

# 月球与行星科学多模态大模型的概念及应用初探

雷丹泓<sup>1</sup>, 刘建忠<sup>1,2\*</sup>, 朱凯<sup>1</sup>, 刘敬稳<sup>1</sup>, 任曼<sup>1,3</sup>, 张圣<sup>1,3</sup>, 张莉<sup>1,3</sup>

1. 中国科学院 地球化学研究所, 月球与行星科学研究中心, 贵阳 550081; 2. 中国科学院 比较行星学卓越创新中心, 合肥230026; 3. 中国科学院大学, 北京100049

**摘要:** 自2022年底一系列大型语言模型(大模型)发布以来, 以大模型为研究热点的新一代人工智能、(AI)技术在各行各业引起高度重视, 随即出现的专业大模型更将重塑千行百业。月球与行星科学领域作为科学技术前沿的代表, 建立该领域的专业大模型是今后的必然趋势。本文在多模态大模型的基础上首次提出了月球与行星科学多模态大模型的概念, 并从其定义、设计及应用场景等细节展开详细阐述和讨论。模型旨在帮助更多人认识月球并了解月球, 同时协助更多学者处理并分析月球与行星数据, 促进挖掘月球与行星大数据间的隐式或非线性关系, 主动发现并解决未知科学问题, 率先推动大数据驱动科学发现这一新的研究范式在行星科学领域落地生根。

**关键词:** 月球与行星科学; 多模态大模型; 大数据; 人工智能

**中图分类号:** P184 **doi:** 10.19658/j.issn.1007-2802.2023.42.113

## A first look at the concept and application of multimodal large models for lunar and planetary science

LEI Dan-hong<sup>1</sup>, LIU Jian-zhong<sup>1,2\*</sup>, ZHU Kai<sup>1</sup>, LIU Jing-wen<sup>1</sup>, REN Man<sup>1,3</sup>, ZHANG Sheng<sup>1,3</sup>, ZHANG Li<sup>1,3</sup>

1. Center for Lunar and Planetary Sciences, Institute of Geochemistry, Chinese Academy of Science, Guiyang 550081, China;

2. CAS Center for Excellence in Comparative Planetology, Hefei 230026, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** Since the release of a series of Large Language Model (LLM) at the end of 2022, the new generation of Artificial Intelligence (AI) technology LLM has become a research hotspot and has attracted great attention, and the ensuing domain-specific LLM will definitely reshape thousands of industries. The field of lunar and planetary science is a representative of the frontiers of science and technology, and the establishment of domain-specific LLM in this field is an inevitable trend for the future. This article proposes the concept of Lunar and Planetary Multimodal Large Language Model (LP-MLLM) for the first time, and elaborates and discusses the details of its definition, design, and application scenarios. LP-MLLM aims to help more people know the Moon and understand it, as well as assist more scholars to process and analyze the lunar and planetary big data, promote the excavation of implicit or non-linear relationships among different lunar and planetary data, proactively discover and solve unknown scientific problems, all of which are leading the way in driving the new research paradigm of big data-driven scientific discovery to take root in lunar and planetary science.

**Key words:** lunar and planetary science; multimodal large language model; big data; artificial intelligence

收稿编号: 2023-159, 2023-7-10 收到, 2023-7-28 改回

基金项目: 中国科学院战略性先导科技专项项目(XDB41000000); 中国科学院重点研发计划项目(2022FYY0503100); 中国地质调查局地质调查项目(DD20221645)

第一作者简介: 雷丹泓(1997—), 女, 硕士, 助理工程师, 研究方向: 行星遥感、人工智能. E-mail: leidanhong@mail.gyig.ac.cn.

\*通信作者简介: 刘建忠(1968—), 男, 博士, 研究员, 研究方向: 行星地质学. E-mail: liujianzhong@mail.gyig.ac.cn.

### 引用此文:

雷丹泓, 刘建忠, 朱凯, 刘敬稳, 任曼, 张圣, 张莉. 2024. 月球与行星科学多模态大模型的概念及应用初探. 矿物岩石地球化学通报, doi: 10.19658/j.issn.1007-2802.2023.42.113

Lei D H, Liu J Z, Zhu K, Liu J W, Ren M, Zhang S, Zhang L. 2024. A first look at the concept and application of multimodal large models for lunar and planetary science. Bulletin of Mineralogy, Petrology and Geochemistry, doi: 10.19658/j.issn.1007-2802.2023.42.113

## 0 引言

人工智能(artificial intelligence, AI),作为20世纪以来最为深刻和最具影响力的科技革命,其研究始于1955年著名科学家艾伦·纽厄尔和奥利弗·塞弗里奇合作开发的首个人工智能程序LogicTheorist(于文轩等, 2023)。次年“人工智能之父”约翰·麦卡锡(John Mc Carthy)首次提出了“人工智能”一词,并确定了其未来的研究方向和目标。自此,人工智能正式成为一个研究领域(崔雍浩等, 2019;于文轩等, 2023)。随后一系列的人工智能成果相继问世,如机器定理证明、跳棋程序等(李红霞, 2007)。但由于计算能力的限制,20世纪70年代中期人工智能迎来了首个低谷期。而在1980年代出现的专家系统(朱祝武, 2011)和人工神经网络(吕伟等, 2018),则为人工智能带了第二次研究热潮,但很快专家系统便陷入了技术瓶颈,也再一次使人工智能跌入了低谷。但这一阶段的另一成果——人工神经网络,却取得了革命性的进展,这也使得未来大规模的神经网络训练成为了可能,也为人工智能的第三次兴起和爆发奠定了基础。2011年,IBM公司的自然语言问答计算机Watson(Ferrucci, 2012),2012年的新一代卷积神经网络AlexNet(Krizhevsky et al., 2012)以及2016年Google的AlphaGo(Silver et al., 2016)等人工智能成果和产品的出现,正式揭开了以深度学习为技术核心的第三次人工智能热潮。

人工智能问世后,人类一直在探索如何让机器掌握语言智能能力。而深度学习的发展,也进一步让语言建模从传统的统计语言模型(Statistical Language Model, SLM)(Jelinek, 1997; Gao and Lin, 2004)发展为神经语言模型(Neural Language Model, NLM)(Bengio et al., 2000; Mikolov et al., 2015)。近年来,通过大规模语料库的训练,基于迁移网络(Transformer)架构(Vaswani et al., 2017)的预训练语言模型(Pretrained Language Model, PLM)在解决各类自然语言处理(Natural Language Processing, NLP)任务上表现出了强大的能力,将NLP领域的研究提升到了一个新的发展阶段,这也激发了后续大量语言模型的研究(Peters et al., 2018; Devlin et al., 2018)。随着研究不断的深入,科研人员发现扩大预训练模型能够进一步提高模型的推理能力,并表现出了不同的行为和惊人的“涌现能力”(Wei et al., 2022a),此后研究界正式将这些大型预训练模型命名为“大型语言模型”(Large Language Model, LLM)(Shanahan, 2023; Wei et al., 2023; Zhao et al., 2023)。

大型语言模型也称为大模型,是现今大数据和AI

时代的产物,其发展速度更是日新月异,也是各国新一轮技术竞争的核心领域。大模型由数百亿甚至更多参数的神经网络所组成,不仅实现了对人类语言的仿真输出,更实现了对人类复杂语言系统的理解,能够执行一系列复杂的语言任务(何哲等, 2023)。根据应用范围,大模型实际又可分为通用大模型和专业大模型两种。通用大模型在各领域和学科中具有强大的泛化能力,而专业大模型则是通用大模型在专业知识基础上进行微调(Fine-tuning)所得到的,并在特定领域内的任务解决上有着更好的表现力(Wu et al., 2023)。目前,各国各业争相发布了自己的通用大模型,包括中国的华为盘古、百度文心一言、阿里通义千问、科大讯飞星火,以及国外的ChatGPT、LLaVa等模型。其中以GPT-3.5为基础进行微调得到的单一模态大模型ChatGPT(Brown et al., 2020),一经问世便产生了巨大反响。在2023年3月,OpenAI进一步在GPT-3.5的基础上发布了GPT-4多模态大型语言模型(Multimodal Large Language Model, MLLM)及ChatGPT Plus聊天机器人(OpenAI, 2023a)。以GPT-4为代表的多模态大模型,进一步扩展了大模型的视野和能力范围,突破了大型语言模型仅限于文本训练的局限性,解锁了多模态输入,再次刷新了大模型在低资源和零数据场景下的语言理解和生成能力,进一步加速了AI与千行百业的融合进程。

尽管通用大模型在各领域上普遍具有较好的表现力,但若直接用其来解决特定领域的复杂问题,通常会得到许多片面的回答和错误的信息(表1),究其原因主要是不同领域知识的多样性和复杂性以及训练样本的有限性所致(Ling et al., 2023)。最近,越来越多的团队对通用大模型在特定领域专业化方面进行了研究和实践。目前生物医学(Gu et al., 2021; Jin et al., 2023; Li et al., 2023; Zhang et al., 2023)、金融(Wu et al., 2023)、教育(InternLM Team, 2023)、交通(Zheng et al., 2023)等领域已逐步公布了相应领域的专业大模型,以弥补通用大模型在垂直方向上的短板。

而深空探测领域作为高新科技一直以来的代表和前沿,也开展了大模型的研究工作,中国探月与航天工程中心正在与百度共同建立“航天-百度·文心”大模型。月球探测是我国深空探测的首要任务之一,“数字月球”(孙鹏举等, 2022)的建设和“元月球”(任曼等, 2023)概念的提出,将为月球探测任务的高效实施和月球科学的快速发展提供智慧化的数据支持与技术保障。在经历了功能设计、技术研究、框架搭建和完善等过程后,“新一代数字月球云平台”应运而生并逐渐成型,本着集科学研究、工程实施和科普教育为一体的理

表1 通用大模型关于月球与行星科学领域问题的回答实例

Table 1 The example of the General Language Model's answer in the field of lunar and planetary science

通用大模型关于月球与行星科学领域问题的回答	
模型:	问:南极-艾肯盆地的形成对月球演化的影响?
Chatgpt(基于GPT-3.5)	<p>答:南极-艾肯盆地是月球上最大的撞击盆地之一,形成于大约38亿年前。它的形成对月球演化产生了广泛而重要的影响。首先,南极-艾肯盆地的撞击事件释放了巨大的能量,导致了岩石的熔融和挥发物的释放。这个撞击事件产生的巨大能量释放形成了一个巨大的撞击坑,深达13 km,直径约为2500 km。撞击事件将大量的岩石和尘埃抛射到空中,并形成了一个巨大的撞击坑环。这些抛射物落回月球表面,覆盖了广泛的区域,形成了许多撞击坑和山脉。其次,南极-艾肯盆地的形成引发了岩石圈的扰动和熔融。撞击事件产生的能量使地壳熔融,并形成了岩浆。这些岩浆从地下涌出,形成了月球上许多火山活动和熔岩平原。这些火山活动和熔岩平原的形成改变了月球表面的地貌,并对月球的地质演化产生了深远的影响。此外,南极-艾肯盆地的撞击事件释放了大量的热能,导致了月球内部的地热活动。这种地热活动可能导致了地壳的变形、火山活动的增加以及热液的产生。这些地热活动对月球的地质演化和地表特征的形成产生了重要的影响。</p> <p>总的来说,南极-艾肯盆地的形成对月球演化产生了多方面的影响。它改变了月球表面的地貌,形成了撞击坑和山脉,同时也引发了火山活动和熔岩平原的形成。此外,它还导致了地热活动的增加,对月球的地壳变形和地热特征产生了影响。这些影响对我们理解月球的演化历史和地球外行星的撞击过程都具有重要意义(OpenAI, 2023b)。</p> <p>评价:回答内容较详细,层次划分也比较清晰。但是该回答在某些知识点上却出现了明显的错误,同时对于一些关键知识点的回答也没有给出出处,整体回答有明显漏洞且不够严谨。首先南极-艾肯(SPA)盆地的形成时间是43亿年前左右,而回答里却说是38亿年前;其次这句话“这些抛射物落回月球表面,覆盖了广泛的区域,形成了许多撞击坑和山脉。”后半句描述不够严谨,应该是“形成了许多二次撞击坑和山脉”,回答里还将“月亮”描述为了“地壳”;另外回答里也未提及撞击事件对月球正面的影响,主要是SPA撞击事件与风暴洋形成之间的关系</p> <p>从上述的回答细节中,可知ChatGPT关于月球科学问题的回答存在多处明显错误且在细节描述方面不够严谨,同时也存在一定的局限和片面性。一方面可能是由于模型在获取数据来源时并未对其正确性进行分辨,导致回答内容中出现了明显的错误。另一方面可能是由于模型的训练数据比较广泛而对于专业领域内的知识不够深入,造成了目前答案的片面性。</p>
模型:	问:南极-艾肯盆地的形成对月球演化的影响?
必应(基于GPT-4.0)	<p>答:南极-艾肯盆地是月球上最大、最古老的撞击盆地之一,也是太阳系最大的撞击盆地之一。它的形成对月球演化产生了重要影响。据中国科学院地质与地球物理研究所的张金海等人的研究,南极-艾肯盆地的形成可能是由于一颗直径约为500 km的小行星撞击月球而形成的。该撞击事件可能导致了月球内部物质的混合和熔融,从而影响了月球的演化。(源:与必应的对话, 2023/7/4)</p> <p>评价:该回答比较简洁,同时也给出了相应的出处,一定程度上保证了回答的可信度。然而必应的大部分回答内容都在描述南极-艾肯盆地和其成因,而对于提问的核心即其对月球演化的影响,却只提供了有限的假说。在模型的回答里只给出了以张金海学者为代表的观点,并未提及假说和理论,也尚未提及SPA撞击事件溅射物对月球表面物质分布的影响。整体回答效果比较片面,这可能是由于模型数据库在专业知识上比较有限,缺少足够的训练数据所造成的。</p>

念,目前数字月球已实现了对月球科学数据的数字化采集、检索、表达和初步应用等功能。打造智慧化的数字月球云平台一直是数字月球努力的方向,无论是其海量的多源数据,还是其用户至上的理念均与多模态大模型的关键核心高度吻合。在当今AI大模型生态的背景下,结合数字月球与多模态大模型,发展月球与行星科学多模态大模型,是将AI落实到行星科学领域的必然趋势。专业模型的建立能够有效为月球与行星科研工作者以及数字月球用户提供更加精准、可信的数据检索、知识回答及科学数据分析和挖掘等服务,真正实现大数据驱动科学发展的目标。

本文以大模型和数字月球为理论基础,首次提出月球与行星科学多模态大模型概念、框架及应用场景,并对模型训练数据和核心技术等细节进行深入讨论。

### 1 月球与行星科学多模态大模型定义

随着大数据和AI的快速发展,以及数据信息化和智能化程度的不断加深,科学研究正在从实验科学(第一范式)、理论科学(第二范式)和计算科学(第三范式),向数据密集型科学的第四范式——“大数据驱动科学发现”的新范式转型(图1)。第四范式强调通过对

大规模数据的挖掘和分析发现新科学问题和认识,突破了以往科学研究中的学科对立、数据质量有限等局限,并促进了不同学科之间的相互相容,重建了科学预测的可能性(米加宁等,2018)。特别是在数据科学和自然语言处理领域中,大型语言模型从已知文本中生成未知文本的能力完美诠释了第四范式,这也正是大数据驱动科学发现的最好实例。

月球与行星科学多模态大模型(lunar and planetary multimodal large language model, LP-MLLM),是建立在数字月球云平台之上的嵌入式多模态大模型,支持从文献、期刊、报告等文本数据中自动挖掘并获取有用知识,同时也支持从影像、光谱、雷达等多源探测数据中学习知识并发现问题,实现数字月球的智能化分析和挖掘功能,完成数字月球云平台的愿景。该模型具备多模态多任务的处理能力,为数字月球云平台用户提供更专业的目标检测、算法生成、图文交互、知识挖掘、研究引导等多样化服务,以协助用户进行更高效的知识获取和数据分析任务,挖掘数据中潜在的因果关系,发现并解决科学问题。LP-MLLM是在第四范式和大模型驱动下所生成的产物,其中多模态大模型和月球与行星大数据是其关键核心。建立LP-MLLM是

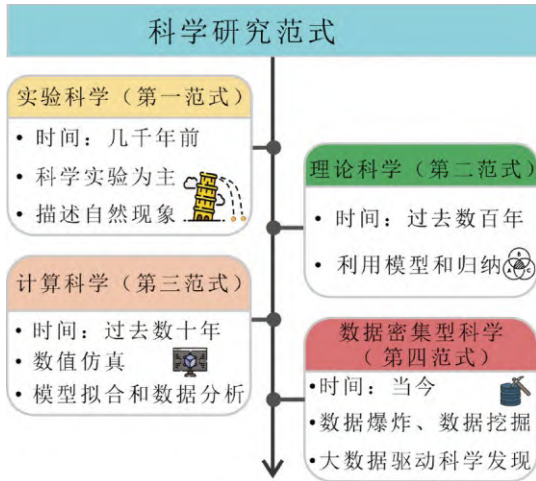


图1 科学研究范式的发展(修改自Hey et al.,2009;邓仲华和李志芳,2013)

Fig.1 The development of scientific research paradigm (modified after Hey et al.,2009,Deng and Li,2013)

月球与行星科学领域的必然发展趋势,也是实现大数据驱动科学发现在行星科学领域落地生根的必要途径。

## 2 月球与行星科学多模态大模型设计

### 2.1 多模态数据集 (Multimodal Dataset)

数据、计算能力和算法是大模型的三大关键要素,而数据集则是所有模型训练的前提,因此高质量的数据集是保障日后模型精确度的关键(Yuan et al.,2022)。在本文中,多模态数据集是LP-MLLM建立和训练的必要因素之一,除了现有的数字月球数据库,多模态数据集将进一步在其他行星探测数据和科学数据(包括火星、水星、金星等行星)的基础上建立。目前,数字月球云平台的数据总量已超过1 PB(孙鹏举等,2022),并集成了大量的月球多模态数据,包括相关期刊文献、技术报告、平台说明书等文本类数据,以及一系列非文本数据,如多媒体科普视频、算法程序代码、

影像、光谱以及雷达等探测数据。对于其他行星的探测数据与科学成果数据,主要包括国内天问一号对火星的探测数据,国际上其他国家火星、水星、金星等行星的探测数据,以及相关矿物元素的含量分布图、地质图及文献期刊等成果数据(表2)。通过对上述数据进行有效筛选、重复数据删除(Lee et al.,2022)、清洗以及不同模态数据的统一(Zhu et al.,2022;Li et al.,2022;Zhang et al.,2023)等数据处理工作,将得到适用于LP-MLLM训练的高质量数据集。

### 2.2 模型组成

LP-MLLM的主要组成包括多模态大型语言模型和多模态知识库两部分。

2.2.1 多模态大型语言模型(MLLM) 在月球与行星科学领域中,很多科学问题和应用场景中的数据都是多模态混合的,因此建立能够同时处理并返回多种模态数据的大模型,是数字月球通向智慧化的重要一步。本文提出的LP-MLLM拟建立在开源的多模态大模型算法基础上,通过“预训练+微调”的形式来增强其在月球与行星科学领域上的知识理解能力和推理能力。在预训练阶段,模型利用未标注的多模态数据集进行无监督的预训练,学习多模态数据之间的潜在关系和表示。而在微调阶段,模型需要在月球与行星科学相关任务的文本、图像、光谱等多模态数据的标注集上,对模型的部分网络层进行微调。然而随着AI技术的发展,大模型的参数也越来越多,传统的微调方法已经不再适用,LoRA(Low-Rank Adaptation)(Hu et al.,2021)、P-tuning v2(Liu et al.,2022)、Freeze等高效的参数微调方法,则解决了现阶段模型参数量太多的问题,同时保证了微调的效果。其中,LoRA可以与其它微调方法相结合,使微调更具资源效率和可访问性。它通过只训练在模型中新增的网络层参数和高阶矩阵秩的分解方法,大大减少了微调的参数量、节省了模型的训练成本,并且还能取得与全模型微调方法类似的

表2 多模态数据集的组成

Table 2 The composition of Multimodal Dataset

内容	数据描述	数据量
数字月球数据库——月球探测数据	嫦娥系列工程CE-1/2/3/4,国外月球探测任务包括Apollo、Clementine、Lunar Orbiter、Lunar Respector、Galileo、GRAIL、LRO、SELENE和Chandrayaan-1等搭载的各种探测载荷所获取的探测数据	约1056 TB
数字月球数据库——科学成果数据	全月1:250万地质图、岩石类型分布、月球构造纲要图、撞击坑分布、美国月球地质图、粗糙度图、矿产资源、太阳能资源、氦3分布、月表温度、空间环境数据、宇宙射线数据、月球地名点数据、控制点数据、月球物质标准光谱数据、基础地形数据、历史图片、月球期刊文献、科研报告、视频音频数据、算法程序代码及其说明文件、数字月球云平台建设代码、技术报告以及相关使用手册等数据	约93 TB
其他行星探测和科学成果数据	国内天问一号火星探测任务,国外火星、水星、金星等探测任务包括Mars Express、Mars Reconnaissance Orbiter、ExoMars、2001 Mars Odyssey、Mars Global Surveyor、Viking Orbiter 1、Viking Orbiter 2、MESSENGER、Magellan等搭载的各种探测载荷所获取的探测数据。以及科学成果数据,包括火星地质图、热辐射成像系统(TES)矿物含量分布图(包括赤铁矿、斜长石、高钙辉石、石英、橄榄石等矿物)、山体阴影、岩石丰度、水星地质图、表面元素和矿物含量、金星地质图、撞击坑分布以及相关期刊文献、视频音频数据等	约600 TB

效果,也更符合月球与行星科学多模态大模型在微调阶段的需求。

多模态大模型将大型语言模型作为“大脑”来执行多模态任务,拥有惊人的涌现能力。它能够接受多种类型的训练数据,不仅可以将文本数据进行编码,还可以将其他类型的数据编码到相同空间,这也使得模型能够通过单一的机制实现所有类型数据的计算。多模态大模型可具体分为三个关键技术和一个实现方式:多模态指令微调、多模态上下文学习、多模态思维链以及 LLM辅助的视觉推理(Yin et al.,2023)。

(1)多模态指令微调(Multimodal Instruction Tuning, M-IT):是指通过指令的形式对多模态任务进行描述并生成相应的多模态指令数据(图2),以此来微调多模态大模型,从而提高模型在多模态数据上的理解和执行能力(Wei et al.,2022b;Liu et al.,2023)。它的基本形式主要包含三个元素:指令、多模态输入及回答。



图2 多模态指令数据的概念示例

Fig.2 The conceptual example of M-IT

(2)多模态上下文学习(Multimodal In-Context Learning, M-ICL):上下文学习不同于传统从大量数据中学习隐式关系的监督学习范式,它的关键思想是从类比中学习(Dong et al.,2023)。目前上下文学习与大型语言模型相结合的方法,在许多零样本条件下的任务上取得了良好的成绩。而如今在多模态大模型的背景下,上下文学习的应用范围已从原始的文本数据扩展到了更多类型的数据,我们将之称为多模态上下文学习(图3)。

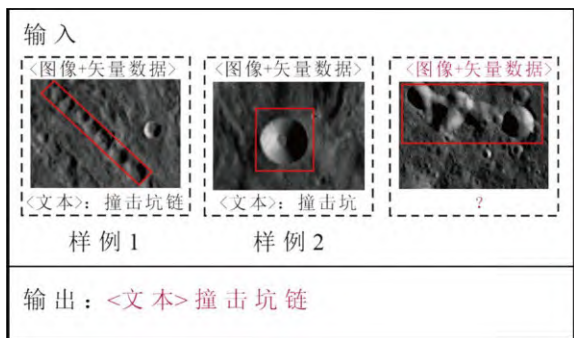


图3 多模态上下文学习的概念示例

Fig.3 The conceptual example of M-ICL

(3)多模态思维链(Multimodal Chain-of-Thought, M-CoT):思维链(CoT)的核心目标是促使大模型不仅仅只输出最终答案,同时还需要将相应答案的推理过程进行返回(Yin et al.,2023)。受NLP的启发,单模态的思维链已逐渐向多模态思维链进行扩展,通过语言文本类数据和其他多模态数据的结合,模型可以对答案进行更深入的推导和反演(图4)。

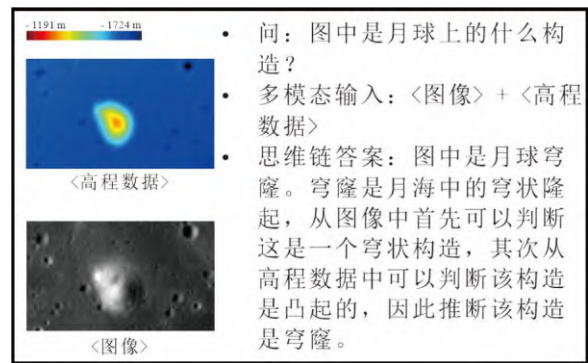


图4 多模态思维链的概念示例

Fig.4 The conceptual example of M-CoT

(4)LLM辅助的视觉推理(LLM-Aided Visual Reasoning, LAVR):这是一种利用大型语言模型来增强视觉推理能力的技术,该实现方式通过调用各种多模态数据与工具以辅助多模态大模型执行更加复杂的任务。随着多模态数据类型的不断增加,多模态大型语言模型的感知能力还会进一步在“视觉”基础上得到拓展。

2.2.2 月球与行星多模态知识库 传统的知识库往往需要人工或机器通过大量文本分类、文本摘要及信息提取等任务进行建立,尤其对于不同类型的任务,更需要设计专门的数据挖掘算法。而对于LP-MLLM,传统的知识库建立方法便不再适用。

月球与行星多模态数据集是建立多模态知识库的基础,多模态数据集包含了多样的月球与行星数据,如矢量、栅格和泛空间等数据。不同模态数据的存储类型、处理方式有所不同,在一定程度上会影响模型训练的效率 and 准确性。因此,为了建立高质量的月球与行星专业知识库,多模态数据间的模态对齐(Modality Alignment)是必不可少的(Yin et al.,2023)。模态对齐具体是指将不同模态的数据,如影像、光谱、雷达、视频等数据统一编码到相同的表示空间,并将不同任务统一为相同的任务范式,从而实现以相同的模型框架及模型参数来处理多模态任务。目前用于处理多模态任务的模型有:Transformer、CLIP(Radford et al.,2021)、ALIGN(Jia et al.,2021)等,其中Transformer模型中的注意力(Attention)机制能够适应各种类型的

数据(Huang et al., 2023),这一算法架构也使得多模态数据能够在模式对齐上表现得更加出色。因此在LP-MLLM中,拟将Transformer的解码器作为多模态数据输入的通用接口,为多模态数据集提供模态对齐的工具,进而建立多模态知识库。

### 2.3 模型框架与软硬件系统设计

本文提出的LP-MLLM具体框架结构如图5所示。它的基本组成模块主要包含多模态大模型与多模态知识库两部分。当面对复杂的专业领域问题时,模型能够通过跨模态编码器将不同模态的数据进行特征对齐,映射到统一的表示空间即多模态知识库。通过将多模态大模型在专业知识库的基础上进行微调,以实现模型的专业知识增强效果。

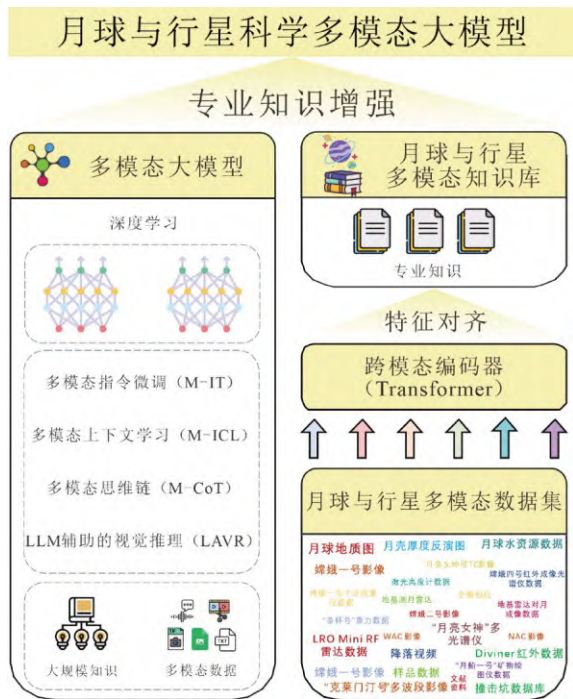


图5 月球与行星科学多模态大模型框架

Fig.5 The framework of LP-MLLM

尽管在微调阶段对基座的多模态大模型参数进行了冻结和不训练,但由于大模型巨大的参数量,训练一个月球与行星科学多模态大模型仍需大量的计算资源和算力。因此,将月球与行星科学多模态数据集上云到专属的数据存储空间,能够在如阿里云、腾讯云、Google Cloud等云计算平台提供的高性能AI计算资源基础上,实现月球与行星科学多模态大模型的高效训练,进而建立月球与行星科学多模态大模型(图6)。

### 3 月球与行星科学多模态大模型的应用场景

自20世纪60年代以来,人类开展了一系列月球探

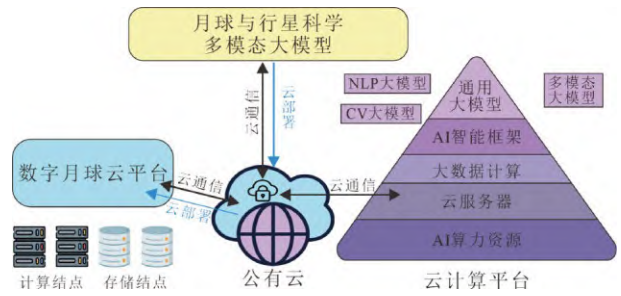


图6 月球与行星科学多模态大模型软硬件系统设计

Fig.6 The software and hardware design of LP-MLLM

测工作,为解决月球形成和演化、月表资源分布以及表面和深部物质组成和性质等问题提供了大量数据。随着探测能力的大幅度提升,月球与行星的探测数据也在不断增加。尽管有如此多的可用数据,但由于数据增长的速度远超过科学家的分析速度,导致目前有许多月球与行星科学的重要问题尚未解决(Varatharajan et al., 2021)。

大模型在各领域和各任务上,对于自然语言处理、图像分析、知识推理等方面表现出的强大能力,推动了AI在解决科学问题方面的发展,为上述问题的解决带来了希望。本文所提出的月球与行星科学多模态大模型旨在为用户提供友好的人机交互模式,并支持多模态数据的输入和多模态任务的执行。模型在月球与行星科学多模态知识库的基础上,能主动发现新的潜在关系和科学问题,指引并协助用户发现问题、解决问题。以目前大模型的技术发展和交互模式为背景,本文对LP-MLLM的应用场景进行了初探(图7),具体包括以下几个方面:

(1)文本回答:模型能够针对用户关于月球与行星科学的问题给出专业的解释和回答;

(2)目标识别:模型支持图像的输入,能够根据用户的输入需求,实现目标对象的自动识别任务;

(3)图文交互:模型支持图像和文本数据的并列输入,能够根据用户具体输入的图像数据和文本指令,给出相应的图文响应;

(4)多模态科学数据分析:模型支持多模态科学数据的输入和多模态任务的执行,能够分析用户当前的需求,对目标数据进行挖掘和分析;

(5)驱动科学发现:目前许多科学研究依赖于传统的物理建模或实验来分析、解释探测仪器的记录数据,这些方法可能会带来一定的偏差。而AI对数据强大的挖掘能力能在一定程度上能解决这些偏差,并发现新的潜在过程(Moseley, 2022)。模型支持用户与之进行深入交流,能够从模型本身的多模态知识库、历史用户的数据日志以及当前用户的输入数据中学习,

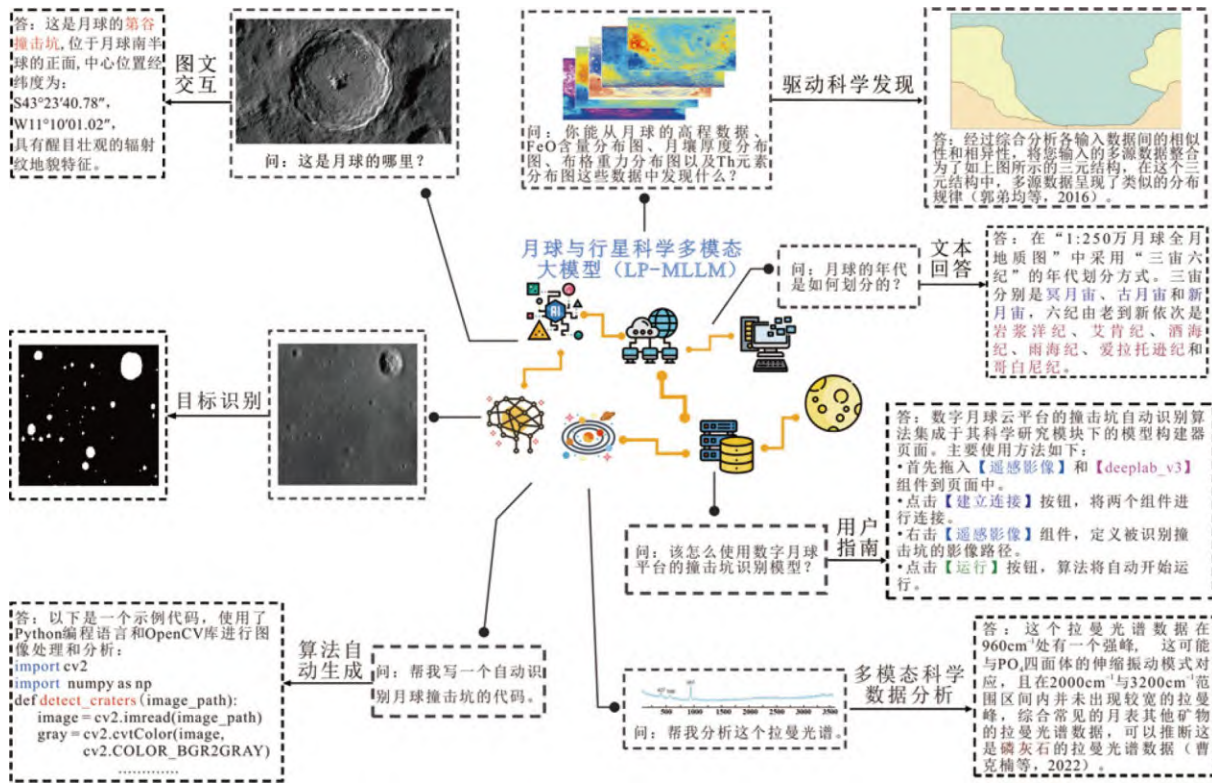


图7 月球与行星科学多模态大模型的应用场景示意图

Fig.7 The schematic diagram of application scenarios for LP-MLLM

并发现数据中新的隐式和非线性关系,提供更加自动化和高效的推理算法,指引用户发现并解决科学问题;

(6)算法自动生成:模型能够根据用户对算法的描述,生成相应的模型代码程序;

(7)数字月球云平台用户指南:模型可以为数字月球云平台提供专业的用户使用指南,一定程度上减少用户学习如何使用平台的时间成本。

#### 4 结论

大型语言模型的横空出现拉开了AI能够主动思考和解决问题的序幕。以ChatGPT为代表的生成式AI大大降低了命令计算机解决科学问题的门槛,减少了数据处理和分析的人工和时间成本,极大地推动了大模型垂向深入发展的进度。尽管目前大模型的表现并不是十分完美的,它们在某些领域和专业问题上仍存在一定的缺陷。但随即问世的多模态大模型,突破了大模型在跨模态任务上的处理瓶颈,解决了大模型在数据单一性和任务局限性方面存在的问题,这也意味着相较于单模态大模型,多模态大模型对于不同研究领域的数据拥有更好的包容性,可以处理并执行更多类型的数据和任务,使得多模态大模型具备了更强的数据分析和推理能力,因此各行业及研究领域加速了对多模态大模型开发和应用的进程。

本文在多模态大模型和数字月球云平台的基础上,首次提出了月球与行星科学多模态大模型,从模型的定义、组成、应用场景以及数据集等进行了详细阐述和讨论。模型旨在利用月球与行星科学多模态大模型,协助用户进行更高效和精确的专业知识获取和数据分析任务,引导用户发现数据的隐式和非线性关系,更新目前数字月球云平台的人机交互模式,在平台现有的二三维视觉交互基础上,增加语言的第四维度交互模式,推进AI在行星科学研究领域的深度应用,实现大数据驱动科学发现率先在行星科学落地生根,期望能够搭建基于自然语言交互的智慧化数字月球云平台。

**作者贡献声明:** 雷丹泓,图表制作,撰写文章;刘建忠、朱凯、刘敬稳,论文框架构建;任曼、张圣、张莉,数据整理。

**利益冲突声明:** 作者保证本文无利益冲突。

**致谢:** 感谢阿里云智能集团姜凯威提供的技术支持!

#### 参考文献 (References):

Bengio Y, Ducharme R, Vincent P. 2000. A neural probabilistic language model. In: Proceedings of the 13th International Conference on Neural Information Processing Systems. Denver, CO: MIT Press, 893-899

Brown T, Mann B, Ryder N, Subbiah M, Kaplan J D, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G,

- Henighan T, Child R, Ramesh A, Ziegler D M, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. 2020. Language models are few-shot learners. arXiv preprint, arXiv: 2005.14165.
- 曹克楠, 董明潭, 余振兵, 肖倩, 王心怡, 钱煜奇, 李毅恒, 汪在聪, 何琦, 巫翔, 宗克清, 胡兆初, 肖龙. 2022. 一种以极低的样品消耗同时测定嫦娥五号月壤粒度和矿物组成的新方法. 中国科学: 地球科学, 52(9): 1726–1736 [Cao K N, Dong M T, She Z B, Xiao Q, Wang X Y, Qian Y Q, Li Y H, Wang Z C, He Q, Wu X, Zong K Q, Hu Z C, Xiao L. 2022. A new method for simultaneous determination of grain size and mineral composition of Chang'e-5 lunar soil with extremely low sample consumption. *Scientia Sinica (Terrae)*, 52(9): 1726–1736(in Chinese)]
- 崔雍浩, 高聪, 陈懿奇, 郝建业. 2019. 人工智能综述: AI的发展. 无线电通信技术, 45(3): 225–231 [Cui Y H, Shang C, Chen S Q, Hao J Y. 2019. Overview of AI: Developments of AI techniques. *Radio Communications Technology*, 45(3): 225–231 (in Chinese with English abstract)]
- 郭仲华, 李志芳. 2013. 科学研究范式的演化: 大数据时代的科学研究第四范式. 情报资料工作, (4): 19–23 [Deng Z H, Li Z F. 2013. The evolution of scientific research paradigm: The fourth paradigm of scientific research in the era of big data. *Information and Documentation Services*, (4): 19–23 (in Chinese with English abstract)]
- Devlin J, Chang M W, Lee K, Toutanova K. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding.: arXiv: 1810.04805. <https://arxiv.org/abs/1810.04805.pdf>.
- Dong Q X, Li L, Dai D M, Zheng C, Wu Z Y, Chang B B, Sun X, Xu J J, Li L, Sui Z F. 2022. A survey on In-context learning.: arXiv: 2301.00234. <https://arxiv.org/abs/2301.00234.pdf>.
- Ferrucci D A. 2012. Introduction to “This is Watson”. *IBM Journal of Research and Development*, 56(3.4): 1: 1–1: 15
- Gao J F, Lin C Y. 2004. Introduction to the special issue on statistical language modeling. *ACM Transactions on Asian Language Information Processing*, 3(2): 87–93
- Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X D, Naumann T, Gao J F, Poon H. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare*, 3(1): 2
- 郭弟均, 刘建忠, 籍进柱, 刘敬稳, 王庆龙, 欧阳自远. 2016. 月球的全球构造格架初探. 地球物理学报, 59(10): 3543–3554 [Guo D J, Liu J Z, Ji J Z, Liu J W, Wang Q L, Ouyang Z Y. 2016. Preliminary study on the global geotectonic framework of the Moon. *Chinese Journal of Geophysics*, 59(10): 3543–3554 (in Chinese with English abstract)]
- 何哲, 曾润喜, 秦维, 郑磊, 张宏, 张效羽, 马亮, 翟云, 潘云龙, 李晓方, 彭云, 张楠. 2023. ChatGPT等新一代人工智能技术的的社会影响及其治理. 电子政务, (4): 2–24 [He Z, Zeng R X, Qin W, Zheng L, Zhang H, Zhang X Y, Ma L, Zhai Y, Pan Y L, Li X F, Peng Y, Zhang N. 2023. Social impact and governance of ChatGPT and other new generation artificial intelligence technologies. *E-Government*, (4): 2–24(in Chinese)]
- Hey T, Tansley S, Tolle K. 2009. Jim Gray on eScience: A Transformed Scientific Method. In: Hey T, Tansley S, Tolle K, ed., *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, Washington, xvii–xxxii
- Hu E J, Shen Y L, Wallis P, Allen-Zhu Z, Li Y Z, Wang S A, Chen W Z. 2021. LoRA: Low-rank adaptation of large language models.: arXiv: 2106.09685. <https://arxiv.org/abs/2106.09685.pdf>.
- Huang S H, Dong L, Wang W H, Hao Y R, Singhal S, Ma S M, Lv T C, Cui L, Mohammed O K, Patra B, Liu Q, Aggarwal K, Chi Z W, Bjorck J, Chaudhary V, Som S, Song X, Wei F R. 2023. Language is not all you need: Aligning perception with language models.: arXiv: 2302.14045. <https://arxiv.org/abs/2302.14045.pdf>.
- InternLM Team. A multilingual language model with progressively enhanced capabilities. [2023-06-07]. <https://github.com/InternLM/InternLM-tech-report>
- Jelinek F. 1997. *Statistical methods for speech recognition*. Cambridge, MA: MIT Press
- Jia C, Yang Y F, Xia Y, Chen Y T, Parekh Z, Pham H, Le Q V, Sung Y, Li Z, Duerig T. 2021. Scaling up visual and vision-language representation learning with noisy text supervision.: arXiv: 2102.05918. <https://arxiv.org/abs/2102.05918.pdf>.
- Jin Q, Yang Y F, Chen Q Y, Lu Z Y. 2023. GeneGPT: Augmenting large language models with domain tools for improved access to biomedical information. ArXiv: arXiv: 2304.09667v3.
- Krizhevsky A, Sutskever I, Hinton G E. 2012. ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, Nevada: Curran Associates Inc., 1097–1105
- Lee K, Ippolito D, Nystrom A, Zhang C Y, Eck D, Callison-Burch C, Carlini N. 2022. Deduplicating training data makes language models better. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 8424–8445
- Li C Y, Wong C, Zhang S, Usuyama N, Liu H T, Yang J W, Naumann T, Poon H, Gao J F. 2023. LLaVA-med: Training a large language-and-vision assistant for biomedicine in one day.: arXiv: 2306.00890. <https://arxiv.org/abs/2306.00890.pdf>.
- Li H, Zhu J G, Jiang X H, Zhu X Z, Li H S, Yuan C, Wang X H, Qiao Y, Wang X G, Wang W H, Dai J F. 2023. Uni-perceiver v2: A generalist model for large-scale vision and vision-language tasks. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2691–2700
- 李红霞. 2007. 人工智能的发展综述. 甘肃科技纵横, 36(5): 17–18 [Li H X. 2007. A review of the development of artificial intelligence. *Scientific & Technical Information of Gansu*, 36(5): 17–18(in Chinese)]
- Ling C, Zhao X J, Lu J Y, Deng C Y, Zheng C, Wang J X, Chowdhury T, Li Y, Cui H J, Zhang X C, Zhao T J, Panalkar A, Cheng W, Wang H Y, Liu Y C, Chen Z Z, Chen H F, White C, Gu Q Q, Pei J, Zhao L. 2023. Domain specialization as the key to make large language models disruptive: A comprehensive survey.: arXiv: 2305.18703. <https://arxiv.org/abs/2305.18703.pdf>.
- Liu F X, Lin K, Li L J, Wang J F, Yacoob Y, Wang L J. 2023. Aligning large multi-modal model with robust instruction tuning. arXiv preprint, arXiv: 2306.14565.
- Liu X, Ji K X, Fu Y C, Tam W L, Du Z X, Yang Z L, Tang J. 2021. P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks.: arXiv: 2110.07602. <https://arxiv.org/abs/2110.07602.pdf>.
- 吕伟, 钟臻怡, 张伟. 2018. 人工智能技术综述. 上海电气技术, 11(1): 62–64 [Lü W, Zhong Z Y, Zhang W. 2018. Overview of AI technology. *Journal*



- of Shanghai Electric Technology, 11(1): 62–64 (in Chinese with English abstract)]
- 米加宁, 章昌平, 李大宇, 林涛. 2018. 第四研究范式: 大数据驱动的社会科学研究转型. 社会科学文摘, (4): 20–22 [Mi J N, Zhang C P, Li D Y, Lin T. 2018. The fourth research paradigm: Big data driven transformation of social science research. Social Sciences Digest, (4): 20–22 (in Chinese with English abstract)]
- Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S. 2010. Recurrent neural network based language model. In: Interspeech 2010, 2(3): 1045–1048
- Moseley B. 2022. Physics-informed machine learning: From concepts to real-world applications. Doctoral Thesis. Oxford: University of Oxford
- OpenAI. 2023. GPT-4 technical report.: arXiv: 2303.08774. <https://arxiv.org/abs/2303.08774.pdf>.
- OpenAI. How did the formation of the South Pole-Aiken Basin affect the evolution of the Moon? [2023-07-04]. <https://chat.openai.com/share/52a2e223-ee14-420b-a822-3db9decf5853>
- Peters M E, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L. 2018. Deep contextualized word representations. ArXiv e-Prints.: arXiv: 1802.05365.
- Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I. 2021. Learning transferable visual models from natural language supervision.: arXiv: 2103.00020. <https://arxiv.org/abs/2103.00020.pdf>.
- 任曼, 刘建忠, 王俊涛, 朱凯, 雷丹泓, 孙鹏举, 李雄耀. 2023. 从数字月球到元月球. 矿物岩石地球化学通报, 42(1): 233–239 [Ren M, Liu J Z, Wang J T, Zhu K, Lei D H, Sun P J, Li X Y. 2023. From digital Moon to Metamoon. Bulletin of Mineralogy, Petrology and Geochemistry, 42(1): 233–239 (in Chinese with English abstract)]
- Shanahan M. 2022. Talking about large language models.: arXiv: 2212.03551. <https://arxiv.org/abs/2212.03551.pdf>.
- Silver D, Huang A, Maddison C J, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587): 484–489
- 孙鹏举, 刘建忠, 王俊涛, 雷丹泓, 任曼. 2022. 数字月球云平台设计. 矿物岩石地球化学通报, 41(1): 135–142 [Sun P J, Liu J Z, Wang J T, Lei D H, Ren M. 2022. Design of digital lunar cloud-based platform. Bulletin of Mineralogy, Petrology and Geochemistry, 41(1): 135–142 (in Chinese with English abstract)]
- Varatharajan I, Angerhausen D, Antoniadou E, Bickel V, D'Amore M, Faragalli M, López-Francos I, Maiti A, Potter R, Shneider C, Shukla S, Soboczenski F, Srishankar N, Watkins R. 2021. Artificial intelligence for the advancement of lunar and planetary science and exploration. 53(4): 222
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, Polosukhin I. 2017. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems.; 6000–6010
- Wei J, Bosma M, Zhao V Y, Guu K, Yu A W, Lester B, Du N, Dai A M, Le Q V. 2021. Finetuned language models are zero-shot learners. ArXiv e-Prints., arXiv: 2109.01652.
- Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, Yogatama D, Bosma M, Zhou D, Metzler D, Chi E, Hashimoto T, Vinyals O, Liang P, Dean J, Fedus W. 2022. Emergent abilities of large language models. arXiv preprint, arXiv: 2206.07682.
- Wei J, Wang X Z, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E, Le Q, Zhou D. 2022. Chain-of-thought prompting elicits reasoning in large language models. ArXiv e-Prints., arXiv: 2201.11903.
- Wu S J, Irsoy O, Lu S, Dabrovski V, Dredze M, Gehrmann S, Kambadur P, Rosenberg D, Mann G. 2023. BloombergGPT: A large language model for finance. ArXiv e-Prints.: arXiv: 2303.17564.
- Yin S K, Fu C Y, Zhao S R, Li K, Sun X, Xu T, Chen E H. 2023. A survey on multimodal large language models. ArXiv e-Prints.: arXiv: 2306.13549.
- 于文轩, 马亮, 王佃利, 韩志明, 谢新水, 叶林, 文宏. 2023. “新一代人工智能技术ChatGPT的应用与规制” 笔谈. 广西师范大学学报(哲学社会科学版), 59(2): 28–53 [Yu W X, Ma L, Wang D L, Han Z M, Xie X S, Ye L, Wen H. 2023. Discussion on the application and regulation of ChatGPT, the new generation artificial intelligence technology. Journal of Guangxi Normal University (Philosophy and Social Sciences Edition), 59(2): 28–53 (in Chinese with English abstract)]
- Yuan S, Zhao S, Leng J H, Xue Z, Zhao H Y, Liu P Y, Gong Z, Zhao W X, Li J Y, Tang J. 2022. WuDaoMM: A large-scale Multi-Modal Dataset for Pre-training models.: arXiv: 2203.11480. <https://arxiv.org/abs/2203.11480.pdf>.
- Zhang K, Yu J, Yan Z L, Liu Y X, Adhikarla E, Fu S Y, Chen X, Chen C, Zhou Y Y, Li X, He L F, Davison B D, Li Q Z, Chen Y, Liu H F, Sun L C. 2023. BiomedGPT: A unified and generalist biomedical generative pre-trained transformer for vision, language, and multimodal tasks.: arXiv: 2305.17100. <https://arxiv.org/abs/2305.17100.pdf>.
- Zhao W X, Zhou K, Li J Y, Tang T Y, Wang X L, Hou Y P, Min Y Q, Zhang B C, Zhang J J, Dong Z C, Du Y F, Yang C, Chen Y S, Chen Z P, Jiang J H, Ren R Y, Li Y F, Tang X Y, Liu Z K, Liu P Y, Nie J Y, Wen J R. 2023. A survey of large language models.: arXiv: 2303.18223. <https://arxiv.org/abs/2303.18223.pdf>.
- Zheng O, Abdel-Aty M, Wang D D, Wang Z J, Ding S X. 2023. ChatGPT is on the Horizon: Could a Large Language Model be Suitable for Intelligent Traffic Safety Research and Applications? : arXiv: 2303.05382. <https://arxiv.org/abs/2303.05382.pdf>
- Zhu J G, Zhu X Z, Wang W H, Wang X H, Li H S, Wang X G, Dai J F. 2022. Uni-perceiver-MoE: Learning sparse generalist models with conditional MoEs.: arXiv: 2206.04674. <https://arxiv.org/abs/2206.04674.pdf>.
- 朱祝武. 2011. 人工智能发展综述. 中国西部科技, 10(17): 8–10 [Zhu Z W. 2011. A summary of the development of artificial intelligence. Science and Technology of West China, 10(17): 8–10(in Chinese)]

( 本文责任编辑: 刘莹; 英文审校: 肖保华 )