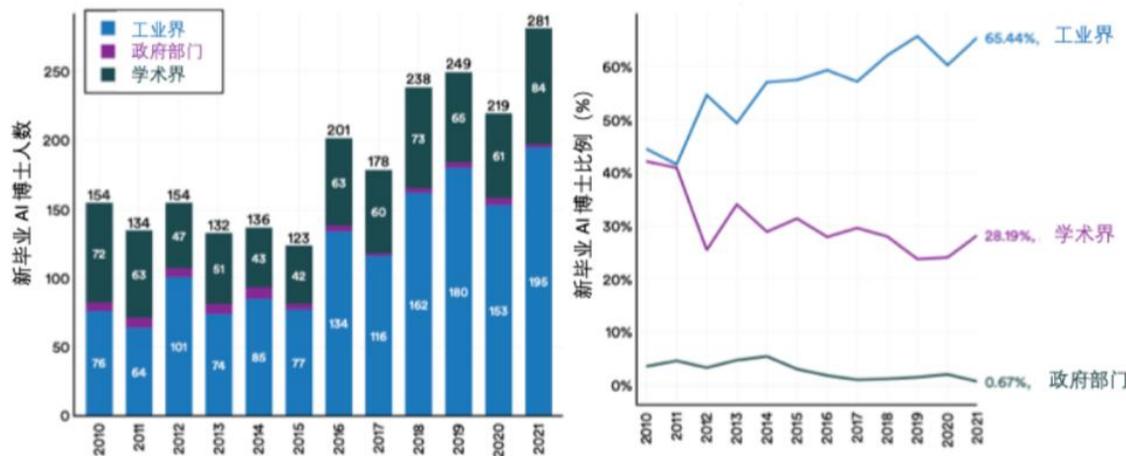
该领域来自美国的论文作者最多（占总数的 54.2%），中国作者数量在美国和英国之后。



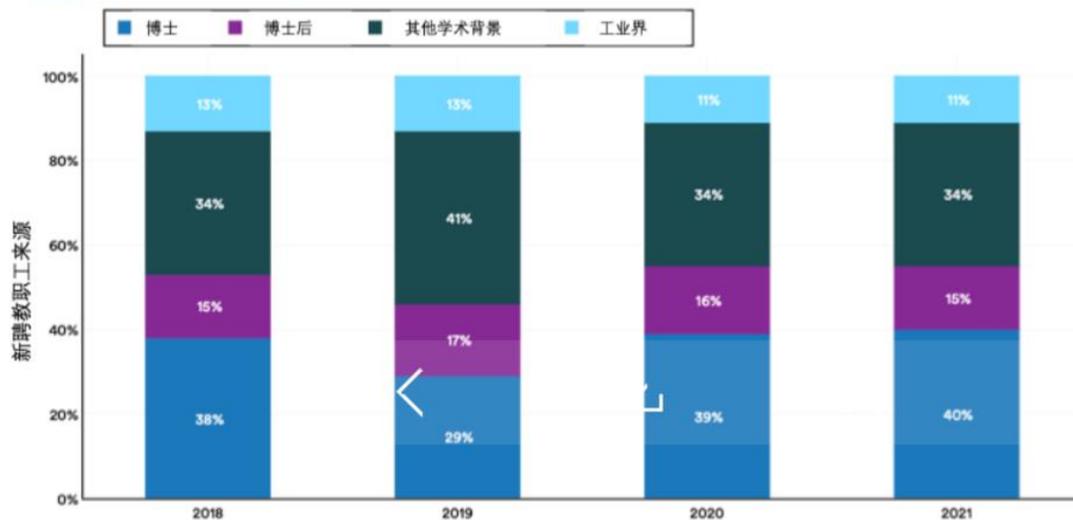
来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

## 学术界对博士的吸引力不足

越来越多的博士选择了工业界。2011 年，北美 AI 专业博士毕业生在工业界（40.9%）和学术界（41.6%）的工作比例大致相同。2021 年，65.4% 的人在工业界工作，比在学术界工作的 28.2% 多一倍。



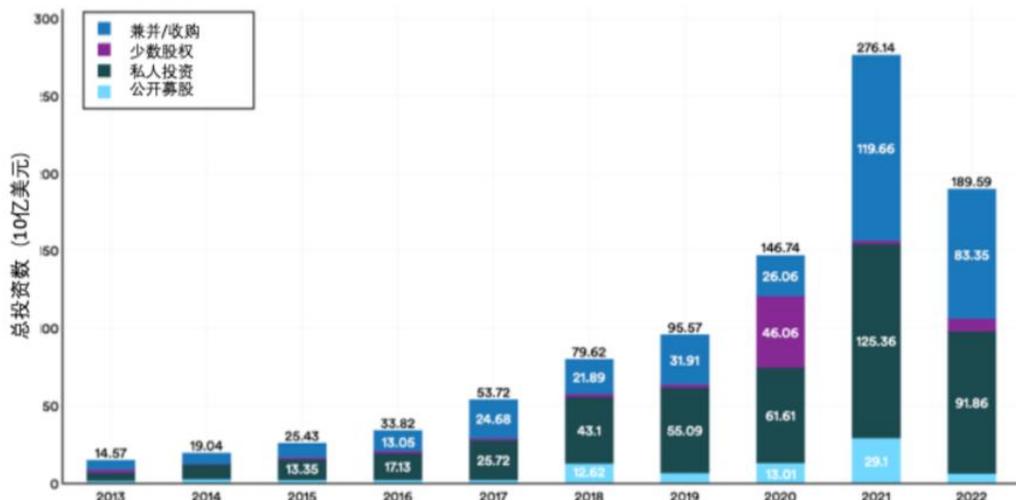
2021 年，北美计算机和信息领域新聘用的教职工有 40% 是没有博士后经历的毕业生，仅有约十分之一来自工业界。



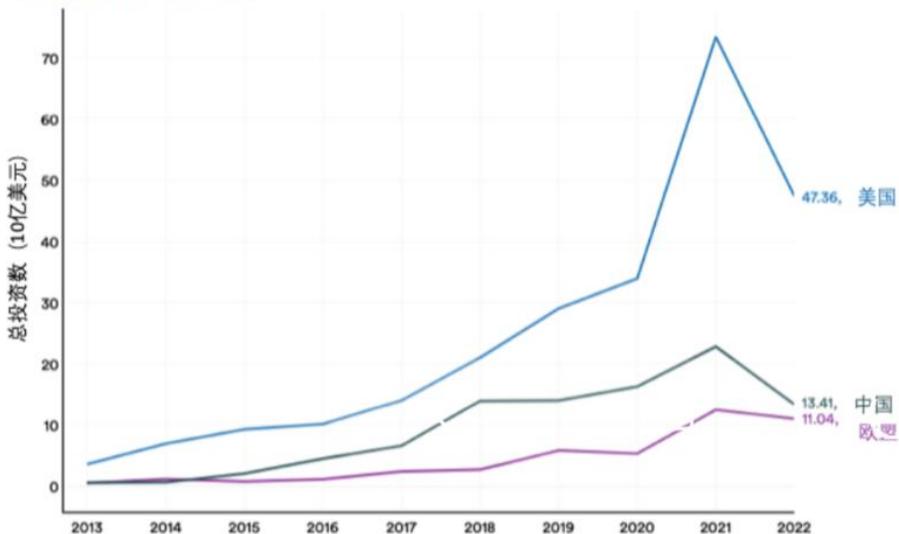
来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

# 谁在投资？

私人投资是全球 AI 投资的主要力量之一。2022 年的私人投资总量为 **919 亿美元**，较 2021 年**下降 26.7%**。



2022 年，美国在 AI 领域的私人投资为 **470 亿美元**，大约是排名第二的中国（130 亿美元）的 **3.5 倍**。

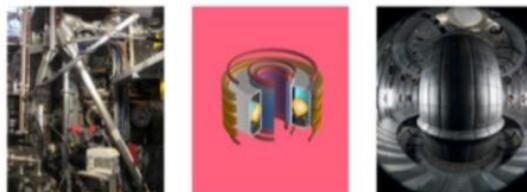


来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

# 是万能助手，还是闯祸精？

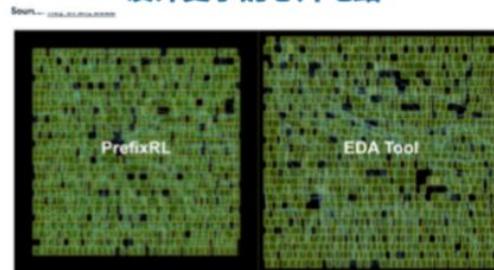
· Deepmind 强化学习算法**管理核聚变**

Source: DeepMind, 2022



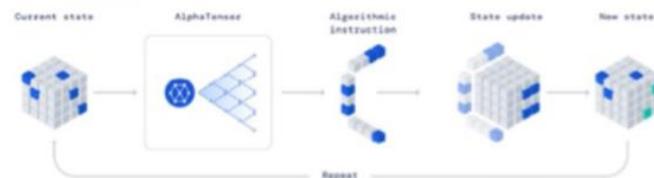
· 深度强化学习模型 PrefixRL

设计更小的芯片电路



· AlphaTensor 发现矩阵运算新算法

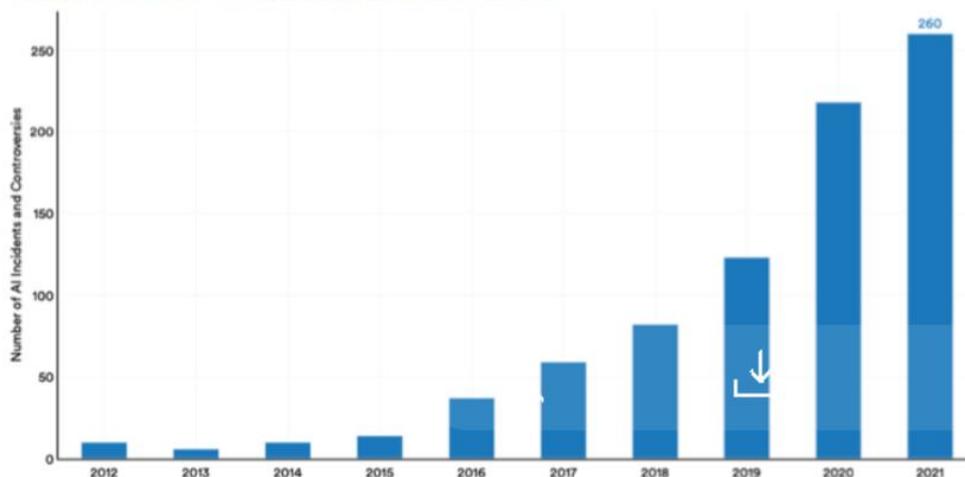
Source: Fawzi et al., 2022



· Zero-Shot 生成式 AI 开启新抗体设计的大门

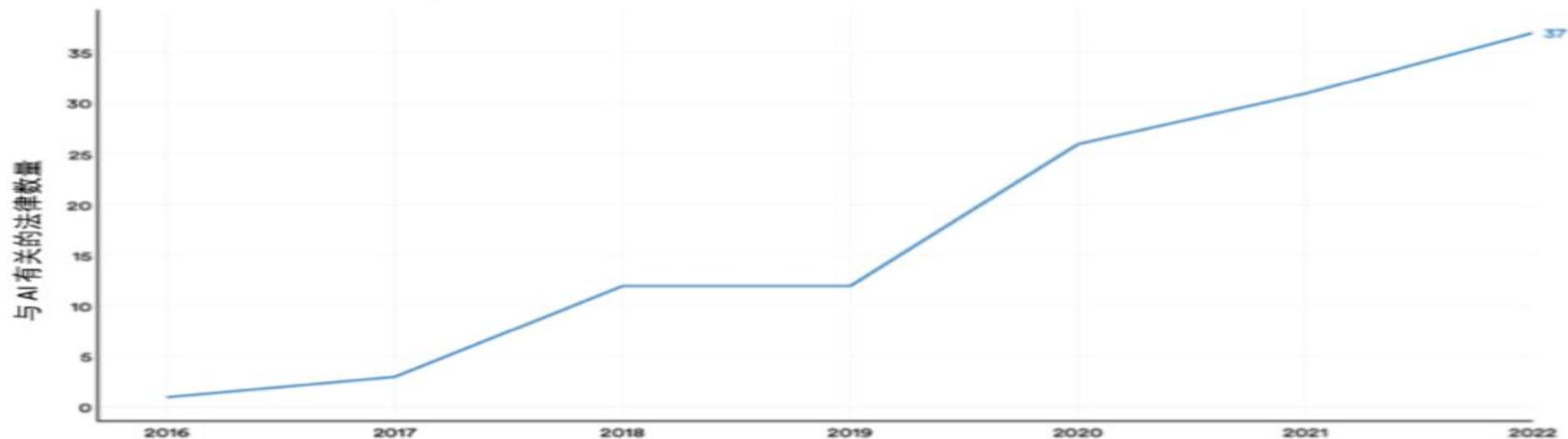
.....

2021 年，AI 带来的**事故和争议**是 2012 年的 **26 倍**。AI 与现实世界的交融程度越来越高，人们对 AI 被滥用的认识也逐渐深刻。

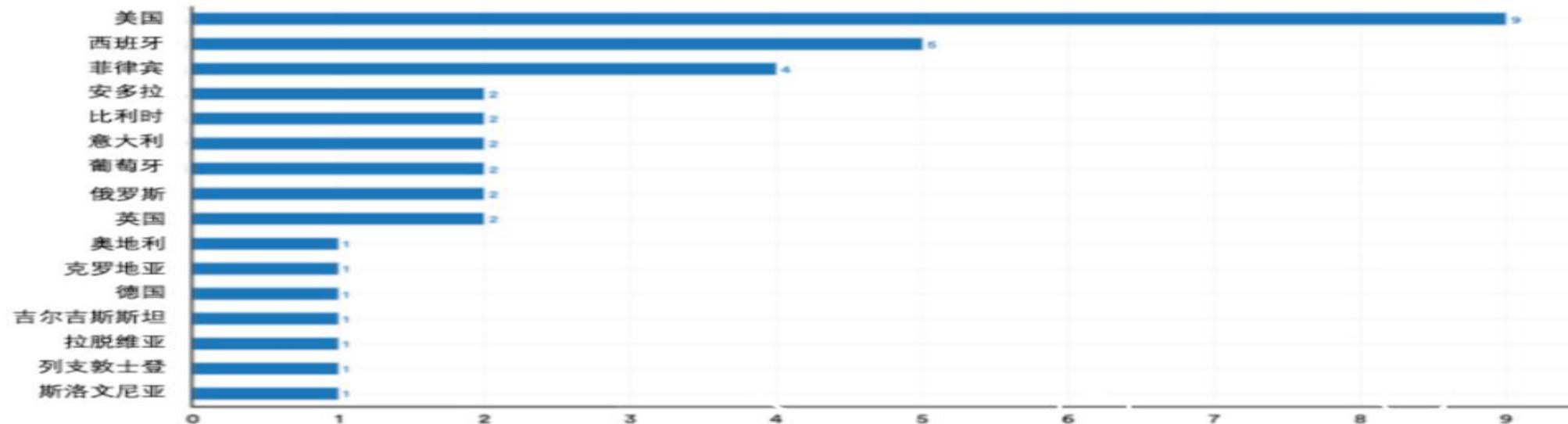


来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.

全球与 AI 有关的法律数量越来越多。2016 年接受调查的 127 个国家仅有 1 项相关法律，2022 年这一数字涨到了 37 项。



2022 年全球共通过了 37 项与 AI 有关的法律。其中美国通过了 9 项，数量最多。



来源：“The AI Index 2023 Annual Report,” Stanford University, April 2023.



数据

开放共享 | 全链条 | 隐私安全

AI基础设施

AI INFRASTRUCTURE



算力

智算中心 | 异构计算 | 弹性大规模集群



算法

开源开放 | 高效 | 流水线  
大模型生产及部署应用

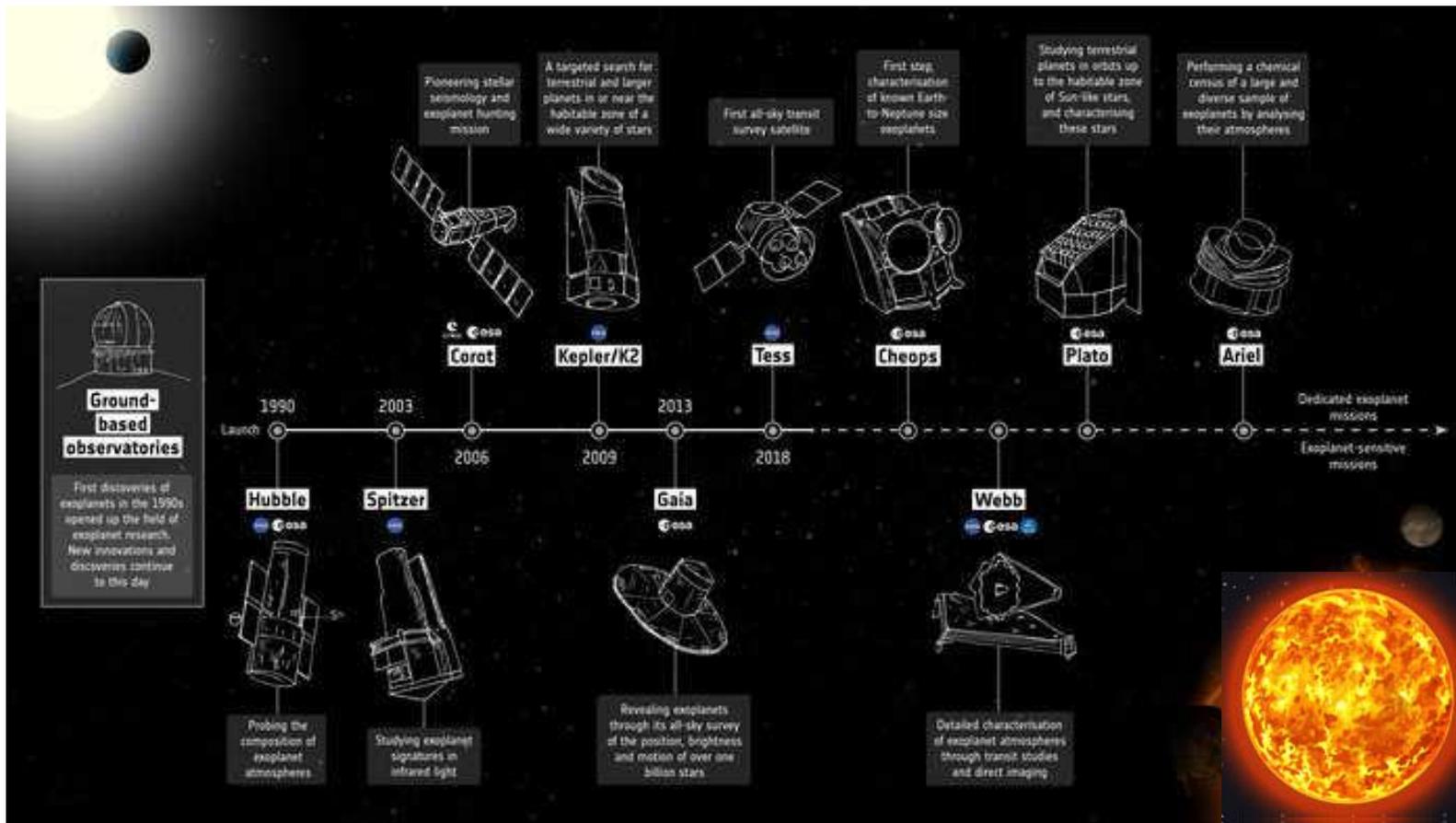
# 天文学呈现新面貌：全波段、多信使、时域、大数据

特点：

“广”

“深”

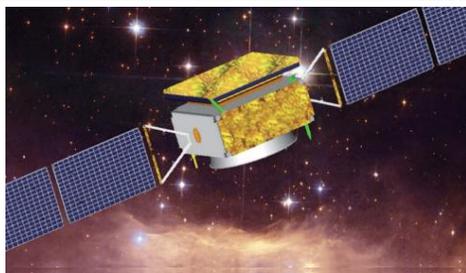
“快”



LAMOST



FAST



悟空



慧眼



爱因斯坦探针



中国空间站望远镜



# Sloan Digital Sky Survey

望远镜: 2.5m  
 $r < 22.5$

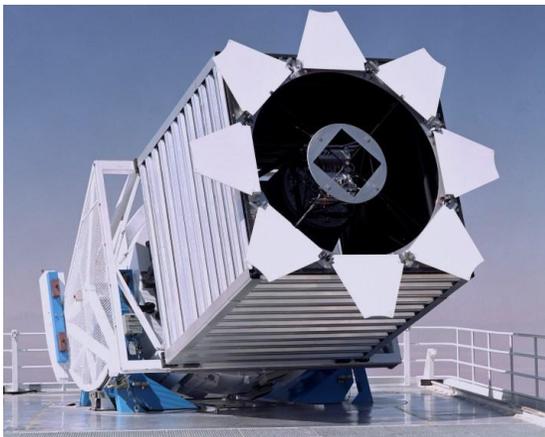
覆盖面积:  $> 14500$ 平方度

天测精度:  $0.1''$

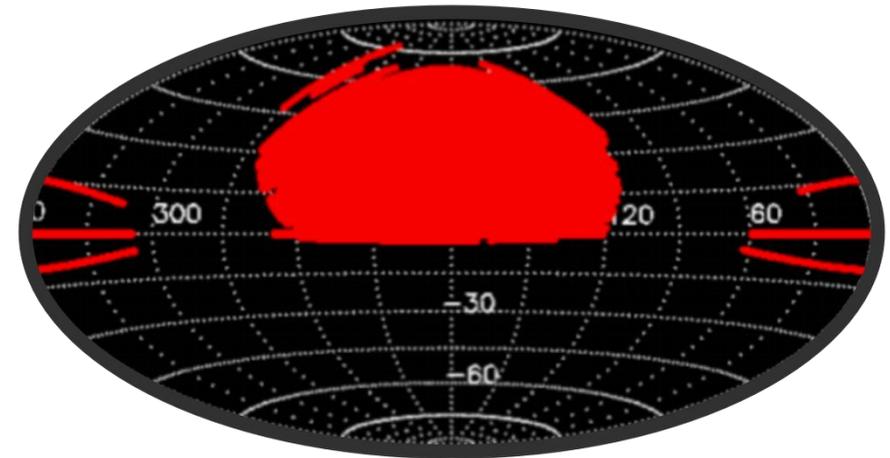
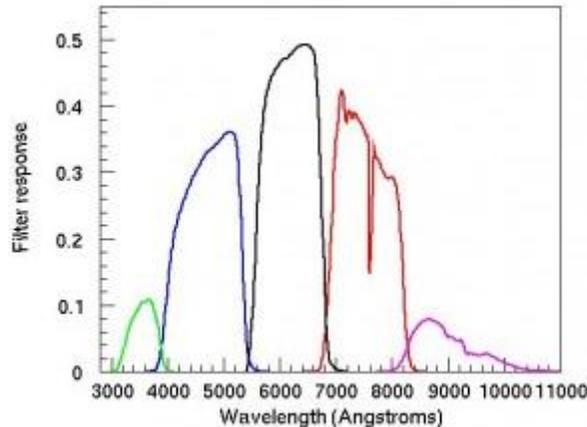
极限星等:

图像源数:  $> 460M$

光谱源数: Millions



u 3551Å  
g 4686Å  
r 6166Å  
i 7480Å  
z 8932Å



*Human Genome < 1 GB  
Human Memory < 1 GB (?)  
1 TB ~ 2 million books*

10年运行数据量:  $\sim 10TB$



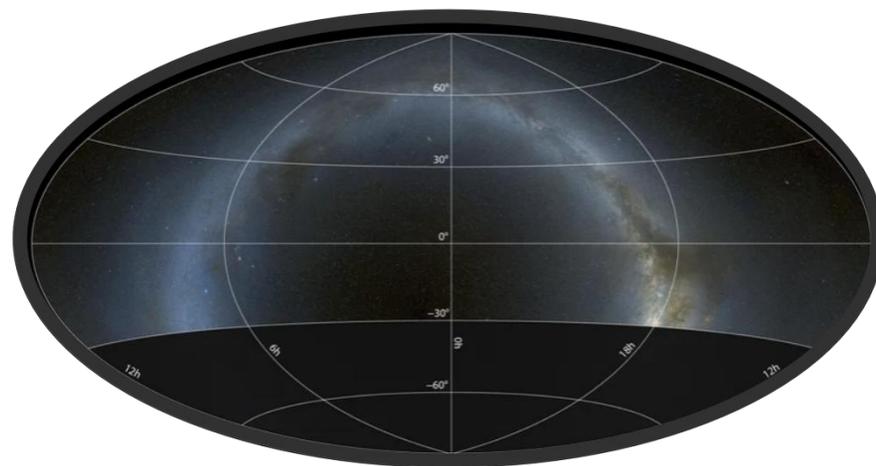
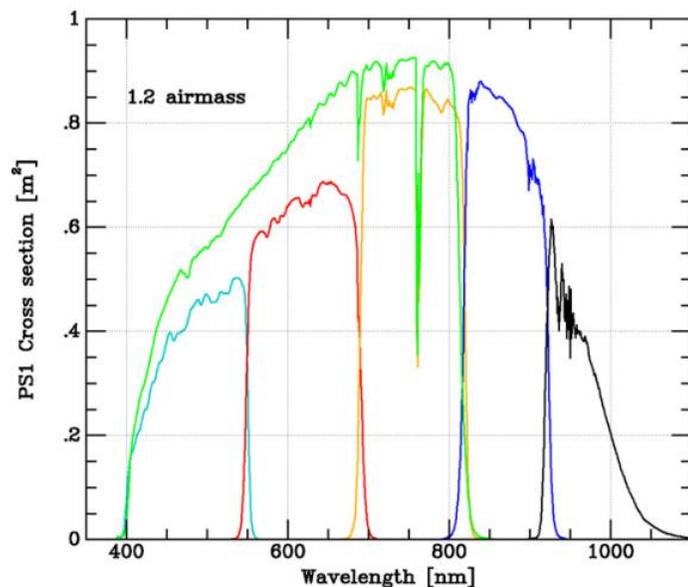
# Pan-STARRS Panoramic Survey Telescope and Rapid Response System

望远镜: 1.8m 覆盖面积: 30000平方度 天测精度: 50mas 极限星等:

$r < 23$



Filter	Mean Wavelength
g	4866
r	6215
i	7545
z	8679
y	9633



每晚约700GB



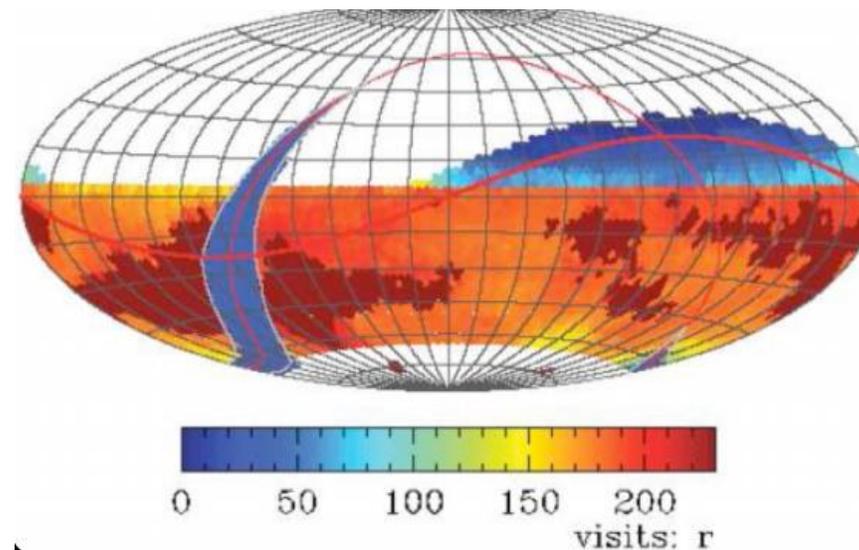
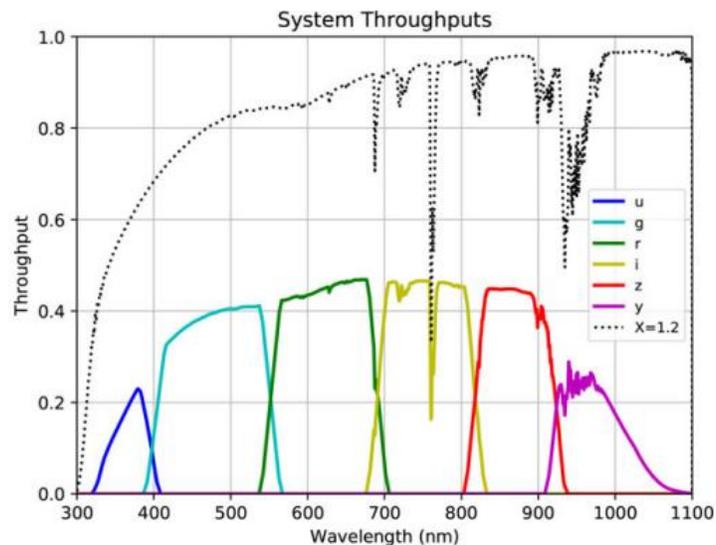
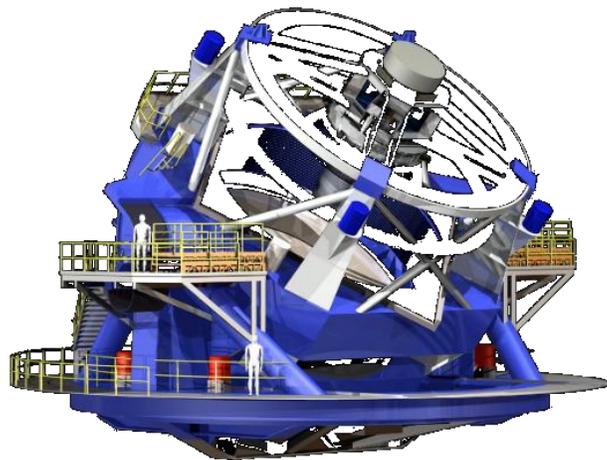
# LSST: A Deep, Wide, Fast, Optical Sky Survey

望远镜: 8.4m  
( $<27.5@10\text{yr}$ )

覆盖面积:  $>18000$ 平方度

天测精度: 10mas

极限星等:  $r < 24.5$



3.2Gpix camera

30sec exp/4sec rd

源数: 37Billion

巡天频率: 3天

时长: 10年 (825 revisits)

每晚约15TB

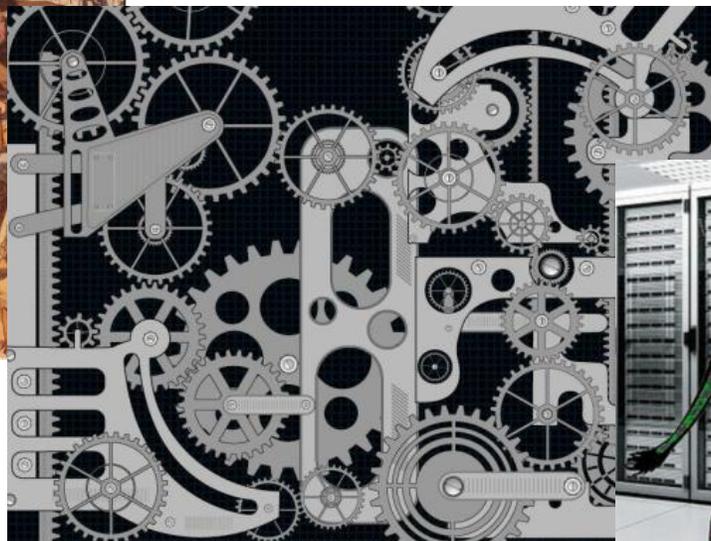


# 天文学：数据日益丰富的学科

“手工时代”



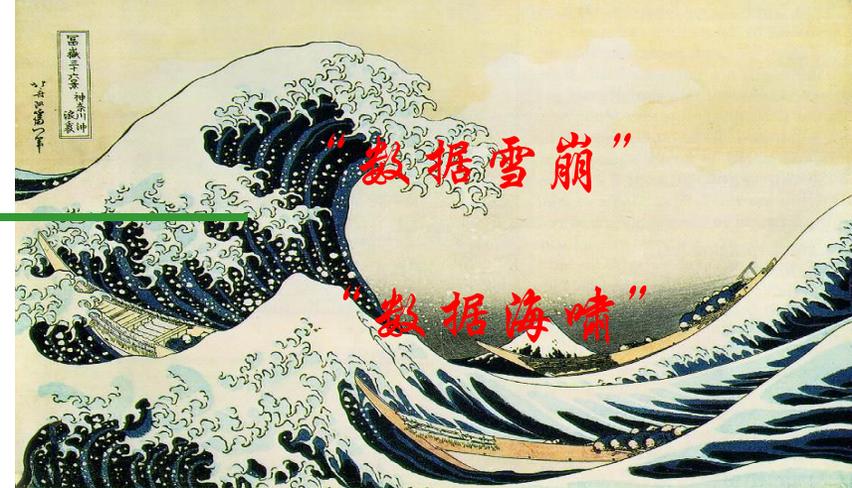
“工业时代”



“大数据时代”

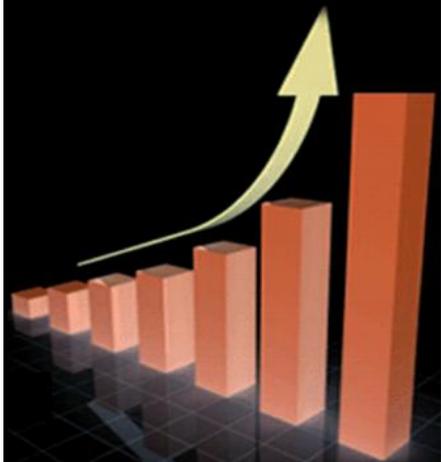


“智能时代”

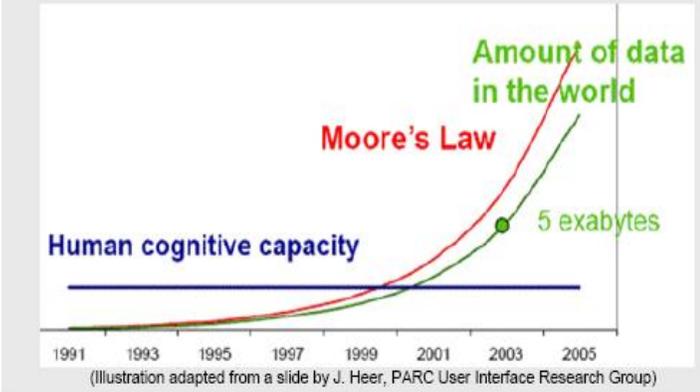


观测技术和信息技术的进步带动天文学的发展

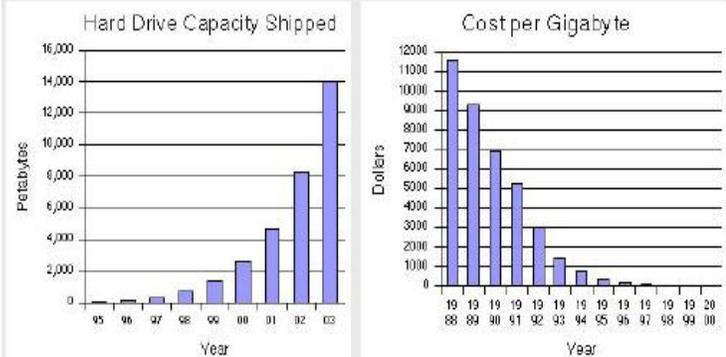
人力 → 电力 → 算力



数据指数增长  
每1.5年涨一倍

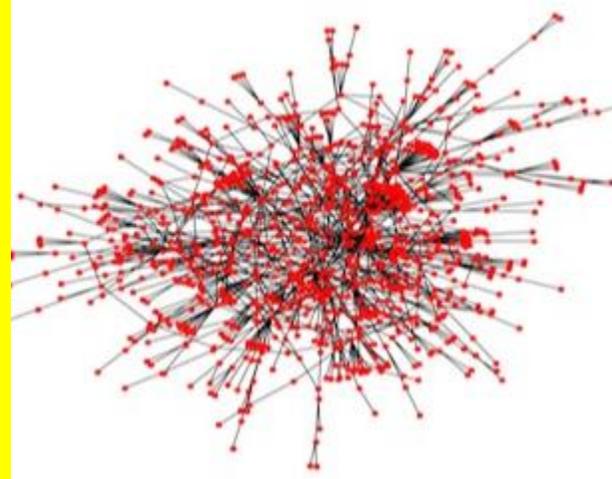


数据复杂性增长  
海量性、空间性、多波段性、异构性、分布性、非线性、高维性、时域性、缺值性、……



数据的变化

数据贫穷	→	数据过剩
数据集	→	数据流
静态	→	动态、演化
任意时刻	→	实时分析和发现
集中	→	分布
数据者所有	→	领域所有



# 科学研究的四个阶段

■ 第一范式

实验或测

量

■ 第二范式

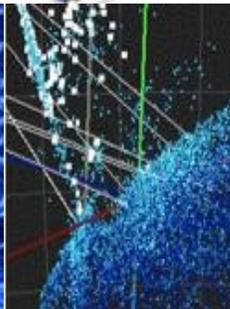
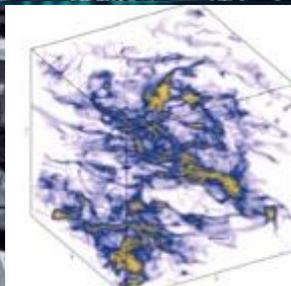
理论分析

■ 第三范式

数值模拟

■ 第四范式

数据密集



Data

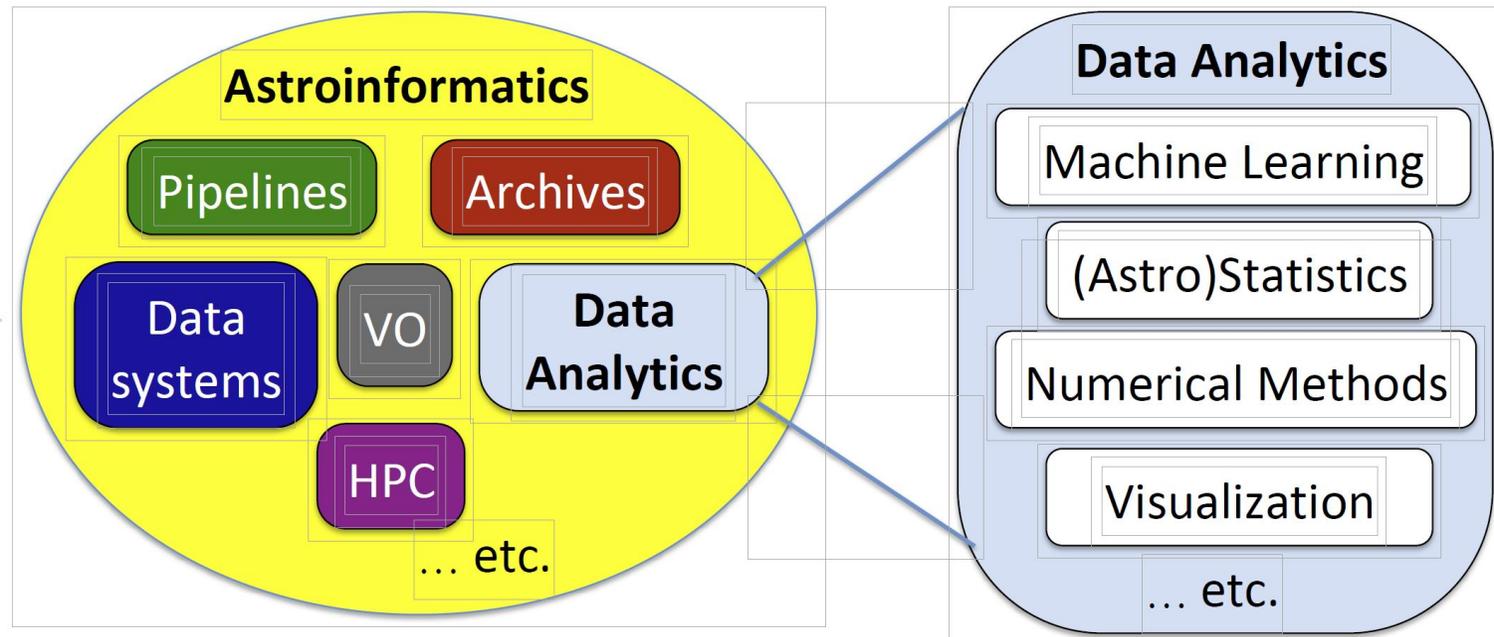
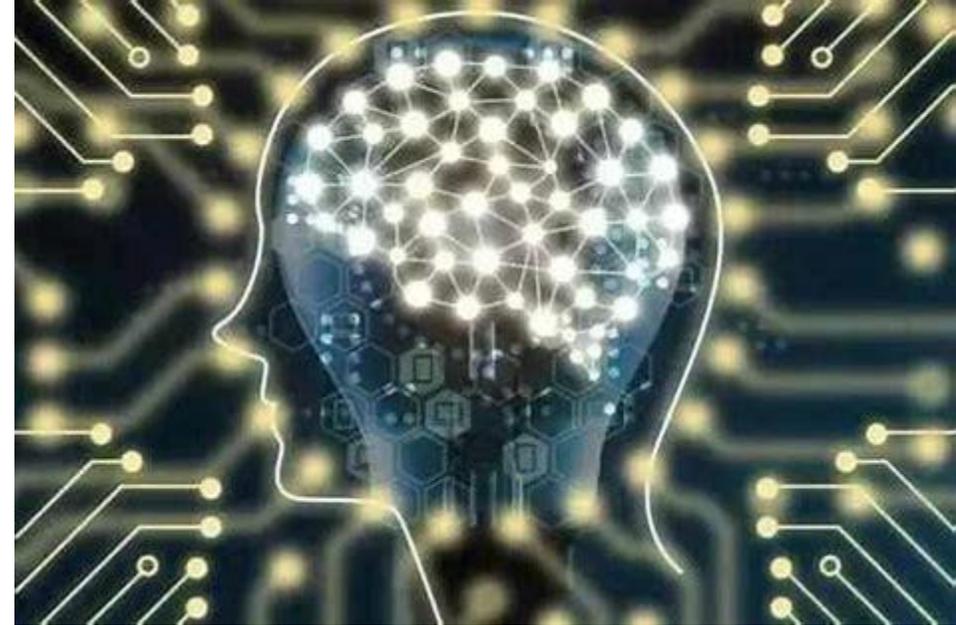
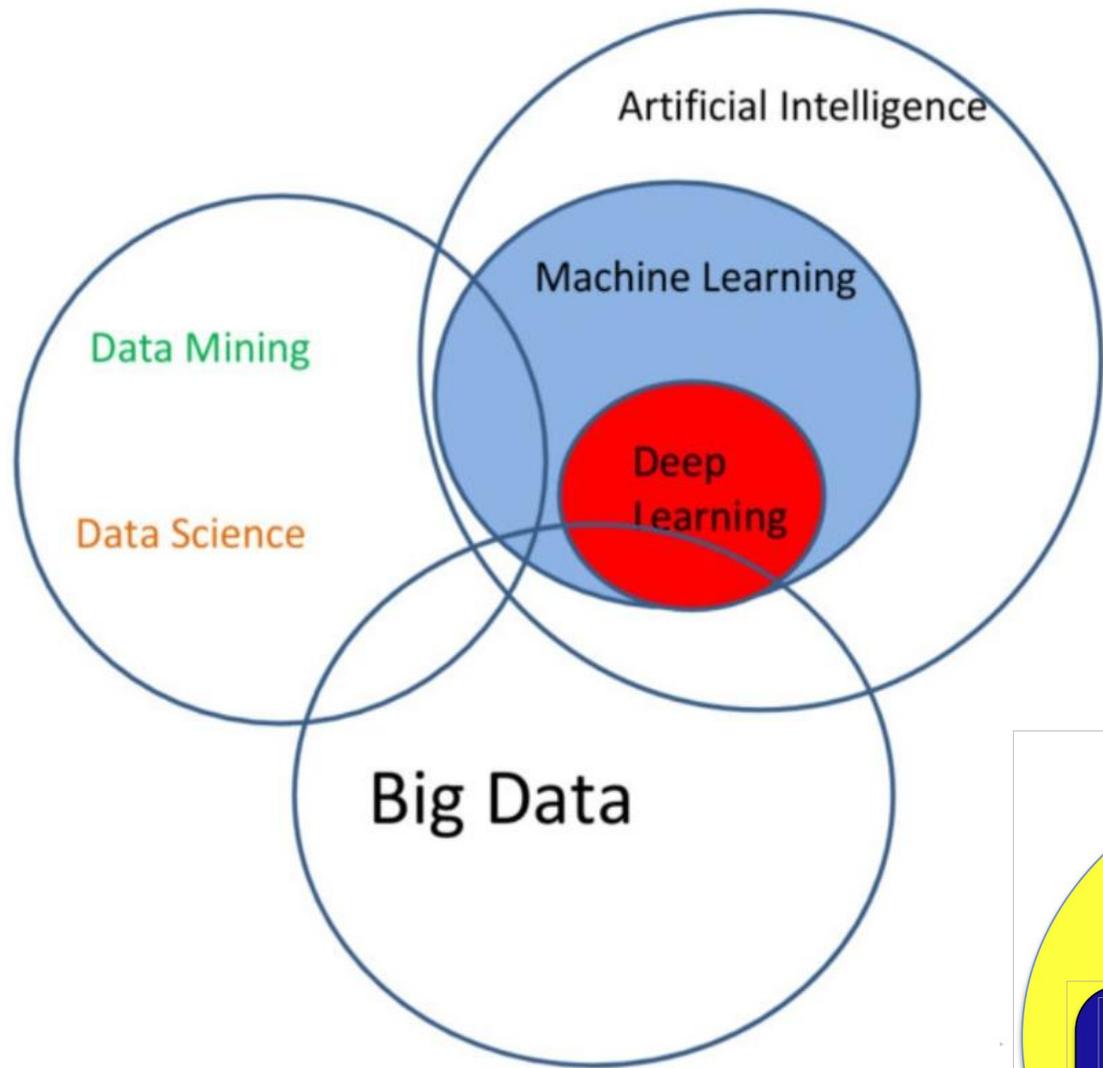
Astronomy  
entered  
new  
era:

**“Ask Not What Data You Need To Do Your Science,  
Ask What Science You Can Do With Your Data.”**



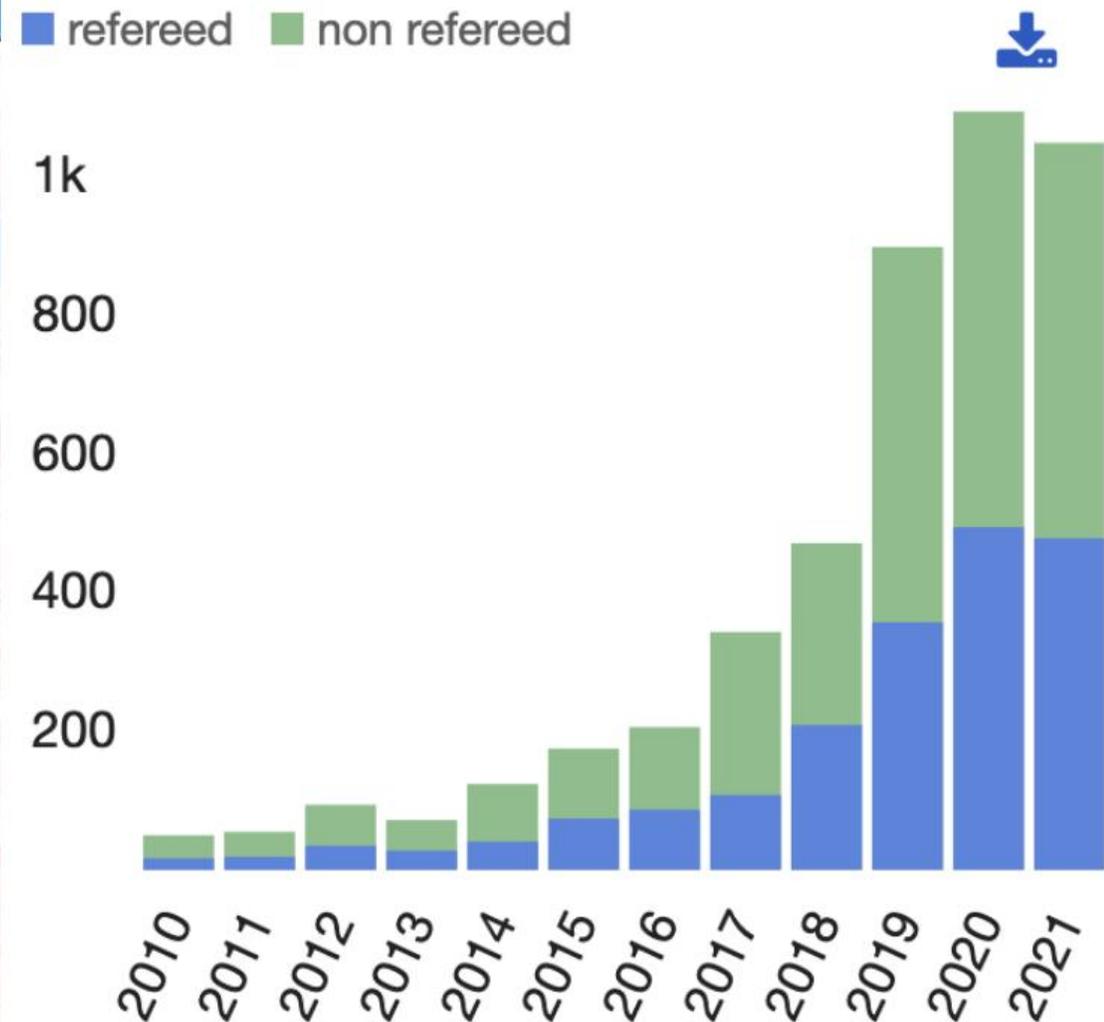
## The era of surveys...

- Standard: “What data do I have to collect to (dis)prove a hypothesis?”
- Data-driven: “What theories can I test given the data I already have?”



# ML use in astronomy

Astronomy papers in ADS containing "Artificial Intelligence" or "Machine learning" or "Deep learning" in the abstract.



# 机器学习



机器学习是一门通过编程让计算机从数据中进行学习的科学（和艺术）。

机器学习是一个研究领域，让计算机无须进行明确编程就具备学习能力。

——亚瑟·萨缪尔（Arthur Samuel），1959

一个计算机程序利用经验E来学习任务T，性能是P，如果针对任务T的性能P随着经验E不断增长，则称为机器学习。

——汤姆·米切尔（Tom Mitchell），1997

- 例子：垃圾邮件过滤器
  - T：识别垃圾邮件
  - E：用户标注的垃圾邮件
  - P：正确识别百分比

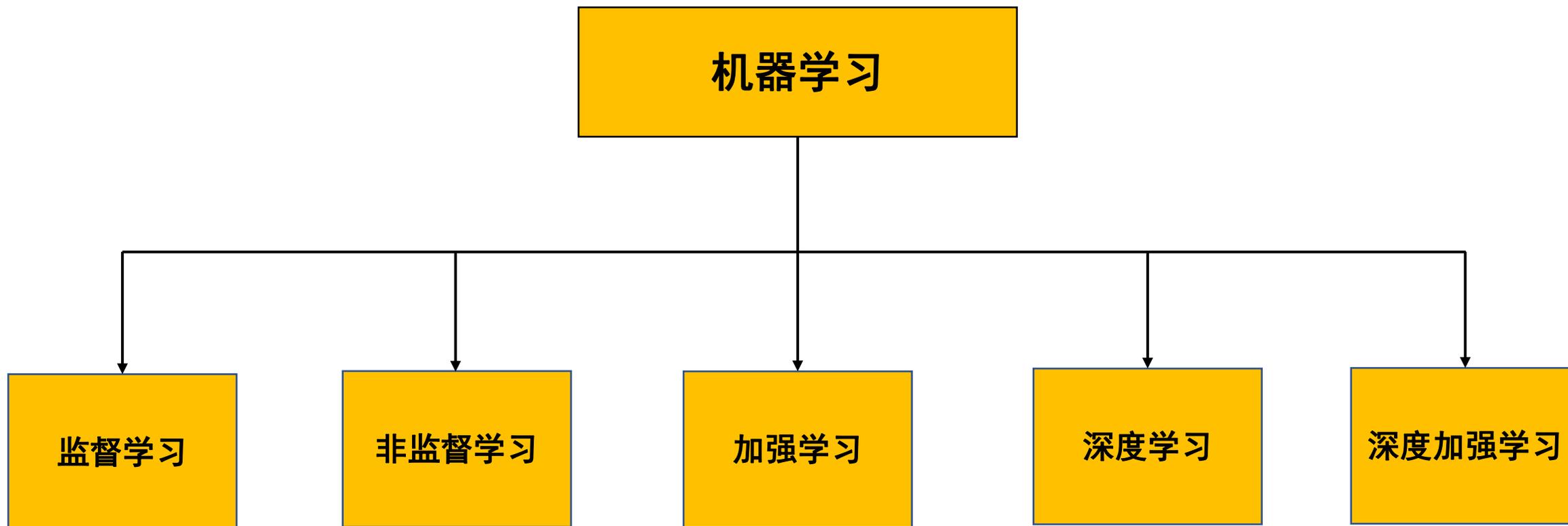


- 它是人工智能核心，是使计算机具有智能的根本途径。





# 机器学习分类



# 机器学习

## 模型方法

- 朴素贝叶斯
- 逻辑回归
- 线性回归
- KNN
- 决策树
- Boosting
- SVM (支持向量机)

## 神经网络

- 单个神经元
- 感知机
- 自编码器与受限玻尔兹曼机

## 多层神经网络

- 浅层
  - 经典BP前馈神经网络
- 深层
  - DNN (深度神经网络)
  - RNN (循环神经网络)
  - CNN (卷积神经网络)

## 深层与浅层的区别

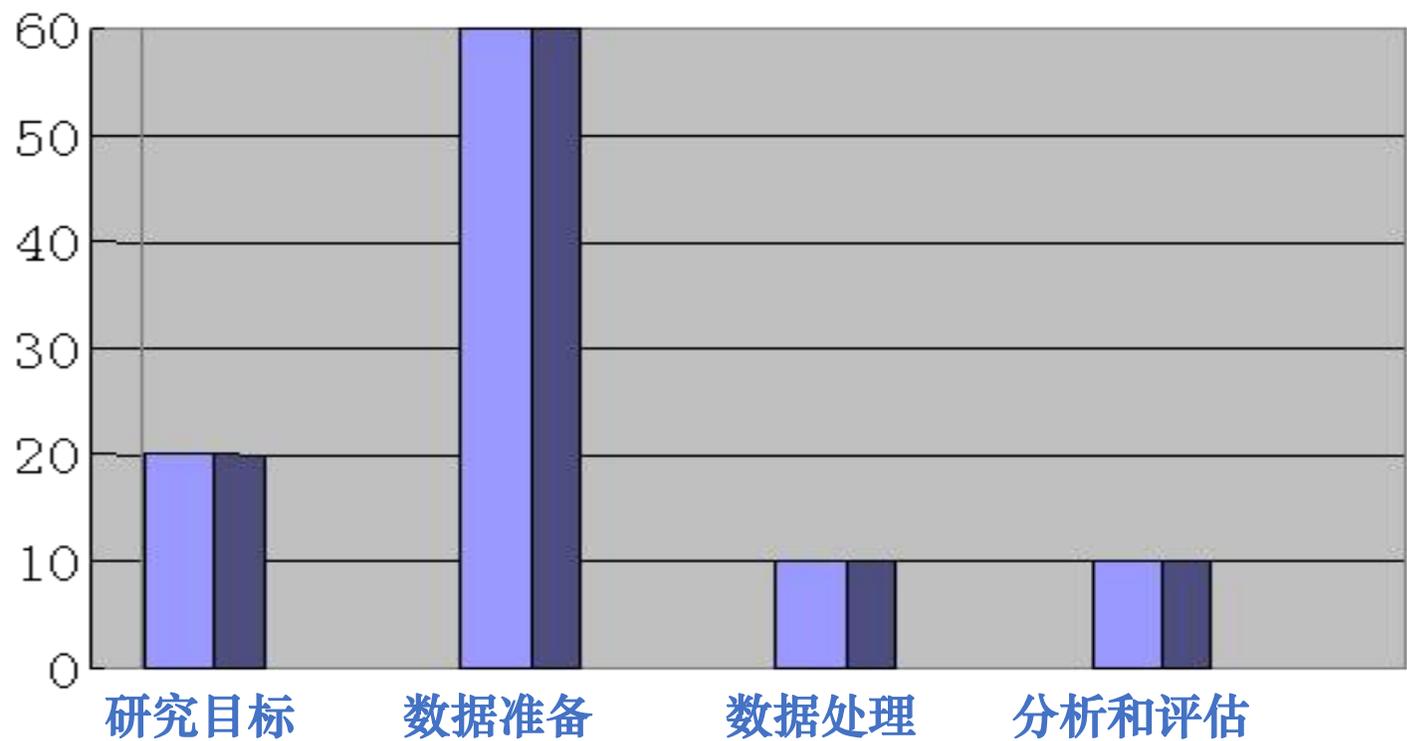
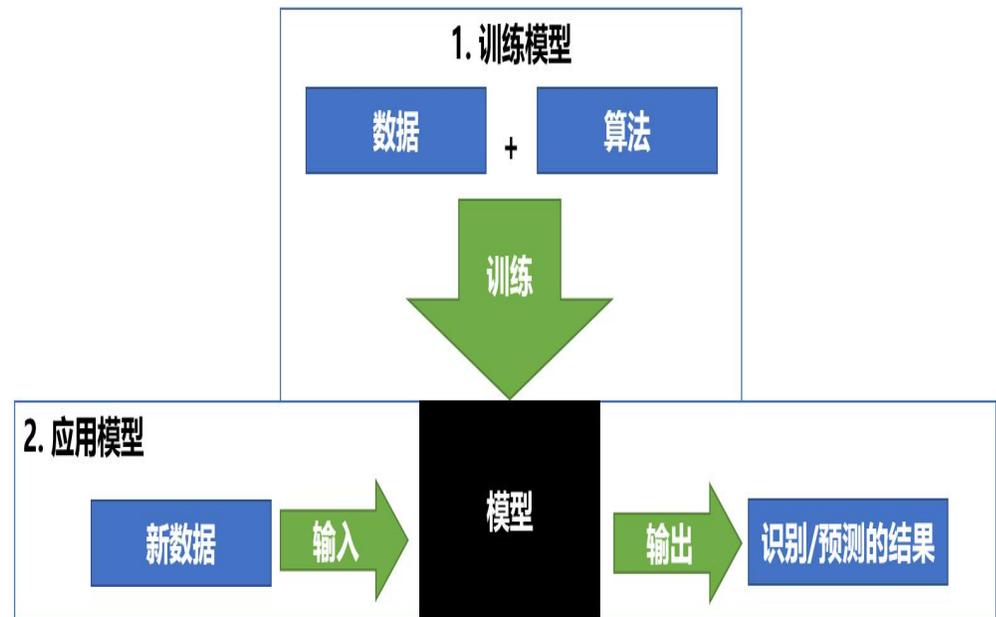
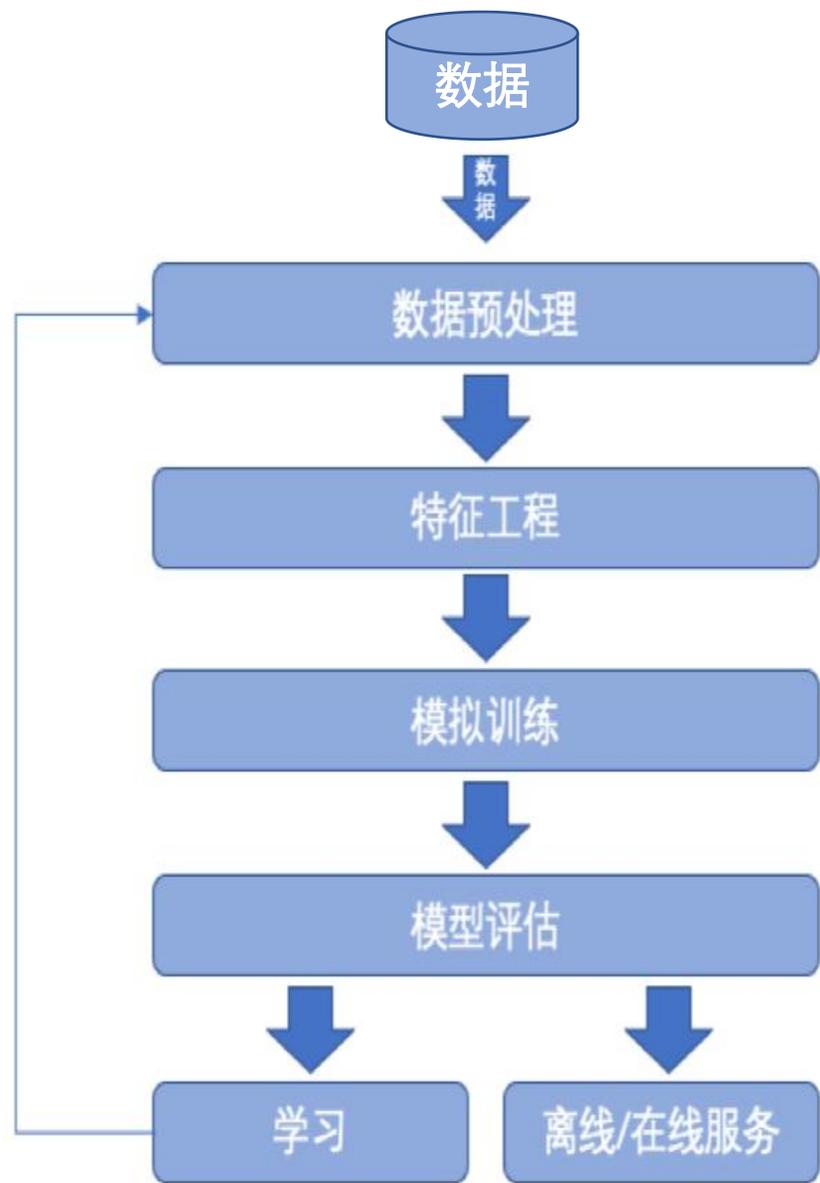
- 神经网络层数
- 有无“特征提取”预训练

## 学习方式

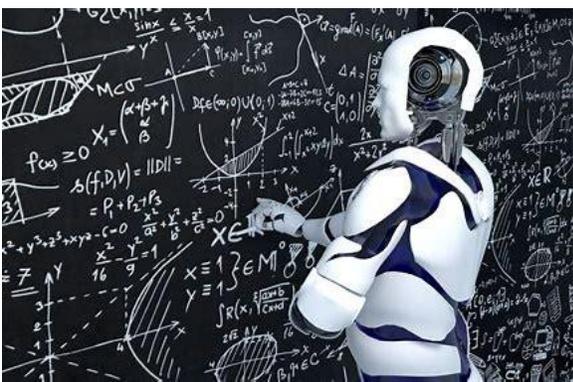
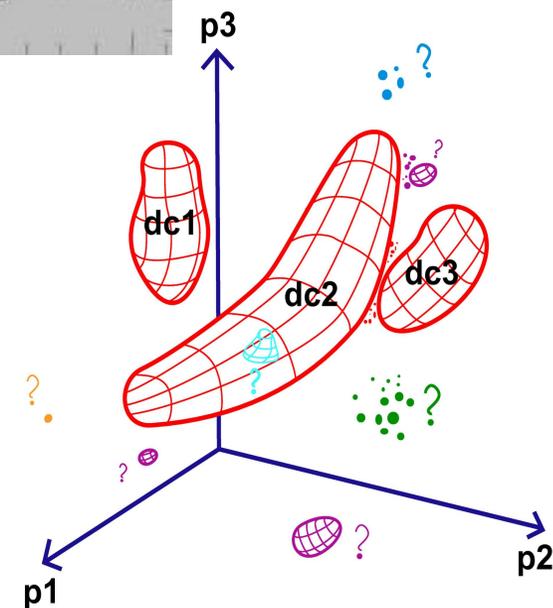
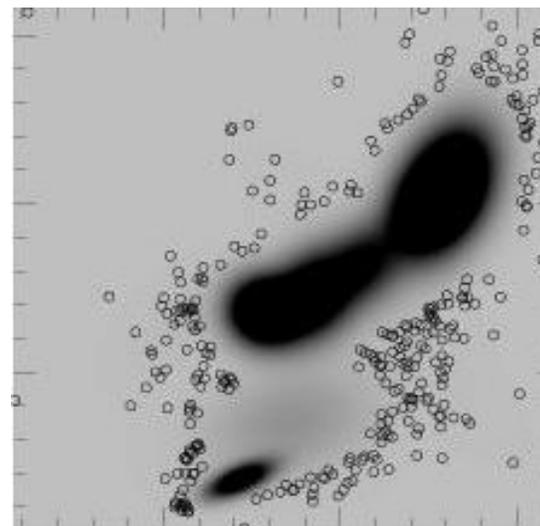
- 监督学习
  - 回归
  - 分类
- 非监督学习
  - 聚类

## 处理过程

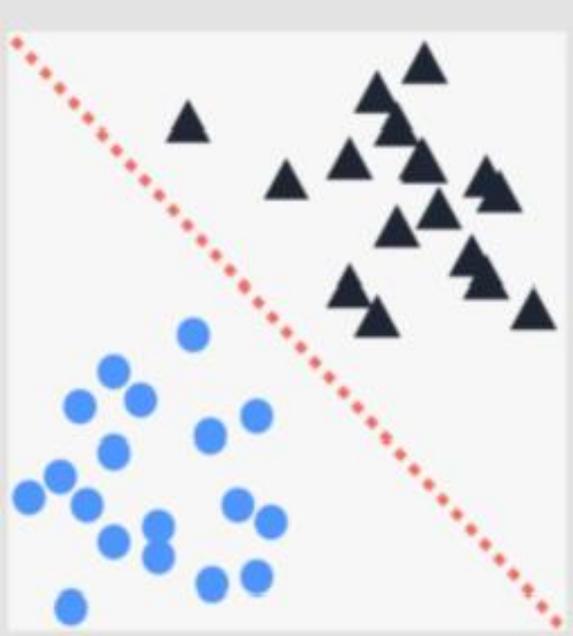
- 输入数据
- 训练模型
- 验证
  - 误差
  - 泛化性
  - 拟合与过拟合



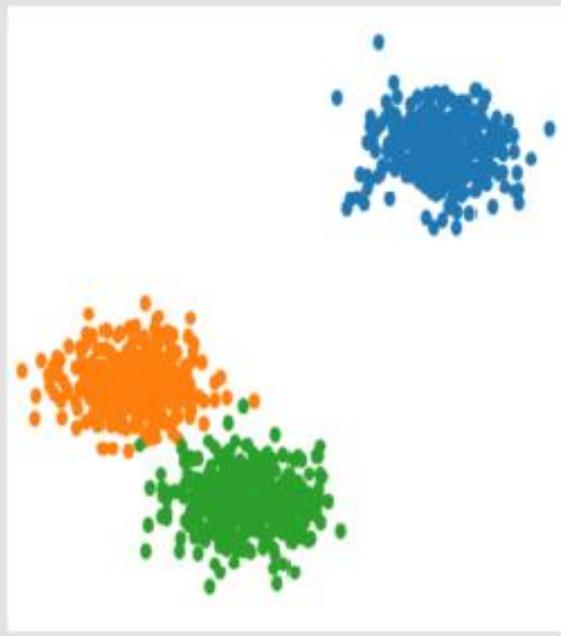
- 数据总结
- 分类分析
- 回归分析
- 聚类分析
- 关联规则分析
- 序列模式分析
- 依赖关系分析
- 异常检测
- 模式分析或统计分析
- 其它



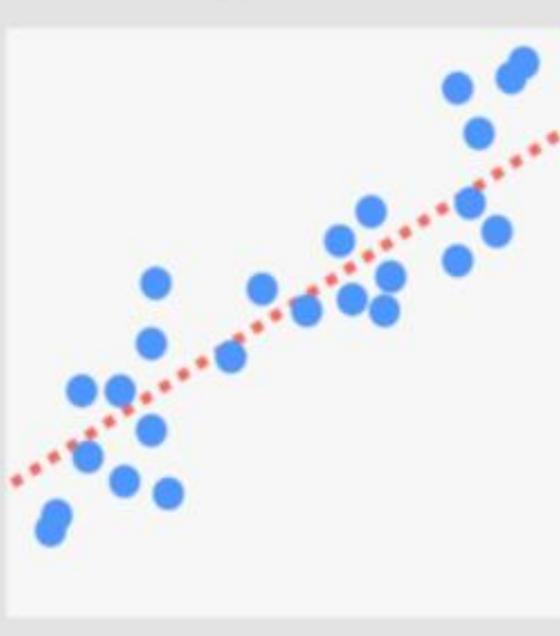
# Classification



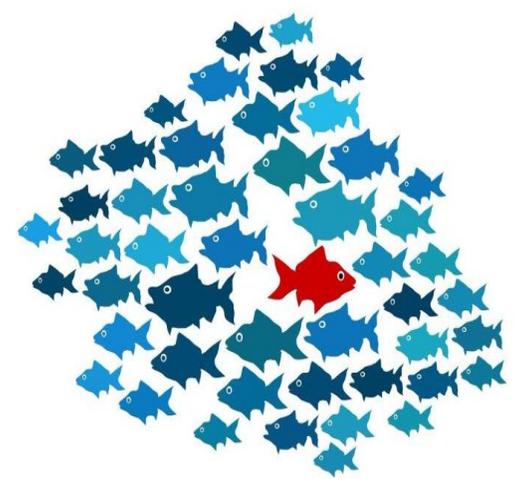
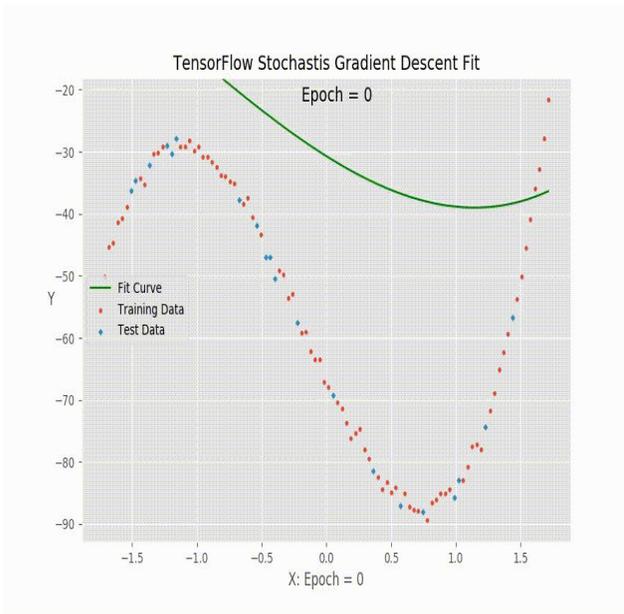
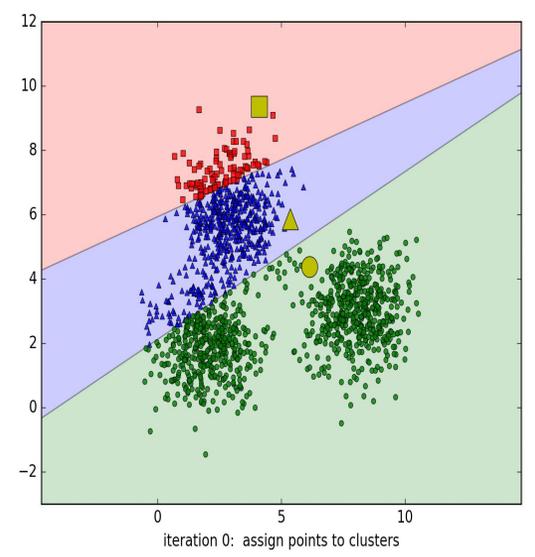
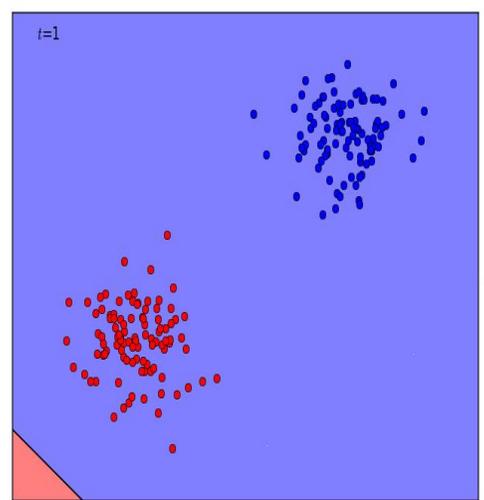
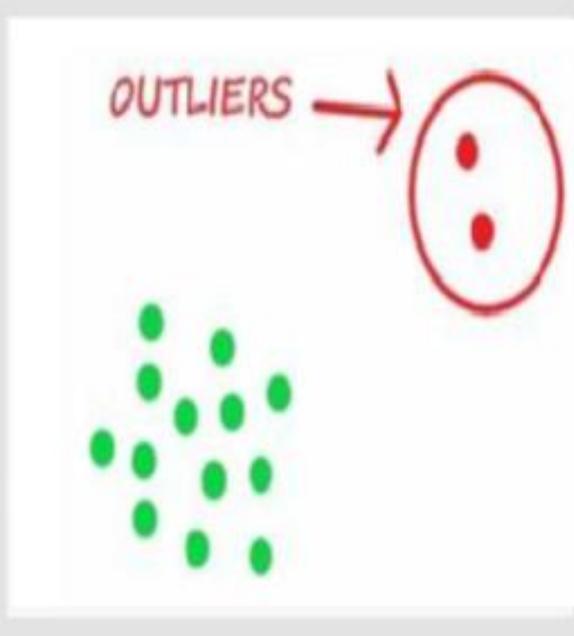
# Clustering



# Regression



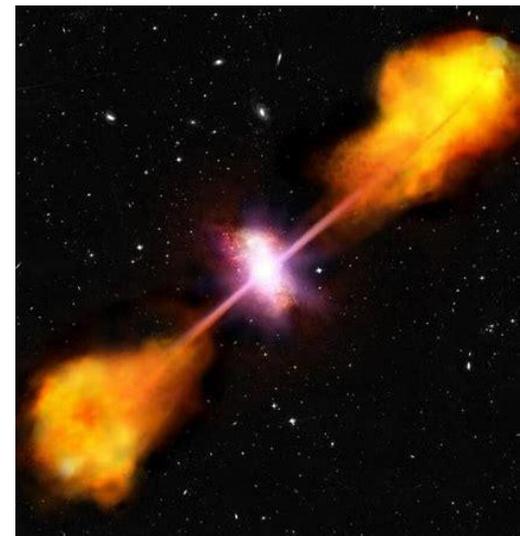
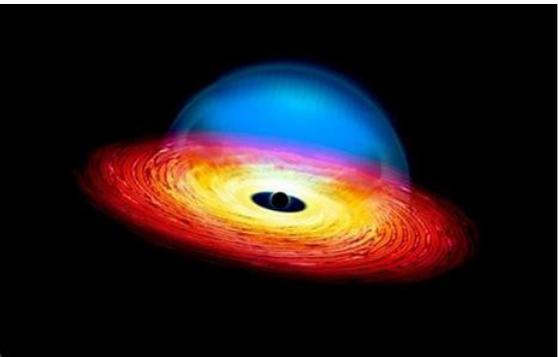
# Outlier



# 天文学应用

---

天文学是一个数据密集型的科学领域，天文学家需要处理各种不同类型的数据，如天体图像、光谱数据、时间序列数据等。随着天文学数据量的不断增加，AI、ML、DL在天文学中变得越来越重要。



# 应用

## 1: 数据分类和目标识别

天文学家面临着大量的天文图像和数据，其中包含各种天体和天文现象。人工智能技术，如深度学习，可以用于分类和识别天体，如星系、恒星、行星、射电源等，从而更好地理解宇宙中的各种天体；可以帮助自动化和加速数据处理过程。

## 2: 天文信号处理

人工智能可以帮助处理和分析天体信号，例如天体信号检测、抑制天体噪声、提高探测灵敏度等。天文观测数据通常受到噪声和伪影的影响，这可能干扰对天体的准确分析和研究。人工智能技术可以应用于数据降噪和伪影去除，帮助提取出天体信号并改善数据质量。

## 3: 数据挖掘和模式识别

天文学家需要从大规模的数据集中挖掘出有价值的模式和趋势。人工智能技术可以应用于数据挖掘和模式识别，帮助发现天文现象、异常事件和新的科学见解。

## 4: 数据拟合和参数估计

对于复杂的物理模型和天文现象，人工智能可以用于拟合数据和估计模型参数。通过机器学习算法，可以实现更准确的模型拟合，从而提供更精确的参数估计和科学推断。

## 应用

### 5: 数据处理流程的自动化

人工智能技术可以用于自动化数据处理流程，从数据获取、预处理、分析到结果生成。这有助于提高数据处理的效率和准确性，并减少人为错误的影响。

### 6: 光谱分析和特征提取

天文学中的光谱数据包含了丰富的信息，可以用于研究天体的化学组成、温度、速度等性质。人工智能技术可以应用于光谱分析和特征提取，帮助识别和解析不同的谱线、特征和结构。

### 7: 数据关联和匹配

天文学家常常需要将不同天体目录或观测数据进行关联和匹配，以进行比较和交叉验证。人工智能技术可以应用于数据关联和匹配任务，帮助自动识别和关联不同数据源中的天体，并提供更全面的数据分析。

### 8: 引力透镜分析

引力透镜现象提供了研究暗物质和宇宙学的重要线索。人工智能技术可以应用于引力透镜数据的分析和模拟，帮助识别和研究引力透镜系统，并提供对暗物质分布和宇宙结构的更深入理解。

# 应用

## 9: 天文图像处理 and 重建

天文图像通常受到大气湍流、仪器噪声和其他干扰的影响，这可能降低图像的分辨率和质量。人工智能技术可以应用于天文图像处理 and 重建，以提高图像的清晰度、对比度和分辨率，从而揭示更多细节和结构。

## 10: 数据可视化和探索

人工智能可以应用于天文数据的可视化和交互探索，以帮助天文学家更直观地理解和分析数据。通过机器学习和数据挖掘技术，可以构建交互式工具和可视化平台，使天文学家能够以更直观的方式探索数据，并发现新的关联和模式。

## 11: 天体物理学的模拟和预测

人工智能可以用于天体物理学的模拟和预测。例如，科学家可以使用深度学习算法来模拟黑洞合并事件。可以应用于星系演化的模拟和预测，通过学习已知星系演化的规律和模式，预测未来的星系演化过程，帮助揭示宇宙中星系的形成和演化机制。

## 12: 天文数据压缩和降维

天文学数据通常非常庞大，处理和存储的成本较高。人工智能技术可以应用于数据压缩和降维，将数据表示为更紧凑的形式，同时保留重要的信息和特征，以减少数据存储和传输的需求。

# 应用

## 13: 数据异常检测和异常值分析

天文数据中可能存在异常的观测或测量，这可能是由仪器故障、噪声或其他因素引起的。人工智能技术可以应用于异常检测和异常值分析，帮助识别和排除异常数据，确保数据的准确性和可靠性；帮助研究人员识别和研究异常的天文现象，如爆发事件、超新星爆发等。

## 14: 天文事件预测和预警

人工智能可以应用于天文事件的预测和预警，如超新星爆发、伽马射线暴、引力波等。通过对历史数据的学习和模式识别，人工智能可以提前预测天文事件的发生，并及时发出预警信息。

## 15: 数据共享和合作

人工智能技术可以促进天文数据的共享和合作，通过自动化数据处理和标注，加快数据共享的过程。同时，人工智能还可以帮助研究人员进行数据集成和合并，实现不同数据源的交叉验证和共同分析。

## 16: 天体物理参数估计

人工智能可以应用于天体物理参数的估计，如恒星质量、有效温度、金属丰度、重力加速度、年龄等；星系和类星体的测光红移。通过学习已知的天体参数和观测数据之间的关系，人工智能可以推断未知天体的物理参数，提供更准确的天体特性估计。

## 应用

### 17: 引力波信号检测和分类

引力波是由大质量物体（如黑洞或中子星）在宇宙中快速移动而产生的扰动。人工智能可以应用于引力波数据的分析，帮助检测引力波信号并将其分类，从而加深对引力波源的理解。

### 18: 天文数据的可靠性评估

天文观测数据可能受到各种误差和系统偏差的影响。人工智能可以应用于评估天文数据的可靠性，并提供数据质量的度量和准确性评估，帮助研究人员在数据分析中考虑到这些误差和不确定性。

### 19: 天体成像和图像处理

天文图像中可能存在模糊、噪声和其他失真，这可能影响对天体的观测和分析。人工智能技术可以应用于天文图像的增强、恢复、去噪、分割和重构等，改善图像质量，揭示更多细节和结构。

### 20: 天文数据的快速搜索和匹配

随着天文观测数据的增加，快速搜索和匹配相关数据变得至关重要。人工智能可以应用于天文数据的索引、搜索和匹配，帮助天文学家快速访问和利用大规模数据集，促进科学研究和发现。

## 应用

### 21: 变星分类和光变曲线分析

人工智能可以应用于变星的分类和光变曲线的分析。通过学习不同类型的变星和它们的光变行为，人工智能可以帮助自动分类变星，并提取光变曲线中的特征，以揭示变星的物理机制和演化过程。

### 22: 天文数据的自动化注释和标记

天文学家需要对大量的天文数据进行注释和标记，以便后续的分析研究。人工智能可以应用于天文数据的自动化注释和标记，帮助识别和标记不同的天体、结构和特征，提高数据处理的效率和准确性。

### 23: 天体目标推荐和筛选

根据天文学家的研究兴趣和需求，人工智能可以应用于天体目标的推荐和筛选。通过分析天文数据和研究历史，人工智能可以帮助推荐潜在的有趣目标，并根据特定的标准和要求进行筛选。

### 24: 天文数据的时间序列分析

天文学中存在大量的时间序列数据，如光变曲线、射电脉冲等。人工智能可以应用于时间序列数据的分析，帮助识别周期性信号、异常事件和趋势，从中提取有用的信息和洞察。

# 应用

## 25: 天文学知识图谱构建

人工智能可以应用于天文学知识的整合和构建知识图谱。通过自动化的文本分析和知识抽取，人工智能可以从大量的天文学文献和数据库中提取关键信息，并构建出结构化的知识图谱，以支持天文学领域的知识管理和发现。

## 26: 天体图像识别和分类

人工智能可以应用于天体图像的识别和分类，帮助自动识别不同类型的天体，如星系、星团、星云等，并进行分类研究。可以用于对星系图像进行分类和识别。例如，天文学家可以使用卷积神经网络来自动识别星系的形状和类型，从而更好地理解星系的演化和性质；科学家可以使用支持向量机等算法来对星系进行分类，并根据其形态、颜色和亮度等特征来研究星系的演化和性质。

## 27: 天体运动模拟和轨道预测

人工智能可以应用于天体运动模拟和轨道预测，帮助研究人员模拟和预测天体的运动轨迹，包括行星、彗星、卫星等，有助于了解它们的物理性质和行为。

## 28: 暗物质分布模拟和研究

人工智能可以应用于暗物质分布的模拟和研究，帮助研究人员理解宇宙中暗物质的分布、演化和相互作用。

## 应用

### 29: 引力透镜天体搜索

人工智能可以应用于引力透镜天体的搜索和识别，帮助研究人员发现新的引力透镜系统，并进一步研究宇宙的引力透镜效应。

### 30: 特殊天体搜寻

如：FRB、双星系、豌豆星系、环星系、变脸AGN等。

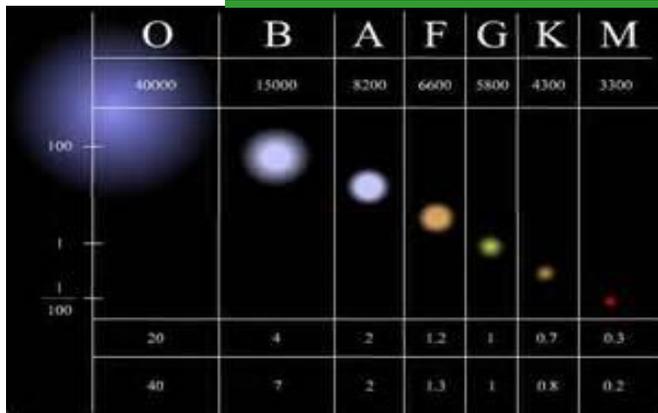
### 31: 宇宙探索和探测

人工智能可以帮助控制太空探测器和卫星，以收集和分析宇宙数据，例如在探测黑洞、探测暗物质等方面。

### 32: 宇宙射线探测

人工智能可以帮助处理和分析宇宙射线数据，以了解宇宙的物理过程和演化。

# 天文的应用：恒星光谱分类



恒星光谱序列 (温度序列) :  
O B A F G K M L T

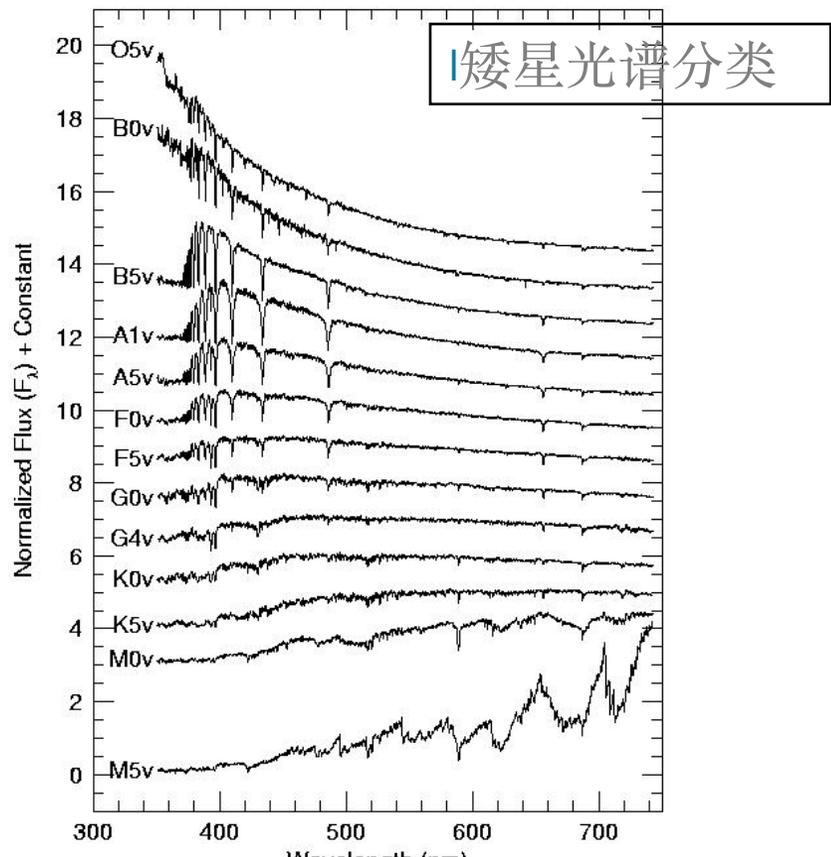
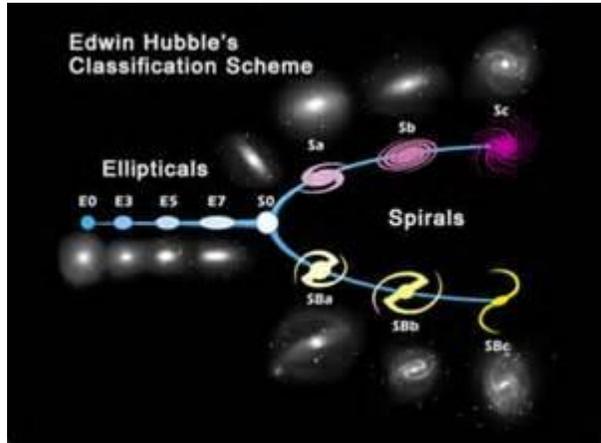


Table 16.1 The Spectral Sequence

Spectral Type	Example(s)	Temperature Range	Key Absorption Line Features	Brightest Wavelength (color)	Typical Spectrum
O	Stars of Orion's Belt	>30,000 K	Lines of ionized helium, weak hydrogen lines	<97 nm (ultraviolet)*	O
B	Rigel	30,000 K–10,000 K	Lines of neutral helium, moderate hydrogen lines	97–290 nm (ultraviolet)*	B
A	Sirius	10,000 K–7,500 K	Very strong hydrogen lines	290–390 nm (violet)*	A
F	Polaris	7,500 K–6,000 K	Moderate hydrogen lines, moderate lines of ionized calcium	390–480 nm (blue)*	F
G	Sun, Alpha Centauri A	6,000 K–5,000 K	Weak hydrogen lines, strong lines of ionized calcium	480–580 nm (yellow)	G
K	Arcturus	5,000 K–3,500 K	Lines of neutral and singly ionized metals, some molecules	580–830 nm (red)	K
M	Betelgeuse, Proxima Centauri	<3,500 K	Molecular lines strong	>830 nm (infrared)	M

\* All stars above 6,000 K look more or less white to the human eye because they emit plenty of radiation at all visible wavelengths.

# 天文的应用：星系形态分类



elliptical galaxy



lenticular galaxy



normal spiral galaxy



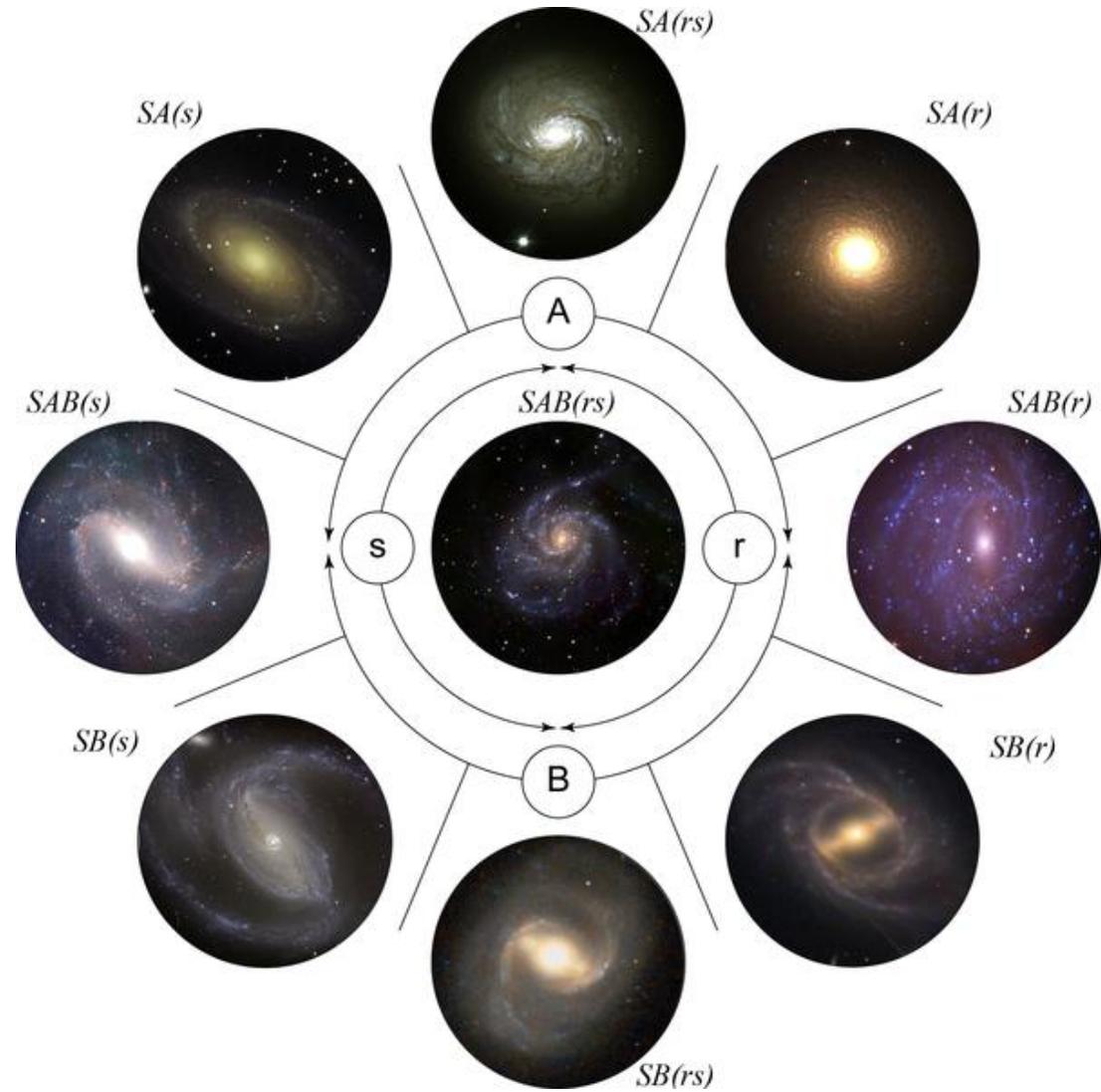
type I irregular galaxy



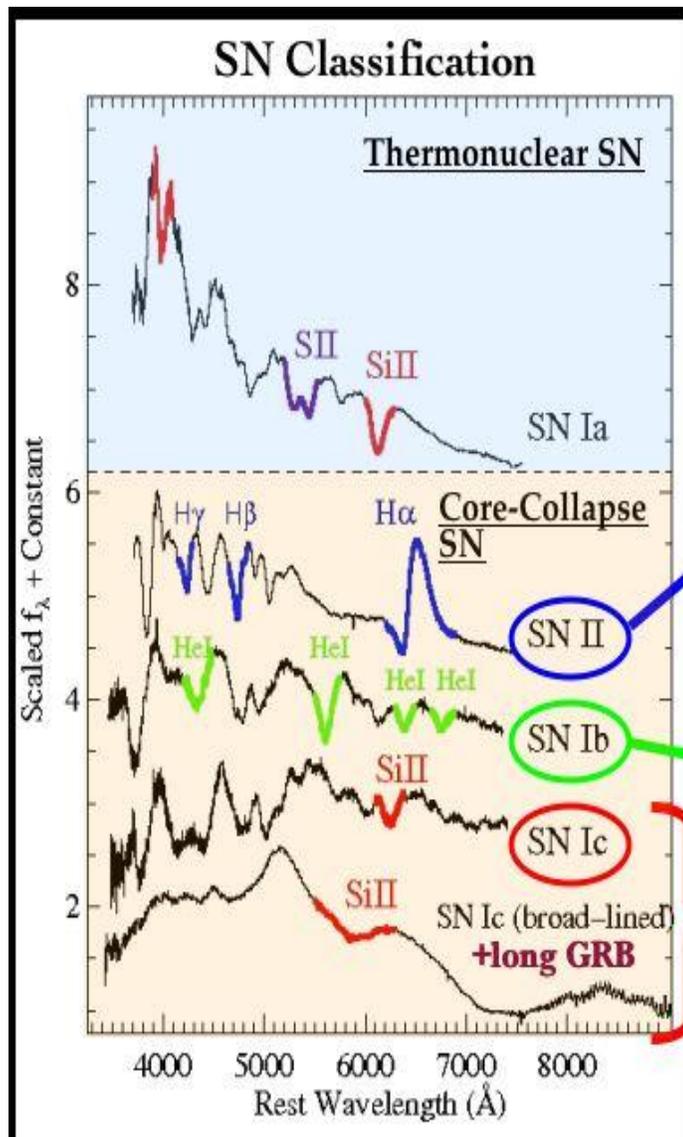
barred spiral galaxy



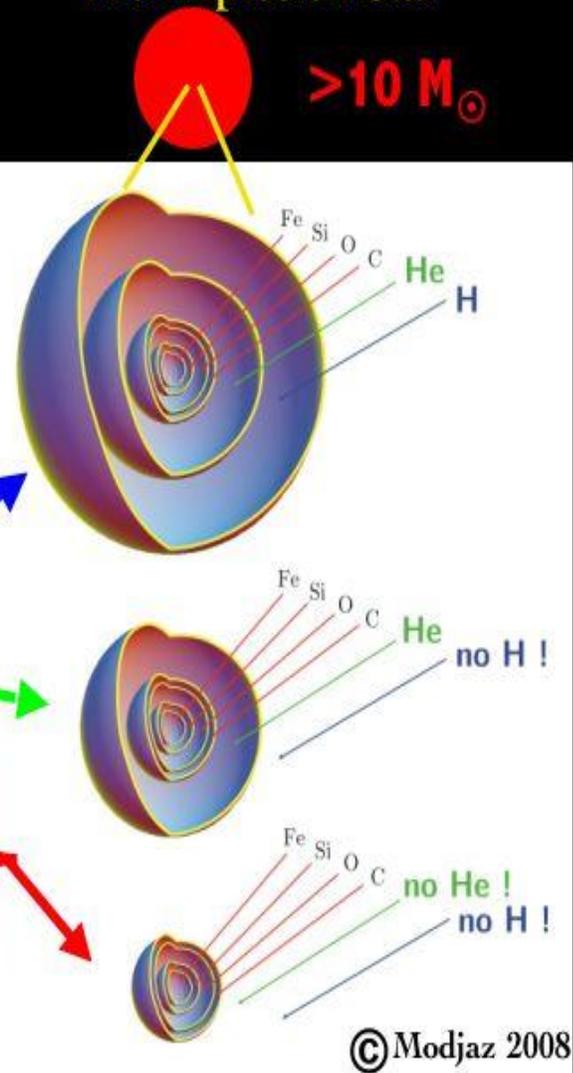
type II irregular galaxy



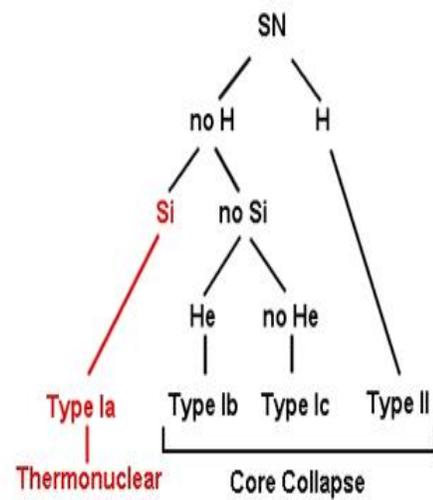
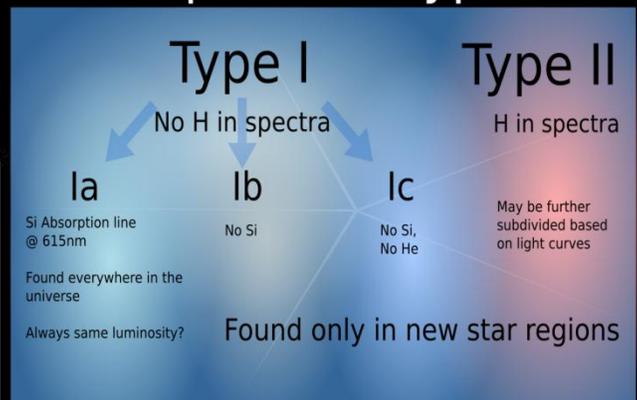
# 天文的应用：超新星分类

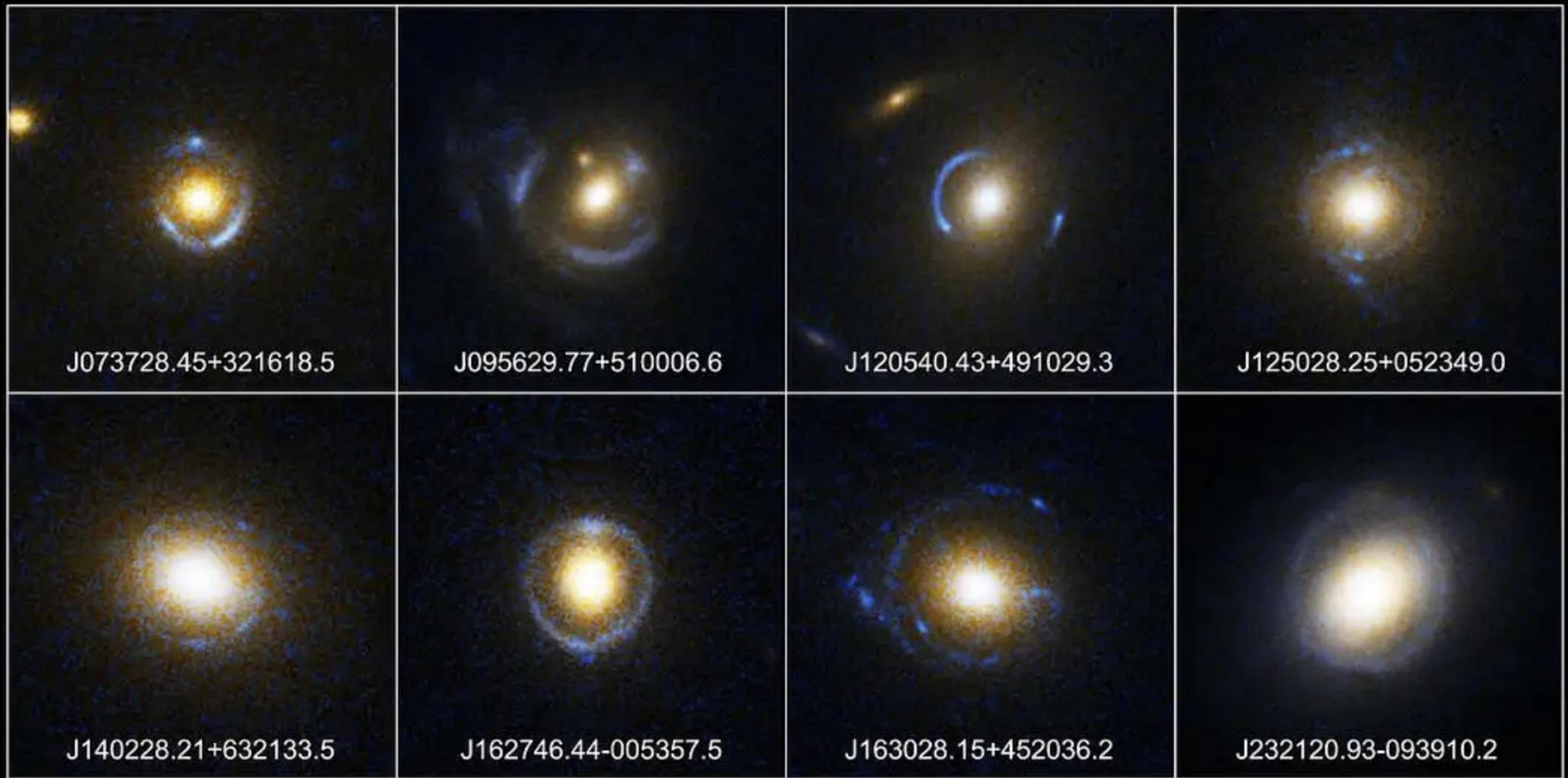


### Pre-Explosion Star



### Supernova Types





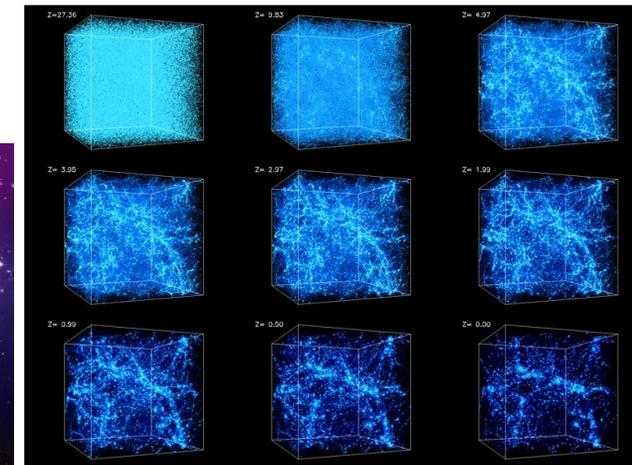
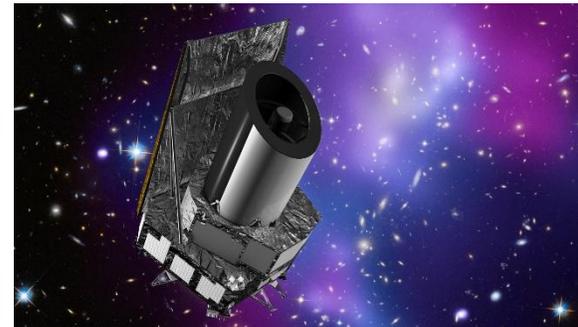
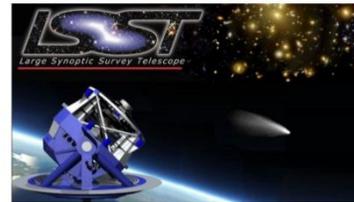
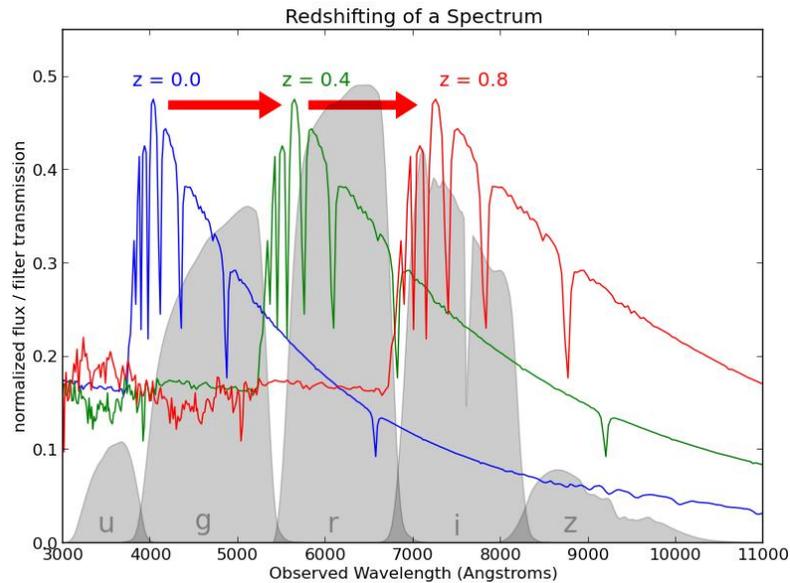
**Einstein Ring Gravitational Lenses**  
*Hubble Space Telescope • Advanced Camera for Surveys*

# 天文的应用：测光红移预测

- 光谱数据的获得比图像数据获得更费时、费力，尤其大型望远镜的时间难以获得
- VST, HST, Pan-STARRS, KiDS, DES, LSST, SKA, EUCLID

有助于：

- 确定宇宙中的暗物质和暗能量
- 限制宇宙学参数
- 研究弱引力透镜
- 重建宇宙大尺度结构
- 天体分类



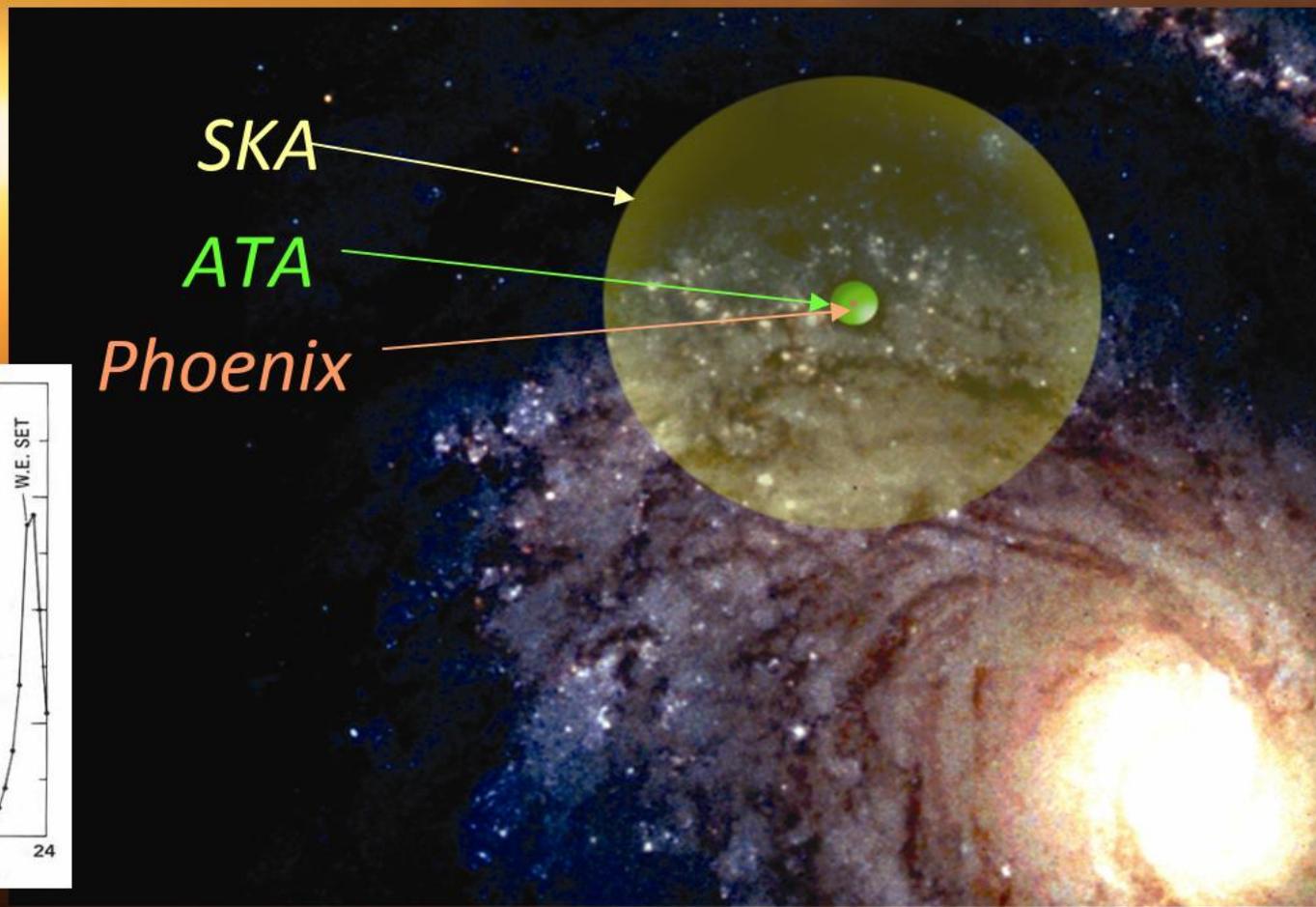
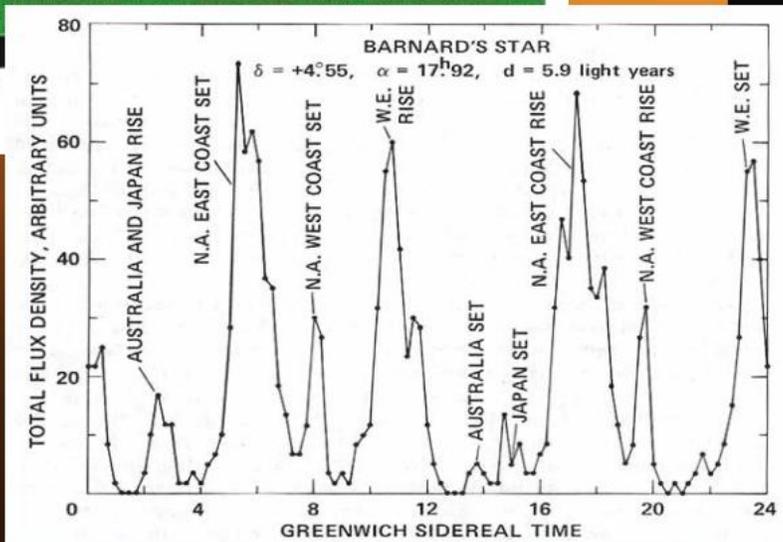
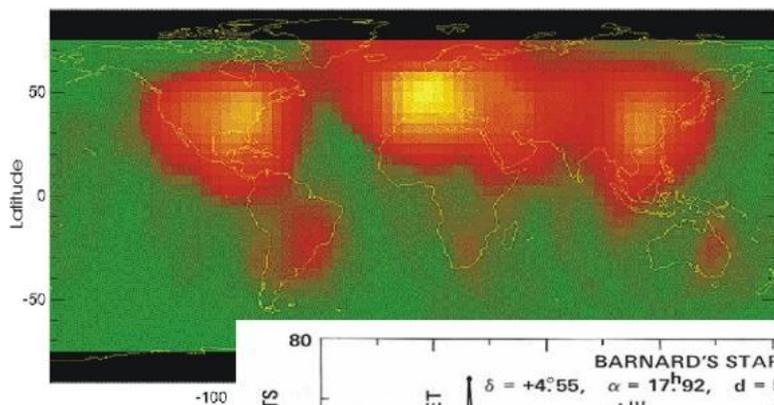
多波段交叉认证



# Are we alone? – is anyone out there?

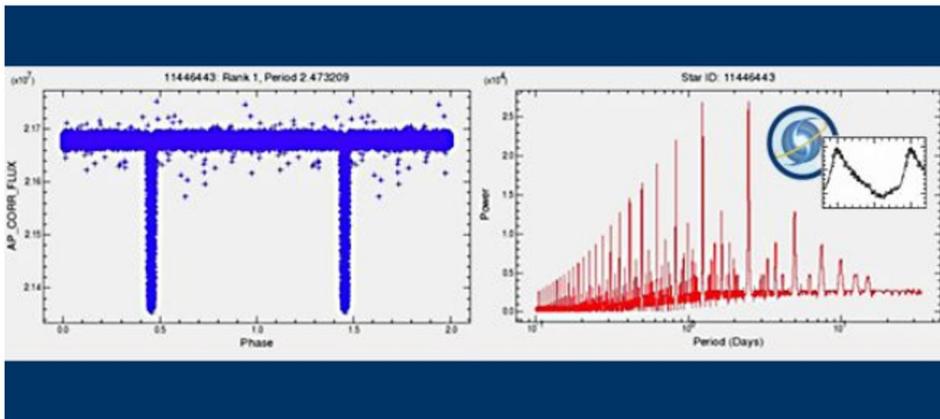
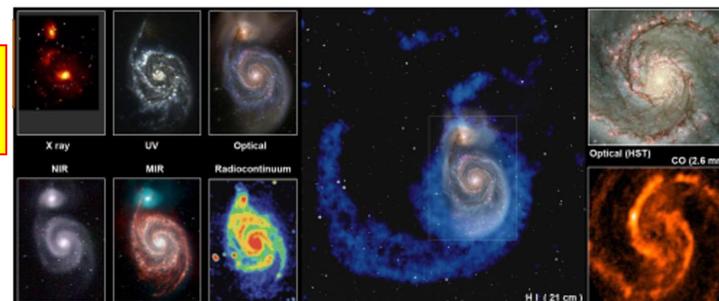
## The Search for Extra-terrestrial Intelligence

TERRESTRIAL INTERFERENCE



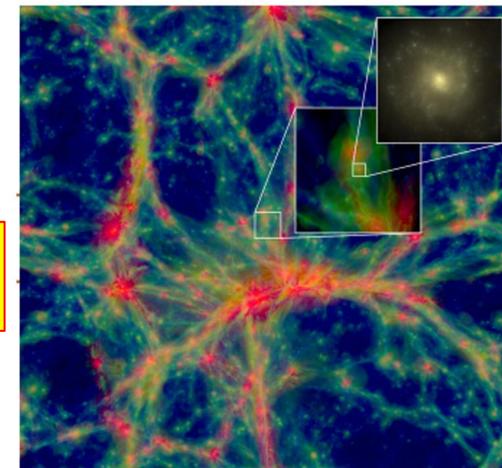
# 重要的天文发现可能来自于:

1. 来自多波段数据的交叉相关



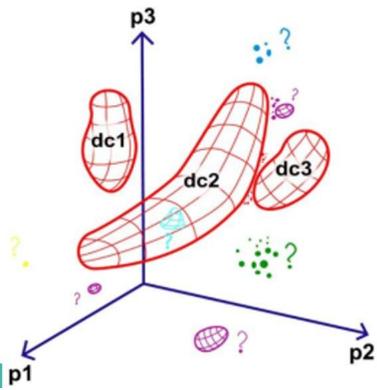
2. 观测源的时域特性

3. 数值模拟与实测数据对比研究



4. 在高维 (大于等于3维) 观测空间中寻找模式或趋势

A Generic Machine-Assisted Discovery Problem:  
Data Mapping and a Search for Outliers



# 人工智能与天文学

---

这些应用展示了人工智能在天文数据处理中的多样性和广泛性。

通过利用人工智能技术，天文学家可以更好地处理、分析和理解海量的天文数据，发现新的天体现象、加深对宇宙的理解，并推动天文学领域的研究和探索。

人工智能技术的发展为天文学研究提供了强大的工具和方法，加速了数据驱动科学的进展，并推动了天文学领域的新发现和突破。

随着人工智能技术的不断发展和创新，将会有更多新的应用出现，帮助天文学家更好地利用和理解天文数据，揭示宇宙的奥秘，推动天文学领域的前沿研究。

# 智能时代天文大数据分析挑战

- 数据量大 (获取、存储、传输、分析等)
- 样本不完备、非平衡
- 缺少标注数据
- 特征工程: 特征选择、特征抽取、降维
- 未知统计分布
- 时序数据 (不规则取样、不同波段)
- 异方差、删失数据、缺值数据
- 不可靠的数据质量 (未知的系统或随机误差)
- 算法和模型的选择与优化 “no one size fits all”
- 模型的可解释性、可扩展性、健壮性
- 可视化 (机器学习的各个阶段)
- 人机交互

瓶颈: 不在获得数据, 而在如何有效地利用 AI、ML、DL 从数据中挖掘出有用的、高价值的信息和知识

# 人工智能有哪些缺陷

---

- 1.数据偏差和过拟合问题:** 人工智能的算法需要大量的数据来进行训练和学习,但是如果数据集存在偏差或噪声,或者模型出现过拟合,那么将会导致预测结果的不准确性。
- 2.解释性问题:** 某些人工智能模型比较复杂,难以解释模型的内部逻辑和决策过程。这使得人们难以理解模型的推理过程,从而限制了人们对模型的信任和可接受度。
- 3.缺乏创造性:** 虽然人工智能可以在很多领域进行优秀的表现,但是它缺乏人类的创造力和想象力。人工智能的行为通常是固定的,无法像人类一样自由地产生新的想法和概念。
- 4.伦理问题:** 人工智能的发展也带来了一些伦理问题,例如人工智能对就业和职业的影响,隐私和数据安全问题,以及武器化和战争问题等。
- 5.没有通用的智能:** 目前的人工智能技术主要是针对特定的任务和问题,还没有实现通用的人工智能,这也是人工智能发展的一个重要瓶颈。

这些缺陷和局限性是人工智能发展过程中需要解决的重要问题,需要通过更加广泛和深入的研究和讨论来逐步解决。

# 人工智能能战胜人类吗

---

目前，人工智能还没有达到完全超越人类的水平。虽然在某些特定的任务上，人工智能已经超过了人类的表现，但是在复杂的任务、创造性的任务、情感和人际交往等方面，人类仍然具有独特的优势。

未来，人工智能的发展可能会给某些职业带来变革，但是人工智能与人类的关系不应该是对立的，而应该是互补的。人工智能可以为人类提供更多的助力，帮助人类更好地完成任务和创造价值，但是在决策和管理领域，人类的智慧和道德判断仍然是不可替代的。

因此，人工智能与人类应该是一种合作关系，而不是竞争关系。我们需要在推进人工智能技术发展的同时，也应该关注其对人类社会的影响，制定相应的政策和规范，确保人工智能的应用是安全、可靠、公正和可控的。

# 大数据、智能、合作共赢的时代

---

- 多学科合作
- 跨界合作
- 跨平台合作
- 虚拟组织
- 培养面向AI的下一代天文学家

合作

拥抱人工智能

拥抱大数据





君子生非异也，善假于物也。

《荀子·劝学》

未来打败你的会是使用AI的人

“如果我看得更远，那是因为我站在巨人的肩膀上。(If I have seen further it is by standing on the shoulder of Giants.)”(Newtown,I. 1676)