

主流大语言模型呈现中华文化符号： 智能拟态环境的建构^{*}

—
何
苑
张
洪
忠
—

【内容摘要】 由大模型建构的“智能拟态环境”基于海量数据和计算进行“拟真”，影响着人们对现实世界和网络虚拟世界的认知。在以 ChatGPT、Bard、文心一言和智谱清言等为代表的国内外主流大模型上，中华文化仍呈现出以传统人文和古代科技要素为统领的“符号图景”，在现当代中国风貌和科技成就上存在“数据空洞”。“中华文化符号”的呈现在国内外大模型和不同大模型之间存在明显差异，同时表现出语料欠全面、偏态化曝光、文化多样性较匮乏和既有刻板印象的“表征趋同”现象等问题。

【关键词】 智能拟态环境；大语言模型；文化表征；中华文化符号；数据空洞
DOI:10.19997/j.cnki.xdcb.2024.12.013

一、问题的提出

(一) 基于大语言模型的“智能拟态环境”的概念提出

通过接入新闻媒体平台、参与影视艺术创作、生成个性化广告文案、内嵌搜索引擎和智能手机系统等多种方式，大模型正渗透到用户的社会文化生活中，“人—大模型对话”成为新的信息获取模式。国内大模型领域同样呈现百花齐放景象，如百度公司 2023 年 2 月发布的“文心一言”、阿里巴巴达摩院开发的“通义”大模型、科大讯飞的“星火认知大模型”、字节跳动发布的“豆包”和腾讯开发的“混元大模型”等。这些大模型在移动端和 PC 端均有布局，已经初具用户规模。同时，基于大模型底层技术支撑的垂类应用产品也开始大量涌现。

沃尔特·李普曼 (Walter Lippmann) 在《公众舆论》中提出了“拟态环境”的概念，即人们会基于既有认知、共同想象和类比参照等对复杂多变的外部环境进行拟真性重构，从而

认识世界。^① 报纸、电视等大众媒介承担着建构“拟态环境”的渠道角色。在人们缺乏直接经验时，大众媒介往往会对象征性信息、要素进行筛选、加工和重组，再以“缩略图景”的形式展现给人们。在过去很长时期内，大众媒介的“拟态环境”实质上是由编辑、记者等信息收集、选择与重组来建构的人造拟态环境。

大模型时代的拟态环境则是通过人工智能的机器学习方式对海量信息进行整理与学习后输出形成的，其本质是一种“智能拟态环境”的建构方式。大模型基于海量互联网语料进行预训练，从互联网众人生产的数据中自我学习，寻找并挖掘知识的潜在结构及规律，并能够不断融合用户输入数据进行学习。“智能拟态环境”是对互联网众人知识数据的学习，且随着大模型与用户的交互，“智能拟态环境”也在不断更新迭代。

由此，本研究将“智能拟态环境”定义为：在通过机器学习对海量网络数据进行“计算”基础上建构的拟真虚拟空间。从当前技术发展阶段

* 本文系国家社科重大项目“增强中华文明海外网络传播力影响力研究”(项目编号:23ZDA094)的研究成果。

来看,“智能拟态环境”是基于 AI 大模型的现实场景再造。对于大模型和相关垂类应用的用户来说,“智能拟态环境”会影响其对现实世界和网络虚拟世界的认知,甚至唤起行为反应。从大模型技术角度看,“智能拟态环境”的建构受到多种因素影响,如大模型训练中的数据筛选过滤机制、用户对大模型的信任程度^②、用户的时间、精力和认知处理能力,以及算法媒介“以过简信息叙述事件所导致的事实扭曲”^③等。

(二)“智能拟态环境”下中华文化传播面临的不确定性问题

通过大众媒介建立的“拟态环境”并不是对真实世界的镜像再现,而是会形成“超越真实世界”的图景。^④同样,尽管大模型的“智能拟态环境”是基于其庞大的底层数据库形成的,也并不能够实现对真实世界的完全投影,并且目前不同大模型之间还存在表现差异。如有研究发现,大模型在复杂任务环境中可能做出“奉承”或“篡改代码规则”等欺骗性行为。^⑤从模型训练来看,基于文化、语言差异的语料分布倾斜,针对特定任务优化的微调算法等,都可能导致不同程度的“区间性偏差”(Intersectional bias),进而形成“文化偏见”并向用户输出。^⑥此外,根植在原生文化价值中的在地化差异也容易造成不同大模型之间的“价值鸿沟”。^⑦尽管已有研究指出国内外主要大模型之间在文化背景、价值对齐等层面存在明显差异,这种文化偏见又会通过交互过程“投射”给用户^⑧,但在市场应用层面,当前基于大模型技术的文化传播应用仍在不断涌现。从 GPT-4 到 Sora,从文心一言到智谱清言、可灵大模型等,国内外大模型应用已开始形成汇聚规模化用户的商业平台,对社会信息和文化传播生态的影响也愈发明显。

大模型在驱动人类社会内容生产和传播范式变革的同时,也在重塑“智能拟态环境”中的国际传播新格局,中华文化的对外传播面临诸多不确定性。一方面,传统的文化形成逻辑和演进路径已受到人工智能技术的颠覆。研究发

现,基于大模型的智能应用能够提升内容生产效率和质量,进而推动生产力和劳动市场变革。^⑨与此同时,大模型生成内容在信息说服和影响用户认知方面更具潜力。^⑩对海量数据的捕捉、理解、自监督学习和微调策略是支撑大模型进行复杂推理和“人机交互”的基础。而当数据成为新的内容生产要素,网络异构空间中的内容准入、分布、流通和存续法均由算法缔造,必然会对传统的文化传播、治理和生态体系形成挑战。^⑪

另一方面,大模型对训练、建模语料的强依赖性和技术“强者”所属的文化价值理念可能进一步加剧国际舆论体系中的既有对华偏见。如有学者发现,ChatGPT 在进行人机互动时倾向将中国价值观与“文化孤立”和“民族中心主义”相联系。^⑫同时,基于不同底层语料库训练的大模型在内容输出上也存在偏差。^⑬在相关领域,我国科技主体起步晚于美国,在核心技术研发、落地应用、迭代等方面仍存在短板。由计算编程语言壁垒、中文优质内容在全球网络数据库中缺失等因素导致的“数据空洞”(Data Void),都可能造成文化层面的意义偏离(Deviation)和误读。^⑭

“中华文化”指向相对宽泛和抽象的概念,既可以体现为物态文化、制度文化、行为文化和心态文化等形态,也可以从具象识别符号、文化知识符号、价值观念符号等角度进行区分。^⑮在此基础上,本研究力图探讨中华文化在大模型中被建构为什么样的“智能拟态环境”?中外大模型中展现的中华文化符号图景是否有差异?具体研究问题如下:

RQ1:当前主流大模型描绘出的“中华文化符号图景”为何?具体是通过哪些象征符号和文化代表类型来呈现的?

RQ2:基于不同大模型的“智能拟态环境”是否会建构出差异化的“中华文化符号图景”?国内外大模型在对中华文化象征符号的呈现方式上存在哪些共性和差异?

表1 大模型描绘的“中华文化象征符号”
所属一级、二级类目

二、研究方法步骤

研究综合采用在线问答、内容分析法开展研究分析。首先是确定拟研究的大模型平台。在文献回顾和对大模型发展现状进行前期调研的基础上,研究选取了在国内外较为主流的4个大模型应用进行调研和在线问答。其中,国内平台包括百度公司的“文心一言”大模型应用和智谱清言的“AI-ChatGLM”对话语言模型,海外平台为OpenAI公司的ChatGPT和Google公司的Bard大模型应用。

其次是制定提问和数据收集规则。在对联合国教科文组织《保护世界文化和自然遗产公约》、联合国《世界文化多样性宣言》、我国关于文物保护、文化市场分级分类管理的相关法律法规以及国内科研机构关于中华文化“海外网络传播力”和“中华文化符号国际传播”等相关文献进行梳理的基础上,制定统一的提问规则和数据统计格式。

再次是进行数据清理、去噪与编码。研究公开招募10名实验人员,经培训后随机分配与不同大模型进行对话实验,询问其认为“最具有代表性的中华文化符号”,并对模型作答结果进行记录和汇总,共获得3008个原始词条(其中包含ChatGPT生成的内容832条、Bard生成的内容680条、文心一言生成的内容651条以及智谱清言生成的内容845条)。研究者进一步对收回的有效数据进行汇总、去重和标准化处理,即将文字表述不同但指代对象相同的词结构化为格式统一且不存在歧义的“最小语义单元”(如将“紫禁城”“明清皇宫”“北京故宫”“故宫”等统一用“故宫”表示)。再由5名经过培训的人工编码员进行独立编码和词条收敛,直至形成饱和且互斥的文化符号类目。共获得15个一级类目和64个二级类目,如表1所示。

最后一步是数据分析与比较。按照语料库来源、文化类目和分模型维度对所获数据进行统计分析和差异、共性规律总结。

序号	一级类目	二级类目
1	文化、自然遗产与生态资源	文化遗产/人文生态资源 文化自然混合遗产/资源 自然遗产/生态资源 国家文物保护单位
2	动物	神话传说动物 濒危动物 吉祥物 国宝
3	植物	观赏性植物 农作物
4	体育竞技	传统武术 传统棋牌运动 奥运比赛项目
5	科学与技术	天文历法航空 数学 地理/交通 医药学 冶炼铸造建筑 农林渔牧生物 织造纺染编绣 传播/通信/出版 军事
6	饮食文化	茶文化 酒文化 特色菜系 代表性食品/菜肴 节令食品 饮食礼仪习俗
7	手工艺品	雕刻塑造 服装配饰 纸艺
8	代表人物	政治人物 哲学家 科学家 文学家/艺术家 演员 神话传说人物
9	节庆习俗	节日 庆祝/纪念活动 重大仪礼 仪式物品
10	语言文学历史	语言文字 体裁形式 文学作品
11	书画艺术	书画用具 绘画形式 印鉴篆刻 传统书法
12	舞乐曲艺表演艺术	乐器 音乐类型 表演艺术 戏曲曲艺 经典作品

(续表)

序号	一级类目	二级类目
13	哲学宗教 思想体系	哲学思想理念 宗教信仰 象征物/图腾 经典著作
14	政治理念 体制制度	中共党史 外交与国际关系 民族政策 历史帝制 国家象征 政治理念
15	其他	—

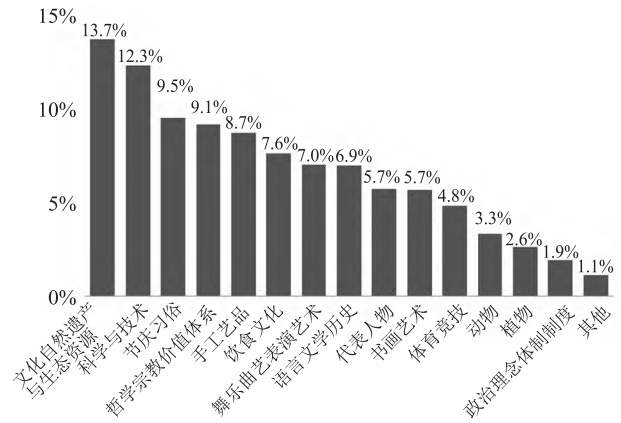


图1 大模型呈现的中华文化符号类型分布及占比

三、大模型呈现的“中华文化符号图景”概览

研究首先对大模型中的代表性“中华文化象征符号”进行了频次和类目分布统计。鉴于全部 3008 个标准化词条的被提及几率符合幂律分布规律,即出现频次最高的前 100 个词条占整体的 75.5%,前 50 个词条占整体的 50.8%,因此主要分析比较整体占比高、排名相对靠前的内容,具体发现如下。

(一)文化、自然遗产和古代科技是大模型世界的主要“中华文化象征符号”

首先从文化类目分布上来看,文化、自然遗产与生态资源类的词条在大模型提及的代表性“中华文化象征符号”中占比最多,其次是科学技术、节庆习俗和哲学宗教思想体系等。需要注意的是,尽管“科学与技术”类词条整体占比比较大,但其中占绝对比例的均为古代科技成就(如“传统医学”“中药”“针灸”“印刷术”“火药”“造纸术”和“指南针”等),代表数学、生化、电子通信等理工类自然科学技术的内容较少,具体如图 1 所示。

随后是大模型提及最多的前 100 个高频词的分布特征。由表 2 可见,“书法”“儒家思想”“故宫”“京剧”“长城”“道家思想”“佛教”“中国龙”“中国画”“春节”等人文要素出现几率排名靠前。而涉及科学技术的内容(如“传统医学”“中药”“印刷术”“造纸术”“针灸”“指南针”“火药”等)的出现几率整体较低,散布在前10~50 名之间。

表 2 大模型提及最多的 100 个中华文化符号的出现几率*

序号	文化符号	出现几率	序号	文化符号	出现几率
1	书法	95.0%	51	风水	35.0%
2	儒家思想	92.5%	52	丝绸之路	35.0%
3	故宫	90.0%	53	景泰蓝	35.0%
4	京剧	90.0%	54	黄河	35.0%
5	长城	87.5%	55	中式园林	35.0%
6	道家思想	85.0%	56	舞龙	32.5%
7	佛教	80.0%	57	春联	32.5%
8	中国龙	77.5%	58	昆曲	32.5%
9	中国画	75.0%	59	诗词	32.5%
10	春节	75.0%	60	唐诗	32.5%
11	剪纸	70.0%	61	重阳节	30.0%
12	兵马俑	67.5%	62	赛龙舟	30.0%
13	传统医学	67.5%	63	泰山	30.0%
14	中国结	65.0%	64	粽子	30.0%
15	端午节	62.5%	65	老子	30.0%
16	太极拳	60.0%	66	刺绣	30.0%
17	孔子	60.0%	67	牡丹花	30.0%
18	旗袍	57.5%	68	宋词	30.0%
19	莫高窟	55.0%	69	红色	27.5%
20	围棋	55.0%	70	竹子	27.5%
21	中秋节	55.0%	71	西游记	27.5%
22	中药	52.5%	72	墨	27.5%
23	功夫	52.5%	73	阴阳哲学	27.5%
24	瓷器	52.5%	74	天安门	27.5%
25	风筝	50.0%	75	梅花	25.0%
26	武术	50.0%	76	麻将	25.0%
27	饺子	50.0%	77	扇子	25.0%
28	汉字	50.0%	78	孟子	25.0%
29	印刷术	47.5%	79	砚台	25.0%
30	玉器	47.5%	80	颐和园	25.0%
31	凤凰	45.0%	81	二十四节气	25.0%
32	大熊猫	45.0%	82	汉服	25.0%
33	茶道	45.0%	83	甲骨文	25.0%
34	灯笼	45.0%	84	筷子	25.0%
35	皮影戏	45.0%	85	水墨画	25.0%
36	黄山	45.0%	86	八卦	25.0%
37	青花瓷	45.0%	87	北京烤鸭	25.0%

(续表)

序号	文化符号	出现几率	序号	文化符号	出现几率
38	丝绸	42.5%	88	陶瓷	25.0%
39	脸谱	42.5%	89	茶	25.0%
40	月饼	42.5%	90	张家界	22.5%
41	舞狮	42.5%	91	天坛	22.5%
42	造纸术	40.0%	92	紫砂壶	22.5%
43	针灸	40.0%	93	红包	22.5%
44	红楼梦	40.0%	94	水浒传	22.5%
45	元宵节	40.0%	95	二胡	22.5%
46	福禄寿喜	37.5%	96	太极	22.5%
47	指南针	37.5%	97	唐装	22.5%
48	长江	37.5%	98	印章	22.5%
49	象棋	37.5%	99	火锅	22.5%
50	火药	35.0%	100	菊花	20.0%

*说明:表中的百分比代表对应词条在任意提问者使用同一指令向任意大模型提问时被提及的几率,值越大表示出现几率越高(如某一词条在所有提问者与所有大模型的对话中均被提及,其出现几率为100%)。上述词条中,共81个被四个大模型共同提及。

(二)大模型呈现的“中华文化符号”由古代人文要素统领,自然科学和近现代发展成就存在“数据空洞”与“断代”

研究进一步将大模型提及最多的前100个代表性中华文化符号按年代、所属类目及出现频次进行了统计和可视化分析,如图2所示。

从时间分布来看,1840年前的古代人文要素占据绝对比重,近现代内容极少。尤其现代科技元素(如“长征火箭”“太空计划”“天眼望远镜”“高铁”等)仅零星被个别大模型提及。这反映出当前国内外大模型对中国近现代发展历程和科技成就普遍缺乏观照,相关语料的完备性和展现逻辑有待优化,存在一定的“数据空洞”和“断代”现象。

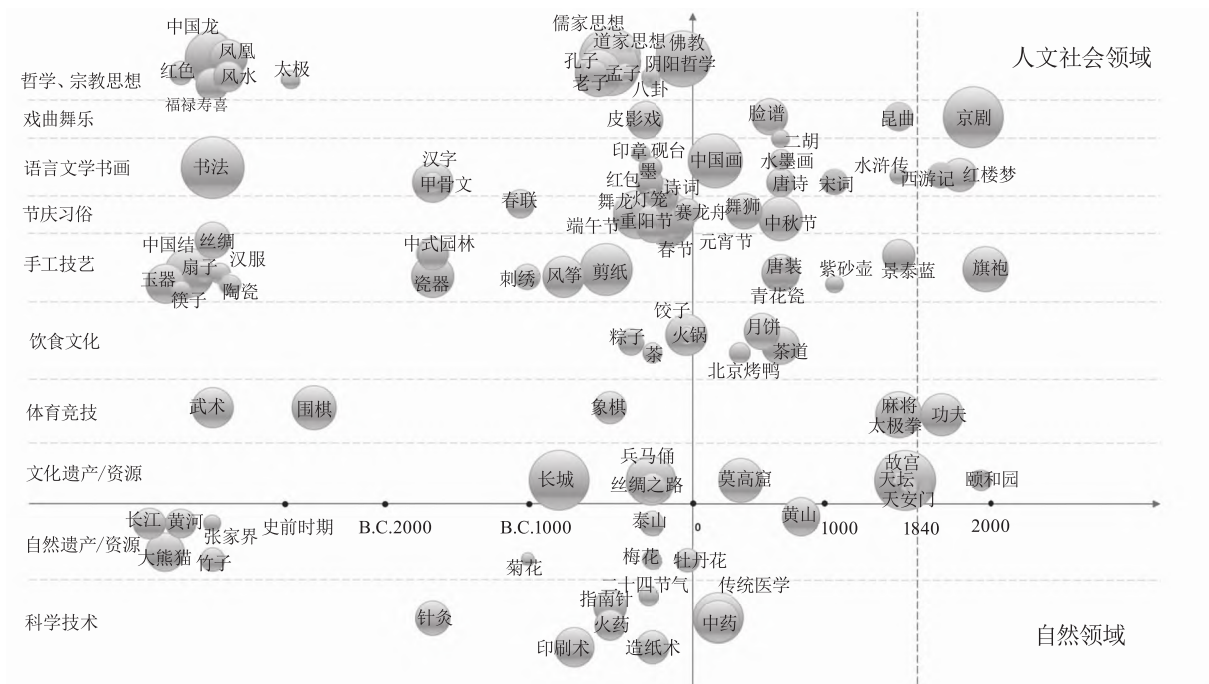


图2 整体出现频次最高的前100个词条分布情况¹⁶

从词条的频次和类型分布来看,被大模型提及最多的高频词主要集中在哲学宗教和文化遗产领域,语言文学、手工艺品和自然资源方面的词条虽然类型丰富,但只有书法、国画、京剧等词重复出现的几率较高。

四、中外大模型上的“中华文化符号”分布差异

研究随后对三种情况下的“中华文化符号”分布进行了比较,即(1)仅被国内大模型提

及(2)仅被国外大模型提及和(3)被国内外大模型共同提及,具体发现如下。

(一)国内外大模型在呈现民间手工艺、节庆习俗和动物等“文化符号”时存在最多共识,政治理念等方面差别较明显

从类目占比上来看,国内外大模型在书画艺术、体育竞技、动物、手工艺品和节庆习俗5个方面共同提及的内容较多,在政治理念体制制度、代表人物和特有植物上存在的差别最大。由差异性检验的结果可知,国内外大模型侧重呈现的维度存在显著差异($Df = 30, \chi^2 = 222.81, p < 0.001$),仅有“手工技艺”(如丝绸、木雕、风筝、中国结等)“动物”(如濒危及珍稀保护动物、动物图腾等)和“节庆习俗”(如中秋节、元宵节、春节、年画、贴春联等)三个类目的词条未被“区别对待”(百分比同质性事后比较 95% 置信区间分别为 $< 0.017, -0.019 >$, $< 0.003, -0.019 >$, $< 0.015, -0.025 >$),具体如图3所示。

具体来看,国外大模型在体育竞技、饮食文化、植物(如梅兰竹菊)、代表人物、政治理念和体制制度等范畴提及的内容多于国内大模型,涵盖的词条类型丰富但分布零散。国内大模型

更侧重呈现书画艺术、文化/自然遗产与生态资源、哲学宗教体系、语言文学历史、舞乐曲艺表演和科学与技术等方面的词条。

从单一词条的出现几率来看,国内外大模型之间同样存在差异。如表3所示,在“亮相”最多的前50个词条中,国外大模型对构成中华文化思想的道家、儒家和佛教思想曝光程度相当,而国内大模型则更强调儒家思想。值得注意的是,在科技领域只有传统中医中药和印刷术是国内外大模型共同提及的内容。

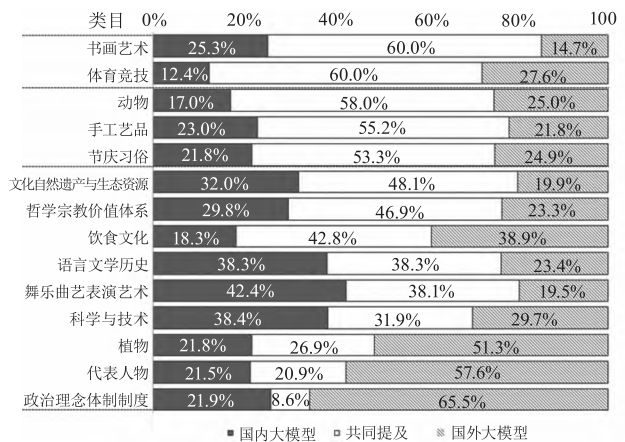


图3 国内外大模型呈现的文化符号类目差异¹⁷

表3 国内外大模型提及几率最高的前50个词条*

序号	国内大模型	出现几率	国外大模型	出现几率	序号	国内大模型	出现几率	国外大模型	出现几率
1	京剧	100.0%	书法	100.0%	26	脸谱	60.0%	黄山	50.0%
2	儒家思想	95.0%	故宫	95.0%	27	风筝	55.0%	武术	50.0%
3	书法	90.0%	道家思想	95.0%	28	中秋节	55.0%	指南针	50.0%
4	剪纸	90.0%	儒家思想	90.0%	29	武术	50.0%	火药	50.0%
5	故宫	85.0%	佛教	90.0%	30	丝绸	50.0%	茶道	50.0%
6	长城	85.0%	长城	90.0%	31	舞狮	50.0%	青花瓷	50.0%
7	中国龙	80.0%	京剧	80.0%	32	围棋	50.0%	旗袍	50.0%
8	皮影戏	80.0%	春节	75.0%	33	凤凰	45.0%	灯笼	50.0%
9	春节	75.0%	中国画	75.0%	34	功夫	45.0%	剪纸	50.0%
10	中国画	75.0%	中国龙	75.0%	35	饺子	45.0%	凤凰	45.0%
11	道家思想	75.0%	莫高窟	60.0%	36	月饼	45.0%	象棋	45.0%
12	兵马俑	75.0%	大熊猫	60.0%	37	汉字	45.0%	瓷器	45.0%
13	传统医学	75.0%	太极拳	60.0%	38	丝绸之路	40.0%	风筝	45.0%
14	中国结	70.0%	功夫	60.0%	39	黄山	40.0%	老子	45.0%
15	佛教	70.0%	中国结	60.0%	40	长江	40.0%	福禄寿喜	45.0%
16	中药	65.0%	端午节	60.0%	41	印刷术	40.0%	天安门	45.0%
17	玉器	65.0%	兵马俑	60.0%	42	茶道	40.0%	梅花	40.0%
18	旗袍	65.0%	围棋	60.0%	43	青花瓷	40.0%	中药	40.0%
19	端午节	65.0%	传统医学	60.0%	44	元宵节	40.0%	月饼	40.0%
20	孔子	65.0%	造纸术	55.0%	45	灯笼	40.0%	元宵节	40.0%
21	莫高窟	60.0%	印刷术	55.0%	46	红楼梦	40.0%	舞狮	40.0%

(续表)

序号	国内大模型	出现几率	国外大模型	出现几率	序号	国内大模型	出现几率	国外大模型	出现几率
22	太极拳	60.0%	饺子	55.0%	47	孝道	40.0%	红包	40.0%
23	针灸	60.0%	中秋节	55.0%	48	黄河	35.0%	宋词	40.0%
24	景泰蓝	60.0%	汉字	55.0%	49	牡丹花	35.0%	唐诗	40.0%
25	瓷器	60.0%	孔子	55.0%	50	二十四节气	35.0%	红楼梦	40.0%

* 说明:表中百分比代表该词条在任意提问者使用任意提问指令向任意大模型提问时出现的几率,数值越大则表示出现几率越高。

(二)国内大模型对传统曲艺、手工艺品和中医的曝光力度更大,国外更“青睐”大熊猫和四大发明

研究分别对国内、国外大模型提及几率最高的前100个词条的频次分布差异进行了比较。通过引入火山图(Volcano Plot)来展示前100个高频词在国内外大模型中的频次分布差异,这一方法在基因表达等生物信息学领域多用于分析高维度数据。¹⁸其中,以国内/国外大模型为分组变量,以 $p = 0.05$ 为显著性水平;横轴为分组变量的对数变化处理,表示某个文化标签在两个分组的频次相对高低;纵轴为 p 值负对数,数值越高,表示显著性越强(即 p 值越小)。

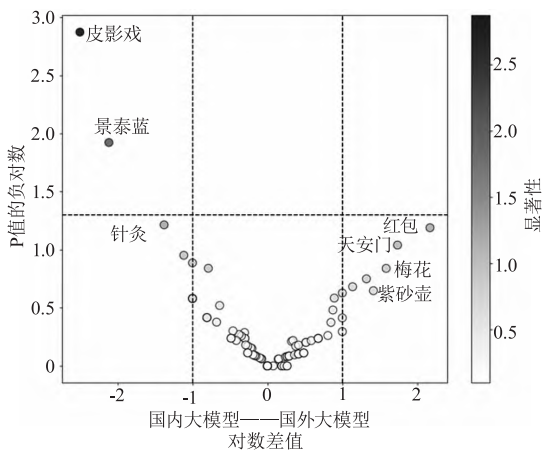


图4 国内外大模型提及的前100个高频词差异比较¹⁸

由图4所示的数据结果可见,国内外大模型在对针灸、景泰蓝、皮影戏和红包这四个符号的呈现上存在显著差异($p < 0.05$)。具体来看,针灸、景泰蓝、皮影戏在国内大模型中提及的频次显著高于国外大模型;而红包在国外大模型中提及的频次显著高于国内大模型。此外,“天安门”“梅花”“紫砂壶”等词条被国外

大模型提及的频次均远超出国内大模型,尽管未达到统计显著性阈值(横轴取值位于 $[-1, 1]$ 区间之内),但从分布上也体现出国外大模型对中华文化的符号呈现差异。

五、差异化的“智能拟态环境”建构:不同大模型对中华文化符号的呈现角度不一、声量悬殊

(一)ChatGPT和Bard占据大模型“中华文化图景”的半壁江山,国内大模型贡献的内容丰富但不占“声量优势”

从数量上来看,由ChatGPT(27.7%)、Bard(22.6%)、文心一言(21.6%)和智谱清言(28.1%)提供的词条占比差异并不大。但在进行分层比较后可见,在“文化/自然遗产与生态资源”“科学与技术”“哲学宗教思想体系”和“节庆习俗”这四大构成“中华文化象征图景”半壁江山的优势类目中(整体占比44.6%),绝大多数的词条均来自ChatGPT和Bard,由国内大模型贡献的内容占比较小,具体如图5所示。

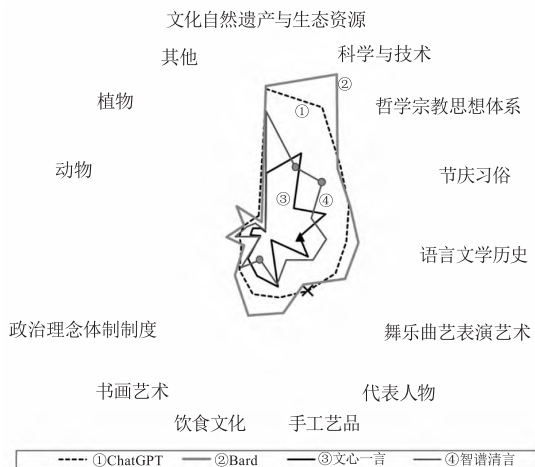


图5 各个大模型上呈现的文化符号类型占比

对各个大模型上呈现和缺失的内容比较揭示了更加明显的反差。如图 6 所示,ChatGPT 和 Bard 上展示的文化类目整体多于缺失的内

容,而文心一言和智谱清言的比例分布则正好相反。这体现出国外大模型在呈现中华文化“数字图景”时占据更大的“声量”优势。

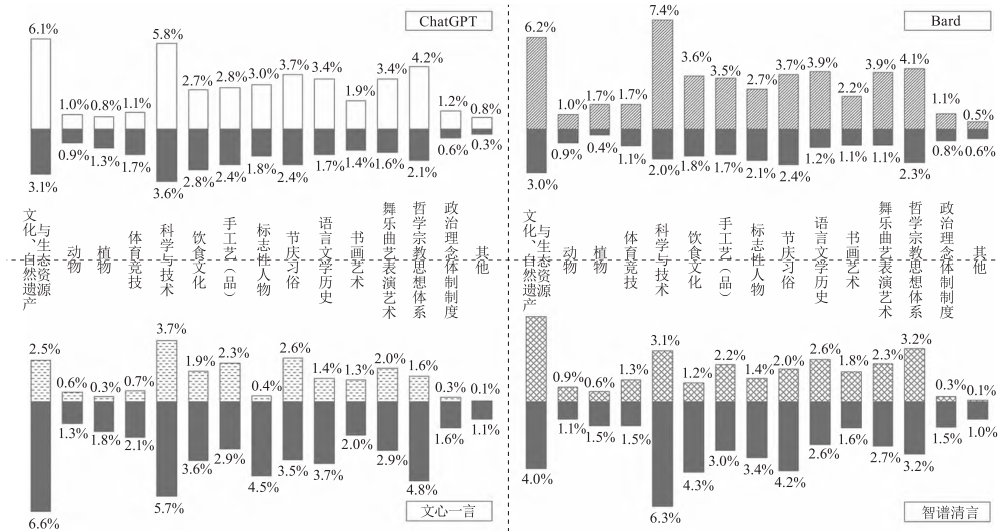


图 6 各个大模型中呈现/缺失的文化符号类型比较²⁰

国内方面,文心一言与智谱清言呈现的词条主要在“语言文学历史”(-0.275**)和“哲学宗教思想”(-0.229**)方面存在差别;智谱清言对科学技术、哲学宗教思想、语言文学和饮食文化方面内容的呈现与 ChatGPT(-0.158**)、Bard(-0.076**)和文心一言(-0.080**)显著不同。值得注意的是,国外大模型更倾向展示类型词(如“中医”),而国内大模型则提供了更具体和体现地方特色的词条(如“针灸”“神农本草”等)。但这些词条的语义颗粒度更细且重复率低,因此在“数据驱动”的大模型世界里反而变得不够“惹眼”,其“光华”容易被计算法则掩盖。

(二)中华文化符号的差序分布显现,仅三分之一高频词被不同大模型共同提及

在将各个应用的活跃用户规模纳入考量后,不同大模型对“中华文化符号”的差序式呈现更加明显,数据可视化分析结果如图 7 所示。

具体来看,在国外大模型中,ChatGPT 对特定元素的“曝光”力度远高于其他模型(如儒释道思想、中国龙、书法、国画、春节、茶道、长城、故宫、太极拳等)。Bard 在“政治制度和理念”

方面提供了最多内容(如“中国梦”“熊猫外交”“文化软实力”“改革开放”等),而其他三个大模型在这一范畴存在明显空白。同时,Bard 还贡献了有关传统节日、自然景观资源和动物图腾等的丰富词条。

在国内大模型中,文心一言在传统手工艺品和传统中医药方面贡献了更多词条。节庆习俗方面,不止提及了传统节日名字,还有具体的文化实践内容(如“舞龙”“舞狮”“春联”等)。并且,文心一言是唯一侧重呈现“珠算”这一数学成就的大模型。智谱清言在所有被测大模型中拥有的活跃用户规模最小,但在戏曲曲艺和语言文学方面(如“相声”“脸谱”“孝道”“印章”“墨”“砚台”等)提供的词条内容最为丰富具体。

综合比较图 7 的四个象限可见,在所有高频词条中,仅有三分之一的内容被国内外不同大模型共同提及。可见,即使在号称中立、客观的数字计算条件下,国内外大模型之间以及不同大模型之间在“中华文化是什么”以及“哪些符号可以代表中华文化”等问题上也并未达成

“共识”。此外,当前 ChatGPT 和 Bard 的全球用户规模可以从数量上“碾压”国内大模型,并且这两个大模型自推出以来已经内嵌主流浏览器、办公软件、搜索引擎、社交媒体等应用,渗透到千万用户的生活和工作场景当中。在此背景

下,不同大模型间存在的文化符号分布差异在真实世界中也必然会被“指数级”放大。这种由开发者技术差异和市场占有率不均衡导致的“偏态化”曝光可能会在无形中塑造出新的对华“刻板印象”。

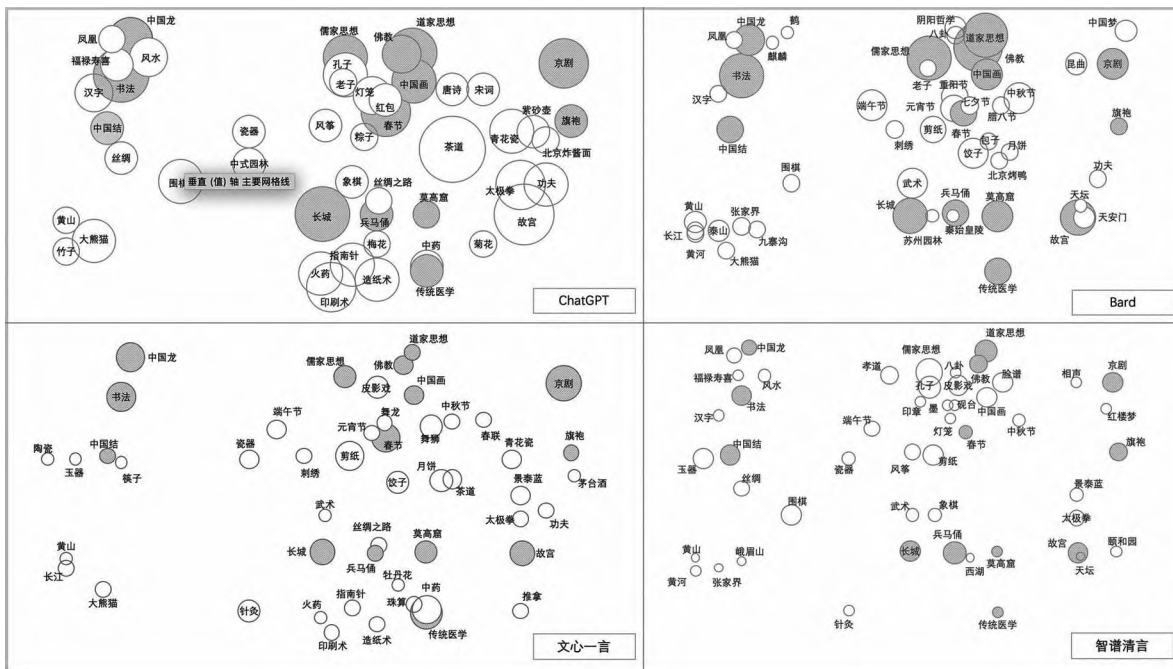


图7 不同大模型上的高频词呈现差异^[2]

(三)四个大模型均存在语料欠全面和长尾词条出现几率不稳定的问题

从数据全面性和代表性上来看,一方面是被测大模型呈现的3000多个符号分布存在明显的头部效应,另一方面,剩余长尾词条的出现几率和频次表现出较强的“偶然性”和“不稳定性”。如“福建土楼”“蒙古包”“藏族唐卡”“楚辞”“庙会”“苗族银饰”等200多个元素仅在个别大模型中出现;“文庙”“春分”“长征火箭”等内容仅被提及过1次,表现出较强的随机性。而在大模型提及代表性人物(如“毛泽东”“孙中山”“邓小平”“周恩来”等领袖人物、“李白”“华佗”“岳飞”等历史名人)、政治理念(如“熊猫外交”“文化自信”“文化软实力”“一带一路”“改革开放”等)等内容时也存在类似的问题。

六、结论与讨论

不同大模型在以文本形式呈现中华文化代表符号时既具有一定共性,也存在区域性差异和模型间差异。

首先,基于海量数据汇聚的大模型“智能拟态环境”正显现出“表征趋同”(representational convergence)的特征。从宏观视角来看,尽管大模型形成于对大规模、多样化数据资源(参数数量、数据点数量、运算次数)的训练和学习,但仍存在符号集中度高、多元性匮乏的现象,即整体符合幂律分布,长尾词数量多但分布零散,词条展示的偶然性等问题,并且指向越具体的词输出几率越不可控。这一发现恰好回应了新近研究中有关“柏拉图表征假说”(Platonic

Representation Hypothesis)的担忧:深度学习的本质是要收敛出全局中最高效直接的最优解,而在不同 AI 模型都在尝试“拟真”的情况下,对底层数据共性的统计、描述和计算就成了概括复杂现实世界的捷径。²²因此,即使处在不同神经网络结构下,大模型系统中也会不可避免地出现趋于相同的数据点表达,对那些面向“通用型”任务场景的大模型架构来说更是如此。在“智能拟态环境”建构的视角下,鉴于当前公众尚未充分意识到人工智能在多样性和偏见问题上可能带来的风险,这种“趋同性”极有可能导致文化要素的“再中心化”,加剧“回音壁”效应,不利于文化多样性发展。²³

其次,当前主流大模型对话应用建构出了一个偏向中国古代历史和传统人文风貌的“智能拟态环境”,弱化了中国的现当代发展历程和科技成就,甚至在部分要素上存在明显的“数据空洞”。同时,国内外大模型对“中华文化”的拟真模式存在显著差异。国外大模型对个别文化符号存在“偏态化”曝光的“聚光灯”效应。国内大模型呈现的文化符号虽然更丰富多元,但在整体网络空间中的“声量”却远不及国外模型。在不同底层训练数据库、模型算法和应用场景等因素的加持下,更容易造成中华文化在互联网世界中的“显式”和“隐式”符号正则化差异。如果说传统媒体时代存在媒介建构的“拟态环境”与现实世界之间的差异,那么在“智能拟态环境”时代,这种差距不但没有被弥合,甚至还出现了基于不同大模型系统的内部“拟真世界”割裂。在当下大模型科技主体的技术水平、底层设施部署、落地应用情况和市场占有率悬殊明显的情况下,由数字世界的偏态“分布式记忆”(distributed memory)造成的文化路径偏移或将成为值得重视的新议题。²⁴

因此,如何建构中华文化的“智能拟态环境”已成为新一代人工智能技术背景下一个非常现实的命题。一方面,要从宏观层面出发,高度关注大模型的培育与技术迭代,积极开发更多面向中文语料库倾斜的基础大模型。据斯坦

福大学以人为本人工智能研究院发布的《2024年度人工智能指数报告》显示,当前美国已领先中国、欧盟和英国,成为顶级人工智能模型的主要来源国。²⁵从全球视野看,要抓紧培育出具备一定比重的中文数据的基础大模型,更加全面、客观地展现中华文化符号要素和文化图景,积极参与全球基础大模型竞争,中华文化在 AI 时代的国际影响力才有技术保证。

另一方面,要立足微观层面,加快大模型在中华文化传播领域的垂类应用开发。相较于基础大模型,特定领域、特定功能的垂类大模型才是大模型技术实现社会化应用和被用户广泛接受的主要途径。因此,站在普通用户的视角来看,应从 AI 智能体、提示工程等技术路径出发,开发充分展现中华文化魅力的跨文化传播大模型垂类应用,拓展用户规模,以提升中文优质语料等关键要素在底层数据库的增长与流通效率,进一步提升中华文化在大模型世界中的影响力。

为此,可对目前处于快速发展阶段的我国大模型开发主体提出如下建议:一是丰富中华文化相关语料数据库,提升中文语料和现代化元素在大模型数据集中的比重与权重。一方面,要在大模型开发中重视中华文明符号体系的独特性、独创性和区分度。建立中华文明符号标识体系元数据库,从文明交流互鉴的视角呈现现代、开放、文明、和谐的东方形象。另一方面,要在加大中文内容向不同国家、不同科技主体旗下大模型的输出力度的同时,灵活地以其他语言作为载体,输出中华文明内容至大模型。建议吸收和借鉴西方基于英语这一表音语言大模型的算法优势,开展汉语这一表意语言大模型的算法创新。为此,可发挥中国 10 亿级的互联网使用者优势,广泛吸纳社交媒体等的中文数据内容,建立涵盖古代汉语、现代汉语等优质中文数据的大模型语料库,从底层出发,并从训练、投喂和学习机制上突破定势,展现古今兼容、文理兼备的大国形象。

二是研发中华文化特色大模型垂类应用与

智能体产品,向内向外推广中华文明标识。其一,基于大模型的强大算力优势,开发适配行业需求的中文垂类大模型产品或应用,加快大模型产品的价值落地,倒逼垂直领域大模型技术迭代。其二,利用和发挥好全球对中华文明的已有认知,用好存量。一方面,开发和设计以中华文明标识人物、技艺、艺术为对象的大模型产品,将全球公认的中华文明标识纳入大模型产品开发体系。另一方面,针对现代科学、科学智能等相对薄弱领域,加强大模型应用和开发。将科学智能(AI for Science)作为发展大模型的重点和难点,开发和设计中国本土的大模型智能体产品,释放云端大数据的竞争活力,推动多元化创意、技术、需求、场景的碰撞与融合,在将大模型不断运用于各种实际场景的同时完成对大模型本身的提档升级。

三是增强产品视觉效应,优化基于大模型的多模态中国风格内容生成。“智能拟态建构”时代的文化影响力与传播力增强,尤其需要重视中国风格、中国气派的符号体系、审美体系的建构。一方面,要在大模型的开发中强化中华文明符号的独特魅力、创新性和辨识度。建立中华文明符号标识体系元数据库,将代表性符号标识提炼、推广;从文明交流互鉴的视角在建筑、服饰、食品等领域突出中华文明的独特性。另一方面,重视多模态大模型产品的开发推广。扶持体现中国审美的生成式人工智能产品的发展,鼓励开展相关领域人工智能大赛,激励企业打造比肩海外主流大模型应用的生成式产品;引导提示工程师与大模型产品的互动,提升视觉类大模型产品的中国审美品质,向世界传递中华文化的古典审美、现代审美。

四是鼓励国内大模型对接已“出海”且具有

影响力的文化生产主体,拓宽中华文化符号体系的有效展示渠道和体系。将文化符号的展示渠道和形成体系建设作为科技企业的发展战略重点,将国内大模型产品对接到网络文学、短视频、游戏等已经“出海”并且具有国际影响力的文化生产渠道中来,从消费端促进中华文化的海外传播。建议将“大模型+网络文学、大模型+短视频、大模型+游戏”作为提升中华文化影响力传播力的重要输出渠道。一方面,针对不同语种开发大模型+网络文学和游戏领域产品,提升不同语种翻译、阅读网络文学的速率,以文化浸润全球读者和玩家。另一方面,吸收和借鉴TikTok、ReelShort等软件应用的经验,开发具有中国风格的大模型+短视频产品,提升内容生产的能力和效率,注重情感叙事、情感共鸣,以网络文化为重要抓手,呈现立体、可爱的中国形象。

五是将大模型应用纳入中华文化的传播力影响力评估体系。当前的媒介实践和传播力影响力评估仍以推特、脸书等国外社交媒体为主,尚未将大模型应用纳入评估和考核体系。在大模型应用成为未来发展趋势的背景下,我国的文化传播力影响力建设也应与时俱进,将主流媒体大模型应用充分纳入现有的评价、考核体系。此战略手段的具体实施路径有两条,第一条是促进科技主体向主流媒体的赋能,在技术学习和一线实践中提升主流媒体新闻工作者的大模型应用能力,培养其成为中华文化大模型传播的“提示工程师”。第二条是将不同媒介渠道主动加入大模型产品的开发和推广过程中来,推出一批富有中华文化底蕴和内涵的大模型产品,同时开辟新的传播阵地,推进“借船出海”工程,助力文化传播力影响力提升。

【责任编辑:杨旭东】

注释:

- ① [美]李普曼:《舆论》,常江、肖寒译,北京大学出版社2018年版,第14页。
- ② 张洪忠、任昊炯:《超越“第二自我”的人机对话——基于AI大模型应用的信任关系探讨》,《新闻大学》,2024年第3期,第50页。
- ③ 彭兰:《新媒体时代拟态环境建构的变化及其影响》,《中国编辑》,2022年第12期,第4页。
- ④ Baudrillard, J. *Simulacra and Simulation*. Ann Arbor: University of Michigan Press, 1994. p. 1. D' Alessandro, S. *From the Pseudo-Environment to the Meta-Verse. Recontextualising Lippmann's Thought*. Cambio-Rivista Sulle Trasformazioni Sociali, vol. 13, no. 25, 2023. p. 265.
- ⑤ Denison, C., MacDiarmid, M., Barez, F. et al. *Sycophancy to Subterfuge: Investigation Reward Tampering in Language Models*. <https://arxiv.org/abs/>

- 2406.10162.2024-06-17.
- ⑥⑬ Navigli, R., Conia, S. and Ross B. *Biases in Large Language Models: Origins, Inventory, and Discussion*. Journal of Data and Information Quality, vol. 15, no. 2, 2023. p. 10;11, p. 10;7.
- ⑦ 胡正荣、闫佳琦:《生成式人工智能的价值对齐比较研究——基于2012—2023年十大国际新闻生成评论的实验》,《新闻大学》,2024年第3期,第13页。
- ⑧ Geirhos, R., Jacobsen, J. H., Michaelis, C., et al. *Shortcut Learning in Deep Neural Networks*. Nature Machine Intelligence, no. 2, 2020. p. 667.
- ⑨ Noy, S., Zhang, W. *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*. Science, vol. 381, no. 6654, 2023. pp. 2-4.
- ⑩ Spitale G, Biller-Andorno N, Germani F. *AI Model GPT-3 (dis) Informs us Better than Humans*. Science Advances, vol. 9, no. 26, 2023. p. 2.
- ⑪ Guy, J. *Digital Technology, Digital Culture and the Metric/Nonmetric Distinction*. Technological Forecasting and Social Change, vol. 145, 2019. p. 60.
- ⑫ 相德宝、吴艳:《中华文化价值、价值偏向及传播策略研究——以国际生成式人工智能平台为研究对象》,《对外传播》,2024年第6期,第58页。
- ⑬ Golebiewski, M. & Boyd, D. *Data Voids: Where Missing Data Can Easily be Exploited*. Data & Society. <https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf>. 2019-10-29. Bakke, A. *Everyday Googling: Results of an Observational Study and Applications for Teaching Algorithmic Literacy*. Computers and Composition, vol. 57, 2020. p. 4.
- ⑭ 刘涛、张媛媛:《通往数字人文的游戏之路:游戏叙事中的传统文海符号再现及其程序修辞机制》,《南京社会科学》,2023年第11期,第130页;陆朦朦、崔波:《网络文学海外传播中华文化的多模态叙事与认同引导》,《出版广角》,2024年第11期,第26-27页。
- ⑮ 说明:图2中的X轴为词条指代的文化符号出现的年代,Y轴为一级类目分布(Y的取值大小不具备实际数值意义,仅体现分类差异;仅在Y轴以正、负象限区别相应词条属于人文社会领域或自然科学范畴),气泡大小代表词条被提及的频次,即面积越大出现越多。诞生年代过于久远的事物(超出公元前4000年)统一归入“史前时期”,X轴的刻度为非等比呈现。
- ⑯ 说明:图3中以各个文化类目为比较总体,即“仅被国内大模型提及的内容+仅被国外大模型提及的内容+共同提及的内容=100%”。鉴于分属“其他”类的词条整体占比过小(1.1%),图中未对此分类上的差异进行比较。
- ⑰ Wodrich M. D., Sawatlon B, Busch M, et al. *The Genesis of Molecular Volcano Plots*. Accounts of Chemical Research, vol. 54, no. 5, 2021. p. 1108.
- ⑱ 说明:图4中的X轴数值代表同一词条在国内大模型上所占比重与其在国外大模型上所占比重的差,即位于Y轴左侧上半部分的词条被国内大模型呈现的几率大于国外,位于数轴右侧下半部分则表示被国外大模型提及的几率大于国内,位于Y轴中段的8个词条(X值为0)在国内外大模型上所占的比重相差不大。
- ⑲ 说明:图6中的百分比为每个大模型中呈现/缺失的内容在全体词条中的占比,水平线上方为被展示的种类,水平线下方为该模型相对整体模型而言缺失的内容。
- ⑳ 说明:图7中X轴为词条指代的文化符号出现的年代,Y轴为一级类目分布(Y的取值大小不具备实际数值意义,仅体现分类差异)。每个大模型为一个象限,气泡的面积大小代表加权后的词条曝光频次(加权系数为该大模型的活跃用户规模),用灰色标识的气泡为四个大模型共同提及的内容。
- ㉑ Huh, M., Cheung, B., Wang, T., Isola, P. *The Platonic Representation Hypothesis*. ArXiv, abs/2405.07987, 2024-07-25.
- ㉒ 匡恺、刘力铭:《生成式人工智能的风险特征及语义图像:基于网络社区讨论的计算文本分析》,《河北大学学报》(哲学社会科学版),2024年第6期,第155-156页。
- ㉓ Roth, S., Clark, C., Trofimov, N., et al. *Futures of A Distributed Memory. A Global Brain Wave Measurement (1800-2000)*. Technological Forecasting and Social Change, vol. 118, 2017. pp. 307-310.
- ㉔ HAI. 2024 AIIndex Report. <https://aiindex.stanford.edu/report/>. 2024-04.

(作者何苑系河北大学新闻传播学院副教授,河北大学伊斯兰合作组织研究中心、跨文化研究中心研究员;张洪忠系北京师范大学新闻传播学院教授、博士生导师)

(上接第26页)

- ㉕⑳ Eemeren, F. H. V. *The Strategic Maneuvering in Argumentative Discourse: Extending the Pragma-Dialectical Theory of Argumentation*. Amsterdam: John Benjamins Publishing Company. 2010.
- ㉖ Dillard, J. P. *A Goal-Driven Model of Interpersonal Influence*. In J. P. Dillard (ed.). *Seeking Compliance: The Production of Interpersonal Influence Messages*. Scottsdale, AZ: Gorsuch Scarisbrick. 1990. pp. 41-56; Dillard, J. P. *Explicating the Goal Construct: Tools for Theorists*. In J. O. Greene (ed.). *Message Production: Advances in Communication Theory*. Mahwah, NJ: Erlbaum. 1997. pp. 47-69; Dillard, J. P. *Goals-Plans-Action Theory of Message Production: Making Influence Messages*. In L. A. Baxter and D. O. Braithwaite (eds.). *Engaging Theories in Interpersonal Communication: Multiple Perspectives*. London: Sage. 2008. pp. 65-76.
- ㉗ 资料来源: <https://www.klm.com.cn/information/corporate/business-integrity>, 2024-10.
- ㉘⑳ 鞠实儿:《广义论证的理论与方法》,《逻辑学研究》,2020年第1期,第1-27、14页。
- ㉙⑳⑳⑳ Coombs, W. T. *Ongoing Crisis Communication: Planning, Managing, and Responding (2nd edition)*. London: Sage. 2007.
- ㉚⑳⑳⑳ Beonit, W. L. *Apologies, Excuse, and Accounts: A Theory of Image Restoration Discourse*. Albany: State University of New York Press. 1996.
- ㉛ Schultz, D., Patti, C. H. and Kitchen, P. J. *The Evolution of Integrated Marketing Communications: The Customer-driven Marketplace*. London: Routledge. 2013.
- ㉜ Allport, G. W. and Postman, L. *The Psychology of Rumor*. New York: Henry Holt. 1947.
- ㉝ Heide, A. D. E. *Disaster Planning*. New York: Free Press. 1996.
- ㉞ Regester, M. *Crisis Management: What to do When the Unthinkable Happens*. London: Hutchinson Business. 1989.
- ㉟ Fink, S. *Crisis Management: Planning for the Inevitable*. New York: American Management Association. 1986.
- ㊱ Covello, V. T. and Allen, F. *Seven Cardinal Rules of Risk Communication*. Washington, DC: U. S. Environmental Protection Agency. 1988.
- ㊲ Cowley, R. A. *A Total Emergency Medical System for the State of Maryland*. Maryland State Medical Journal, vol. 24, no. 7, 1975. pp. 37-45.
- ㊳ Scher, S. J. and Darley, J. M. *How Effective are the Things People Say to Apologize? Effects of the Realization of the Apology Speech Act*. Journal of Psycholinguistic Research, vol. 26, no. 1, 1997. pp. 127-140.

(作者陈彦瑾系暨南大学新闻与传播学院讲师;鞠实儿系广东外语外贸大学外国语言学及应用语言学研究中心、中山大学逻辑与认知研究所教授、博士生导师)