

传统大厂：百度推出文心一言

- 国内对大模型投入最早的企业，人力/资金/数据/应用多方面储备最领先，生态链建设覆盖芯片-框架-模型-应用，积淀深厚于国内其他家
- 应用层从单一的自然语言理解延伸到多模态，包括视觉、文档、文图、语音等多模态多功能，开放生态合作伙伴超过650家

团队

领导团队：CTO王海峰任项目总指挥，带队其他高管还包括百度集团副总裁吴甜（同时担任深度学习技术及应用国家工程研究中心副主任）、百度技术委员会主席吴华等人

- 王海峰
 - 基本信息：哈尔滨工业大学本、硕、博，博士1999年毕业于，师从ACL终身成就奖获得者、哈工大教育部语言语音重点实验室主任李生，中国工程院院士、北大信息科学技术学院院长高文等
 - 研究方向：机器翻译、自然语言处理、搜索技术、语音技术、机器学习与数据挖掘、推荐及个性化等
 - 学术成就：曾主持及参与了多个产品的研发，已申请中国、美国及日本专利30余项，已发表学术论文超过160篇，已授权或公开的专利申请300余项，作为负责人承担国家核高基重大专项、863重大项目等，并正在承担国家重点研发计划
 - 任职经历：先后在微软中国研究院、东芝等国际顶级科技公司任职；2010年加入百度，2020年王海峰将原AIG（AI技术平台体系）、TG（基础技术体系）和ACG（百度智能云事业群组）整体整合为人工智能体系（AI Group，AIG）其整体负责
- 吴甜：2006年加入百度，目前整体负责百度AI技术平台和智能云AI产品，领衔研发“飞桨深度学习平台（PaddlePaddle）”，百度文心大模型的背后，正需要飞桨深度学习平台的支撑
- 吴华：2010年加盟百度，中科院自动化研究所博士，机器翻译、自然语言处理、机器学习、智能交互
- 团队架构：负责文心一言的团队，主要由TPG（技术中台事业群）、MEG（移动生态事业群）两大部门协同，前者负责技术攻坚，后者进行搜索、内容产品承接

历史沿革

- 2019年3月，提出知识增强的语义理解框架ERNIE，在深度学习的基础上融入知识，同时具备持续学习能力，曾一举登顶全球权威数据集GLUE榜单，首次突破90分大关，刷新榜单历史，为国内首家在此阶段提出布局大模型框架的企业
- 2021年12月，百度与鹏城自然语言处理联合实验室发布全球首个知识增强的千亿AI大模型－ERNIE 3.0 Titan
- 2022年11月30日，百度集团在WAVE SUMMIT+2022深度学习开发者峰会带来了文心大模型的最新升级，包括新增11个大模型
- 2023年2月7日，百度正式官宣将在3月上线百度版ChatGPT“文心一言”（英文名为ERNIE Bot），将多项主流业务如搜索与文心一言整合，于3月16日14时在北京总部召开新闻发布会；3月16日起正式开放预约，搜索“百度智能云”进入官网，可申请加入文心一言云服务测试

技术

- 底层模型及架构：新一代知识增强大语言模型，在ERNIE及PLATO系列模型的基础上研发，关键技术包括有监督精调、人类反馈的强化学习、提示、知识增强、检索增强和对话增强
- 算力：百度阳泉智算中心算力规模可达4EFLOPS，从23年2月开始布局新机房为文心一言提供支持，XPU采用昆仑芯；为了支持千亿参数模型的高效分布式训练，飞桨研发了4D混合并行技术
- 参数：发布会未公布
- 数据量：万亿级网页数据、数十亿搜索数据和图片数据、百亿级语音日均调用数据，以及5500亿事实的知识图谱等

商业化

- 产品形态
 - 落地基于大模型架构多模态交互AI聊天机器人“文心一言”，展示了在文学创作、商业文案创作、数理推算、中文理解、多模态生成五个使用场景中的综合能力
 - 提交注册Searchat、百度百晓生、百度晓搜、Chatflow等多个商标名称，百度搜索全线整合文心一言后，或可能提供文本、语音、图像等多模态的搜索能力和AI生成能力；产品升级大概率将搭载百度APP此前推出的AI数字人度晓晓，提供更智能的AI对话式搜索
 - 同时自研深度学习框架飞桨，目前已开源了140+的模型，亦包含工业级的预训练模型，已可适配22种芯片型号
- 应用场景：目前已经有包括互联网、媒体、金融、保险、汽车、企业软件等行业的超650家企业宣布加入百度“文心一言”生态
- 客户体系：文心一言合作包括彩讯股份、远光软件、银之杰、国光电器、视觉中国、北信源、东软集团、宇信科技、汉得信息、致远互联、航天宏图等

传统大厂：华为开发盘古大模型

- 华为较早进入大模型领域，提出CANN、MindSpore框架、ModelArts平台等不同层级的协同优化，为盘古大模型提供了强大的背后支撑
- 目前To B端应用较广泛，主要应用于智能舆情及智能营销（NLP大模型），To C端正在布局，未来将上传盘古大模型至AI资产共享社区，或针对开发者推出系列课程

团队

- **带头人-田奇**：华为云人工智能领域首席科学家、IEEE Fellow，国际欧亚科学院院士，长期致力于计算机视觉、多媒体信息检索及AI领域研究
 - **基本信息**：本科毕业于清华大学电子工程系，硕士毕业于美国德雷塞尔大学，后赴美国伊利诺伊大学香槟分校学习，师从Thomas S. Huang教授，获得博士学位
 - **过往经历**：2002年至2019年期间，田奇历任过美国德克萨斯大学圣安东尼奥分校计算机系助理教授、副教授和正教授；2018年加入华为诺亚方舟实验室，出任计算视觉首席科学家，2020年3月担任华为云人工智能领域首席科学家
 - **学术成就**：在计算机视觉及多媒体方向顶级期刊及会议如IEEE TPAMI、IJCV、TIP、TMM、CVPR、ICCV、ECCV、ACM MM上发表文章超过650篇，谷歌学术引用次数超过30,700次
 - **获奖情况**：于2017年获UTSA校长杰出研究奖；2018年入选国家级领军人才创新项；22年获有“中国智能科学技术最高奖”之称的“吴文俊人工智能杰出贡献奖”
 - **社会头衔**：为国家自然科学基金海外杰青、中国科学院海外评审专家、IEEE Fellow（2016）、国际欧亚科学院院士（2021）
- **团队规模**：3位华为天才少年、20多位博士、30多位工程师和50多位来自C9高校的参与者

历史沿革

- 2020年开始布局大模型，同年3月田奇先生加入，开始组建团队并且进行方向梳理，8月迎来了核心专家的加入，随后在9月开始推动盘古大模型的立项，11月盘古大模型在华为云内部立项成功，也完成了与合作伙伴、高校的合作搭建
- 基于昇腾AI，与鹏城实验室（提供算力调用）、循环智能合作，华为于2021年4月发布了鹏城盘古大模型，包括中文语言NLP、视觉CV、多模态、科学计算四个大模型，是业界首个千亿级生成和理解中文NLP大模型
- 2022年11月发布预训练大模型白皮书

技术

- **底层模型及架构**：在模型结构上，盘古采用Encode和Decode的架构，保证了其在生成和理解上的性能
- **算力**：华为联合鹏城实验室联合打造“鹏城云脑II”，是我国首个国产E级AI算力平台；盘古大模型在训练时调用了超过2,000块昇腾910进行了两个多月的训练，可以提供超过640P FLOPS的FP16算力
- **参数**：千亿级参数量-NLP（盘古大模型整体为2,600亿参数量）；30亿参数的视觉（CV）预训练模型
- **数据量**：40T中文文本数据量（截止2021年5月统计）

商业化

- **产品形态**：华为已发起了智能遥感开源生态联合体、多模态人工智能产业联合体，智能流体力学产业联合体等，华为将以联合体的模式把科研院所、产业厂商等结合起来，更好地让大模型产业形成正向的闭环回路
- **应用场景**：目前盘古大模型已经应用于科学计算、气象、物流、工业质检等领域，未来在交通、智慧城市、安平、营销、金融等多个领域均有应用机会
- **客户体系**：本次“盘古”服务将借助华为运营商身份的优势，推动汽车、生活、能源、健康等行业数字化发展；目前，已经在100多个行业场景完成验证，有23个类别、9,500多个产品介入“盘古”服务
 - 在物流场景，盘古大模型协助浦发银行构建了“物的银行”-浦慧云仓

传统大厂：腾讯推出混元大模型

- 动手慢+团队小+目前未开放，更像是内部特殊项目组
- 强调目前高质量小团队作战，目前没有依靠开放来换取外部数据，潜在内部有大量高质量数据优势
- 强调TEG（AI Lab挂靠的部门）/PCG/CSIG的赞助支持，同时也包括了微信和互娱的人员

团队

- 参与团队¹：联合腾讯内部多方团队构建大参数语言模型，共18人，Owner为张正友，PM分别为俞栋、王迪、刘田，现该项目有至少7位组长、7位Sponsor
- Owner张正友：擅长方向偏计算机视觉、立体视觉、3D重建
 - 学历背景：1985年本科毕业于浙江大学信息与电子工程系，1987年硕士毕业于在南锡第一大学（现洛林大学）计算机专业，1990年获巴黎第十一大学计算机科学博士学位
 - 工作背景：曾就职于微软研究院，带领视觉团队从事大量学术研究，在Windows、Office、Xbox、Kinect、Skype for Business、Office Lens等诸多微软产品里都有其团队的贡献；2018年加入腾讯，带领创建了腾讯机器人实验室Robotics X，2021年1月，腾讯宣布张正友成为**腾讯首位17级研究员/杰出科学家**
 - 能力体现：在国际顶尖会议和杂志上发表论文250余篇，论文引用次数51,000多次，有超200项专利；其于1999年论文提出的摄像机标定法“张氏标定法”在全世界被普遍采用，获2013 IEEE Helmholtz时间考验奖
- PM团队
 - 俞栋：TEG-AI Lab副主任，负责强化学习和多轮对话，自动语音识别、语音处理和自然语言处理等，发明了100多项专利，曾获得IEEE、NAACL的多个论文奖项，谷歌学术引用量5余万次
 - 王迪：负责大模型、预训练，现为**腾讯TEG-AI平台部的助理总经理**，2008年7月校招加入腾讯，16年采访得知当时他与团队正在给公司内部其他BG的业务产品提供“搜索引擎检索系统”相关模块的开发
 - 刘田：商业化组长
- 7位组长：Ying Shan（PCG-ARC Lab杰出科学家/主任）、李迪夫、康战辉、陶阳宇、王涛、王波、Daniel

历史沿革

- 2022年4月，腾讯首次对外披露HunYuan研发进展，HunYuan集CV、NLP、多模态理解能力于一体，先后在MSR-VTT，MSVD等五大权威数据集榜单中登顶，实现跨模态领域的大满贯
- 2022年12月，混元AI大模型团队推出了万亿中文NLP预训练模型HunYuan-NLP-1T，实现在中文语理解能力上的新突破
- 2023年2月27日，腾讯针对类ChatGPT对话式产品已成立“混元助手（HunyuanAide）”项目组

技术

- 底层模型及架构：腾讯太极机器学习平台自研的训练框架AngelPTM
- 算力：基于腾讯太极机器学习平台
- 参数：万亿
- 数据量：1.7TB；22年11月推出HunYuan-NLP 1T大模型，仅用256卡在一天内即可完成万亿参数大模型的训练，整体训练成本仅为直接冷启动训练万亿模型的1/8

商业化

- 产品形态：在广告内容理解、行业特征挖掘、文案创意生成等方面的应用
- 应用场景：HunYuan先后支持了包括微信、QQ、游戏、腾讯广告、腾讯云等众多产品和业务，通过NLP、CV、跨模态等AI大模型，不仅为业务创造了增量价值而且降低了使用成本
- 客户体系：目前进度以内部为主，未来计划To B；目前**HunYuan-NLP-1T大模型已成功落地于腾讯广告、搜索、对话等内部产品**并通过腾讯云服务外部客户

传统大厂：阿里布局通义大模型

- 21年发布M6后放缓节奏，2023年2月重启，全力投入生成式AI
- 现阶段以达摩院研究主导
- 大模型较早，现已实现十万亿参数级，算力消耗低，对外商业化布局进展稍慢

团队

领导团队

- 周靖人：达摩院副院长及阿里云智能 CTO；哥伦比亚大学计算机科学博士，浙江大学兼职教授，IEEE Fellow，现任阿里集团资深副总裁，曾负责阿里云计算平台，淘宝和支付宝搜索、推荐、广告等数据智能业务；研究方向包括云计算、大数据、人工智能算法及应用，多次担任 VLDB，SIGMOD，ICDE 等国际顶级会议程序委员会主编、主席，在顶尖国际期刊和会议上发表论文超百篇，并拥有几十项技术专利
- 司罗：达摩院语言技术实验室负责人；2014 年成为阿里人工智能科学家阵营的一员，带领阿里 NLP 团队取得多项重要成果，美国普渡大学计算机系的终身教授，并先后获得美国国家科学基金会成就奖、雅虎、谷歌研究奖等
- 李永彬：毕业于清华大学，达摩院资深算法专家，研究方向为Conversational AI & 预训练大模型，近年来围绕对话预训练、人机对话交互、人机协同交互、多模态对话等方向发表30+篇顶会论文
- 林俊暘：智能计算实验室算法专家，2019年毕业于北京大学，侧重于多模态预训练及其应用，曾在自然语言处理、机器学习等领域以第一作者和共同作者的身份在多个顶级会议包括ACL、KDD、EMNLP、NeurIPS、ICML等发表多篇论文；擅长超大规模多模态预训练基础研发，以及其相关下游应用如图像生成、自然语言生成、跨模态检索等，自然语言生成及预训练相关工作已经落地到阿里推荐业务场景中
- 团队背景：阿里达摩院M6团队致力于认知智能方向，研发了AliGraph、M6、洛犀等较为有影响力的人工智能开源平台和系统，发表顶级会议、期刊文章超过100篇，美国和中国专利超过30项；获2019年世界人工智能大会最高奖卓越人工智能引领者（Super AI Leader，简称SAIL奖），2020年国家科学技术进步奖二等奖和杭州市领军型创新团队，2021年电子学会科学技术进步奖一等奖

历史沿革

- 2021年开始在AI大模型领域加码投入，开发出百亿参数的超大规模中文多模态预训练模型M6（对标中文GPT3）及被称为“中文版GPT-3”的语言大模型PLUG
- 2022年下半年，阿里巴巴达摩院发布“通义”大模型系列，核心模型通过“魔搭”社区向全球开发者开源开放，降低了AI的应用门槛
- 23年2月23日，阿里张勇宣布全力投入生成式AI大模型建设，向外界确认了正在研发类ChatGPT产品，已进入内测阶段

技术

- 底层模型及架构：GPT3.0开源版；阿里版ChatGPT为基于其在2022年9月发布的“通义”大模型系列融合升级而来，通义统一底座中的单一M6-OFA模型（业界最大的中文多模态预训练模型），在不引入任何新增结构的情况下，可同时处理图像描述、视觉定位、文生图、视觉蕴含、文档摘要等10余项单模态和跨模态任务
- 参数：十万亿以上；据信源表示有水分；对比参数规模1,750亿的GPT-3模型，只需其1%的算力，运行可提速10倍
- 算力：阿里云在张北和乌兰察布建设有两座超级智算中心，建设规模分别为12 EFLOPS和3 EFLOPS AI算力
- 数据量：预训练数据量1.9TB图像和292GB文本

商业化

- 产品形态
 - 类ChatGPT产品内测版本已被提前曝光，不仅能通过文字与人对话，甚至还具备“AI绘画”等多模态任务能力；据透露，新产品可实现的功能包括知识问答、AI绘画、代码生成、小说续写、文案撰写、写诗作词等
 - 22年11月推出“魔搭”社区 ModelScope，首批合作机构包括澜舟科技、智谱 AI、深势科技、中国科学技术大学等多家科研机构，首批开源模型超过 300 个，中文模型占比超过 1/3，SOTA 模型超过 150 个，有 10 多个前沿的大模型，覆盖了计算机视觉、语音、语义、多模态等多个领域，覆盖了 60 多个通用任务
- 应用场景：服装等外观设计、推荐理由生成、图片描述生成、智能及视觉问答、图文检索、新闻分类、金融领域内的观点抽取及事件点评
- 客户体系：服务内部近50个部门，并在阿里云对外200+产品中投入使用，如每平每屋、斑马智行、犀牛制造等

传统大厂：京东布局ChatJD

- 基于言犀人工智能应用平台宣布将整合过往产业实践和技术积累，推出产业化ChatJD
- 作为辅助京东主营业务的工具，由原Chat AI迭代，**当前仍处于布局及起步阶段**，对外合作、商业化尚未见明显进展

团队

领导团队

- **何晓东**：现任京东集团副总裁、京东AI研究院执行院长，京东大模型研究团队的带头人
 - **学历背景**：于1996年获得清华大学学士学位，1999年获得中国科学院硕士学位，2003年获得美国密苏里大学哥伦比亚分校博士学位，研究方向主要聚焦在人工智能领域
 - **能力积累**：曾担任微软雷德蒙德研究院深度学习技术中心的首席研究员和负责人，研究主要集中在人工智能领域，包括深度学习，自然语言处理，语音识别，计算机视觉，信息检索和多模态智能；2018年从微软离开加入京东，现任京东集团副总裁、IEEE FELLOW、智能客户服务事业部总裁，京东人工智能研究院常务副院长，深度学习及语音和语言实验室的负责人，在深度学习、自然语言处理、语音识别、计算机视觉及信息检索等方面均有重要贡献
- **吴友政**：京东科技语音语言算法部负责人
 - **基本情况**：2006年中科院博士毕业后，先后在日本国立信息通信研究机构、英国爱丁堡大学、索尼中国研究院从事自然语言处理相关研究工作
 - **能力积累**：工作主要聚焦自然语言处理、人机对话、语音识别、机器翻译等前沿技术研究和产品研发工作，并取得了诸多技术突破，累计发表顶级国际会议和期刊论文30余篇，曾获得语音识别（IWSLT2012、IWSLT2013）和自然语言处理（QuAC2021，DROP2022，Multiwoz2022）领域多项国际比赛的冠军，与团队密切合作打造了言犀人工智能应用平台

历史沿革

- 京东从2020年开始研究AI文本生成，用以支持京东商城商品页面描述的生成，期间，由于NLP技术的整体消沉，研发受阻，但后来京东又重新重视，推出了言犀大模型
- 2021年，何晓东与原京东AI掌门人周博文（已于2021年11月离职）带队，在京东言犀篇平台上研发了领域性模型K-PLUG该模型与京东的电商场场景紧密结合，从电商领域特定知识中学习，能够自动生成商品文案
- 2023年，京东云旗下言犀人工智能应用平台推出产业版ChatGPT，命名为ChatJD

技术

- **底层模型及架构**：言犀大模型，言犀在京东内外部产业实践中衍生了许多技术成果，尤其是在自然语言处理和多模态交互技术领域，先后斩获了WikiHop（2019）、HotpotQA（2021）、QuAC（2020，2021）、DocRED（2021）、DROP（2022）6项国际AI竞赛冠军
- **算力**：暂无公开信息
- **参数**：**预计参数量达千亿级**
- **数据量**：暂无公开信息

商业化

- **产品形态**
 - ChatJD智能人机对话平台，即自然语言处理中理解和生成任务的对话平台，包括内容生成、人机对话、用户意图理解、信息抽取、情感分类这五种应用
 - 在通用型 Chat AI 方向，京东云已经打造了包括京东智能客服系统、京小智平台商家服务系统、智能金融服务大脑、智能政务热线，言犀智能外呼、言犀虚拟主播等系列产品和解决方案
- **应用场景**：聚焦零售、金融，得益于京东云在零售与金融领域10余年真实场景的深耕与沉淀，已拥有4层知识体系、40多个独立子系统、3,000多个意图以及3,000万个高质量问答知识点，覆盖超过1,000万种自营商品的电商知识图谱，更加垂直与聚焦
- **客户体系**：京小智将作为京东布局ChatGPT重要产品之一，未来将向产业端持续拓展

传统大厂：快手正开展LLM研究

- 目前看到新闻称要布局LLM做与业务场景融合的产品如虚拟人/智能客服等，无其他进展信息
- 目前定位为专项项目，由MMU团队支持研发

团队

- 领导团队
- 王仲元：快手技术副总裁、MMU&Y-tech 负责人
 - 基本情况：曾担任 Facebook 公司 Research Scientist，负责 Facebook 产品 NLP Service；在 Facebook 之前，曾任微软亚洲研究院的主管研究员，负责微软研究院知识图谱项目和对话机器人项目，后任美团点评高级研究员、高级总监、搜索与 NLP 部负责人；2020 年 10 月加盟快手，负责多媒体内容理解部
 - 能力评价：多年来专注于自然语言处理、知识图谱及其在文本理解方面的研究，在国际顶级学术会议如 VLDB、ICDE、IJCAI、CIKM 等发表论文数十篇
 - 张富峥：快手 MMU NLP 中心负责人
 - 基本情况：主要负责 NLP 及相关多模态技术在快手场景下的建设和应用，发表论文 60+ 篇，谷歌引用 7800+；张富峥博士入选了斯坦福大学发布的「全球前 2% 顶尖科学家榜单」以及「2022 年人工智能全球最具影响力学者榜单 AI 2000 中的数据挖掘领域排名前 30 的最具影响力学者」
 - 研发团队架构：快手 MMU 是负责快手短视频搜索系统以及短视频理解的 AI 中台，业务覆盖 OCR，ASR，分词，NER 等基础 AI 算法，短视频分类，标签体系建设等中台技术以及短视频搜索等系统服务；目前主导大规模语言模型 LLM 研究专项

历史沿革

- 23年2月，快手确认正在开展大规模语言模型LLM相关的研究，并启动了相应专项，覆盖LLM模型训练、文案自动创作与生成、对话系统开发等领域

技术

- 底层模型及架构：暂无公开信息
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息

商业化

- 产品形态：暂无公开信息
- 应用场景：快手在LLM领域的研究，更聚焦于用户体验的提升，以及和业务场景的融合；例如，对话系统在快手中的主要应用包括虚拟人、电商智能客服、智能助理等人机交互或人人交互场景
- 客户体系：暂无公开信息

传统大厂：小米布局AI大模型

- 2023年2月高管团队提及小米在AI大模型领域早有尝试，但现有产品形态及技术参数尚不明确
- AI实验室带队进行大模型相关研发

团队

- 王斌：小米AI实验室主任，NLP首席科学家
 - 学历背景：本科、硕士毕业于武汉大学，博士毕业于中科院计算所
 - 能力积累：加入小米之前在中科院从事NLP及信息检索的研发工作，曾为中科院研究员、博导，研究方向为信息检索与自然语言处理，目前主要负责小米AI实验室的技术研发和落地工作
 - 代表成果：小米AI实验室的搜索、推荐、广告等技术支持了小米商城、有品、新零售、广告推荐等业务，利用多模态技术自动生成广告创意，成功支持广告投放，机器写作基于最新的预训练生成技术在对联、诗歌生成领域达到接近人类水平的效果，应用到小爱同学节日运营传播

历史沿革

- 2016年初正式成立小米AI实验室，研究方向包括计算机视觉、声学、语音、NLP、知识图谱、机器学习等多个领域
- 23年2月9日，小米投资者会议上宣称早就在AI大模型等方面有尝试，并且是多路并行（多技术路线）；小米关注基于硬件的AI技术支撑，将语音交互作为基础能力进行构建，优势体现在应用场景多样化、数据丰富度高；在自然语言处理核心技术上，团队重点打造了搜索、推荐、广告、多模态、机器翻译、人机对话和文本创作等方向

技术

- 底层模型及架构：暂未公开信息
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息，但表明小爱已经拥有庞大的高质量数据支撑，会让小米在大模型方面的进展更迅速

商业化

- 产品形态：暂无公开信息
- 应用场景：全栈自研的小爱同学是典型的大模型落地场景，现有结果表明通过应用大模型能够获得比原来更好的效果，小爱对话、机器人亦是落地场景；此外，通过引入更先进的大模型能力，包括探索多模态能力，对智能座舱交互体验提升会有很大的助力
- 客户体系：暂无公开信息

传统大厂：字节计划今年年中推出大模型

▪ 目前公司内部架构/产品体系/进展仍未官方公布

团队

- 领导团队**
- **朱文佳**：加入字节跳动前，曾在百度担任搜索部主任架构师，2015年进入头条，被称为“头条里算法技术的Top 3人选”，2019年成为今日头条CEO；2021年2月调任成为Tik Tok产品技术负责人，向Tik Tok CEO 周受资汇报
 - **李航**：字节跳动人工智能实验室总监、ACL Fellow、IEEE Fellow、ACM 杰出科学家；硕士毕业于日本京都大学电气工程系，后在东京大学取得计算机科学博士学位；先后就职于 NEC 公司中央研究所（任研究员）、微软亚洲研究院（任高级研究员与主任研究员）、华为技术有限公司诺亚方舟实验室（任首席科学家），主要研究方向包括自然语言处理、信息检索、机器学习、数据挖掘等
 - **团队背景**：语言大模型团队由字节搜索部门牵头，搜索业务部门、AI Lab和AML（应用机器学习）团队均调任部分人马支援，**目前团队规模在10人左右**，图片大模型团队由产品研发与工程架构部下属的智能创作团队牵头；团队负责人的间接和直接汇报对象，则均为 TikTok 产品技术负责人朱文佳

历史沿革

- 2023年2月公布大模型团队建立，**预计2023年年中推出大模型**；同期推出AI绘画/ AI作曲“海绵乐队”，字节的优化主要是训练过程，在云服务架构上，通用的算力部署估计会有1.5-2.5倍的效率提升
- 据久谦库内部分专家信源了解，字节倾向于外包训练大模型

技术

- **底层模型及架构**：暂无公开信息
- **算力**：字节旗下的企业技术服务平台火山引擎，于2021年正式进军基础云市场（IaaS层），为大模型所需的算力奠定了基础
- **参数**：暂无公开信息
- **数据量**：暂无公开信息

商业化

- **产品形态**：产品分别在语言和图像两种模态上发力，探索方向主要为与搜索、广告等下游业务的结合
- **应用场景**：暂无公开信息
- **客户体系**：暂无公开信息

传统大厂：网易伏羲自研玉言系列大模型

- 网易已积累了丰富的预训练模型训练和工程优化经验，虽在模型层研发优势不及腾讯、华为等，但在应用层可与自身音乐、教育、游戏等板块进行协同
- 未来将加大大模型及AIGC应用的研发投入力度（150亿元），目前自研大模型主要服务于集团业务，其他预训练模型B端已有合作成功案例

团队

领导团队

- 胡志鹏：网易高级副总裁、网易雷火事业群总裁
 - 基本信息：甘孜康定人，出生于1978年，从小学直到高中他的成绩一直名列前茅，1997年时参加高考，以608的高分考入浙江大学计算机学院后顺利保研
 - 工作经历：2004年进入微软亚洲研究院，加入到新技术转换组，逐渐成长为了计算机图形引擎专家；2007年加入网易，目前牵头网易伏羲的大规模预训练相关工作
- 吕唐杰：网易伏羲AI研究&技术落地负责人

历史沿革

- 2018年启动GPT（生成式预训练）模型研究，网易目前已自研数十个超大规模预训练模型，覆盖自然语言、文图跨模态、文音跨模态、智能抓取、情绪感知等多种模型领域
- 2022年，网易连续四个季度加码研发，持续推进AIGC等前沿技术自研突破，在游戏、教育、音乐等领域表现出巨大应用潜力，全年研发投入超150亿元，研发投入强度达15.6%

技术

- 底层模型及架构：采用Prefix-LM模型
- 算力：网易伏羲与浪潮信息紧密合作，通过采用AI算力服务模式，解决了超大规模参数大模型所需训练算力需求，同时还获得了高质量中文语言数据集的有力支持
- 参数：暂无公开信息
- 数据量：110亿，网易自有高质量数据、各类百科、新闻、社区问答、wiki、小说等等

商业化

- 产品形态：超大规模预训练模型赋能企业内部板块
- 应用场景：未来相关技术可应用于智能制造、智慧学习、数字文旅、数字娱乐、数字音乐等数字产业
- 客户体系：目前体系外客户合作并非基于中文文本预训练大模型“玉言”系列，其他超大规模预训练模型与总台网络春晚及中建八局合作成功，预训练大模型“玉言”系列预计将应用于网易内部游戏《逆水寒》手游中的智能NPC系统

科研院所：智源研究院储备较多大模型研发技术人才

- 由北京智源人工智能研究院牵头，智谱AI等多家机构共同研发悟道2.0，参数规模达到1.75万亿个，可以同时处理中英文和图片数据，现悟道已开源
- 已与美国、小米、快手、360、搜狗、寒武纪、第四范式、推想科技、好未来、小冰科技等达成合作共建模型生态

团队

- **黄铁军**：北京智源人工智能研究院院长、北京大学信息科学技术学院教授，主要研究方向为视觉信息处理与类脑计算，国家杰出青年科学基金获得者
- **张宏江**：北京智源人工智能研究院理事长、美国国家工程院（NAE）外籍院士，曾任金山软件CEO，微软亚太研发集团CTO，微软亚洲工程院院长，微软亚洲研究院副院长，并成为微软第一批“杰出科学家”
- **曹岗**：智源研究院常务副院长，博士，研究员，负责智源运行管理和成果转化工作，曾任北京市科学技术委员会高新技术产业化处副处长、文化科技处处长、电子信息与装备制造处处长
- **唐杰**：智源研究院学术副院长、清华大学计算机系长聘教授、系副主任；主要研究方向为智能数据挖掘、预训练模型、社交网络和知识图谱，智源研究院“悟道”大模型项目总负责人
- **林咏华**：智源研究院总工程师，负责智源智算平台、开源平台及产业生态建设；曾任IBM中国研究院院长、IBM全球杰出工程师、IBM全球技术研究院核心领导成员，具有多年人工智能（视觉分析）、系统架构、云计算、智能硬件领域的技术研发和研发管理经验，具备面向产业落地先进技术、打造技术生态的丰富经历
- **团队背景**：智源学者94人，分别来自北大、清华、中科院等高校院所与旷视、京东等优势企业

历史沿革

- 北京智源人工智能研究院（Beijing Academy of Artificial Intelligence, BAAI）是落实“北京智源行动计划”的重要举措，在科技部和北京市委市政府的指导和支持下，由北京市科委和海淀区政府推动成立
- 在2018年11月14日举行的2018中国（北京）跨国技术转移大会开幕式上，智源研究院正式揭牌
- 20年5月，智源创新中心成立，助力11个明星项目孵化
- 20年8月，智源算力大平台上线，1000P算力软硬件体系
- 20年10月，启动悟道大模型21年3月，智源才发布了“悟道1.0”，这是由智源研究院学术副院长、清华大学教授唐杰领衔，来自北大、清华、中科院等单位的100余位AI科学家团队联合攻关完成21年9月，建立10个AI开放实验室
- **融资历程**：暂无融资

技术

- **底层模型及架构**
 - **FastMoE**：技术打破国外技术瓶颈，为首个支持PyTorch框架的MoE系统；**GLM2.0**：更通用的预训练模型；
 - P-tuning2.0算法**：极大拉近少样本学习和全监督学习的差距，少样本学习能力强；**CogView**：文本生成图像的新框架，将VQ-VAE和Transformer进行结合，表现SOTA，在MS COCO FID指标上性能优于DALL·E等模型
- **算力**：暂无公开信息
- **参数**：1.75万亿
- **数据量**：4.9TB

商业化

- **产品形态**：悟道2.0（核心）、天宝1.0、九鼎智算平台、FlagAI开源平台
- **客户体系**
 - 联合浪潮信息、昆仑芯科技、壁仞科技、天数智芯、燧原科技、摩尔线程等硬件企业，清华大学、北京大学、中科院计算所优势团队，共建“北京国家新一代人工智能创新发展试验区AI开放生态实验室”，共同开展基础软硬件适配、评测等工作
 - 智源研究院与新华社、美团、小米、快手、搜狗、360、寒武纪、第四范式、好未来、智谱华章等22家“悟道”大模型产业生态战略合作企业进行了签约，签约企业涵盖行业应用企业、IT龙头企业、中小创新企业等
 - 悟道2.0已经与美团、小米、快手、360、搜狗、寒武纪、第四范式、推想科技、好未来、小冰科技等达成合作共建模型生态

新兴挑战者：澜舟科技布局轻量级模型孟子

- 基于大厂研发模型处于企业内部赋能阶段，以及中小企业受限于大模型预训练高成本的特点，**决定做轻量化大模型+垂直领域大模型，深耕算力而非疯狂堆参数的发展战略**
- 目前已华为云、传神语联网、中文在线等企业达成合作，并与北大、清华等10所高校进行产学研的联璧

团队

- **带头人-周明**：世界级AI科学家，自然语言处理领域先驱者，**创新工场首席科学家**
 - **基本信息**：1991年就从NLP重镇哈工大拿到了博士学位，其后进入清华任职任教，并在1999年被微软亚洲研究院创始院长李开复挖走，后来不久成为NLP研究组负责人，于2020年离开，已是MSRA副院长多年，2020年加盟创新工场，担任创新工场首席科学家
 - **科研成果**：参与带队的NLP成果，光入选ACL的论文就超过100篇，相关技术还被应用到了Windows、Office、Azure、微软小冰等世界级产品中
 - **学术成就**：在NLP领域的影响力也位居世界前列，为过去十年来发表NLP论文最多的科学家之一，据谷歌学术最新数据，其文章被引用总量超过3.4万

历史沿革

- 2021年1月澜舟在创新工场开始孵化澜舟公司
- 2021年6月10日公司正式成立，7月就推出轻量化模型—孟子；基于孟子模型作为核心技术，推出了“孟子认知服务引擎”，包括搜索、生成、翻译、对话等功能，还能提供针对垂直行业的SaaS产品
- **融资历程**
 - 天使轮（2021）：创新工场
 - Pre-A轮（2022）：联想创投，斯道资本，创新工场，融资体量近亿元人民币

技术

- **底层模型及架构**：基于Transformer架构，可支持以及快速替换GPT及BERT等多种预训练模型
- **算力**：算力比较低，用16块GPU、训练两三周就能训练出一个孟子轻量化模型
- **参数**：不追求模型参数规模绝对大，探索轻量级模型的实用案例
- **数据量**：数百G，涵盖互联网网页、社区、新闻、电子商务、金融等领域的高质量语料训练

商业化

- **产品形态**：目前澜舟科技认知智能平台年度产品涵盖4大主题：AIGC（智能创作）平台、机器翻译平台、孟子大模型、金融NLP平台，共计20余项产品服务
- **应用场景**：可高效地接入搜索、生成、翻译、对话等业务场景，主要面向营销、金融、文化创意等行业
- **客户体系**：包括华为、中文在线、传神、数说旗下容徽、同花顺及华夏基金等
 - **中文在线**：2023年2月16日公布签订战略合作协议，双方将在AIGC领域拓展合作，AI辅助创作功能已向中文在线旗下17K文学平台全部作者开放
 - **传神**：合作“任度”机器翻译引擎，以孟子模型作为支撑技术，覆盖20余个垂直领域，在传神平台上支持数以千计翻译员
 - **容徽**：于2021年年底成功上线首个合作产品“contentnote智能文案”，以AI助力文案生成
 - **华为云/同花顺/华夏基金**：签署战略合作协议，携手创造“产业+AI”时代的业务价值

新兴挑战者：智谱华章合作推出悟道2.0

- 由清华大学计算机系知识工程实验室Aminer成果转化而来，联合开发预训练语言模型，参与悟道2.0研发
- 现已有一定数量政企客户

团队

- 领导团队**
- **董事长刘德兵**：中科院计算所博士，师从高文院士，清华数据科学研究院科技大数据研究中心副主任，兼任Thomson北京研究院高级研究员，清华计算机系高级工程师
 - **CEO张鹏**：毕业于清华大学计算机科学与技术系，研究领域包括文本数据挖掘和语义分析、知识图谱构建和应用等
 - **学术成果**：ICML、ISWC等顶级会议上发表10+篇文章
 - **参与项目**：欧盟第七合作框架跨语言知识抽取、国家863计划、Aminer、XLORE
 - **李涓子**：清华大学长聘教授，清华大学人工智能研究院知识智能研究中心主任，中国中文信息学会知识与语言计算专业委员会主任
 - **参与项目**：基于语义链接的跨语言知识图谱XLORE、研究者社会网络挖掘和服务系统Aminer
 - **曾获奖项**：北京市科技进步一等奖、人工智能学会科技创新一等奖、王选新闻科学技术进步一等奖等
 - **首席科学家唐杰**：ACM Fellow、IEEE Fellow，清华大学计算机系教授，获国家杰青、王选杰青奖
 - **学术成就**：发表论文300余篇，获ACM SIGKDD Test-of-Time Award
 - **参与项目**：悟道、AMiner
 - **首席顾问张钹**：中科院院士，1958年毕业于清华大学自动控制系，同年留校（计算机系）任教至今，已发表200多篇学术论文及5篇中英文版专著或章节

历史沿革

- 成立于2019年，由清华大学计算机系知识工程实验室的技术成果转化而来
- 22年8月，智谱AI与清华大学KEG共同研发的大规模中英文预训练语言模型GLM-130B正式发布
- **融资历程**
 - **战略投资（19/08）**：获数千万人民币，投资方为中科创星、清华控股、清华产业基金
 - **A轮（21/09）**：获数亿人民币，由达晨财智、华控基金、将门创投、南京图灵、北京达凡、通智投资、枣庄通智、荣品投资、财智创赢、凌云光等机构联合投资
 - **B轮（22/09）**：获数亿人民币，由君联资本和启明创投联合领投，华兴资本担任独家财务顾问

技术

- **底层模型及架构**：大规模中英文预训练语言模型GLM-130B及悟道2.0超大规模预训练模型
- **算力**：暂无公开信息
- **参数**：万亿
- **数据量**：暂无公开信息

商业化

- **产品形态**
 - **Aminer**：收录全球1+亿位学者、38万家机构发表的3.3亿篇论文、1.1亿篇专利、280万个科研项目，建立含40个学科800万个知识概念和11亿条关联信息图谱
 - **大规模中英文预训练语言模型GLM-130B**：已开源开放
 - **聊天机器人XDAI和chatGLM**
- **客户体系**：To B
 - **大模型**：曙光智算、QingCloud、凌云光、创泽智能机器人、达阔、北京广播电视台
 - **AMiner**：搜狗、有道、华为、清华大学、浙江大学、哈尔滨工业大学、CSDN
 - **智慧人才**：清华大学、NSFC、中国科学技术协会、中国计算机学会、中国电子学会、北京市科学技术委员会、深圳市科技创新委员会、CSIG、阿里巴巴
 - **知因分析**：中华人民共和国科学技术部、中国工程院、中国船舶工业集团公司、中国水利水电科学研究院、清华大学出版社

新兴挑战者：美团前大佬离职成立光年之外，产品未知

- 王慧文创立，目前公司处于招兵买马阶段
- 2023年2月14日，公开信息表示北京智源人工智能研究院副院长刘江已决定加入，李志飞亦有加入打算

团队

创始人王慧文

- 工作经历：曾担任美团网联合创始人、副总裁，前人人网联合创始人；从2004年创办的第一个项目开始，王慧文便跟随王兴，从校内、饭否、海内、到如今的美团，其为王兴创业以来最忠实的伙伴
- 其他信息：2023年2月13日，王慧文在朋友圈中发言“关于人工智能到现在为止，我只有一个判断，必须参与”、“5,000万美元带资入组，不在意岗位、薪资和title，求组队”，其对于此次人工智能创新项目十分看重；其发布“英雄帖”时，李志飞就坐在他的对面，两天之后，李志飞（有搜索、语义分析及工程相关经验）便表示自己一定要参与

历史沿革

- 公司是一家由美团联合创始人王慧文创办的人工智能公司，致力于打造中国版OpenAI
- 融资历程
 - 目前估值2亿美元，个人不占股，但其出资的5,000万美元占股25%，剩余的75%用于邀请顶级研发人才
 - 下轮融资已有顶级VC认购2.3亿美元，据传“顶级VC”至少包括源码资本和真格基金，但目前还没有任何人和王慧文的新公司签订正式投资文件，（第二轮）估值也还未确定

技术

- 底层模型及架构：暂无公开信息
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息

商业化

- 产品形态：暂无公开信息
- 应用场景：暂无公开信息
- 客户体系：暂无公开信息

新兴挑战者：AWS首席科学家李沐及导师成立Boson.ai，产品未定

▪ 目前仅有创始人背景信息，官网正建设中，具体产品形态未定

团队

领导团队

▪ 创始人Alex Smola：担任CEO；亚马逊云科技（AWS）“参数服务器之父”

– 基本信息：2013年1月至2017年10月，Alex Smola是美国卡耐基梅隆大学（CMU）的教授，在此期间与李沐携手创办了数据分析算法公司Marianas Labs，其担任CEO，李沐担任CTO，Marianas Labs曾获得几十万美元投资，最后成功将其卖给了一家上市公司；2016年7月，Alex Smola教授加入亚马逊做了两年零9个月的主管，随后于2019年3月在AWS担任副总裁级别的科学家（职级为Distinguished Scientist/VP）；2023年2月选择离职

▪ 创始人李沐

– 学历背景：2008年毕业于上海交通大学计算机系，大学期间曾在微软亚洲研究院担任实习生
– 工作经历：在参与创办Marianas Labs及加盟AWS之前，曾在百度担任过深度学习研究院（IDL）的主任研发架构师一职；2016年7月，李沐紧随其博士生导师先以兼职身份加盟AWS成为高级应用科学家，直接向其老师Alex Smola教授汇报；9个月后，选择留在AWS，并成为公司高级科学家，随后于2020年10月成为AWS资深高级科学家
– 学术成就：以第一作者的身份在计算机科学领域的大会以及期刊上发表了多篇论文，包括以下方向：跨度理论（FOCS）、机器学习（NIPS、ICML）、应用层面（CVPR、KDD）、操作系统（OSDI）

历史沿革

▪ 23年3月7日消息，亚马逊首席科学家李沐已离职创业，其导师另即一位前亚马逊AI大牛Alex Smola与其一同创业，创业内容与大模型应用有关，方向是面向所有人的大模型；目前Boson.ai的GitHub仓库页面由李沐编写

▪ 融资历程

– 宣传融资顺利，并未有实际融资信息公布

技术

- 底层模型及架构：暂无公开信息
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息

商业化

- 产品形态：暂无公开信息
- 应用场景：暂无公开信息
- 客户体系：暂无公开信息

新兴挑战者：前搜狗高层创立百方众智宣称入局，处于雏形阶段

▪ 原搜狗CEO王小川创立人工智能技术公司，据报道王小川将入局AI大模型创业，仍处于初期筹备阶段

团队

王小川

- **基本信息**：1978年出生于四川成都，1996年获得国际奥林匹克信息学竞赛金牌，进入清华大学计算机科学与技术专业，拥有工学学士、工学硕士，以及EMBA学位
- **任职经历**：进入搜狐后，在2004年主导推出搜狗搜索，并旋即在27岁那年成为彼时中国互联网界最年轻的高管；参与共建了清华大学天工智能计算研究院，并担任联席院长，他同时也是清华大学计算机学科顾问委员会委员。这都并非虚职 - 他推动搜狗和清华大学等学术机构合作，做了很多实际的尝试。比如当年包含1.3亿原始网页的SogouT开放数据库，就以免费形式对科研机构开放，希望从语料的质量上推动国内技术的发展，也成为很多NLP研究者和从业者最常用的数据库之一

历史沿革

- 2022年6月，王小川设立了人工智能技术公司北京百方众智信息技术合伙企业（有限合伙），占股80%，2023年2月20日报道，王小川个人表示将回归创业做AI大模型，表示在“快速筹备中”

融资历程

- 暂未融资

技术

- 暂未开启技术研发阶段

商业化

- 王小川正在与产业界和学术界的一些重要人士进行密集沟通，目前并没有未公开进程

新兴挑战者：循环智能助力盘古研发

- 预训练NLP大模型开发及智能客服SaaS服务商
- 主营搭建基于AI的销售策略生成与执行辅助系统，提高企业销售效率带来业绩增长，积累多家保险、银行、房产、汽车等行业头部客户
- 大模型开发方面与华为云联合开发“盘古”大模型，具备一定技术壁垒及研发实力

团队

领导团队

- **联合创始人杨植麟**：曾效力于Google Brain和Facebook AI
 - 31岁，本科毕业于清华大学计算机系，本科期间师从唐杰教授，博士毕业于卡内基梅隆大学计算机学院，师从苹果AI研究负责人Ruslan Salakhutdinov和谷歌首席科学家William Cohen
 - 2019年作为第一作者，与谷歌、卡内基梅隆大学联合发表深度学习和自然语言处理模型XLNet，2018年，作为第一作者，与图灵奖得主杨乐昆（Yann LeCun）合作发表论文“GLoMo: Unsupervised Learning of Transferable Relational Graphs”；与图灵奖得主约书亚·本吉奥（Yoshua Bengio）合作发表论文“HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering”
- **联合创始人&CTO张宇韬**：32岁，2019年加入清华大学计算机系，曾获获得吴文俊人工智能科技进步一等奖
- **联合创始人&COO揭发**：国内大数据创业公司GrowingIO早期核心团队成员，曾带领团队获得近亿元的年营业额
- **联合创始人&CEO陈麒麟**：卡内基梅隆大学计算机系硕士，曾在美国参与创业连续创业者
- **团队情况**：目前团队人数超过70人，其中80%为工程师

历史沿革

- 2016年成立，至2018年为公司初创期，重点是打磨核心能力，即研发初代语音和语义生产系统
- 2019年取得世界级技术突破，一方面联合谷歌发表Transformer-XL、XLNet模型引领NLP，另一方面公司升级行业增强型语音及语义生产系统
- 融资历程
 - 天使轮（2018）：华山长盈；PreA轮（2019）：金沙江创投、靖亚资本、华山长盈
 - A轮（2019）：真格基金领投、金沙江创投、靖亚资本、华山长盈，约千万美元
 - B轮（2020）：红杉中国、金沙江创投、万物资本，1,200万美元
 - 战略投资（2021）：博裕资本（领投）红杉资本中国（领投）、金沙江创投、靖亚资本、真格基金、万物资本、义柏资本（财务顾问），3,800万美元

技术

- **底层模型及架构**：公司主要业务是运用NLP、语音、多模态、大模型等人工智能技术打造“销售科技”方案，主要算法为transformer-XL、XLN儿童、PANGU、GPT3、P-tuning、Pathways、MoE、Bert
- **算力**：暂无具体数据，由鹏城实验室提供算力支持
- **参数**：1,100亿密集参数
- **数据量**：40TB，目前日均处理上亿次对话，累计生产了超过10,000个面向各行业业务流程的AI数据模型

商业化

- **产品形态**：“盘古”零样本NLP平台，会话洞察&客户洞察、实施辅助、对话采集、挖掘+建模、智能销售助手
- **应用场景**
 - 采用SaaS订阅制的收费模式，借助无需行为数据的成单预测方案，为企业提供转化率更高的精准销售策略，为业务人员提供实时辅助系统
 - 成熟品推向更多行业，拓展更多小领域，例如虚拟人带货，将广泛接入线上、线下等更广阔的销售场景
- **客户体系**：循环智能已经与两家头部股份制商业银行中国工商银行、招商银行达成了合作；在保险、教育、房产等领域签约了招商信诺、众安保险、泰康人寿、水滴、轻松筹、新东方在线、51Talk、万科、自如、我爱我家、上汽等行业标杆客户

新兴挑战者：聆心智能专注研发大模型陪伴AI机器人

- 由清华大学计算机科学与技术系副教授黄列民教授带队，团队拥有强大硬核的产品和算法背景，是一家具备自研对话大模型能力的陪伴式AI机器人提供商
- 成立一年内推出自研中文对话大模型OPD及情绪支持对话系统，研发速度较快，未来应用场景多为AI智能交互
- 现已和游戏厂商及车厂达成合作

团队

领导团队

- 黄民烈：清华大学计算机科学与技术系教授，对话系统领域的国际顶级学者，国家杰出青年基金项目获得者
 - 学历背景：本科就读于清华大学工程物理学，博士毕业于计算机科学与技术专业
 - 研究领域：人工智能、机器学习理论与方法，包括深度学习、强化学习等；自然语言处理技术与方法，如语言理解、语言生成、语言匹配与推理，具体应用包括自动问答、阅读理解、对话系统、情感分析等，在对话系统领域有杰出成就，并研发了ConvLab、EVA、OPD等对话框架和大模型
 - 论文成果：曾获得汉王青年创新奖、微软合作研究奖（Microsoft Collaborative Research Award）、IJCAI-ECAI 2018杰出论文奖、CCL 2018最佳系统展示奖、NLPCC 2015最佳论文奖，2016、2017年两次入选PaperWeekly评选的最值得读10/15篇NLP论文之一；已超过60篇CCF A/B类论文发表在ACL、IJCAI、AAAI、WWW、SIGIR、EMNLP、KDD、ACM TOIS等国际顶级或主流会议及期刊
 - 社会任职：清华大学计算机系智能技术与系统实验室副主任，中文信息学会自然语言生成与智能写作专委会副主任，CCF学术工作委员会委员秘书长；曾担任多个国际顶级会议的领域主席或高级程序委员，如AAAI 2019、IJCAI 2019、IJCAI 2018（杰出SPC）、IJCAI 2017、ACL 2016、EMNLP 2014/2011、IJCNLP 2017等，长期担任ACM TOIS、TKDE、TPAMI、CL等顶级期刊的审稿人
- 产品合伙人：产品和技术双背景，曾深度参与Google Assistant项目
- 团队背景：研发团队核心成员均来自于清华大学、卡内基梅隆大学、哥伦比亚大学等国内外顶尖高校及科研机构，预计吸纳具有移动社交APP经验的年轻产品经理

历史沿革

- 2021年成立，作为国内最早从事大模型研发的团队之一，聆心智能依托自身的核心技术，打造了超拟人大模型，将Model as a service这一商业逻辑融入到数字IP、游戏NPC拟人化、数字员工等垂直场景，完成小说创作、交互陪伴式学习助手、定制化教师、面试、绘本读物交互式阅读等功能
- 融资历程
 - PreA轮（2023）：无限基金SEE Fund
 - 天使+轮（2022）：千万元，连星尚智私募基金（领投）、图灵创投、北京智谱华章科技有限公司
 - 天使轮（2021）：清控资管集团

技术

- 底层模型及架构：生成式大模型及中文对话大模型OPD（OPD采用UniLM8架构）
- 算力：暂无公开信息
- 参数：数百亿
- 数据量：70GB，数据均来自公开可爬取、可访问数据源

商业化

- 产品形态
 - AI情绪疗愈机器人Emohaa：是目前唯一全AI模式的情绪支持对话系统，主针对18-45岁年龄区间人群
 - AI乌托邦：PC端、小程序，雏形阶段用户可选择与梅西、马斯克、于谦、爱因斯坦、甄嬛等跨次元跨时空名人聊天，也可与程序员、心理咨询师、算命先生等群体属性的角色交流，还可通过文字描述“捏出”自己想要的AI朋友
- 客户体系：已于一些游戏厂商和车厂落地深度合作
 - 与高端豪华电车品牌Beyonca合作，打造了新一代智能驾舱的贴心助手基于聆心智能的对话生成技术为品牌定制开发AI角色，通过合作粉丝基数庞大的知名IP，也能对聆心智能的AI互动平台起到引流作用

新兴挑战者：MiniMax推出三模态AI大模型

- 前商汤副总裁创业项目，宣传首家三模态AI大模型，**自研模型已1+年**，推出TO C虚拟人产品GLOW，现注册用户已近百万量级，TO B产品计划23年对外开放
- 技术参数暂不明确

团队

领导团队

▪ 联合创始人- 闫俊杰

- **基本信息**：2015年博士毕业于中科院自动化所，其在深度学习和计算机视觉领域发表顶级会议及期刊论文100余篇，Google Scholar引用超过10,000次
- **工作经历**：前商汤科技副总裁、通用智能技术负责人，曾担任商汤研究院副院长，期间负责搭建深度学习的工具链和底层算法，以及通用智能的技术发展，并搭建了商汤的人脸识别和智慧城市相关的技术体系

▪ 技术合伙人- 杨斌

- **基本信息**：中科院自动化所硕士，2016年加拿大多伦多大学攻读博士学位
- **工作经历**：海外留学期间，先后作为创始团队成员在Uber ATG研究院、自动驾驶初创公司Waabi等团队供职，拥有多年研发经验，对基于数据驱动的端到端系统问题颇有研究

▪ 核心成员- 周或聪：商汤科技早期员工之一，曾经在商汤科技研究院带领算法团队

▪ 团队背景：团队人数已过百；核心技术研发成员均来自全球知名高校和全球顶尖科技公司，拥有世界顶尖自然语言处理、语音、计算机视觉、计算机图形学等工业界和学术界经验，拥有多项全球领先的人工智能领域研究成果，具有上百个全球发明专利，1/3的团队成员拥有世界顶尖技术实验室的博士学位

历史沿革

- 2021年11月公司成立，以大模型作为主要的研发方向，形成了文本到视觉（text to visual）、文本到语音（text to audio）、文本到文本（text to text）三大模态的基础模型架构
- 2021年11月，首款基于大模型技术的自研产品Glow上线，主打情感陪伴类AI社交，具有自由聊天及剧情任务多种模式，目前注册用户已近数百万量级
- 融资历程
 - 天使轮融资（21/12）：云启资本独家
 - Pre-A轮融资（22/7）：明势资本领投

技术

- **底层模型及架构**：在Transformer基础上自研；通过内容生成与用户反馈这样一种“user-in-the-loop”的方式，来不断迭代基础大模型的AI能力
- **算力**：暂无公开信息
- **参数**：暂无公开信息
- **数据量**：暂无公开信息

商业化

- **产品形态及应用场景**：**首款大模型技术产品Glow的基本玩法为在应用中创建“智能体”，即虚拟AI机器人**，主打情感陪伴类AI社交，有自由聊天和剧情任务多种模式
- **客户体系**
 - To C方向，其首款AI虚拟聊天社交软件Glow推出四个月已经积累数百万用户
 - To B方向，MiniMax计划在23年对外开放API

新兴挑战者：启元世界宣布构建千亿级参数预训练大模型

- 自研模型，构建出千亿级参数规模的预训练大模型，专注游戏品类，提供AI产品及平台服务，对外宣称22年达千万级营收
- 阿里团队背景

团队

- 领导团队
- 创始人CEO袁泉：阿里算法技术的P10科学家
 - 基本信息：西安交通大学计算机科学与技术专业
 - 工作经历：早年于IBM研究院从事与机器学习个性化推荐相关算法研究；前阿里认知计算实验室资深总监、淘宝推荐算法团队创始人，2012年加入阿里任职5年，期间曾从0-1搭建淘宝天猫推荐团队并实现落地
 - 联创CTO龙海涛
 - 工作经历：前阿里搜索广告架构负责人，10年前曾是IBM研究院最年轻科学家，对互联网广告、电信、动力等多个行业的零碎架构颇有建树；在阿里期间从事搜索广告业务的架构设计，主导了离线系统、在线引擎和搜索内核的升级换代，其近年来完成阿里妈妈搜索广告的重构
 - 团队背景：核心团队来自阿里、腾讯、字节、Netflix等高科技公司，聚集了香港科大、北大等众多名牌大学的博士和硕士，并拥有伯克利、CMU、UCL等知名机构的特聘参谋；**员工规模约200人**

历史沿革

- 2017年公司成立，是一家专注于人工智能领域的科技公司，拥有深度学习、强化学习、云计算等自主研发的核心技术和产品，同时公司屡次获得世界级AI比赛的冠军
- 2020年6月，启元世界自主研发的AI智能体在《星际争霸II》人机大战中以2:0击败了中国星际职业冠军级选手，刷新了人机对抗的历史最高成绩，和DeepMind一同成为了全球唯二的两家能够用AI击败星际顶级选手的通用人工智能公司
- 融资历程
 - 天使轮融资（18/04）：莉莉丝、高榕资本
 - A轮融资（21/07）：嘉御资本、高榕资本、GGV纪源资本、莉莉丝游戏，3亿元人民币

技术

- 底层模型及架构
 - 大规模深度强化学习：经过5年技术研发，积累一整套深度强化学习算法框架及引擎系统，能够高效调度起数十几万CPU核、数百张GPU卡进行大规模学习及演化训练，解决10的26次方复杂动作空间决策问题
 - 认知决策大模型：**构建千亿级参数规模**的预训练大模型，重点支持语言和行为的多模态多任务功能

商业化

- 产品形态
 - AI玩家/AI角色/AI决策：针对各类游戏、社交元宇宙、数字人等
 - AI设计师：数值设计、内容生成
 - AI Beings：为各类游戏、数字场景、仿真环境提供开发、训练、评估工具
- 应用场景：目前主要集中在游戏、虚拟世界及元宇宙
- 客户体系：2023年2月28日，智源研究院在北京发布FlagOpen（飞智）大模型技术开源体系，并宣布与启元世界结为开源体系生态合作伙伴，共建“认知决策大模型开放实验室”，客户项目包括《三国志战略版》、《三国志战棋版》、神武回合制游戏、《斗地主》、《Free City》、《汉末霸业》、《Unciv》

新兴挑战者：元语智能推出定位ChatGPT中文版的ChatYuan

- 初创型功能型对话大模型开发商，目前推出基于自研PromptCLUE模型的ChatYuan对话产品，上线小程序及网络版，**目前产品并未有合作B端客户或落地项目**
- 成立仅1年初创公司，**目前团队及产品仍处于雏形阶段**

团队

- 领导团队
- **元语智能创始人、CEO徐亮**：CLUE的发起人，此前在AI+RPA厂商实在智能担任算法合伙人
 - **张轩玮为联合创始人、CTO**：也是CLUE联合发起人及副会长，毕业于北京大学，此前在爱奇艺从事NLP大模型和机器翻译等相关工作
 - **朱雷为联合创始人，COO**：曾在平安和阿里达摩院从事NLP产品设计相关工作，在AR独角兽Rokid负责算法应用产品相关工作
 - **团队规模**：现在在扩大规模的过程中，**今年预计会达到30人的规模**，之前是小团队在加速迭代的过程

历史沿革

- 2022年成立（西湖心辰（杭州）科技有限公司持股10%，徐亮持股50%）
- 2022年10月1日发布PromptCLUE1.0模型& API – 首个中文支持的零样本开源模型；目前全平台下载量4.4W+，API调用量10w+/天
- 2022年11月发布PromptCLUE1.5模型&API – 支持更多任务的零样本开源模型
- 2022年12月20日发布ChatYuan – 国内首个基于百亿参数大模型的功能对话产品，支持小程序和网页
- 2023年1月12日发布ChatYuan大模型 – 首个功能对话开源模型供免费下载，全平台已有数万下载量
- 2023年2月3日发布ChatYuan API – 可供API调用的功能对话模型
- 融资历程
 - 天使轮已经推进到非常靠后的阶段，下一轮已在筹划中

技术

- **底层模型及架构**：基于T5架构自研PromptCLUE开源模型，利用大量多轮对话数据、结构微调、参数量升级，结合人类反馈和强化学习训练出大模型ChatYuan能够支持理解类、抽取类、生成类等所有类型的中文任务，并在此基础上构建了ClueAI的API服务
- **算力**：暂无公开信息
- **参数**：百亿
- **数据量**：600-800GB

商业化

- **产品形态**：小程序及网页版ChatYuan对话产品，但是目前已经无法打开，显示小程序系统更新维护中
- **应用场景**：现小程序仍为to C，未来计划暂无公开信息
- **客户体系**：暂无公开信息

新兴挑战者：原子回声研发表格预训练大模型

- 主营产品为数据分析基础设施，自研表格预训练大模型“万卷”，产品面向数据分析师，现阶段产品、技术指标及商业化进展的公开信息较少
- 团队主要为清华系，年轻及资历较浅

团队

- 张俊祺：原子回声创始人兼首席执行官
 - 学历背景：清华大学计算机科学与技术系2013级本科，计算机科学与技术系2017级博士
 - 任职背景：曾于新加坡国立大学做访问学者，并就职于华为诺亚方舟实验室和摩根大通投资银行团队
 - 所获奖项：荣获北京市优秀毕业生、国家奖学金、ACM CIKM最佳论文奖，为大会三十多年历史上第三次华人获该奖项
 - 期刊论文：发表国际顶级会议和期刊论文六篇
- 团队背景：团队成员均来自清华大学等知名高校和知名企业

历史沿革

- 成立于22年3月，原子回声致力于打造全球领先的数据分析基础设施，构建下一代全流程AI数字分析师
- 融资历程
 - 种子轮（22/09）：奇绩创坛，数百万人民币

技术

- 底层模型及架构：暂无公开信息
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息

商业化

- 产品形态：公司自研表格预训练大模型“万卷”，结合少样本学习技术、自然语言检索技术与可解释机器学习技术，使得产品在不同领域及不同行业达到高度通用的效果
- 应用场景：产品面向数据分析师，无需机器学习背景，开箱即用，帮助企业发现更多数据价值，让一切决策从表格开始
- 客户体系：暂无公开信息

新兴挑战者：深言科技推出新一代智能文本信息处理引擎

- 大规模预训练模型技术开发商，孵化自清华大学自然语言处理实验室（THUNLP）和北京智源人工智能研究院（BAAI）
- 目前推出WantWords和WantQuotes两款搜索引擎产品，已有商业化订单

团队

领导团队

- 创始人/ CEO 岂凡超
 - 基本信息：清华大学计算机系博士，电子信息科学与技术、工商管理双学士，已在国际顶级刊物发表论文30余篇，申请发明专利十余项，曾获国家奖学金、北京市优秀毕业生、全国研究生人工智能创新大赛总冠军等荣誉，拥有顶级风投和硅谷创业公司实习经历
- 联合创始人/ COO 李满翔
 - 基本信息：清华大学电子系博士，电子信息科学与技术、工商管理双学士，曾获北京市优秀共青团员、清华大学一二·九辅导员、校辩论赛冠军，拥有2家技术创业公司实习经历
- 首席科学家孙茂松：创始人/ CEO 岂凡超导师、欧洲科学院院士孙茂松教授
- 团队背景：公司孵化自清华大学人工智能研究院、清华大学自然语言处理实验室、北京智源人工智能研究院，核心成员来自大模型国家队-智源悟道大模型研发团队，此前研发的产品已有数十万用户

历史沿革

- 2019年，清华大学自然语言处理实验室（THUNLP）研制出一款名为WantWords的反向词典，能够根据语义描述查找词语，可用于写作辅助等多个场景
- 2021年11月，此款应用在微博爆红，成千上万用户涌入WantWords，多次挤爆了服务器
- 2022年3月公司成立，旨在打造基于大规模预训练模型的新一代智能文本信息处理引擎，涵盖AIGC文本生成、信息抽取聚合、语义检索等功能，为数亿脑力劳动者和数千万信息密集型组织重塑信息处理全流程
- 2022年5月，基于语义的名言名句检索系统WantQuotes上线，短短两三周的时间已获得近30万用户；截至目前，WantWords和WantQuotes已有近200万用户
- 融资历程
 - 天使轮融资（22/08）：英诺天使基金，奇绩创坛

技术

- 底层模型及架构：基于大规模预训练模型的新一代智能文本信息处理引擎，涵盖AIGC文本生成、信息抽取聚合、语义检索等功能
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息

商业化

- 产品形态：目前公司产品包括可以根据意思搜索词语的“WantWords反向词典”，以及根据意思查询句子的“WantQuotes据意查句”通过文本摘要、语义检索、信息抽取等技术在信息爆炸时代提高信息获取的效率，同时通过可控文本生成、素材自动推荐、文本校对和风格检查实现文本内容创作的提质增效
- 客户体系：正式运营3个多月，产品用户数已达数百万，部分产品还受到央视等媒体报道；同时已达成数百万元订单，受到多个行业的标杆客户的认可

新兴挑战者：秘塔科技推出中文版ChatGPT 对话写作猫

▪ 自研LLM大模型MetaLLM，专注于法律领域翻译及AI辅助写作，目前产品商业化进程可控

团队

领导团队

▪ CEO 闵可锐

- 工作经历：前玻森数据CTO/ 联合创始人，猎豹移动首席科学家；其曾在微软亚洲研究院IM及VC组担任实习研究员，参与微软搜索引擎、稳健主成分分析方法的理论与应用研究；在美国谷歌参与AdSense基于内容广告建模组点击率预测，改进原有预测模型；其还曾担任上海科学技术委员会互联网话题识别与跟踪系统设计及实现项目的算法设计师，以及ACM国际大学生程序设计竞赛科学委员会委员，负责解决技术难题首席科学家孙茂松

▪ COO 王益为

- 基本信息：本科毕业于北京大学，后于伦敦大学（UCL）及哥伦比亚大学法学院获法学硕士

- 工作经历：前方达律师事务所资深律师，未来法律联合创始人，曾代表阿里巴巴/ 搜房网/ 中国英利/ 蓝港互动等公司，为他们的IPO提供法律服务

历史沿革

▪ 2018年4月公司成立，致力于运用人工智能技术为用户打造得心应手的生产工具，大幅提升工作效率，让用户享受科技的便利

▪ 2019年，公司推出了首款产品“秘塔翻译”，这一专注于法律、公文类文书的中英文AI翻译软件，目前已在中国各大顶尖律师事务所、仲裁中心、数百家企业和其他法律相关机构中广泛使用

▪ 2023年2月27日，国内秘塔科技自研的LLM大模型产品-对话写作猫，3月7日宣布推出基于自研大模型的中文版ChatGPT「对话写作猫」

融资历程

- Pre-A轮融资（2021）：源渡创投、丰元资本、明势资本

- 天使轮融资（2018）：明势资本、思满科技、小豹AI音箱、个人投资者、猎豹移动

技术

▪ 底层模型及架构：秘塔团队在Transfomer的基本框架上进行了多项自主模型创新，[自研LLM大模型MetaLLM](#)，并非调用OpenAI API，从零开始处理数据/ 训练，并非借助开源模型，训练花费不足1亿美金

▪ 算力：暂无公开信息

▪ 参数：百亿

▪ 数据量：暂无公开信息

商业化

▪ 产品形态

- 秘塔写作猫：AI写作辅助工具

- 秘塔翻译：专为法律人训练的机器翻译系统

▪ 应用场景：错别字校对、多语种法律文件翻译、OCR识别、智能合同、写文案、写方案、写广告语、写辞职信、文生图等各类任务

▪ 客户体系：“秘塔翻译”这一专注于法律、公文类文书的中英文AI翻译软件，目前已在中国各大顶尖律师事务所、仲裁中心、数百家企业和其他法律相关机构中广泛使用，[秘塔写作猫目前已经上线](#)

新兴挑战者：心识宇宙开发脑启发AI大模型

▪ AI模型公司：宣传拥有基于大模型为底座的脑启发AI系统，首个AI角色生产引擎，未公布具体技术参数，客户目前主要为虚拟人企业，感觉大模型技术能力暂未体现

团队

- 陶芳波：创始人，数据挖掘和人工智能技术专家
 - 工作经历：先后在美国微软研究院，Facebook Research，NASA，阿里达摩院神经符号实验等从事研发工作，其研发成果应用到NASA，波音公司等美国核心企业和部门；商业引用层面曾搭建Facebook大规模内容理解平台，日均服务超过2亿用户；在阿里期间带领打造市场第一款商用光子认知系统
 - 社会头衔：还担任EMNLP、CIKM、ACL、SIGMOD、WWW等领域顶会的评审专家
- 团队背景：目前有来自于清华、北大、港大、伊利诺伊大学等校的具有十余年经验的团队，并且有Facebook、Google、阿里巴巴、字节跳动、微软研究院等世界科技企业和研究机构的高管背景

历史沿革

- 2022年1月在杭州成立，是一家专注于通用人工智能技术的创业公司；公司自主研发以大模型为底座的脑启发AI模型（Unified Mind Model）系统，通过模拟大脑宏观架构，提供有思考、情感、记忆及服务的AI能力
- 2022年11月，团队发布MindOS内测版本，开放给少数B端客户
- 融资历程
 - 天使+轮（22/11）：银杏谷资本领投，璞跃中国（Plug and Play China）、天使轮投资方线性资本超额跟投、红杉中国种子基金跟投

技术

- 底层模型及架构：大模型为底座的脑启发AI模型（Unified Mind Model）系统
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息

商业化

- 产品形态：MindOS，一个AI角色生成引擎，通过简单的填写配置、拖拽上传，就能完成一个具备专业知识、记忆和人格的AI角色
- 应用场景：目标是赋能品牌、应用开发者和个人打造属于自己的智能虚拟人，将与3月发布升级版,新版将会有更强的对话可控性和复杂场景解析能力等
- 客户体系：2023年1月，MindOS就获得了百万元订单；拥有VBS拟人的次世文化、拥有顶尖虚拟人IP柳夜熙的创壹科技都与其达成深度合作，还有数家媒体公司也成为心识宇宙的战略渠道，多家海外企业纷纷前往寻求商业合作

新兴挑战者：西湖心辰已实现类ChatGPT产品开发

- 发展迅速的大规模预训练模型开发初创公司，创始人背景较佳，产品布局全面，推出聊天机器人、AI画图等产品
- 目前推出类ChatGPT文本生成产品“心辰Chat”，To C 及 To B端各产品商业化进程顺利

团队

- **蓝振忠：NLP 自然语言处理领域预训练语言模型「ALBERT」第一作者**
 - **基本信息：**西湖大学助理教授、博士生导师，本科毕业于中山大学，后到卡内基梅隆大学读博，研究方向为计算机视觉与多媒体分析，长期致力于研究自然语言处理，计算机视觉及深度学习结合与应用
 - **学术成就：**曾在Google主导研发了大模型ALBERT（学术引用4,300+），在大模型领域颇有影响力，与AI风云人物们建立了紧密合作，如Transformer作者、Character.ai创始人Noam Shazeer
 - **能力积累：**此前在谷歌AI研究所工作，研发成果被应用于谷歌新闻、谷歌助手等多个拥有亿级以上用户的产品；2020 年受聘于西湖大学，创办深度学习实验室并担任博士生导师，而后迅速带领团队研发出了AI心理咨询师「小天」、智能写作平台 FRIDAY 以及 AI 绘画产品「盗梦师」
- **张岳博士：学术带头人**
 - **基本信息：**2003 年毕业于清华大学计算机专业，获得学士学位；2006 年毕业于牛津大学计算机专业，获得硕士学位；2009 年毕业于牛津大学计算机专业，获得博士学位；2010 年3月-2012 年6月在剑桥大学计算机专业从事博士后研究，2012 年7月-2018 年8月在新加坡科技与设计大学担任助理教授；2018 年9月全职加入西湖大学，担任终身副教授
 - **学术成就：**在2012 年-2020 年期间，张岳博士在NLP领域的顶会发表的论文数量是全球第三，仅次于周明和Yoshua Bengio，长期开展自然语言处理方面的研究，在ACL、EMNLP等自然语言处理顶级会议发表过多篇高水平论文
- **团队组建：**初步组建了一支高人才密度的团队，覆盖了大模型的架构、训练、推理等关键环节，吸纳了来自Google、Meta、Amazon等公司的高管和科研专家；**目前正在招募CEO**

历史沿革

- 蓝振忠2020年6月回国、加入西湖大学后，又继续基于大模型开发了辅助心理咨询的聊天机器人“聊会小天”，并创立了西湖心辰
- 2022年8月快速推出基于text2image模型的AI画图产品“造梦日记”，目前积累了上百万C端用户，数十家B端用户
- 2023年1月推出“Hey Friday”
- 2023年2月推出类ChatGPT文本生成产品“心辰Chat”
- **融资历程**
 - 天使轮（2021）：凯泰资本，西湖科创直投基金，蓝驰创投

技术

- **底层模型及架构：**RIO模型，是国内最早商业化落地的经过Instructional Learning的大模型
- **算力：**若旨在追赶AGI与ChatGPT，**现有算力资源不够支撑，需要更多资源**
- **参数：**暂无公开信息
- **数据量：**暂无公开信息

商业化

- **产品形态**
 - 辅助心理咨询的聊天机器人“聊会小天”
 - AI画图产品“造梦日记”
 - 自研文本生成产品“Hey Friday”
 - 类ChatGPT“心辰Chat”
- **应用场景：**AI绘画、AI写作、AI心理咨询等平台级人工智能应用
- **客户体系：**西湖心辰的产品矩阵包含小天、造梦日记、Hey Friday，合计用户量超过140万C端用户、数十家B端用户

新兴挑战者： AI 图像生成技术服务商Tiamat推出MorpherVLM

▪ AI 图像生成技术服务商自研大模型深耕细分赛道，目前商业化已起步

团队

- 团队背景：创始人Eric，团队主要孵化自中科院上海科技大学，团队核心成员来自于中科院上海科技大学信息学院以及伦敦艺术大学，在AI图片引擎、数字孪生等方面有着深厚的技术积淀

历史沿革

- 成立于2021年，作为国内第一批入局AI生成领域的技术团队，致力于通过自研的算法、底层引擎、平台为个人用户和企业客户提供创意生产、广告、数字藏品等场景形式下的服务
- 22年上半年，在小红书、抖音等社交媒体上开始发布中文Text-to-image的产品，进行内部测试，单个用户在一个月的时间内就可以生成三千到一万张图像；内测高活跃用户数千人、月度新增数万人，社区用户涵盖国内设计、艺术创作、多模态技术等头部博主，有超过三分之一的用户有持续付费的意愿，C端商业化能力可观

融资历程

- 天使轮（22）：DCM独家领投，数百万美元
- A轮（23/02）：老股东DCM和绿洲资本联合领投，近千万美元

技术

- 底层模型及架构：自研的MorpherVLM通过异构的视觉编码-解码网络结构，并引入基于用户反馈的强化学习（RLHF）和细粒度的提示-隐变量对齐技术，提高了模型对图像多尺度信息的建模能力
- 算力：暂无公开信息
- 参数：Tiamat 文本到图像模型，约数十亿参数；创始人表示不停地把模型的参数量级进行扩张并不是最重要的指标
- 数据量：暂无公开信息
- 其他：新模型每天支持生成10-14万张图片，单位生成速度控制在几秒到两分钟之内

商业化

- 产品形态：现以TO C为主，20%的用户群体是小B端的设计师，60%是泛设计行业人群
- 应用场景：AI绘画生成
- 客户体系：「TIAMAT」已经与知名科幻IP、央视网、《时尚芭莎》、《嘉人》等一线时尚大刊建立长期合作关系，并与众多头部游戏公司，特别是持有故宫文化相关IP的非遗文化IP机构知造局等开展了战略合作；22年Q4，有几百万元合同订单

垂类玩家：音视频上市企业联汇科技布局多模态预训练大模型

- 上市公司，主营媒体行业数字音视频技术应用，现由赵天成博士带队研发，自研发布OmModel多模态预训练大模型&OmVision视觉认知服务平台
- 已发布大模型及开放平台产品，[大模型实现百亿参数规模](#)

团队

- **首席科学家赵天成**
 - **基本信息：**博士毕业于卡耐基梅隆大学计算机学院，主要从事多模态机器学习、自然语言处理等领域研究；曾主持包括国家科技部重点专项等重大科研项目，并且主持研发多模态预训练大模型OmModel与业界首个基于预训练大模型的低代码人工智能算法开发平台OmVision Studio
 - **学术成就：**在顶级人工智能会议与期刊发表论文40余篇，获得最佳论文奖2次，获得20余项美国和中国发明专利，曾担任国际顶级会议ACL，EMNLP、AAAI与核心期刊的审稿人和领域主席
 - **业务能力：**基于在卡内基梅隆大学的长期研究，赵天成博士带领团队创建了拥有自主知识产权的视觉语言大规模预训练模型OmModel，目前联汇科技的OmModel已经完成了基于超过数十亿字符、近十亿图片和视频数据的大规模预训练，同时在包括目标检测、行为识别、跨模态检索等多项重要任务上展现出超强的零样本泛化能力和小样本学习能力

历史沿革

- 1994年公司成立，是一家致力于数字音视频技术的研究和应用的高科技企业，专注于广播行业的数字技术的研究和应用，为国内外广播电台机构提供专业的一站式整体解决方案
- 2019年，开始致力于大模型研究
- 2021年，正式发布OmModel1.0
- 2022年，OmModel2.0通过工信部信通院正式检测
- 2023年，OmModel3.0即将发布

技术

- **底层模型及架构：**OmModel多模态预训练大模型、Om-PLM文本语义理解AI算法、Om-DOF视觉异常检测技术、Om-VQL视觉语言AI算法
- **算力：**暂无公开信息
- **参数：**百亿级
- **数据量：**暂无公开信息
- **其他信息：**已经完成在上亿图片、数万小时视频和50亿字符上的大规模多模态数据

商业化

- **产品形态：**Om-PLM文本语义理解AI算法/ Om-DOF视觉异常检测技术/ Om-VQL视觉语言AI算法
 - OmModel多模态预训练大模型，具有20,000多种目标、300多种属性、200多种行为的强大通识能力，只要对识别目标灵活搭配组合，就可以生成千万种不同功能的应用
 - 模型识别精准度提高40%，训练数据成本降低90%，开发周期由平均3个月缩短为10天，[可有效覆盖大量长尾应用场景](#)，实现亿级多模态数据秒级匹配，匹配速度在GPU环境提高5倍以上、在CPU环境提高300倍以上，推理速度提高5倍以上，大幅度降低部署成本
- **应用场景：**OmModel拥有全场景视觉认知与语义人机交互两大特点，发布OmVision视觉认知服务平台，打造具有行业领先性的人工智能算法服务中心，提供“零代码、零样本、零基础”的AI算法生产SaaS服务
- **客户体系：**暂无公开信息

垂类玩家：毫末智行推出汽车级ChatGPT

- 汽车级ChatGPT玩家
- 国内最早一家进行汽车及人工智能技术研究公司，且现阶段已合作20余款车，积累较多驾驶数据，有利于模型训练，号称为全球首个自动驾驶认知大模型

团队

领导团队

- 张凯董事长
 - 基本信息：长城汽车车辆安全工程研究院院长，长城汽车技术副总工程师，长城智能驾驶系统开发部部长，飞的科技有限公司经理执行董事，长城研究院副总工程师，获得中国汽车工业科技进步奖一等奖2次（2012年、2015年），《自动化学报》发表论文2篇
- 顾维灏首席执行官CEO
 - 基本信息：毕业于北京交通大学计算机应用技术专业，于2003年加入百度，历任百度MP3和视频搜索主架构师、百度语音识别技术负责人、百度地图副总经理、百度智能汽车事业部总经理等职位，2021年加入毫末智行
- 甄龙豹首席交付官CIO、侯军首席运营官COO
- 团队情况：目前拥有北京、保定、上海三城研发中心，团队规模超500人

历史沿革

- 从2019年成立，其前身是长城汽车智能驾驶前瞻部，融合来自科技公司自动驾驶人才，由长城汽车孵化而出；截止2021年底，营收将超过1个亿
- 2023年2月17日，毫末智行宣布将自动驾驶认知大模型（人驾自监督认知大模型）正式升级为“DriveGPT”，在自动驾驶模型算法方面，全面运用ChatGPT的模型和技术逻辑
- 融资历程
 - Pre-A轮（2021）：首钢基金领投，美团、高瓴创投等资本跟投，数亿元人民币
 - A轮（2021）：美团、高瓴创投、高通创投、首程控股、九智资本等，近10亿元
 - A+轮（2022）：中银集团投资有限公司领投，首程控股跟投

技术

- 底层模型及架构：Transformer大模型，借鉴了ChatGPT的实现思路，采用人类反馈强化学习（RLHF）技术，还拥有视频自监督大模型、3D重建大模型、多模态互监督大模型、动态环境大模型
- 算力：MANA OASIS算力可达0.67 EFLOPS（每秒6.7x1,017次浮点运算），和字节跳动旗下火山引擎的合作
- 参数：参数规模可对标GPT-2的水平，大约7.74亿
- 数据量：引入真实人驾接管数据

商业化

- 产品形态：车辆自动驾驶系统搭载
- 应用场景：车载语音和自动驾驶领域，未来计划用在汽车研发
- 客户体系：截止2022年底，毫末HPilot搭载车型接近20款，用户辅助驾驶行驶里程突破2,500万公里，2023年毫末城市NOH预计会在国内100个城市落地

垂类玩家：云从科技宣布布局预训练大模型

- 上市公司，以人脸识别技术为基础打造智慧化生态圈，提供人工智能服务
- 提前布局预训练大模型研究及实践，模型细节信息较少

团队

- 领导团队**
- **创始人周曦先生**
 - 获中国科学技术大学学士学位和硕士学位、美国伊利诺伊大学（UIUC）博士学位，入选中科院“百人计划”，研究员，博士生导师
 - 曾任中国科学院重庆绿色智能技术研究院信息所副所长、智能多媒体技术研究中心主任，云从科技总裁，2013年入选重庆市青年拔尖人才
 - 2015年，周曦率领团队成立专攻人工智能领域的云从科技，该公司主要研发人脸识别技术
 - **研发总监何洪路先生**
 - 何洪路曾先后担任NVIDIA技术带头人及软件部高级经理和Facebook信息和用户画像（Signals+Identity）团队技术带头人等关键职位
 - **刘琼博士**
 - 曾担任ACM Multimedia（A+类会议）2017的大会主席，IEEE ISM2014的程序主席和其他多个会议的领域主席，他还发起了非盈利的旧金山湾区多媒体论坛，论坛在14次活动中吸引了1,600多名与会者
 - 曾担任ACMSIGMM最佳博士论文委员会（2018、2019）委员，美国国家自然科学基金多媒体提案评审委员会委员，加州大学尔湾分校及加州大学圣塔芭芭拉分校博士论文指导委员会委员等重要角色

历史沿革

- 成立于2015年，初创于中国科学院重庆绿色智能技术研究院，是一家专注计算机视觉与人工智能的高科技企业，技术运用于智慧金融、智慧治理、智慧出行、智慧商业等重点行业市场
- 22年5月上市股票代码为688327，作为第一家在科创板上市的AI平台企业
- 近期也表示已提前布局“预训练大模型+下游任务迁移”的技术趋势，陆续在NLP、OCR、机器视觉、语音等多个领域开展预训练大模型的实践并持续加大投入

技术

- **底层模型及架构**：NLP、OCR、机器视觉、语音等预训练大模型人机协同操作系统
- **算力**：暂无公开信息
- **参数**：暂无公开信息
- **数据量**：暂无公开信息

商业化

- **产品形态**：陆续在NLP、机器视觉、语音等多个领域开展预训练大模型的实践，已经在城市治理、金融、智能制造等行业应用中体现价值布局人机协同领域，打造了像人一样思考和工作的人机协同操作系统
- **应用场景**：主要产品已广泛应用在物联网、移动互联网、银行、安防、交通等场景
- **客户体系**：云从科技与中国联通大数据有限公司、中国民航管理干部学院、金山云、浪潮集团、中国邮政集团公司软件开发中心分别建立战略合作关系；云从与包括重庆大学、重庆邮电大学、江西财经大学、英国华威大学、华南理工大学、中国人民公安大学、西南财经大学、上海交通大学等高校建立联合实验室研究人工智能基础技术，并在美国硅谷、UIUC建立设有研发中心

垂类玩家：拓尔思基于大模型的AIGC关键技术实现问答式AI

- 上市公司，Saas服务商，提前布局专业领域大模型研发，提供人工智能和大数据技术领域的通用产品、行业产品和SaaS化的软件和数据智能服务
- 于23年2月公告将突破基于大模型的AIGC关键技术，在问答式AI等领域实现成功应用

团队

- 总经理施水才先生
 - 1989年毕业于西安电子科技大学计算机应用专业，获硕士学位，2017年获清华大学五道口金融学院EMBA学位；1998年晋升教授
 - 曾获国家科技进步二等奖，上海市科技进步一等奖，2012年获全国优秀科技工作者称号
 - 1989年毕业分配至北京信息工程学院，在学院科研机构中文信息处理研究中心从事科研工作
 - 1993年拓尔思前身易宝北信成立时，加入易宝北信公司，历任天津红日药业股份有限公司工程部经理、总经理、董事长兼总经理
 - 现任天津红日药业股份有限公司副董事长兼总经理职务，兼任北京信科互动科技发展有限公司董事、北京拓尔思信息系统有限公司执行董事兼总经理、北京金信网银金融信息服务有限公司董事长等
 - 任北京信息科技大学和南京大学兼职教授，中国中文信息学会副理事长，中国软件行业协会副理事长，中国语音产业联盟副理事长、中国非结构化数据管理系统标准工作组副组长，中国计算机学会和中国电子学会大数据专家委员会委员，中关村大数据产业联盟副理事长，科技部863计划“中文信息处理”重点项目专家组成员，中国计算机学会大数据专家委员会委员等

历史沿革

- 成立于1993年，是中文全文检索技术的始创者，领先的大数据、人工智能和数据安全产品及服务提供商
- 2011年在深圳证券交易所创业板上市，公司净资产近30亿元，拥有员工2,000余人，在全国设有4大区域中心31家分公司
- 有40+发明专利、1,000+软件著作权，成功服务一批国家级重要信息化项目建设，成为国家企业信用信息公示系统、信用中国、国家知识产权检索系统、GA云搜索等国家级基础设施的数据管理和检索引擎
- 公司自主研发的TRS大数据平台、TRS人工智能平台和安全一体化平台，已成为行业领先的数字化基座，同时，公司已建立了大规模的数据和知识资产，正在全面构建云和数据服务生态（SaaS&DaaS）
- 公司的数智产品与服务已被国内外10,000余家企业级用户广泛使用，尤其在企业数字化转型、数字政府、网络内容安全、融媒体、金融科技等领域实现了深度应用

技术

- 底层模型及架构：以“专业大模型+领域知识数据”为核，以NLP技术突破来推动更多AIGC商业落地；公司也将以预训练大模型、In-Context Learning等技术为基础，将公司自主研发的文本生成、交互式生成、跨模态生成、大模型与外部知识库的融合等功能作为研发重点
- 算力：暂无公开信息
- 参数：暂无公开信息
- 数据量：暂无公开信息

商业化

- 产品形态：AI平台服务、自然语言处理平台、OCR识别、智能对话机器人等
- 应用场景
 - 人工智能领域：国内最早从事NLP研发的企业之一，在NLP、知识图谱、OCR、图像视频结构化领域都具备自主可控的底层技术，处于行业领先地位
 - 大数据领域：具有全球领先的智能检索技术，现已将领先优势拓展到数据采集、治理、分析挖掘等大数据核心技术的全生命周期，所有底层技术均自主可控
 - 数据安全领域：子公司天行网安是国内最早从事网络安全和数据交换的企业，发明了国内第一台安全隔离网闸，在数据视频交换、单向导入等方面实力雄厚
- 客户体系：暂无公开信息

垂类玩家：云通信上市企业梦网科技致力于Chatbot研发

▪ 上市公司，主营云通信服务，规划产品chatbot，现正在使用LLM模型建模

团队

- **余文胜**
 - **学历背景**：江西大学数学系本科毕业，曾就读于清华大学EMBA总裁班
 - **工作经历**：1992年在澳大利亚悉尼KIRBY公司从事电机自动化研究及设计工作；1994年在美国协和集团电信研究院从事软件设计工作；1997年创办珠海亚强电子有限公司，开发系列寻呼软件和各种型号的无线设备，占有国内近90%的铁路寻呼台市场；2001年，创办深圳市梦网科技发展有限公司，2015年10月入职梦网荣信科技集团股份有限公司，现任梦网荣信科技集团股份有限公司董事长兼总经理，目前兼任深圳市梦网科技发展有限公司执行董事兼总经理、珠海亚强电子有限公司执行董事、深圳市梦网控股发展有限公司执行董事兼总经理、深圳市物联天下技术有限公司总经理、深圳市梦网云播科技发展有限公司董事长、辽宁荣信兴业电力技术有限公司董事长、深圳市梦网物联科技发展有限公司董事长、职通线教育科技（北京）股份有限公司董事

历史沿革

- 成立于2001年，二十年来一直专注于移动通信服务，于2015年成功登陆资本市场（股票代码：002123），目前主营中国最大规模之一的企业云通信平台，是中国领先的云通信服务商
- 梦网构建了以“消息云、视讯云、终端云、物联云”四位一体的企业云通信服务生态，为企业提供全方位的云通信服务
- 梦网携手中国移动、中国电信、中国联通及数千家大型知名企业，在云通信领域开展了深入的研发、建设、维护和服务，软硬件产品广泛应用于金融、互联网、商超等领域，为数十万家大中型企业与超十亿用户提供沟通便利，每年实现互动千亿次

技术

- **底层模型及架构**：基于NLP，使用LLM模型建模
- **算力**：暂无公开信息
- **参数**：暂无公开信息
- **数据量**：暂无公开信息

商业化

- **产品形态**：chatbot是5G会话消息服务载体，可实现服务机器人跨平台的数据共享，结合云识图、语音识别等技术，以chatbot为载体，打造了一站式、智能化的5G消息便捷服务；与国外火热的chatGPT相对仅仅是技术实现方案差异，本质一致，同步研究模块化的集成方式，可实现客户已有机器人业务快速导入，实现资源共享
- **应用场景**：主要应用于银行业务咨询、政务办事指南、电商商品推荐、文旅周边服务等丰富的业务场景，可为全行业提供便捷、高效的业务咨询、智能推荐，通过“智能识别+用户交互+多轮会话”，构建智能化、全场景的业务闭环
- **客户体系**：暂无公开信息

垂类玩家：依图科技自研ConvBERT 模型

- 主营为AI技术服务，依托人脸识别技术开发多款产品应用落地于智能公共服务和智能商业，2020年自研预训练语言理解模型ConvBERT入选NeurIPS 2020
- 团队背景扎实

团队

领导团队

- **朱琰博士**：全球杰出人工智能科学家
 - **学历背景**：2008年获加州大学洛杉矶分校（UCLA）统计学博士，师从计算机视觉奠基人Alan Yuille教授（IEEE 院士，师承物理大师霍金）
 - **工作背景**：2008-10年在麻省理工学院（MIT）人工智能实验室担任博士后研究员，同计算摄影学奠基人Bill Freeman教授（IEEE院士）工作；2010-12年在纽约大学（NYU）Courant数学研究所担任研究员，在神经网络奠基人、深度学习鼻祖Yann LeCun教授（美国工程院院士）实验室期间，带领NYU团队获得国际计算机视觉算法竞赛冠军；2012年回国创立依图科技
 - **学术成就**：曾在世界顶级人工智能、机器学习、计算机视觉刊物PAMI、IJCV、NIPS、AISTATS、CVPR等发表10+篇论文
- **林晨曦**：前阿里云资深专家、技术总监
 - **工作经历**：曾在微软亚洲研究院从事机器学习、计算机视觉、信息检索以及分布式系统方向的研究工作；阿里期间，带领百人团队，搭建了国内最大的拥有自主知识产权的飞天分布式云计算操作系统
 - **曾获奖励**：曾获得ACM世界冠军、国际计算机视觉算法竞赛冠军、国际信息学奥赛（IOI）金牌、全球人脸识别算法竞赛冠军、国际人机交互最佳论文提名等众多国际冠军奖项
- **核心团队**：成员具备麻省理工学院、牛津大学、剑桥大学、芝加哥大学等世界顶级学府及全球知名企业机构背景，技术团队囊括中国自主知识产权云计算操作系统开拓者，全球安卓机器学习推荐系统开创者，以及大规模分布式存储、分布式计算、运维、算法等众多技术领域资深产品专家

历史沿革

- 成立于2012年，2020年发表预训练语言理解模型ConvBERT入选NeurIPS 2020
- **融资历程**
 - A轮（14/12）：数百万美元，岳斌、红杉资本中国、高榕资本
 - B轮（16/06）：数千万美元，云锋基金、虞锋
 - C轮（17/05）：3.8亿人民币，洪婧、云锋基金、众为资本、红杉资本中国、高榕资本、真格基金、虞锋、高瓴投资；C+轮（18/06）：2亿美元，工银国际、高成资本、浦银国际
 - 战略融资（18/07）：1亿美元，兴业资管-兴业证券
 - D轮（19/05）：联新资本、上海科创投集团；D+轮（19/05）：红杉资本中国、高榕资本、真格基金
 - 战略融资（20/03）：3,000万美元，润诚产业领航基
 - 战略融资（20/06）：红杉资本中国、高瓴投资

技术

- **底层模型及架构**：ConvBERT 模型、NLP模型、语音识别算法、声纹识别算法、视觉计算、知识推理
- **算力**：暂无公开信息
- **参数**：仅用1/10的训练时间和1/6的参数就获得了与BERT模型一样的精度
- **数据量**：暂无公开信息

商业化

- **产品形态**：主要业务分为智能公共服务和智能商业两大类
 - 智能公共服务包括智能城市、智能医疗，产品包括胸部CT赋能辅助诊断产品、肺癌临床科研智能病种库、上海“一网通管”系统建设、“脸行贵阳”等
 - 智能商业包括智慧金融及智慧零售，产品包括无卡取款机、招商银行网点接入AI识别等
- **客户体系**：目前已为国内30余省、自治区、直辖市及境外10多个国家和地区的800余家政府及企业终端客户提供产品及解决方案
 - **典型合作伙伴**：中国工商银行、中国联通、中国移动、上海交大、新华智云、BIGO LIVE、YY、珍爱网、一条、墨早、SELF格致论道、花椒直播

新兴挑战者：街远科技基于AI大模型开发多轮对话AI产品

▪ SaaS产品已实现商业化，同时自建底层大模型，在研ProductGPT产品支持多轮对话，助于产品创新，成为兼具底层大模型能力和服务具体场景的应用层能力的公司

团队

- **周伯文**：拥有二十多年自然语言生成、对话与交互式人工智能的研究经验
 - **学历背景**：1996年在中国科学技术大学（少年班）获得电子信息工程专业学士学位，1999年在中国科学院取得信息和信号处理专业硕士学位，2003年在美国科罗拉多大学波尔得分校获得电子和计算机工程专业博士学位；清华大学电子工程系长聘教授，2022年5月亦创建并领导清华大学电子系协同交互智能研究中心
 - **工作经历**：此前在美国学习与研究人工智能近20年，为原IBM Research美国总部人工智能基础研究院院长、IBM Watson集团首席科学家、IBM杰出工程师；2017年7月入职京东，担任历任京东集团高级副总裁、京东集团技术委员会主席、京东云与AI总裁、京东人工智能研究院院长
 - **能力积累**：多次直接负责技术产品业务的0-1、1-10、10-100，既有丰富的带领国际化的研究与技术团队不断突破创新的经历，又有管理数千人业务团队、对P&L直接负责的实践
 - **学术成就**：2016年提出“自注意力融合多头机制的自然语言表征机理”，是Transformer架构的核心思想之一，被Transformer、GAT等论文引用超过2,000次；在AIGC领域，提出自然语言生成算法被引用了3,000余次
- **团队背景**：核心团队均来自于IBM、京东、腾讯、阿里巴巴、百度、清华、字节、小米、华为等一线研究、产业和互联网领域头部机构；技术、产品及业务骨干团队的成员则分别来自北京大学、清华大学、中国科技大学、复旦大学、上海交通大学、哈尔滨工业大学、中国人民大学，及美国科罗拉多大学、密歇根州立大学、澳大利亚悉尼大学等国内外知名院校，拥有丰富的产业经验和深厚的技术产品创新能力

历史沿革

- 成立于2021年底，公司以C2B思想为原则，产品企划IPD流程方法为指导，人工智能算法技术为支撑，聚焦企业产品创新，为企业解决产品创新难、产品成功率低、企划效率低的问题；目前为AI大模型及SaaS服务提供商
- **融资历程**
 - 天使轮（23/03）：启明创投领投，经纬创投跟投，数亿元人民币

技术

- **底层模型及架构**：自建底层大模型，基于模型搭建平台化产品；自主研发了全球首个通过生成式人工智能、多轮交互的人机协同，帮助企业创新的数智化平台—**领衔CIP**（ForeSight CIP TM），基于领衔CIP，街远科技进一步打造了多轮对话ProductGPT
- **算力**：暂无公开信息
- **参数**：暂无公开信息
- **数据量**：暂无公开信息，主要来源于公域数据源

商业化

- **产品形态**
 - **领衔CIP**：已经商业化，同时提供多角色系统、人机互动共创等功能，帮助企业实现高效组织创新，提供从机会洞察、产品规划、产品设计、产品验证到产品上市的全流程、全链路的一站式服务，助力企业产品与服务的全周期智慧运营，增强企业的运营效率和盈利能力，助力企业成为数字经济时代的高价值品牌
 - **ProductGPT**：在研，多轮对话类AI产品，能够让客户与AI系统进行交流、共创与协同，但区别在于，此产品是基于领衔CIP研发的，聚焦3C、消费品、美妆等领域
- **客户体系**：服务京东、蒙牛等众多头部客户

科研院所：鹏城实验室为多个大模型的算力提供方

- 2018年3月中央批准设立鹏城实验室，总部位于广东深圳，主要研究方向是网络通信、网络空间及网络智能
- 为“盘古/文心/悟道”等大模型提供算力支持

团队

领导团队

- **理事会：**聂新平、周玉、高文、高圣元、陈秋明、郭子平、李卓文、梁永生、贾兴东、汤暑葵、王立德、黄湘岳、孙福金、李清泉、薛其坤、徐杨生、杨斌、詹启敏、甄良、樊建平、钟海、刘韵洁、方滨兴
- **学术委员会：**潘云鹤、陈左宁、徐杨生、高文、丁文华、于全、王巨宏、方滨兴、刘韵洁、孙凝晖、陆建华、陈纯、陈国良、郑志明、孟丹、赵沁平、俞大鹏、唐小光、谢大雄、甄良、廖湘科
- **人才培养**
 - 2021年，鹏城实验室与高校联合培养博士研究生专项计划正式启动，鹏城实验室首批与北京大学、清华大学、中国科学院大学、哈尔滨工业大学和南方科技大学五所高校联合招收博士生
 - 2022年，在五所高校基础上，新增中山大学、华南理工大学，与七所高校招收普博生、硕博连读生、直博生，同时加入工程硕博培养改革专项试点实验室，现有联培博士生300+人，持续为国家战略实施提供有力的人才支撑

历史沿革

- 成立于2018年3月，由中央批准成立的突破型、引领型、平台型一体化的网络通信领域新型科研机构
 - 宽带通信：**依托“硅光芯片工艺平台”“宽带通信全场景云网创新测试环境”等科学设施与平台，打造网络通信核心能力和技术长板，重点在无线通信、光通信、空天一体化网络、深空通信、水下通信等方向进行布局
 - 新型网络：**依托“鹏城靶场”“鹏城生态”突破新型网络体系架构关键技术，重点在新型网络体系架构、工业互联网等方向进行布局
 - 网络智能：**依托“鹏城云脑”打造全球领先的智算平台，突破高效能云计算服务等关键核心技术，构建自主可控的AI基础软硬件体系；鹏城云脑II系统的AI计算子系统包含 4096 颗昇腾 910AI 处理器，理论上可提供 1EOPS FP16 和 2EOPS INT8 的 AI 运算能力，其通用算力由基于 ARM 架构的 2048 颗 CPU 组成，每个 CPU 提供不少于 0.49T FLOPS 的通用算力，支持了国内多个大模型训练（如盘古/文心/悟道2.0等）

技术

- **底层模型及架构：**首创顺序自回归预训练语言模型ALM，具备MindSpore超大规模自动并行技术，模型基于国产全栈式软硬件协同生态，包括MindSpore、CANN、昇腾910及ModelArts
- **算力：**2,048卡
- **参数：**2,000亿
- **数据量：**80TB，源自开源开放数据集、common crawl网页数据、电子书等

商业化

- **产品形态及应用场景：**
 - 已开发出“鹏程·盘古”“鹏程·大圣”“鹏程·神农”等人工智能预训练大模型，分别聚焦自然语言处理、计算机视觉以及生物医药等领域
 - 鹏程·盘古 α 预训练模型在知识问答、知识检索、知识推理、阅读理解等文本生成领域表现突出，小样本学习能力较强
- **客户体系：**由政府主导，以哈尔滨工业大学（深圳）为依托单位，与北京大学深圳研究生院、清华大学深圳国际研究生院、深圳大学、南方科技大学、香港中文大学（深圳）、深圳先进院、华为、中兴通讯、腾讯、深圳国家超算中心、中国电子信息产业集团、中国移动、中国电信、中国联通、中国航天科技集团等高校、科研院所和高科技企业等优势单位共建

科研院所：粤港澳大湾区数字经济研究院研发多系列开源大模型

- 创新型研究机构
- 致力于在AI基础技术与开源系统，提供足够开放的场域，鼓励各种合作模式，**目前开源6个系列共10个模型，包含4种模型结构**

团队

领导团队

▪ 沈向洋：理事长

-**过往经历**：于1998年底回北京参与创建微软亚洲研究院，并于2004-07年担任院长及首席科学家，为中国培养了众多国际一流的计算机科学家、技术专家和企业家人；后于2013-19年期间担任微软公司全球执行副总裁，主管微软全球研究院和人工智能产品线，并负责推动公司中长期总体技术战略及前瞻性研究与开发工作

▪ 张家兴：带队认知计算与自然语言研究中心

-**学术成就**：在人工智能、深度学习、分布式系统等多领域的顶级会议和会刊上（NIPS、OSDI、CVPR、SIGMOD等）发表论文二十多篇，提交七十余项专利

-**过往经历**：曾就职于百度、微软、阿里巴巴，曾任微软亚洲研究院研究员、蚂蚁金服人工智能资深算法专家、360数科（原360金融）首席科学家

▪ 张磊：带队计算机视觉与机器人研究中心

-**行业地位**：主要研究领域是大规模图像理解中的模式挖掘及物体检测，提出基于搜索的图像标注方法首次将超大规模（二十亿）图像搜索用于计算机视觉的图像理解中，并解决一系列特征匹配、高维索引及超大规模聚类问题，取得一系列的创新成果并在领域长期处于国际领先地位

-**过往经历**：原微软亚洲研究院、微软总部研究院及计算机视觉相关产品部门任首席研究员，IEEE Fellow

历史沿革

- IDEA开源的大模型系列，能够让高校、小公司及部分传统公司，有渠道获取算力资源进行研究
- 融资历程暂无融资

技术

▪ “封神榜”大模型架构

-**二郎神系列**：开源Bert大模型，以Encoder结构为主的双向语言系列模型/ 参数：13亿参数/ 数据量：暂无公开信息

-**余元系列**：Yuyuan-GPT2-3.5B/ 参数：35亿参数/ 数据量：暂无公开信息

-**周文王系列**：与追一科技联合开发，Zhouwenwang-Unified-1.3B大模型/ 参数：开源13亿参Zhouwenwang-Unified-1.3B大模型/ 数据量：暂无公开信息

-**闻仲系列**：以Decoder结构为主的单向语言模型，是一系列强大的生成模型/ 参数：开源35亿参数Wenzhong-GPT2-3.5B大模型/ 数据量：暂无公开信息

-**燃灯系列**：以Transformer结构为主的编解码语言模型/ 参数：开源7.7亿参数Randeng-MegatronT5-770M大模型/ 数据量：暂无公开信息

-**比干系列**：暂无公开信息/ 参数：开源11亿参数Bigan-Transformer-XL-denoise-1.1B大模型/ 数据量：暂无公开信息

商业化

- 暂未进行商业化