E-Commerce Letters 电子商务评论, 2024, 13(2), 986-990
Published Online May 2024 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2024.132118

Hans汉斯

上海大模型产业发展现状、优势及建议

朱嘉琳，常永波，陈俊琰*

中国信息通信研究院华东分院，上海

收稿日期：2024年1月15日；录用日期：2024年2月4日；发布日期：2024年5月10日

摘　要

随着人工智能(AI)技术的快速发展，大模型成为AI产业的一个研究热点。上海作为中国的经济中心和科技创新城市，也正在大力发展大模型产业。本文旨在深入分析上海大模型产业的发展现状和特点，探讨上海在数据、算力、算法等要素方面的发展优势，并提出相应的未来发展建议，以期为上海大模型的发展提供参考和指导。

关键词

上海，大模型，人工智能，数据，算法，算力

The Current State, Advantages, and Recommendations for the Development of the Large Model Industry in Shanghai

Jialin Zhu, Yongbo Chang, Junyan Chen*

East China Branch of China Academy of Information and Communications Technology, Shanghai

Received: Jan. 15th, 2024; accepted: Feb. 4th, 2024; published: May 10th, 2024

Abstract

The rapid advancement in artificial intelligence (AI) has positioned large models at the forefront of academic inquiry. Shanghai, as a key economic and scientific nexus in China, is significantly investing in the development of the large model industry. This paper aims to provide an in-depth examination of Shanghai's large model industry, focusing on its current developmental trajectory and distinctive features. Additionally, it evaluates Shanghai's competitive advantages in data acquisition, computational power, and algorithmic innovation. Finally, we also offer targeted rec-

*通讯作者。

文章引用: 朱嘉琳, 常永波, 陈俊琰. 上海大模型产业发展现状、优势及建议[J]. 电子商务评论, 2024, 13(2): 986-990.
DOI: 10.12677/ecl.2024.132118

ommendations for the industry's future growth, aiming to contribute insights and directives for the evolution of Shanghai's large model industry.

Keywords

Shanghai, Large Model, Artificial Intelligence, Data, Algorithm, Computing Power

1. 上海大模型的产业发展现状

2022 年下半年，OpenAI 发布 ChatGPT，掀起了大模型和 AIGC 在全球范围内的讨论热潮[1] [2]。AI 是上海的三大先导产业之一，上海市相关的创新主体也在快速跟进大模型的发展态势，产业发展呈现出以下特点：

一是大模型发布数量较多。截至 2023 年 11 月，我国目前已有 188 家机构发布了 AI 大模型，上海大模型的数量仅次于北京，排名第二，达到了 22 家[3]。

二是通用 + 垂类大模型并行。通用大模型聚焦基础层，泛化能力较强；垂类大模型更关注垂直领域或任务的解决，专业能力较强[4]。上海 AI 实验室、商汤科技、MiniMax 已发布了通用大模型，并成为国内第一批通过《生成式 AI 服务管理暂行办法》备案的 AI 大模型产品[5]；达观数据、复旦大学、虎博科技等机构也发布了金融、工业、科研等领域的垂类大模型。

三是相关投融资活跃。大模型是资本密集型产业，需要持续的资金投入以支撑高昂的基础设施、训练和研发人才的成本[6]。据不完全统计，目前共计 60 家大模型企业(含相关大模型产品及应用)获得融资，其中上海企业接近 20 家，占比近三分之一，略低于北京(31 家)，但高于全国其他地区。例如，MiniMax 在 2023 年 6 月 1 日完成了超 2.5 亿美元的新一轮融资，估值超 12 亿美元，有望成为 AIGC 领域的“独角兽”。

四是落地场景加快推进。上海目前已拥有丰富的大模型落地应用场景，涉及工业制造、企业服务、个人助理、创意设计、智能搜索、科学研究等各领域。例如，上海 AI 实验室发布的“风乌”大模型仅需 30 秒即可生成未来 10 天的全球高精度预报结果，效率大幅优于传统模型；达阔科技发布的 RobotGPT 与机器人的具身智能相结合，可以完成复杂场景的应用，实现虚实融合。

2. 上海大模型产业的发展优势

数据、算力和算法是人工智能时代下最重要、最活跃、边际回报最高的生产要素[7]。上海之所以成为大模型的创新高地，与其在数据、算力和算法方面的优势密不可分。

2.1. 数据支撑优势

数据是人工智能发展的“燃料”。大模型通常具有数以亿计的参数，这些参数在经过大规模数据集的训练后，能够显著提升模型对于复杂任务的理解和处理能力[8]。因此，大模型的数据集应重点考虑数量、质量、种类等方面。

上海是中国的经济中心，许多科研机构、大型企业总部或分支机构位于上海。这些机构在运营中积

累了业务、消费和市场等领域的丰富数据，为大模型的训练打下了坚实的数据基础。同时，社交、传媒、游戏是上海的优势产业，小红书、哔哩哔哩、喜马拉雅等相关企业积累了大量的图像和音频数据，促进了多模态大模型的进步[9]。

此外，上海市政府还一直致力于推动数据开放和共享。《上海市公共数据共享实施办法(试行)》[10]强调公共数据应当全量上链、上云，充分共享。上海图书馆等公共机构建设了相关的开放数据平台，持续扩大普惠供给。2023 年 7 月 6 日世界人工智能大会开幕式上，上海 AI 实验室还联合中央广播电视总台等十家单位联合发起了“中国大模型语料数据联盟”，旨在通过链接模型训练、数据供给、学术研究、第三方服务等多方机构，联合打造多知识、多模态、标准化的高质量语料数据和生态圈。8 月 14 日，该联盟开源发布“书生 • 万卷”1.0 多模态预训练语料[11]，开源数据总量超过 2TB。

2.2. 算力支撑优势

算力是“加速器”，可以实现人工智能的高性能计算、大规模数据处理和复杂模型的训练。大模型在算力端带来了一些新需求，需要更大的互连带宽、内存容量和内存带宽，以及更小的延迟和更高的稳定性、可靠性[12]。

芯片方面，上海目前拥有完善的芯片产业配套政策、健全的基础设施建设及充足的高技能人才，涵盖云端训练、云端推理和终端推理三大产品领域，智能芯片的发展居于全国前列[13]。根据毕马威中国“芯科技”新锐企业 50 榜单，约六成的企业集中在长三角地区，其中上海就有 17 家[14]。

智算中心方面，自 2022 年起，上海市陆续发布了《新型数据中心“算力浦江”行动计划(2022~2024 年)》《上海市数字经济发展“十四五”规划》《上海市推进算力资源统一调度指导意见》《上海打造未来产业创新高地发展壮大未来产业集群行动方案》等一系列相关政策举措。截至 2022 年底，算力基础设施分布集群化态势显著，在用互联网数据中心达到五十余个，标准机架数四十余万个，在用算力总规模约 8995 Pflops，智算规模约 5675 Pflpos [15]。目前，上海的“一平台、五中心”的智算中心建设格局初步形成，在将来可为城市数字化转型提供坚实的算力底座。

2.3. 算法支撑优势

算法是“引擎”，决定了 AI 系统的学习、推理和决策过程[16]。上海在大模型技术和应用转化方面具有较好的算法基础，部分研究机构和企业在语言模型、图像识别、智能推荐等领域进行了深入研究。例如，复旦大学邱锡鹏教授团队提出了名为 LOMO (低内存优化)的新优化器，并在配备 8 卡 RTX 3090 的单台服务器上，成功微调 65B LLaMA [17]。上海交通大学 MediaBrain 团队和上海 AI 实验室智慧医疗团队等提出了一种基于配准的少样本异常检测框架 RegAD，用于学习多个异常检测任务之间共享的通用模型，无需模型参数调整，仅利用少量正常样本，就可以直接应用于新的异常检测任务[18]。

此外，开源开放是驱动大模型技术创新和产业生态集聚的重要路径之一。从国际上看，Meta、LlaMa2、OpenAI 等相继开源，催生大模型和生成式 AI 创业热潮；从国内看，复旦大学 Moss 为国内最早开源的大模型[19]。上海 AI 实验室“书生 •浦语”开源了 InternLM-7B、InternLM-20B、InternLM-123B 三款模型，并推出了首个面向大模型研发与应用的全链条开源体系。

3. 上海大模型产业的未来发展建议

上海的大模型产业虽然已在语料、算力和算法方面拥有较多的优势，但相比于首都北京和部分发达国家，仍然面临许多挑战和困难。下面从数据、算法、算力、人才四个方面提出未来的相关发展建议。

第一，数据方面，应加快健全数据管理机制，积极提高上海市公共领域存量数据的挖掘、清洗水平，

构建高质量、具有上海特色的大模型数据集。探索推进公共数据授权运营，推动公共数据与市场化数据平台对接，实现数据融合创新应用[20]。充分发挥上海数据交易所的市场配置作用，重点推动专业领域挂牌数据的转化。加强数据安全和隐私保护和对企业内部数据治理的评估辅导[21]。研究数据供给与需求的衔接问题，使高质量数据与大模型间形成更流畅、合理的联动机制。

第二，算法方面，应鼓励和支持研究机构和企业等创新主体加强对通用大模型、垂类大模型、多模态多任务大模型的研发与创新。重点支持全自主研发、大模型复现 + 微创新、国内开源 + 微调等模式下的产品研发，对于基于插件、API 接口调用/集成等提供适当关注[22]。支持加强基础架构研究，力争探索形成与 Transformer 差异路线、符合上海需求的新型基础架构。依托上海国际算法中心，围绕下一代 AI、科学智能等方面开展研究，打造出“场景挖掘－算法创新－工程实现－产业落地”的循环[23]。

第三，算力方面，应建立算力专项补贴，针对重点大模型创新团队研发所产生的算力成本、在沪建设的基于国产高性能芯片的大规模智能计算集群优先给予返点优惠或算力补贴等政策支持，重点面向中小企业提供“算力券”优惠补贴[24]。参考绿电购买模式，进一步探索推动政府打包采购社会智算资源的模式，开放更加优惠的算力供给服务。统筹优化算力部署和调度，深入参与并实施“东数西算”工程，促进云边端算网融合，提供“联接 + 感知 + 计算 + 智能”的算网一体化服务[25]。

第四，人才方面，应加快大模型领军人才引进，给予顶尖科学家更大的研发自主权。建立健全产学研联动融合培养机制，鼓励高校开设与大模型的相关专业和研究方向，推动跨学院、跨学科 AI 合作，并加强企业自主人才培养和人才转型[26]。依托上海 AI 实验室等高水平平台开展大学生大模型使用培训，通过体验营、竞赛等形式，培养大学生国产大模型使用习惯。建议颁布大模型人才专项政策，实现从人才培养、高端人才引进到优质人才保留的完整人才政策，梳理大模型领域人才需求清单，制定引才育才计划[27]。

基金项目

本项研究受到上海市 2023 年度“科技创新行动计划”软科学研究项目(项目编号 23692102200)资助。

参考文献

[1] 郭凯明. 人工智能发展、产业结构转型升级与劳动收入份额变动[J]. 管理世界, 2019, 35(7): 60-77+202-203.

[2] 刘古权, 赵育慧. 大模型为什么不能自主决策? [J]. 企业管理, 2024(1): 104-107.

[3] 余生不设限. 国内外大模型汇总[EB/OL]. https://mp.weixin.qq.com/s/XMOZfp071uf6FC10yeZfgw, 2023-12-11.

[4] 徐月梅, 胡玲, 赵佳艺, 等. 大语言模型的技术应用前景与风险挑战[J]. 计算机应用, 2023: 1-10.

[5] 赵子忠, 王喆. 2023 年国内大模型发展综述与趋势研判[J]. 青年记者, 2023: 1-4.

[6] 于水, 范德志. 新一代人工智能(ChatGPT)的主要特征、社会风险及其治理路径[J]. 大连理工大学学报(社会科学版), 2023, 44(5): 28-34.

[7] 郭华源, 刘盼, 卢若谷, 等. 人工智能大模型医学应用研究[J]. 中国科学:生命科学, 2024: 1-25.

[8] 范德志, 于水. 生成式人工智能大模型助推实体经济高质量发展: 理论机理、实践基础与政策路径[J]. 云南民族大学学报(哲学社会科学版), 2024, 41(1): 152-160.

[9] 李耕, 王梓烁, 何相腾, 等. 从 ChatGPT 到多模态大模型: 现状与未来[J]. 中国科学基金, 2023, 37(5): 724-734.

[10] 上海市政府办公厅. 上海市公共数据共享实施办法(试行) [EB/OL]. https://www.shanghai.gov.cn/nw12344/20230311/18d7ba4ffa69423489889bb7af9d78c5.html, 2023-03-02.

[11] 上海人工智能实验室. 大模型语料数据联盟开源发布高质量多模态语料“书生·万卷” [EB/OL]. https://www.shlab.org.cn/news/5443473, 2023-08-14.

[12] 舒文韬, 李睿潇, 孙天祥, 等. 大型语言模型: 原理、实现与发展[J]. 计算机研究与发展, 2024: 1-12.

[13] 张鹏飞, 田雯, 武振宇, 等. AI 大模型场景下智能计算技术选型分析[J]. 电信工程技术与标准化, 2024, 37(1):

3-7.

[14] 毕马威中国. 中国“芯科技”新锐企业 50 报告(第四届)重磅发布 半导体产业新解析[EB/OL]. https://mp.weixin.qq.com/s/LZuBp-HG12pCafNTWQ_AsA, 2023-11-08.

[15] 杨洁. 夯实底座 上海算力生态持续进阶[N]. 中国证券报, 2023-10-18(A05).

[16] 戎珂, 施新伟, 吕若明. “i7 算”赋能 AI 产业生态可持续发展[J]. 科学学研究, 2024: 1-15.

[17] Lv, K., Yang, Y.Q., Liu, T.X., Gao, Q.H., Guo, Q.P. and Qiu, X.P. (2023) Full Parameter Fine-Tuning for Large Language Models with Limited Resources. http://arxiv.org/abs/2306.09782

[18] Huang, C.Q., Guan, H.Y., Jiang, A.F., Zhang, Y., Spratling, M. and Wang, Y.F. (2022) Registration Based Few-Shot Anomaly Detection. https://arxiv.org/abs/2207.07361

[19] 侯树文, 王春. 复旦 MOSS 距离 ChatGPT 还有多远? [N]. 科技日报, 2023-02-23(002).

[20] 金晶. 欧盟的规则, 全球的标准? 数据跨境流动监管的“逐顶竞争” [J]. 中外法学, 2023, 35(1): 46-65.

[21] 董航, 李慧芳, 陈泱, 等. 大模型时代的隐私保护与内容安全[J]. 通信世界, 2023(21): 42-45.

[22] 江小涓, 靳景. 中国数字经济发展的回顾与展望[J]. 中共中央党校(国家行政学院)学报, 2022, 26(1): 69-77.

[23] 苏竣, 魏钰明, 黄萃. 基于场景生态的人工智能社会影响整合分析框架[J]. 科学学与科学技术管理, 2021, 42(5): 3-19.

[24] 钟新龙, 渠延增, 王聪聪, 等. 国内外人工智能大模型发展研究[J]. 软件和集成电路, 2024(1): 80-92.

[25] 石勇, 寇纲, 李彪. “东数西算”战略与问题的分析研究[J]. 大数据, 2023, 9(5): 3-8.

[26] 江颖, 吴维刚, 郑伟诗, 等. “计算·AI + X”创新型计算机研究生人才培养模式探索[J]. 计算机教育, 2024(1): 51-55.

[27] 孙坦, 张智雄, 周力虹, 等. 人工智能驱动的第五科研范式(AI4S)变革与观察[J/OL]. 农业图书情报学报, 2024: 1-29.