

2026年AIGC 行业现状及发展趋势

AIGC in 2026
Industry Status and development trends

白皮书





引言

2026 年 Vibe Coding 创作者经济正式来临，Vibe Coding 创作者经济将迎来万亿市场规模，到 2030 年将产生 300w 的 Vibe Coding 创作者，将是 OPC 的重要组成部分。

整个过程中将伴随着 2023 年以前的职业被替代，新的职业 (OPC、FED、Context Engineering、Researcher 等) 正在产生，让我们迎接 2026 全民 AI 新范式。

占冰强

AIGCLink 发起人

编写委员会

(排名以首字拼音排序，排名不分先后)

董慧智、吕叶、王晨阳、占冰强、张江达



白皮书简介

——本白皮书由 AIGCLink 研究院编撰，旨在为 AI 从业者、投资人及企业决策者提供 2026 年行业前瞻指南。报告收录了占冰强等行业观察者的 13 个关键趋势判断，发布了“vibe coding 经济”概念，预测 2026 年 OPC (one person company) 和 FDE (Forward Deployed Engineer, 前线部署工程师) 等职业的崛起；还涵盖上下文资产化、腰部 AI 公司崛起、Vibe Coding 新生产力范式、可解释模型与无限上下文等技术演进路径、AI 短剧爆发、组织架构定量化等核心议题。同时，报告以详实数据呈现 AI 商业化与投融资全景，并附 AI 行业上下游公司图谱、开源项目生态图谱及 AIGCLink 全球开源项目 2026 年 1 月增长榜单，为读者提供一份立体、可操作的 AI 发展全景参考。

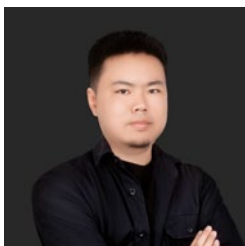


欢迎关注 AIGCLink 公众号
获得更多内容

2026 AI 新增长点：Vibe Coding创作者经济

生产力、生产关系与技术演进的全景洞察

作者介绍



占冰强，AIGCLINK发起人，微软MVP，信通院金融领域大模型评估标准制定组专家，中国合作贸易企业协会出海分会副会长，中关村超互联联盟副秘书长，MCP中国开发者联盟发起人，《Llama大模型实践指南》和《一本书讲透MCP》作者

2025年见证了AI行业的历史性转折。DeepSeek以低于600万美元的训练成本打造出媲美GPT-4的模型，动摇了硅谷“以算力换智能”的根本假设。Y Combinator 2025冬季批次中，25%的创业公司代码库由AI生成超过95%。与此同时，Vibe Coding从一个Twitter热梗演变为重塑软件开发的新范式——92%的美国开发者每天使用AI编码工具，全球41%的代码已由AI生成。

这些数字背后是一场更深层的变革：AI正从“工具”进化为“基础设施”，从“辅助”转向“协作”，从“文本”拓展到“物理世界”。本白皮书基于对2025年AI技术演进、市场格局和应用实践的系统性调研，提出13个关键判断，勾勒2026年AI发展的全景图谱。

核心发现：上下文将成为2026年企业最重要的资产，地位等同于互联网时代的数据资产；AI腰部公司将迎来爆发式增长，复刻互联网垂直网站崛起的路径；世界模型的ChatGPT时刻即将到来，VLA模型正在开启物理AI新纪元；AI短剧在中国率先验证了商业化路径，多模态一致性问题取得突破性进展；组织架构变革从定性进入定量阶段，Token消耗量正在成为衡量AI员工的核心指标。

一、将vibe coding转换为确定性系统

占冰强观点：2026年上下文将成为公司最重要的资产，犹如互联网时代的数据资产，这将成为AI公司未来估值的核心指标之一；同时支持人类上下文和AI上下文，将人类的上下文翻译为AI能理解的上下文将是未来公司人员最重要的能力。。

● 从Vibe Coding到确定性系统

DeepSeek的突破不仅仅是技术成就，更是市场认知的转折点。正如高盛研究报告指出，DeepSeek重新定义了AI基础模型性能和成本的基准，加速了企业AI采用，并在云服务价值链中引发深远变化。中国云计算支出预计在2025年增长15%，这一增长主要由DeepSeek驱动的AI采用推动。

据统计，Vercel 和 Netlify 等平台用户量在 2025 年出现爆发式增长，推动力正是涌入的 Vibe Coding 开发者。Cursor AI 重新定义了集成开发环境，让开发者在自然语言描述和代码之间无缝切换。IBM 将 Vibe Coding 定义为用户用日常语言表达意图、AI 将其转化为可执行代码的开发模式，代表着软件工程从严格手工编码向灵活 AI 驱动的根本转变。

● 2026年企业最重要的资产

如果说数据是互联网时代的石油，那么上下文将成为AI时代的新能源。上下文不仅仅是数据的堆砌，而是数据的组织方式、关联逻辑和语义理解的综合体现。一个拥有优质上下文的企业，意味着其AI系统能够准确理解业务逻辑、用户意图和领域知识，从而产出更精准、更可靠的结果。

Llama 4 的发布标志着大模型上下文能力的又一里程碑——1000万Token的上下文窗口使模型能够一次性处理相当于数十本书的信息量。然而，上下文窗口的扩展带来显著的成本压力。研究表明，Llama4在1000万Token查询时的单次成本可能达到2-5美元。更关键的是，将人类的隐性知识、业务流程和决策逻辑转化为AI能够理解的结构化上下文，将成为企业人员的核心能力。

这意味着2026年的企业估值逻辑将发生根本变化。投资者将开始关注：企业积累了多少高质量的上下文资产？其上下文系统能够支持多少AI用例？将人类知识转化为AI上下文的效率如何？这些指标将与用户数、收入等传统指标同等重要。

二、AI行业大爆发：从To B到腰部公司

占冰强观点：2025年deepseek教育市场后，所有人开始相信ai，引发了to b行业的大爆发，同理：2026年人们的对AI尝鲜过后，开始对交付结果有要求，将会引发垂直细分的腰部Ai公司和to c细分应用的大爆发

● DeepSeek效应：市场教育的临界点

2025年1月，DeepSeek R1模型的发布堪称AI行业的分水岭时刻。这家中国创业公司以不到600万美元的训练成本和2,048块受限的Nvidia H800芯片，打造出性能媲美OpenAI o1的推理模型。消息传出当天，Nvidia市值蒸发近6000亿美元，创下单日最大跌幅记录。

DeepSeek的突破不仅仅是技术成就，更是市场认知的转折点。正如高盛研究报告指出，DeepSeek重新定义了AI基础模型性能和成本的基准，加速了企业AI采用，并在云服务价值链中引发深远变化。中国云计算支出预计在2025年增长15%，这一增长主要由DeepSeek驱动的AI采用推动。

更深远的影响在于开源生态的激活。智谱AI宣布2025年将是开源之年；百度、腾讯、阿里纷纷开放其大模型；零一万物创始人李开复预言，当竞争对手免费且强大时，OpenAI将很难维持其定价体系。这种动态正在从根本上改变AI的商业模式，从封闭的专有模型转向开源生态系统的竞争。

● To B与腰部公司的爆发窗口

回顾互联网发展史，搜索引擎的成熟并未阻止垂直网站和地方门户的崛起，反而为后者提供了流量入口和基础设施。同样的逻辑正在AI领域重现。

2025年，DeepSeek完成了对市场的教育——不是AI是否有用，而是AI如何具体应用。这种认知转变为To B市场打开了闸门。据Counterpoint Research分析师评估，中国的AI采用规模和速度都惊人，而且没有放缓的迹象。高盛预测，到2030年，AI采用将为中国GDP贡献20–30个百分点的增长。

2026年将迎来第二波浪潮：当用户从尝鲜转向要求交付结果时，通用大模型的局限性将充分暴露。企业需要的不是能聊天的AI，而是能理解其业务、遵守其规范、融入其流程的AI解决方案。这正是垂直细分AI公司的机会窗口。

三、AI行业大爆发：从To B到腰部公司

占冰强观点：Lim 公司犹如互联网时代的四大门户，豆包deepseek 等犹如搜索引擎，明年的腰部公司犹如互联网时代的垂直网站、地方门户，对应新时代替代职务性ai和行业性ai

● To B与腰部公司的爆发窗口

互联网早期，新浪、搜狐、网易、腾讯被称为四大门户——它们提供综合性的信息服务，是用户进入互联网的第一入口。今天的大模型公司正在扮演类似角色：OpenAI、Anthropic、Google、Meta构建了AI的基础设施层，用户通过ChatGPT、Claude、Gemini进入AI世界。

更有意思的类比是搜索引擎。豆包、DeepSeek等中国玩家的崛起，让人联想到百度挑战Google的历史。它们通过本地化优势、成本优势和开源策略，在特定市场建立起竞争护城河。正如搜索引擎时代Google和百度可以共存，大模型时代的竞争格局也不会是赢家通吃。

● 两类AI公司的崛起路径

职能性AI：替代标准化、重复性的岗位职能。典型场景包括客服机器人、内容审核系统、数据标注平台、文档处理工具等。这类公司的核心竞争力在于流程标准化和规模化部署能力，商业模式通常是按处理量收费或SaaS订阅。

行业性AI：深度嵌入特定行业的业务流程和专业知识。典型场景包括医疗影像诊断、法律文书审查、金融风险评估、工业质检等。这类公司需要积累深厚的领域知识和行业关系，商业模式更接近传统企业软件，强调深度服务和长期合作。

2025年的中国短剧行业提供了一个生动案例。据启信宝统计，截至2025年10月，国内活跃的短剧相关企业超过35万家。AI+短剧领域在2025年至少发生9起融资事件，总金额超过1亿元。这种爆发式增长正是AI长尾效应的典型体现——大模型降低了内容生产的门槛，释放了海量的创作者能量。

四、AI行业大爆发：从To B到腰部公司

占冰强观点：2026年vibe coding新的工作方式:stitch做产品+gemini做前端+cc做后端+Gpulink 做部署

● 专业化工具链的形成

2026年的软件开发将不再是一个模型打天下，而是根据任务特性调用最适合的工具。一个典型的专业化工作流可能包括：Stitch/v0用于产品原型设计，Gemini处理前端界面生成，Claude Code负责后端逻辑开发，GPUlink等平台完成部署上线。

这种分工反映了AI能力的差异化演进。最有效的方法是像我们使用传统工具一样——用Claude做架构设计，用本地模型做代码补全，用Gemini处理多模态任务。这种组合式AI工作流的兴起，意味着2026年最重要的开发者技能不再是编写代码本身，而是有效地协调AI工具。

● 开发者角色的重新定义

Y Combinator 2025冬季批次的数据令人震撼：25%的创业公司代码库由AI生成超过95%。这个数字在两年前还是不可想象的。它暗示着软件开发正在经历类似制造业自动化的转型——人类从操作者变成监督者和设计ISACA的分析框架清晰地描绘了这一转变：架构师需要超越系统设计，创建安全、可复用的脚手架来引导AI；开发者从编码者转变为审核者和策展人；QA团队不仅验证功能，还要验证AI的逻辑路径；安全和风险团队需要像对待第三方软件一样评估AI生成的代码。

五、AI短剧：多模态一致性的商业验证

占冰强观点：AI短剧大爆发，多模态的一致性初步解决，10秒内的视频能解决，5分钟的长视频一键生成2025将发生

● 技术突破：从10秒到5分钟

2025年是AI视频生成的突破之年。Sora 2、Vevo 3、Kling O1等模型相继发布，将视频生成的质量和时长推向新高度。Kuaishou发布的Kling O1被定位为世界首个统一多模态视频模型，能够保持角色、道具和场景在动态镜头中的特征稳定性。

这些技术进步正在改变内容生产的经济学。据报道，10秒以内的高质量视频已经可以稳定生成；1分钟以上的长视频在时序一致性和画面质量上取得了突破性进展。业内预测，5分钟的一键成长视频将在2025-2026年成为现实。

● 中国AI短剧的商业化验证

中国的微短剧行业成为AI视频生成商业化的最佳试验场。这类内容单集时长从30秒到几分钟不等，竖屏呈现，剧情节奏快、反转多，完美契合移动优先的观看习惯。更重要的是，观众在手机屏幕上观看时，往往不会注意到AI生成的细微瑕疵——这为技术的商业化提供了宝贵的过渡期。

上海出品的AIGC微短剧《灵探》创下了跨平台1000万播放量的记录。这部46集的系列剧，每集约2分钟，总时长超过100分钟，被誉为中国首部总时长超100分钟的AIGC真人风格微短剧。据行业估算，AI驱动的生产方式将成本降低至传统方法的五分之一，制作周期缩短一半以上。

六、大模型zero路径

占冰强观点：RL之后的模型能力提升路径，将是可解释模型，强化学习的策略算法是可解释的

● 强化学习的天花板

DeepSeek R1的成功建立在强化学习（RL）的突破之上。通过基于规则的Group Relative Policy Optimization（GRPO），模型获得了长链思维推理能力，在数学、编程等任务上实现了质的飞跃。然而，RL方法也面临根本性局限：它优化的是结果，而非过程；它提升的是性能，而非理解。

这意味着RL训练出的模型可能在基准测试上表现优异，但其决策逻辑对人类而言仍是黑箱。在高风险场景（医疗诊断、金融决策、自动驾驶），这种不可解释性构成了根本性障碍。监管机构需要知道为什么，而不仅仅是是什么。

● 强化学习的天花板

可解释强化学习（Explainable RL, XRL）正在成为新的研究热点。据ACM Computing Surveys的系统综述，XRL旨在阐明强化学习代理在序列决策环境中的决策过程。最新的研究方向是将Shapley值引入策略解释，在保持模型性能的同时实现了策略的透明化。

展望2026年，可解释模型很可能成为RL之后的下一个能力提升路径。世界模型的想象机制可以被用于生成逐步的、对比性的解释，帮助人类理解AI的决策依据。

七、自我学习：训推一体的推理加速

占冰强观点：模型训推一体的推理加速框架将产生

传统的AI系统将训练和推理视为两个独立阶段：先在海量数据上离线训练，再将固化的模型部署到生产环境。这种分离意味着模型一旦部署，就无法从新数据中持续学习——除非经过昂贵的重新训练流程。

2025–2026年，训推一体的新范式正在形成。核心思想是让模型在推理过程中持续优化自身参数，实现边用边学。Meta发布的V-JEPA 2世界模型展示了这种可能性：它可以在仅62小时的机器人数据上完成二次训练，就能够实现零样本的机器人规划和控制。

这种架构的关键技术包括：参数高效微调（PEFT）使大模型可以在边缘设备上进行轻量级适应；混合专家架构（MoE）允许模型根据任务动态激活不同参数；持续学习框架防止灾难性遗忘。训推一体的商业意义在于：企业可以部署一个会成长的AI系统，它能够从每次交互中积累经验，逐渐适应特定业务场景。

八、无限上下文：技术突破与成本约束

占冰强观点：无限上下文将实现，但成本是制约发展的最大的变量

● 上下文窗口的指数级扩张

大模型的上下文窗口经历了指数级增长。2018年GPT-1仅支持512 Token；2024年Claude 3和Gemini 1.5 Pro将上限推至100万Token；2025年Llama 4 Scout更是宣布支持1000万Token的上下文窗口——相当于一次性处理超过750万个单词，或约7500本书的内容。

技术突破来自多个方向。Google研究者提出的Infini-Attention机制通过压缩记忆替代完整的注意力计算，使模型能够在恒定内存占用下处理100万Token以上的输入。然而，无限上下文仍面临物理和经济限制——研究表明，即使模型能够完美检索证据，大量无关内容的存在也会干扰推理能力。

● 成本：制约发展的核心变量

成本仍是无限上下文商业化的最大障碍。按Llama 4的定价估算，一次满载1000万Token的查询成本在2-5美元之间。行业正在通过多种方式应对成本挑战：提示缓存技术可以复用重复的上下文片段；检索增强生成（RAG）通过精准筛选相关信息来降低输入量；混合专家架构只激活部分参数，降低单次推理的计算成本。展望2026年，无限上下文将实现技术可行性，但成本仍是制约大规模应用的核心变量。最成功的策略不是一味追求最长的上下文，而是精准匹配任务需求与上下文规模，在质量和成本之间找到最优平衡点。者。

九、多模态推理：统一加速框架的诞生

占冰强观点：多模态推理加速框架出现一个适配各类多模态模的框架

多模态推理加速框架出现一个适配各类多模态模的框架

多模态 AI 正在从实验性功能向标准配置演进。OpenAI 的 GPT-4o、Google 的 Gemini、Meta 的 Llama 4 都将多模态理解作为核心能力。然而，多模态推理的计算复杂度远超纯文本模型——需要同时处理视觉编码器、语言模型和跨模态对齐，对硬件和软件框架都提出了严峻挑战。

2025年见证了统一多模态推理框架的兴起。MDPI发表的研究提出了一种统一的、资源感知的推理优化框架，整合了混合精度量化、稀疏注意力和动态专家路由三大技术，能够根据Token复杂度和硬件条件动态调整计算路径。

2026年，适配各类多模态模型的统一推理框架将成为基础设施的标配。开发者不再需要为每种模型单独优化部署，而是可以通过标准化接口调用最适合当前任务的多模态能力。

十、模态类型：感知边界的持续扩展

占冰强观点：模态感知边界正从native的文本、图像、视频扩展到其他类型

AI的模态感知能力正在从原生模态向拓展模态演进。原生模态包括文本、图像、视频——这些是互联网数据的主体，也是大模型预训练的主要来源。拓展模态则涵盖3D空间理解、音频情感识别、传感器数据解析、代码执行等更专业的领域。

GLM-4.5V模型引入了3D旋转位置编码（3D-RoPE），显著增强了对三维空间关系的感知和推理能力。腾讯的HunyuanVideo-Avatar则是专门为情感控制和多角色对话设计的多模态模型，通过专用模块实现面部表情和语音情感的精准同步。

更前沿的探索在于具身智能领域。NVIDIA发布的Cosmos Reason 2是一个开放的推理视觉语言模型，使智能机器人能够像人类一样观察、理解和在物理世界中行动。这种模态扩展的趋势预示着AI将从数字世界的理解者进化为物理世界的参与者。

十一、世界模型：ChatGPT时刻即将到来

● VLA模型：物理AI的基础设施

视觉-语言-动作（VLA）模型是2025年最重要的AI技术突破之一。它将视觉理解、语言指令解析和机器人动作生成统一在一个框架内——给定摄像头图像和自然语言指令，直接输出可执行的机器人动作。

这一领域的竞争异常激烈。Google DeepMind发布的Gemini Robotics基于Gemini 2.0构建，能够完成折纸、玩纸牌等高精度操作。Figure AI的Helix采用双系统架构，首次实现了对人形机器人全身（手臂、手指、躯干、头部）的高频控制。Physical Intelligence公司的 $\pi 0$ 模型展示了跨机器人泛化的可能性——在8种不同的机器人形态上训练后，可以控制单臂、双臂等多种配置。

● 世界模型：理解物理世界的内部表征

世界模型是比VLA更底层的能力——它构建对外部环境的内部表征，并利用这种表征预测未来状态。Meta发布的V-JEPA 2被描述为首个在视频上训练的世界模型，能够实现最先进的理解和预测，以及零样本的规划和机器人控制。

Runway发布的GWM-1是另一个里程碑。它能够生成可探索的环境、对话角色和机器人操作场景，核心能力在于保持空间一致性。正如ChatGPT让公众认识到语言AI的能力，世界模型的ChatGPT时刻即将到来——一个能够在物理世界中自由探索、预测和规划的AI系统，将从根本上改变我们对智能的认知。

Bank of America预测，人形机器人的材料成本将从2025年的约35,000美元降至未来十年的13,000-17,000美元，高盛报告显示2023-2024年人形机器人制造成本已下降40%。

十二、生产关系新范式：Vibe Coding创作者经济

占冰强观点：新的vibe coding创作者经济来临，所想即所得，犹如图文创作者经济

● 从图文创作者到AI创作者

创作者经济正在经历AI驱动的又一次革命。Epidemic Sound的《2025创作者经济未来报告》显示，91%的内容创作者现在使用AI参与创作流程；86%的创作者相信AI对创作者经济产生了积极影响。

更深层的变化在于创作门槛的消解。正如图文时代让人人都是创作者，AI时代正在让人人都是开发者。非技术背景的人员可以通过自然语言描述需求，AI将其转化为可运行的应用、视频、音乐甚至游戏。

Research and Markets的报告预测，AI在创作者经济领域的市场规模将从2024年的33.1亿美元增长至2029年的128.5亿美元，年复合增长率31.1%。

● 所想即所得的新生产力

Vibe Coding 的精髓在于所想即所得——创意与实现之间的距离被压缩到极致。这与图文创作者经济的逻辑一脉相承：博主不需要掌握 InDesign，YouTuber 不需要精通 Final Cut Pro，AI 时代的创作者也不需要精通 Python 或 JavaScript。

Adobe的《2025创作者工具箱报告》揭示了这种转变的规模：76%的创作者表示AI加速了其业务或粉丝增长；81%表示AI帮助他们创作出原本无法完成的内容。当创意不再被技术能力所限制，人类的想象力将成为真正的生产力。

十三、组织架构量化：AI渗透率的新度量

占冰强观点：组织架构变革正从定性进入到定量，考核AI渗透率正在量化，一个人每月能用多少token或算力决定了这个人是否是AI员工的核心指标

● 从定性到定量的组织变革

OpenAI 发布的《2025 企业 AI 状态报告》提供了一个惊人的数据：ChatGPT Enterprise 的周消息量同比增长 8 倍，API 推理 Token 消耗量同比增长 320 倍。这些数字不仅反映了 AI 采用的加速，更预示着一种新的组织度量方式的诞生。

传统的组织效能评估依赖定性指标——团队士气、协作氛围、创新文化。AI时代，这些指标正在被可量化的数据所补充：每个员工每天发送多少AI查询？使用了哪些高级功能？完成了多少新类型的任务？节省了多少工作时间？

报告显示，使用ChatGPT Enterprise的员工平均每天节省40-60分钟；重度用户每周节省超过10小时。更关键的是，75%的用户表示能够完成以前无法完成的新任务——AI不仅提高了效率，更扩展了能力边界。

● Token消耗：AI员工的核心指标

一个员工每月能用多少Token或算力，正在成为衡量其AI化程度的核心指标。OpenAI的数据显示，前沿员工（95百分位）发送的消息量是中位数员工的6倍；前沿企业每个席位的消息量是中位数企业的2倍。

这种分化具有深远的竞争含义。BCG 2025年的研究发现，AI领先企业相比落后企业在生产力、创新和员工满意度上获得3-4倍的优势。高消耗Token的员工和团队，本质上在建立难以复制的能力护城河。

2026年，我们预计将看到更多企业将AI渗透率纳入正式的KPI体系：不仅考核AI工具的部署覆盖率，更考核使用深度、任务类型多样性和业务成果关联度。AI员工将从一个比喻变成一个可操作的管理概念，Token消耗量将成为其核心度量指标。

结论：站在范式转换的十字路口

2025年是AI从新奇走向实用的转折年；2026年将是AI从实用迈向基础设施的关键年

本白皮书提出的13个判断，试图捕捉这场范式转换的关键维度：

在生产力层面，上下文工程取代 Vibe Coding 成为主流，专业化工具链重新定义开发者角色，训推一体和无限上下文技术持续演进。在产业层面，DeepSeek 效应催化市场教育，腰部公司迎来爆发窗口，AI 短剧验证多模态内容的商业路径。在技术层面，VLA 和世界模型推动物理 AI 成熟，多模态推理框架趋于统一，模态感知边界持续扩展。在组织层面，Token 消耗量成为 AI 渗透率的核心指标，组织变革从定性进入定量阶段。

这些判断的共同指向是：2026年，上下文将成为企业最重要的资产，世界模型将迎来 ChatGPT 时刻，AI 创作者经济将重塑生产关系，组织架构将完成量化转型。

对于企业决策者而言，关键行动包括：投资建设高质量的上下文资产；评估并布局垂直 AI 应用机会；建立 AI 渗透率的度量和激励体系；培养将人类知识转化为 AI 上下文的组织能力。

站在范式转换的十字路口，最大的风险不是行动太快，而是观望太久。那些在 2026 年建立起上下文优势、AI 能力深度和组织适应力的企业，将在接下来的十年中持续领跑。

参 考 来 源

● 本白皮书数据基于以下权威来源的公开数据：

OpenAI 《2025企业AI状态报告》；Y Combinator 2025冬季批次数据；MIT Technology Review；Thoughtworks Technology Radar；Goldman Sachs Research；CNBC；South China Morning Post；IBM Think；The New Stack；ISACA；Fast Company；DataCamp；Deloitte Insights；Britannica Money；Wikipedia；ArXiv；Springer；ACM Computing Surveys；Frontiers in Robotics and AI；NVIDIA Newsroom；Runway Research；Meta AI Blog；Physical Intelligence；Epidemic Sound；Adobe；Research and Markets等。

数据截至2026年1月15日。市场数据和预测基于各研究机构的公开报告，具体数值可能随时间更新。

2025年AI投融资情况分析

核心发现：上下文将成为2026年企业最重要的资产，地位等同于互联网时代的数据资产；AI腰部公司将迎来爆发式增长，复刻互联网垂直网站崛起的路径；世界模型的ChatGPT时刻即将到来，VLA模型正在开启物理AI新纪元；AI短剧在中国率先验证了商业化路径，多模态一致性问题取得突破性进展；组织架构变革从定性进入定量阶段，Token消耗量正在成为衡量AI员工的核心指标。

核心发现

维度	美国	中国
融资规模	\$1590亿（占全球79%）	Q1融资181亿元,全年预计超2000起
头部估值	OpenAI \$5000亿, Anthropic \$1380亿	智谱AI上市,月之暗面\$43亿
热门赛道	基础模型、AI基础设施	具身智能、AI应用
投资特点	资本集中、少数巨额交易	「投早投小」、脱虚向实
全阶段布局	放松监管、加速创新	「AI+」深度融合、自主可控

给AI开发者的核心建议

- 技术方向：AI Agent和具身智能是中美两国共同看好的下一波浪潮
- 创业机会：垂直行业应用（医疗AI、工业AI）比通用大模型更具可行性
- 就业市场：AI工程师年薪\$12-20万美元,需求持续旺盛
- 长期趋势：从「模型可用」向「场景好用」转变,应用层机会涌现

一、2025年全球AI投融资概况

● 史无前例的融资规模

2025年全球AI投融资达到了前所未有的高度：

- 总融资额：\$2020-2260亿美元,占全球风险投资的48%-51%
- 较2024年增长：AI占VC比例从34%跃升至约50%
- 交易特征：交易数量下降,但单笔金额大幅上升,呈现「少而精」特点

根据CB Insights数据,2025年Q1 AI融资即达\$666亿美元,较上季度增长51%。全年来看,AI公司累计融资\$2260亿美元,占全部VC融资的48%。

● 资本高度集中的马太效应

2025年AI投资最显著的特征是资本向头部企业的极度集中：

公司	融资额	主要投资者
OpenAI	\$400亿	SoftBank、Microsoft、Founders Fund
Scale AI	\$143亿	Meta领投
Anthropic	\$130亿	Lightspeed、Salesforce、Alphabet
xAI	\$100亿	多轮累计\$220亿

基础模型公司在2025年累计融资\$800亿美元,占全球AI融资的40%。

仅OpenAI和Anthropic两家就获得了全球VC投资的14%。

● 主要投资机构格局

机构	2025年AI布局	代表投资
Andreessen Horowitz (a16z)	募集\$150亿+新基金	OpenAI、Anthropic、Cursor
Sequoia Capital	\$1.5亿投向基础模型	OpenAI、Nvidia、Notion
SoftBank Vision Fund	\$400亿投资OpenAI	OpenAI及AI基础设施

Microsoft、Alphabet、Amazon、Meta四大科技巨头2025年合计承诺在AI技术和基础设施上投入\$3200亿美元。

二、中国AI投融资市场分析

● 融资规模与增长趋势

2025年中国AI投融资市场活跃度创历史新高:

- 全年融资事件: 预计达2062起,创历史新高
- Q1融资额: 241起案例,涉及融资金额181.4亿元
- AI应用融资: 一级市场AI应用融资总额达1070.7亿元,涉及930家公司
- 产业规模: 中国AI产业规模预计达8662亿元

● 投早投小、脱虚向实的投资趋势

1. 投资阶段前移

- 天使轮和A轮各占已披露轮次的近三分之一
- 资本更愿意押注早期创新项目
- 大模型平均单笔融资额从2023年1.73亿元增至2025年2.41亿元

2. 从概念走向落地

- 2025年被普遍认为是「AI应用元年」
- 投资重心从通用大模型转向垂直行业应用
- 核心关注从「模型可用」转向「场景好用」

● 主要投资机构

红杉中国 (HongShan)

- 2025年投资事件中,早期阶段占比达80%
- 重点布局AI和先进制造
- 已投资近30家AI相关公司
- 2025年11月宣布两只总规模约\$9.5亿美元新基金

高瓴资本 (Hillhouse Capital)

- 重点布局硬科技: 芯片、自动驾驶、基础软件
- 是MiniMax最早投资者之一
- 2025年末启动约\$70亿美元新PE基金募资

● 热门赛道分析

具身智能: 2025年最火赛道

- 2025年前5个月融资232亿元人民币 (\$32.3亿),超过2024年全年
- 全年投资案例激增144%至298起
- 融资规模大涨291%至329亿元
- 主要投资者: 京东、美团、阿里巴巴、腾讯、字节跳动

头部企业动态

公司	估值/融资	2025年关键进展
智谱AI	港交所上市,融资超83.6亿元	全球大模型第一股,年收入超1亿美元
月之暗面	\$43亿美元估值	C轮\$5亿融资,Kimi K2超越GPT-5排名
MiniMax	千亿级估值	最高市值超1500亿港元
百川智能	战略转型中	聚焦AI医疗,发布Baichuan-M3
DeepSeek	估值1万亿元人民币	开源模型获全球关注

三、美国AI投融资市场分析

● 全球领先的融资规模

美国在2025年巩固了其AI投资的全球领导地位：

- 总融资额：\$1590亿美元,占全球AI融资的79%
- VC渗透率：美国VC的85%资金流向AI项目
- 交易占比：美国AI交易占全球总交易数的53%
- 硅谷AI初创公司：2025年吸引创纪录的\$1500亿融资

● 巨额融资主导市场

H1 2025表现：

- 美国初创公司融资\$1628亿美元,同比增长76%
- 其中AI初创公司获得\$1043亿美元
- 交易占比：美国AI交易占全球总交易数的53%
- Q1单季度即超\$800亿美元,较上季度增长28%

巨额融资主导市场

机构	特点	2025年动态
Andreessen Horowitz	AI+Crypto双轮驱动	募集\$150亿+新基金,投资165+项目
Sequoia Capital	全阶段布局	预测AI应用将成主战场
SoftBank Vision Fund	大手笔押注头部	\$400亿投资OpenAI,获11%股权
Lightspeed Venture Partners	专注增长期	领投Anthropic
Founders Fund	Peter Thiel旗下	参与OpenAI融资

● 热门赛道分析

生成式AI (Generative AI)

- 2025年上半年融资超\$492亿美元,超过2024年全年
- 占全球VC的11.77%
- 基础模型公司融资\$800亿,占AI总额40%

AI基础设施

- 科技巨头CapEx合计超\$3000亿
- 主要投向: 数据中心、定制芯片、GPU、低延迟云网络
- Stargate \$5000亿AI数据中心计划启动

● 头部企业估值

公司	2025年估值	关键动态
OpenAI	\$5000亿	史上最高私有公司估值,筹备IPO目标\$1万亿
Anthropic	\$1380亿	Q3获\$130亿融资
xAI	\$2300亿	较2025年3月翻倍,Colossus超算扩建
Thinking Machines	\$100亿	前OpenAI CTO Mira Murati创立
Reflection AI	\$80亿	7个月估值飙升15倍

四、中美对比分析

● 投资规模与结构对比

维度	美国	中国
总规模	\$1590亿 (全球79%)	Q1 181亿元,全年2000+笔交易
平均单笔	大额巨笔,资本集中	天使/A轮主导,金额相对分散
独角兽数量	224家,总估值18.97万亿元	150家,总估值12.83万亿元
投资阶段	中后期主导	早期项目占比高

● 投资偏好差异

美国投资特点：

- 重押基础模型和AI基础设施
- 「赢家通吃」心态明显
- 关注AGI和超级智能方向
- 企业级SaaS和B2B应用受青睐

中国投资特点：

- 「脱虚向实」,偏好可落地应用
- 「投早投小」,发掘早期创新
- 具身智能成为差异化突破口
- 垂直行业（医疗、制造、汽车）深耕

五、热门赛道深度分析

● 生成式AI

- 2025上半年融资超\$492亿美元
- 基础模型公司融资\$800亿,占AI融资40%
- OpenAI、Anthropic、xAI三巨头占据主要份额
- 从模型竞赛转向应用落地

● AI Agent

- 2025上半年融资\$28亿美元,全年预计\$67亿美元
- 种子轮公司吸引约\$7亿融资,超200笔交易
- 应用场景: 客户服务、软件开发辅助、工作流自动化

● 具身智能

- 中国2025年前5个月融资232亿元
- 政府工作报告首次列入未来产业培育清单
- 人形机器人: 2026年预计交付10-20万台
- IDC预测2026年中国市场用户支出超\$110亿

● AI基础设施

- 科技巨头2025年合计CapEx超\$3000亿
- 全球生成式AI支出预计达\$6440亿,其中硬件\$3980亿
- Stargate: \$5000亿AI数据中心计划

六、重点垂直行业AI商业化与投融资分析

● 医疗AI: 商业化最具确定性的赛道

- 2025上半年融资超\$492亿美元
- 基础模型公司融资\$800亿,占AI融资40%
- OpenAI、Anthropic、xAI三巨头占据主要份额
- 从模型竞赛转向应用落地

全球医疗AI投融资概况

2025年医疗AI成为AI投资中最活跃的垂直领域之一:

指标	数据
总投资额	超\$180亿美元, 占医疗总投资的46%
数字医疗融资	\$142亿美元, 同比增长35%, 创2022年以来新高
AI医疗占比	AI驱动的数字医疗初创公司获得54%的融资, 较2024年37%大幅提升
交易溢价	AI医疗公司平均交易额比非AI公司高19%
新增独角兽	6家医疗AI公司晋升独角兽

● 重大融资案例：

- Truveda: \$3.2亿美元，构建药物发现数据库
- Aidoc Medical: \$1.5亿美元，加速AI诊断
- Harrison.ai: \$1.12亿美元，AI放射和病理诊断

● AI药物发现

- 2025年AI药物发现融资达\$33亿美元，同比增长27%
- 市场规模预计从2025年\$199亿增长至2034年\$1339亿 (CAGR 23.22%)
- AI设计药物Phase I成功率80-90%，远超传统药物40-65%
- 可降低药物研发成本40%，大幅压缩研发周期

中国AI医疗市场

指标	数据
市场规模	2023年88亿元 → 2033年3157亿元 (CAGR 43.1%)
AI药物研发	未来市场空间有望突破50亿元
政策支持	工信部AI医疗器械创新任务182家单位入选
政策支持	北京计划2027年医疗机构普遍开展AI应用

代表企业：百川智能战略聚焦AI医疗，发布Baichuan-M3医疗大模型

● 金融AI：大模型规模化落地元年

全球金融科技AI投融资

指标	数据
FinTech总融资	\$527亿美元 (5918笔交易)，同比增长27%
AI金融市场规模	2025年\$176亿 → 2034年\$977亿 (CAGR 19.9%)
另一估算	2025年\$300亿 → 2030年\$831亿
美国FinTech	\$251亿美元 (2449笔交易)，全球Top 20中占14席
银行AI采纳率	75%的大型银行 (\$1000亿+资产) 已整合AI

金融AI核心应用场景

- 运营效率：自动化处理、复杂流程优化、决策支持
- 风控与反欺诈：信用评估、贷款违约预测、异常交易检测
- 客户体验：7×24小时智能客服、个性化产品推荐
- 合规监管：自动监控交易、异常检测、模型风险管理
- 保险定价：基于驾驶记录、健康史的个性化定价

中国金融AI市场

2025年被视为中国金融业“大模型规模化落地元年”：

指标	数据
大模型采购项目	587个，涉及金额15.06亿元
银行业占比	项目数量占近50%，采购金额占75.2%
投资转向	从算力类转向应用类（智能体、场景解决方案）
发展重心	从“拥有大模型”转向“用好大模型”
银行AI采纳率	75%的大型银行（\$1000亿+资产）已整合AI

驱动因素：

- 政策支持
- DeepSeek等开源大模型崛起降低门槛
- 市场竞争压力
- 金融消费者对智能服务的期待

● 法律AI：资本布局的新蓝海

全球法律AI投融资

2025年法律科技经历了前所未有的投资热潮：

指标	数据
法律科技总融资	近\$60亿美元
AI法律工具融资	\$43亿美元，同比增长54%（356笔交易）
AI占比	法律科技投资的70%流向AI工具
律所科技投入	技术和知识管理工具支出分别增长9.7%和10.5%
企业法务期望	67%的企业法务期望律所使用前沿技术

重大融资案例：

- Harvey：累计融资超\$8亿美元，法律AI领头羊
- Clio：\$5亿美元单轮融资 + \$10亿收购vLex
- Filevine：两轮共\$4亿美元
- Legora（瑞典）：€7060万欧元B轮
- SpotDraft：\$5400万美元，合同管理平台

法律AI核心应用：

- 合同审查与分析
- 法律研究辅助
- 文档起草与生成
- 合规与风险管理

中国法律AI市场

中国法律科技仍处于技术探索和模式验证阶段:

- 36家代表性企业中55.6%处于天使轮至A+轮
- 正从“互联网+法律”向“人工智能+法律”转变
- 热点：智能合同管理、AI法律工具、智慧司法、电子签名
- 代表企业：秘塔AI等AI原生企业已获显著融资
- 应用场景：咨询、数据检索、合同草拟与审查

垂直行业AI投资机会总结

行业	全球融资规模	中国市场特点
医疗AI	\$180亿+	商业化明确，政策支持强
金融AI	\$527亿 (FinTech)	大模型落地元年，银行主导
法律AI	\$43亿 (增长54%)	早期阶段，空间大

给开发者的垂直行业建议：

1. 医疗AI：门槛较高，需医疗领域知识+AI能力，但商业化路径清晰
2. 金融AI：银行和保险公司需求旺盛，合规要求严格
3. 法律AI：起步阶段，竞争相对较小，适合创业切入

数据来源说明

本白皮书数据主要来源于以下机构的公开报告和数据：

国际研究机构：

- CB Insights – State of AI 2025系列报告
- Crunchbase – AI融资数据追踪
- IDC – 全球AI市场预测
- Gartner – 技术成熟度曲线
- PitchBook – 风险投资数据

中国研究机构：

- IT桔子 – 创投数据库
- 艾瑞咨询 – 行业研究报告
- 36氪研究院 – 新经济研究
- 亿欧智库 – 产业互联网研究
- 胡润研究院 – 独角兽榜单

AI时代全新职业形态发展

2025年标志着生成式AI从技术探索迈向产业落地的关键转折点。本白皮书通过对150余份行业报告、300+企业部署案例以及多个权威研究机构数据的深度分析,系统性地揭示了AI时代三种最具代表性的新兴职业形态:前线部署工程师(FDE)、一人公司(超级个体)以及上下文工程师。

● 核心观点速览

2025年被业界视为生成式AI应用爆发的元年。在这一年, AI技术从实验室的技术狂欢正式迈向产业落地的深水区, 职业市场随之经历了深刻的结构性重塑。本白皮书聚焦三种极具代表性的新兴职业形态, 它们分别代表了企业级落地、个人赋能和人机交互三个维度的核心变革:

职业形态	核心价值主张	关键市场数据
FDE (前线部署工程师)	解决企业AI落地的“最后一公里”难题, 将模型能力转化为商业价值	募集\$150亿+新基金, 投资165+项目
OPC (One-Person Company)	AI智能体赋能的个人商业闭环, “一人成军”	预测AI应用将成主战场
Sequoia Capital	AI系统的信息架构师, 构建动态上下文生态	预测AI应用将成主战场

一、职业全景: AI落地困境催生新职业

● 95%的失败率: 企业AI落地的残酷现实

麻省理工学院NANDA研究计划于2025年发布的重磅报告《The GenAI Divide: State of AI in Business 2025》揭示了一个令人警醒的现实: 高达95%的企业级生成式AI试点项目未能产生可衡量的财务回报或利润影响。

这一结论并非空穴来风。研究团队通过三重验证方法得出上述结论:

- 对300+项公开披露的企业AI项目进行系统性审查
- 与52家组织的代表进行深度访谈
- 收集153位高管的问卷调研数据

报告指出, 失败的根源往往不在于AI模型本身的能力缺陷, 而是组织层面的整合失败。具体而言, 企业面临以

- 流程整合缺失: AI系统与日常业务运营脱节
- 学习能力断层: 系统无法保留反馈或适应 workflows
- 信任建立困难: 高管对“黑盒”式AI决策持谨慎态度
- “影子AI”现象: 员工绕过官方系统私自使用ChatGPT等工具

● 从技术供给到需求侧变革

2024年被称为“模型之年”，以GPT-4、Claude 3、Gemini为代表的大模型能力飞速提升；2025年则是“智能体之年”，AI Agent从概念走向应用。然而，技术供给侧的狂飙突进并未自动转化为企业端的价值落地。

正是这一“落地鸿沟”催生了新的职业形态。企业不再仅仅需要开发模型的研究科学家，更需要能够将模型嵌入业务流程、建立用户信任、实现规模化部署的复合型人才。这一需求直接推动了FDE、OPC模式和上下文工程师的崛起。

二、FDE（前线部署工程师）：企业AI落地的“特种部队”

● 职业定义与起源

FDE (Forward Deployed Engineer, 前线部署工程师) 这一概念最早源于大数据独角兽Palantir。Palantir将FDE定义为“既会写代码，又能深入客户现场解决问题”的复合型人才。与坐在总部写代码的传统工程师不同，FDE驻扎在客户现场，与终端用户并肩工作。

Palantir内部将FDE团队进一步细分为两类：

- Echo团队（嵌入式分析师）：拥有深厚领域知识，负责识别业务痛点并提出10倍级改进方案。这类人才需要“叛逆者”精神，能够看穿现有流程的不足。
- Delta团队（部署工程师）：擅长快速原型设计，追求“解决问题的速度”而非完美架构。他们甚至会编写“用后即焚”的代码来验证假设。
- FDE的核心使命可以用一句话概括：“吃掉客户落地的痛苦，产出可用的产品” (Eat pain and excrete product)。这意味着FDE不仅是技术交付者，更是商业价值的翻译官

● 核心能力模型

基于多家头部公司的岗位要求分析，FDE需要具备以下核心能力组合：

能力维度	核心价值主张
技术基础	全栈开发能力（API集成、前端原型、DevOps）；熟悉LangChain、向量数据库、OpenAI API等AI技术栈
业务理解	快速萃取领域知识（Domain Knowledge）；理解客户业务流程中的“流程滤光片”
沟通协作	与非技术高管有效沟通；建立用户信任；推动组织变革
问题解决	快速原型验证（Workaround）；在不确定性中导航；“碎石路”思维
产品思维	从现场经验中提炼可复用的产品功能；避免沦为纯外包劳动力

● 市场需求与薪资

FDE岗位正经历爆发式增长。根据Indeed Hiring Lab数据，2025年1月至9月间，FDE相关职位数量暴增超过800%。部分行业分析甚至显示同比增长长达1165%，成为AI领域增速最快的职位类型之一。

科技巨头正在激进扩招：

- OpenAI：2024年初仅有2名FDE，目前团队已扩充至约39人，计划年底达到52人
- Anthropic：宣布将应用AI团队规模扩大5倍
- Palantir：持续将FDE作为核心商业模式

薪资方面，FDE通常能获得极具竞争力的报酬。由于他们直接参与解决高价值商业问题（如为企业节省数千万美元成本），薪资结构往往包含与最终业务成果挂钩的绩效激励。

● 典型案例：OpenAI与摩根士丹利

OpenAI FDE团队与摩根士丹利的合作堪称经典案例。技术层面，团队仅用数周即完成了AI系统原型。然而，信任建立才是真正的挑战——FDE团队额外花费4个月时间进行试点和数据标注工作，直到财务顾问们完全信任AI的建议并愿意在实战中使用。

这一案例揭示了FDE的核心价值：

- 构建“确定性护栏”（Deterministic Guardrails），让概率性的AI变得可预测、可验证
- 采用“评估驱动开发”（Eval-driven Development）框架，为用户提供验证AI结果的工具
- 通过长期试点将“Demo”转化为“生产级系统”

三、OPC（One-Person Company）：AI智能体赋能的“超级个体”

● 组织结构的范式转变

2025年AI智能体（AI Agent）的爆发正在重塑传统组织架构。传统企业采用金字塔型（Pyramid）结构，权力自上而下层层传递。而OPC（One-Person Company，一人公司）模式则采用蜘蛛网型（Hub & Spoke）结构：

- 单一决策核心（Hub）：个体是唯一的CEO和战略决策者，处于网络中心
- 数字节点（Spoke）：周围连接无数AI智能体、SaaS工具和外部合作伙伴
- 零损耗执行：指挥官意图直接转化为指令并被100%执行
- 极高弹性：有项目时火力全开，项目结束后成本归零

这一模式将个体从“出卖时间”解锁为“指挥系统”。传统个体户的营收与工作时间成正比，而OPC的营收与人员规模不再挂钩。

● 运作形态与工具生态

这一案例揭示了FDE的核心价值：

- 开发者规模：平台已聚集超过百万活跃开发者
- 智能体数量：发布超过200万个AI智能体
- 可视化搭建：通过“拖、拉、拽”即可构建复杂 workflow
- 多渠道发布：支持一键发布至小程序、API、社区等多种形态

一个典型的OPC工作日场景：

上午9:00 — 策略智能体在10秒内生成一周推广方案

上午10:00 — 绘画智能体（如Midjourney）10分钟产出50张高清图海报

上午11:00 — 视频智能体一键合成配音视频

下午 — RPA工具自动分发内容，客服智能体24小时在线接待

结果：半天完成传统团队一周的工作量

OpenAI《企业AI现状报告》数据显示：

- 结构化 workflow（如Projects）使用量同比增长19倍
- “重度AI用户”每周节省超过10小时
- 前沿员工的消息发送量是普通员工中位数的6倍

● 市场需求与薪资

OPC模式在以下领域展现出强大的变现潜力：

轻资产数字化交付

- 全链路品牌运营：策划→设计→视频→客服全流程自动化
- 24/7智能电商：从流量获取到客户转化的完整闭环

开发者社区与模板

- 智能体商店变现：发布解决特定痛点的Bot获取流量和付费收益
- 爆款模板复制：将个人业务逻辑（SOP）转化为可销售的数字资产

垂直领域专业顾问

- 细分行业AI微调服务
- 轻量化FDE角色：快速搭建企业定制原型

● 边界与局限性

OPC模式在以下领域展现出强大的变现潜力：

- 重资产运营：物流、制造等需要实体资源投入的行业
- 复杂信任建立：大宗交易、B2B销售等需要人与人深度信任的场景
- 战略责任承担：AI可以给出100个方案，但不敢拍板——因为它无需承担后果

OPC的天花板最终取决于人的认知和判断力。真正的超级个体懂得将非核心业务外包给AI，而将建立深度信任的核心业务留给自己。

四、上下文工程师（Context Engineer）：AI系统的“信息架构师”

● 本白皮书数据基于以下权威来源的公开数据：

2025年中期，AI领域发生了一场重要的概念演进：「提示词工程」（Prompt Engineering）正在被「上下文工程」（Context Engineering）所取代。这一转变由两位科技领袖推动并定义：

Shopify CEO Tobi Lütke 将上下文工程定义为：

“提供所有上下文信息，使任务对LLM来说可以合理解决的艺术。”

前特斯拉AI主管 Andrej Karpathy 进一步阐述：

“上下文工程是用恰到好处信息填充上下文窗口的精妙艺术与科学。上下文窗口就像LLM的工作内存

与提示词工程的核心区别：

职业形态	提示词工程	上下文工程
关注范围	单次输入-输出交互	模型所见的全部信息生态
核心思维	如何写出好提示词	如何设计整个信息架构
系统性	单点技巧	预动态系统设计
系统性	获得好回答	构建可靠、可扩展的AI系统

Gartner将上下文工程定义为：“设计和构建相关数据、工作流和环境，使AI系统能够理解意图、做出更好决策并交付符合企业目标的上下文输出——无需依赖手动提示。”

● 核心能力模型

上下文工程师需要管理AI模型理解所需的多种信息类型：

系统组件（System Components）

- 初始行为准则、规则和示例
- 定义模型行为的系统提示

记忆系统（Memory Systems）

- 短期记忆：对话历史
- 长期记忆：跨会话的持久知识、用户偏好

用户消息（User Messages）

- 即时任务或问题
- 用户意图的精准捕捉

检索信息（Retrieved Information）

- 通过RAG（检索增强生成）从文档、数据库、API获取的外部知识
- 确保信息的实时性和准确性

可用工具 (Available Tools)

- 模型可调用的函数或能力定义
- 与外部世界交互的接口

结构化输出 (Structured Output)

- 响应格式规范
- 输出质量控制

● 市场需求与薪资水平

上下文工程师正在成为AI团队中不可或缺的角色，其重要性可比肩机器学习团队中的数据工程师。

薪资水平 (2025年美国市场)：

- 入门级 (0-2年)：\$63,000 – \$95,000
- 中级 (2-5年)：\$120,000 – \$180,000
- 高级 (5年以上)：\$180,000 – \$275,000+
- 头部AI公司 (如Anthropic)：可达\$320,000 – \$405,000

地域溢价：

- 旧金山湾区：+40-50%
- 金融/医疗/法律等垂直领域：+10-20%
- 远程岗位：+10-20%

Gartner建议企业将上下文工程作为战略优先级，任命专门的负责人或团队，并将其与现有AI工程和治理结构整合。预测未来将出现“上下文架构师”(Context Architect)这一新角色，其重要性将与数据工程师在机器学习团队中的地位相当。

五、2026年趋势预判

● 关键词：优化与微调

如果说2024年是“模型之年”、2025年是“智能体之年”，那么2026年极可能成为“优化之年”或“微调回归之年”。通用的AI Agent将无法满足不同需求。企业将逐渐意识到，真正产生差异化竞争优势的不是使用相同的通用模型，而是能够通过微调 (Fine-tuning) 让模型胜任特定领域难题的能力

职业影响：

- 需求从“会用AI”升级为“会定制AI”
- 芯片设计、药物研发、金融建模等专业领域将出现大量微调专家岗位
- 数据标注和评估体系设计能力价值凸显

● FDE标准化：新一代DevOps

FDE将从“特种部队”角色逐渐演变为标准化岗位，成为企业AI战略的核心组成部分——类似于DevOps在软件开发领域的地位演进。

趋势预测：

- 低代码AI平台 (如JitAI) 普及将增强FDE的原型验证能力
- FDE将推动AI从“试点”走向“规模化生产”
- 未来FDE将更多扮演数据科学、工程和业务之间的深度协作桥梁

Gartner报告显示，72%的企业在AI试点后难以实现规模化，这一缺口正是FDE标准化的核心驱动力。

技能鸿沟扩大

OpenAI《企业AI现状报告》揭示了一个重要趋势：“前沿”员工（Top 5%）与普通员工的AI使用效率差距正在迅速拉大。

量化数据：

- 前沿员工的消息发送量是中位数员工的6倍
- 前沿员工能够完成此前无法胜任的新任务类型
- 采用结构化 workflow（Structured Workflows）的比例远高于普通用户

展望：

2026年的职场将不再以“懂不懂AI”划分，而是以“AI使用深度”分层。掌握结构化 workflow、能构建自动化流程的员工，将获得巨大的职业杠杆。



上游基础层

芯片硬件



云计算基础设施



数据服务



中游技术层

基础模型/大模型



多模态 AI



算法框架



AI开发平台



垂直应用

办公协作类



图像生成



音频音乐



客户服务



教育培训类



金融科技类



制造工业类



游戏 AI



编程开发



视频制作



广告投放



医疗健康类



电商零售类



法律



其他



通用 AI

通用 AI 助手



AI Agent



行业	解决方案	作用
通用	tensorflow/tensorflow	深度学习训练与推理, 生产部署
	huggingface/transformers	预训练模型调用与微调, NLP/多模态任务
	pytorch/pytorch	深度学习研究与工程训练, 自定义模型开发
	openai/whisper	语音转文本, 多语言ASR
	opencv/opencv	视觉基础能力, 图像/视频处理
	scikit-learn/scikit-learn	传统机器学习建模, 特征工程
	keras-team/keras	高层训练接口, 快速搭建神经网络
	ggerganov/llama.cpp	本地推理引擎, 边缘端部署
	pandas-dev/pandas	数据清洗与分析, 训练数据准备
	ggml-org/whisper.cpp	本地ASR推理, 端侧语音识别
	apache/spark	大规模数据处理, ML流水线
	deepspeedai/DeepSpeed	分布式训练加速, 训练成本优化
	gradio-app/gradio	快速做模型Demo, Web交互接口
	ray-project/ray	分布式训练/推理, 任务调度
	openai/gym	强化学习环境接口, 仿真基准
	vllm-project/vllmLLM	推理服务, 高吞吐Serving
	google/jax	自动微分数值计算, 高性能训练
	apache/airflow	数据/模型 workflow 编排, 定时调度
	streamlit/streamlit	数据/AI应用展示, 交互式面板
	facebookresearch/detectron2	目标检测/分割训练, CV基准
	explosion/spaCyNLP	流水线, 分词/NER/文本分类
	open-mmlab/mmdetection	检测算法开发与训练, CV框架
	numpy/numpy	数值计算底座, 线性代数
	milvus-io/milvus	向量数据库, 相似度检索
	Lightning-AI/pytorch-lightning	训练工程化, 训练循环抽象
	huggingface/diffusers	扩散生成库, 图像生成
	fastai/fastai	高层训练库, 快速实验
	apache/kafka	实时数据流, 特征/日志管道实时数据流, 特征/日志管道
	microsoft/semantic-kernel	LLM/Agent编排SDK, 多语言
	dmlc/xgboost	结构化数据建模, GBDT
	deepset-ai/haystack	RAG/问答/搜索管线, 生产应用
	qdrant/qdrant	向量数据库, 相似度检索
	mlflow/mlflow	实验追踪与模型管理, MLOps
onnx/onnx	模型交换与推理格式, 跨框架部署	
microsoft/LightGBM	GBDT高性能训练, 结构化数据	
huggingface/peft	参数高效微调, LoRA等	
open-mmlab/msegmentation	语义分割训练, CV框架	

行业	解决方案	作用
通用	chroma-core/chroma	向量存储, RAG原型
	kubeflow/kubeflow	K8s上的ML平台, Pipeline/训练
	alumentations-team/alumentations	数据增强, 视觉训练数据处理
	dmlc/dgl	图神经网络框架, 图学习
	dask/dask	并行计算, 大数据分析
	DLR-RM/stable-baselines3	强化学习算法库, RL训练
	weaviate/weaviate	向量数据库, 检索与RAG
	huggingface/accelerate	分布式训练工具, 训练加速
	openai/spinningup	强化学习教程代码, 算法实现
	autogluon/autogluon	AutoML, 快速训练与集成
	catboost/catboost	GBDT, 类别特征友好
	apache/beam	批流一体数据管道, Dataflow
	bentoml/BentoML	模型服务化, API/部署
	制造 工业 能源	ultralytics/yolov5
ultralytics/ultralytics		ultralytics/ultralytics视觉质检; 缺陷检测
open-mmlab/mmdetection		工业视觉检测; 质检算法工程化
facebook/prophet		能源预测; 电力负荷预测; 需求预测
open-mmlab/mmdetection		表面缺陷分割; 视觉质检
yzhao062/pyod		预测性维护; 异常检测; 设备健康监测
openvino/openvino		工业质检推理; 边缘推理加速
unit8co/darts		能源预测; 设备故障预测; 工业时序
sktime/sktime		时序预测; 设备监测管线
awsml/gluon-ts		能源负荷预测; 需求预测
facebookresearch/Kats		能耗异常检测; 变点检测; 预测
sktime/pytorch-forecasting		能源预测; 预测性维护; 时序深度模型
Nixtla/statsforecast		负荷预测; 需求预测; 快速基线
salesforce/Merlion		设备异常检测; 运维监测平台化
Nixtla/neuralforecast		能源预测; 负荷预测; 工业时序
openvino/open_model_zoo		质检/检测模型库; 工业落地参考
numenta/NAB		预测性维护评测; 异常报警基准
ARISE-Initiative/robosuite		工业机器人仿真; 装配/抓取任务
uber/orbit		能源/需求预测; 不确定性区间
ARISE-Initiative/robomimic		工业机器人示教学习; 策略训练
amazon-science/patchcore-inspection		工业视觉异常检测; 无监督质检
khundman/teleanom		设备故障预警; 多变量异常检测
ai4co/r14co		生产排产; 产线/路径优化
umbertogriffo/ Predictive-Maintenance-using-LSTM		预测性维护; RUL; 设备退化预测

行业	解决方案	作用
制造 工业 能源	pyaf/load_forecasting	电力负荷预测；能源预测
	Azure/Istms_for_predictive_maintenance	预测性维护；剩余寿命预测；设备监测
	awslabs/ predictive-maintenance-using-machine-learning	预测性维护；故障预测；工业IoT
	jolibrain/wheatley	生产调度；工厂任务分配
	ARISE-Initiative/robosuite-benchmark	工业机器人基准评测
	ARISE-Initiative/robosuite-notebooks	机器人学习上手实验
金融与保险	OpenBB-finance/OpenBB	金融数据聚合；投研工作台；量化研究
	freqtrade/freqtrade	自动化交易；策略回测；交易执行
	microsoft/qlib	量化投研平台；特征工程；模型训练/回测
	TauricResearch/TradingAgents	多智能体交易研究；交易决策；工具化投研
	AI4Finance-Foundation/FinGPT	金融大模型；研报/新闻理解；金融NLP
	OpenBB-finance/openbb-cli	命令行投研终端；数据拉取；量化分析
	polakowo/vectorbt	高性能回测引擎；研究；组合实验
	AI4Finance-Foundation/FinRobot	金融Agent；自动化投研；工具链编排
	hudson-and-thames/mlfinlab	金融机器学习工具箱；特征工程；可复现实验
	AI4Finance-Foundation/ElegantRL	强化学习库；可用于交易/投资RL
	AI4Finance-Foundation/FinRL-Trading	交易训练脚手架；数据对接；策略实验
	OpenBB-finance/openbb-ai	投研AI助手；自然语言查询；研究自动化
	AI4Finance-Foundation/FinRL-Meta	市场环境与基准；数据管线；RL Benchmark
	OpenBB-finance/openbb-api	投研能力API化；为应用/Agent提供服务
	hudson-and-thames/arbitragelab	统计套利；均值回归组合；量化研究
	hudson-and-thames/backtest_tutorial	回测入门；实验模板；研究示例
	hudson-and-thames/arbitrage_research	套利研究示例；Notebook实验
	hudson-and-thames/ guide_to_modern_portfolio_optimization	组合优化；资产配置方法
	hudson-and-thames/ definitive_guide_to_pairs_trading	配对交易；统计套利策略
零售与电商	Milvus	商品语义搜索、相似商品召回、向量检索
	Weaviate	电商搜索/推荐向量库、语义检索、召回
	DeepCTRCTR	预估、推荐排序、广告点击率建模
	GluonTS	销量/需求预测、补货预测、时间序列建模
	Implicit	协同过滤、商品推荐（隐式反馈）
	TensorFlow Recommenders	推荐系统训练/评估/部署流程
	qdrant-client	向量检索接入、商品语义检索/推荐（客户端）
	HugeCTRCTR	大规模训练、广告/推荐排序训练加速
医疗健康	google-deepmind/alphafold	蛋白结构预测/生物医学研究
	MIC-DKFZ/nnUNet	医学影像分割SOTA/自动配置
	Project-MONAI/MONAI	医疗影像AI开发框架/训练推理
	google-deepmind/alphafold3	蛋白/生物分子结构预测(新一代)

行业	解决方案	作用	
医疗健康	google-deepmind/alphafold	蛋白结构预测/生物医学研究	
	MIC-DKFZ/nnUNet	医学影像分割SOTA/自动配置	
	Project-MONAI/MONAI	医疗影像AI开发框架/训练推理	
	google-deepmind/alphafold3	蛋白/生物分子结构预测(新一代)	
	deepchem/deepchem	药物发现/分子性质预测工具箱	
	facebookresearch/esm	蛋白语言模型/表征学习	
	Beckschen/TransUNet	医学影像分割(Transformer+UNet)	
	Project-MONAI/tutorialsMONAI	教程与工作流示例	
	allenai/scispaCy	生物医学文本NLP流水线	
	sunlabuiuc/PyHealth	临床预测建模/EHR建模工具箱	
	OpenGVLab/SAM-Med2D	医学图像分割(基于SAM)	
	mlmed/torchxrayvision	胸片/放射影像模型与基线	
	Project-MONAI/MONAILabe	智能标注/主动学习/影像标注平台	
	CogStack/MedCATEHR	概念标注/术语链接(UMLS/SNOMED)	
	bowang-lab/MedSAM2	医学影像分割(升级版MedSAM)	
	Project-MONAI/model-zooMONAI Bundle	模型仓库/复用	
	Project-MONAI/monai-deploy-app-sdk	医疗影像AI应用部署SDK	
	Mauville/MedCLIP	医学影像Caption/CLIP变体	
	VectorInstitute/cyclops	医疗AI模型评估与监控工具箱	
	Project-MONAI/monai-bootcamp	医疗影像AI训练营资料	
	Kaktucs/Symptom-Checker	症状自查/分诊问答原型	
	教育与培训	JushBJJ/Mr.-Ranedeer-AI-Tutor	AI家教 / 学习陪伴 / 个性化辅导
		d2l-ai/d2l-en	深度学习教材+代码 / 课程训练
huggingface/agents-course		AI代理课程 / 工具使用教学	
karpathy/minGPT		最小GPT实现 / 教学用代码	
karpathy/nn-zero-to-hero		神经网络课程Notebook / 从0到1学习	
karpathy/micrograd		反向传播 / 自动求导教学	
huggingface/course		NLP/LLM课程 / 入门到实战	
fastai/course22		课程笔记与实践 / 教学材料	
fastai/fastbook fastbook		配套代码 / 教学	
karpathy/build-nanogpt		从零实现GPT课程 / 讲解+代码	
karpathy/makemore		生成模型教学 / 字符级LM练习	
huggingface/computer-vision-course		视觉课程 / 从CNN到ViT	
theJayTea/WritingTools		系统级写作助手 / 改写润色	
patil-suraj/question_generation		题目生成 / 阅读理解出题	
grammarly/gector		语法纠错 / 作文反馈	
asahi417/lm-question-generation		多语种题目生成 / 数据与模型	

行业	解决方案	作用	
教育与培训	pykt-team/pykt-toolkit	自适应学习评估 / 知识追踪基准	
	hcnoh/ knowledge-tracing-collection-pytorch	知识追踪模型合集 / 预测掌握度	
	CAHLR/OATutor	自适应辅导系统 / 技能掌握度推断	
	GeminiLight/awesome-ai-llm4education	教育领域AI/LLM资源索引	
	pydaxing/ Deep-Knowledge-Tracing-DKT-Pytorch	个性化练习 / 掌握度预测	
	bigdata-ustc/EduCAT	自适应测评 / 题库与测评系统	
	cofe-ai/fast-gector	语法纠错加速版 / 批改效率	
	yxonic/DTransformer	知识追踪Transformer / 诊断式KT	
	dxywill/deepknowledgetracing	知识追踪基线实现	
	bigdata-ustc/EduSim	教育推荐仿真环境 / 离线评估	
	Chang-Chia-Chi/ SaintPlus-Knowledge-Tracing-Pytorch	知识追踪Transformer (SAINT+)	
	WarmingBee/ai-flashcard	知识卡片 / 闪卡生成	
	内容媒体 文娱 广告	AUTOMATIC1111/stable-diffusion-webui	文生图、图生图、批量出图工作台
		comfyanonymous/ComfyUI	节点式生成 workflow、批处理、自动化出图
llyasviel/Fooocus		一键高质量出图、创意图片生成	
coqui-ai/TTS		配音 TTS、多说话人训练、语音合成服务	
upscayl/upscayl		图片超分放大、清晰增强、素材修复	
suno-ai/bark		文本转语音、旁白配音、音效生成	
TencentARC/GFPGAN		人脸修复、老照片增强、人像清晰化	
myshell-ai/OpenVoice		语音克隆、跨语言音色迁移、配音资产化	
RVC-Project/ Retrieval-based-Voice-Conversion-WebUI		变声、翻唱、音色迁移 WebUI	
xinntao/Real-ESRGAN		通用图像增强、超分修复、去噪去模糊	
huggingface/diffusers		生成式媒体推理与训练组件、管线搭建	
svc-develop-team/so-vits-svc		歌声转换、音色转换、翻唱变声	
facebookresearch/audiocraft		音乐生成、音频生成、创作者原型工具	
Sanster/IOPaint		智能抹除、修补、局部重绘	
invoke-ai/Invoke		AI生成创作工作台、队列批量生成、可控编辑	
m-bain/whisperX		字幕时间轴对齐、说话人分离、转写增强	
Mikubill/sd-webui-controlnet		姿态线稿深度等条件控制、可控出图	
sczhou/CodeFormer		人像修复、面部清晰化、低质视频修复	
languagetool-org/languagetool		写作纠错、语法检查、风格建议	
OpenTalker/SadTalker		图片说话、口型驱动、口播视频生成	
ltdrdata/ComfyUI-Manager		ComfyUI 插件管理、模型与节点安装	
Rudrabha/Wav2Lip		视频对口型、配音同步、口型修复	
guoyww/AnimateDiff		文生视频、动图生成、镜头运动控制	
bmaltais/kohya_ssLoRA	训练、扩散模型微调、训练工具链		
rhasspy/piper	本地离线配音、端侧 TTS、轻量部署		

行业	解决方案	作用
内容媒体 文娱 广告	facebookresearch/demucs	人声伴奏分离、混音拆分、音频清洗
	MaartenGr/BERTopic	选题分析、话题聚类、社媒趋势洞察
	google/lightweight_mmm	投放归因、预算分配、效果评估
政务 法务	Unstructured-IO/unstructured	公文/合同/卷宗等非结构化文档解析、分块、结构化输出
	microsoft/presidio	隐私合规：PII识别、脱敏/匿名化
	mindee/doctr	扫描件/证件/表单OCR，用于政务材料数字化
	Layout-Parser/layout-parser	版面分析与结构化抽取 (PDF/图像中的段落、表格、字段)
	kermitt2/grobid	PDF结构化 (标题/段落/引用) 用于法律/政策文档处理
	jphall663/awesome-machine-learning-interpretability	可解释AI与合规治理资源库 (含隐私/政策相关内容)
	pbiecek/xai_resources	XAI/可解释AI资源，用于监管/审计AI系统
	exPredict/lexpredict-lexnlp	合同/法规文本处理：条款、引用、实体抽取
	openai/model_spec	AI治理：模型行为规范与安全策略文档 (政策/合规视角)
	ICLRandD/Blackstone	法律文本NLP：英国判例/文书的实体识别、分类
	chawins/llm-sp	安全与隐私研究资料库，覆盖合规与攻防
	thunlp/CAIL	法律判决预测/法条适用任务基准与代码 (司法辅助)
	TheAtticusProject/cuad	合同审查：合同条款标注数据集与基线
	LeapBeyond/scrubadub	PII清洗与脱敏 (用于政务/法务文本)
	maastrichtlawtech/awesome-legal-nlpLegalNLP	LegalNLP资源导航 (检索/抽取/判决预测等)
	openlegaldata/awesome-legal-data	法律数据集与工具资源库 (法律文本处理)
	thunlp/LegalPLMs	法律领域预训练模型与任务合集 (落地到检索/分类/问答)
	thunlp/TopJudge	司法辅助：罪名/法条/刑期预测模型
	neo4j-product-examples/graphrag-contract-review	合同审查Agent (结合图谱检索)
	china-ai-law-challenge/CAIL2018	判决预测赛题数据与baseline (司法AI)
	PolarisRisingWar/LJP_Collection	法律判决预测模型复现/整理
	PolarisRisingWar/pytorch_ljp	法律判决预测 (多数据集/多模型实现)
	hunlp/QAJudge	司法问答/裁判辅助研究代码
	jiangyanjie/Deep-Learning-Based-Software-Engineering	隐私政策/GDPR合规研究资料与代码索引
	HooRin/AuditWen	审计/合规方向的领域模型与数据 (偏模型)

AI 开源解决方案榜聚焦“可落地的 AI 行业方案”，而非通用大模型本身。榜单从 GitHub 中筛选 真正解决具体行业或业务问题的 AI 项目，并按照 Star 数进行排名数据由AIGCLink研究院监控

排名	项目名称	行业	当前	上月	月增长 Star	增长率 (%)	当前	上月	核心用处关键词
1	ggml-org/llama.cpp	通用	92992	55400	37592	67.86	14477	7900	本地推理引擎, 边缘端部署
2	vllm-project/vllm	通用	67545	34400	33145	96.35	12600	4600	本地LLM推理服务, 高吞吐Serving
3	fastai/fastbook	教育与培训	24341	4800	19541	407.1	9361	3700	fastbook 配套代码 / 教学
4	milvus-io/milvus	通用	42240	31000	11240	36.26	3761	2900	向量数据库, 相似度检索
5	milvus-io/milvus	通用	42240	31000	11240	36.26	3761	2900	向量数据库, 相似度检索
6	chroma-core/chroma	通用	25444	15400	10064	65.35	2000	1300	向量存储, RAG原型
7	apache/airflow	通用	43851	34200	9651	28.22	16275	15800	数据/模型工作流编排, 定时调度
8	streamlit/streamlit	通用	43069	34100	8969	26.3	4027	2900	数据/AI应用展示, 交互式面板
9	qdrant/qdrant	通用	28199	21700	6499	29.95	1989	1400	向量数据库, 相似度检索
10	opencv/opencv	通用	85716	80500	5216	6.48	56481	50000	视觉基础能力, 图像/视频处理
11	weaviate/weaviate	通用	15404	10200	5204	51.02	1177	660	向量数据库, 检索与RAG
12	weaviate/weaviate	通用	15404	10200	5204	51.02	1177	660	向量数据库, 检索与RAG
13	apache/kafka	通用	31708	27300	4408	16.15	14896	14000	实时数据流, 特征/日志管道
14	invoke-ai/InvokeAI	内容媒体 / 文娱 / 广告	26547	22300	4247	19.04	5157	2500	生成创作工作台, 队列批量生成, 可控编辑
15	miflow/miflow	通用	23685	20100	3585	17.84	5157	4400	AI推理引擎, 批处理, ML Ops
16	huggingface/peft	通用	20461	16900	3561	21.07	2150	1900	参数高效微调, LoRA等
17	huggingface/diffusers	通用	32461	31100	3461	11.93	6683	6700	扩散生成库, 图像生成
18	huggingface/diffusers	通用	32461	29000	3461	11.93	6683	6700	扩散生成库, 图像生成
19	Comfy-Org/ComfyUI	内容媒体 / 文娱 / 广告	100235	96900	3335	3.44	11367	11000	节点式生成工作流, 批处理, 自动化出图
20	OpenBB-finance/OpenBB	金融与保险	58638	55441	3197	5.77	5691	5376	金融数据聚合, 投研工作台, 量化研究
21	openai/spinningup	通用	11525	9000	2525	28.06	2423	2500	强化学习教程代码, 算法实现
22	sktime/sktime	制造 / 工业 / 能源	9449	7500	1949	25.99	1752	1900	时序预测, 设备监测管线
23	openvino/toolkit/open_model_zoo	制造 / 工业 / 能源	4342	2600	1742	67	1401	1200	质检/检测模型库, 工业落地参考
24	open-mmlab/mmdetection	制造 / 工业 / 能源	32273	30700	1573	5.12	9840	8700	工业视觉检测, 质检算法工程化
25	open-mmlab/mmdetection	制造 / 工业 / 能源	32273	30700	1573	5.12	9840	8700	工业视觉检测, 质检算法工程化
26	autogluon/autogluon	通用	9783	8300	1483	17.87	1103	1400	AutoML, 快速训练与集成
27	TauricResearch/TradingAgents	金融与保险	28038	26630	1408	5.29	5353	5017	多智能体交易研究, 交易决策, 工具化投研
28	ultralytics/yolov5	制造 / 工业 / 能源	56632	55300	1332	2.41	17399	16300	视觉质检, 缺陷检测, 工业质检检测
29	ultralytics/ultralytics	制造 / 工业 / 能源	51209	49900	1309	2.62	9884	9700	视觉质检, 缺陷检测, 边缘部署
30	openai/whisper	通用	93193	92000	1193	1.3	11653	11500	语音转文本, 多语言ASR
31	dmlc/xgboost	通用	27855	26700	1155	4.33	8832	14500	结构化数据建模, GBDT
32	huggingface/transformers	通用	27855	154000	1082	0.7	31736	31400	预训练模型调用与微调, NLP/多模态任务
33	microsoft/qlib	金融与保险	35595	34572	1023	2.96	5540	6308	量化投研平台, 特征工程, 模型训练/回测
34	xinntao/Real-ESRGAN	内容媒体 / 文娱 / 广告	33907	32900	1007	3.06	4229	5900	通用图像增强, 超分修复, 去噪去模糊

查看更多往期排行榜、全球AI App 实力榜、全球AI项目新势力榜等
可以扫描二维码关注AIGCLink公众号获得



排名	项目名称	行业	当前	上月	月增长 Star	增长率 (%)	当前 Fork	上月	核心用处 关键词
69	AI4Finance-Foundation/FinGPT	金融与保险	18399	18216	183	1	2588	2568	金融大模型; 研报/新闻理解; 金融NLP
70	polakowo/vectorbt	金融与保险	6473	6295	178	2.83	849	852	高性能自动引擎; 研究; 组合实验
71	stefan-jansen/ machine-learning-for-trading	金融与保险	16335	16166	169	1.05	4943	4907	金融机器学习教程代码; 因子研究; 回测
72	bmaltais/kohya_ss	内容媒体 / 文娱 / 广告	11960	11700	160	1.37	1544	1500	LoRA 训练; 扩散模型微调; 训练工具链
73	google-deepmind/alphafold3	医疗健康	7453	7300	153	2.1	1073	1000	蛋白/生物分子结构预测(新一代)
74	openai/gpt4	通用	20149	20000	149	0.74	3855	3800	模型交换与推理格式; 跨框架部署
75	Project-MONAI/MONAI	医疗健康	7747	7600	147	1.93	1401	1400	医疗影像AI开发框架/训练推理
76	microsoft/semantic-kernel	通用	27044	26900	144	0.54	4425	4400	LLM/Agent编排SDK, 多语言
77	facebookresearch/detectron2	通用	33942	33800	142	0.42	7881	7900	目标检测/分割训练; CV基线
78	AI4Finance-Foundation/FinRobot	金融与保险	4913	4772	141	2.95	901	879	金融Agent; 自动化投研; 工具链编排
79	rhasspy/piper	内容媒体 / 文娱 / 广告	10438	10300	138	1.34	893	872	本地离线配音; 端侧 TTS; 轻量部署
80	openvinotoolkit/openvino	制造 / 工业 / 能源	9519	9400	119	1.27	2953	2900	工业质检推理; 边缘推理加速
81	Lightning-AI/pytorch-lightning	通用	30717	30600	117	0.38	3642	3600	训练工程化; 训练循环抽象
82	OpenTalker/SadTalker	内容媒体 / 文娱 / 广告	13517	13400	117	0.87	2595	2600	图片说话; 口型驱动; 口播视频生成
83	keras-team/keras	通用	63715	63600	115	0.18	19676	19600	高层训练接口; 快速搭建神经网络
84	Sanster/IOPaint	内容媒体 / 文娱 / 广告	22835	22500	115	0.51	2400	2400	智能抹除; 修补; 局部重绘
85	sunlabuic/PyHealth	医疗健康	1396	1300	96	7.38	556	547	临床预测建模/EH-HR建模工具箱
86	sunlabuic/bark	内容媒体 / 文娱 / 广告	38895	38800	95	0.24	4682	4700	文本转语音; 旁白配音; 音效生成
87	kubeflow/kubeflow	通用	15394	15300	94	0.61	2596	2600	K8s上的ML平台; Pipeline/训练
88	karpathy/minGPT	培训与教育	23293	23200	93	0.4	3059	3000	最小GPT实现 / 教学用代码
89	MIC-DKFZ/miUNet	医疗健康	7893	7800	93	1.19	2273	2200	医学影像分割SOTA/自动配置
90	facebookresearch/audiocraft	内容媒体 / 文娱 / 广告	22890	22800	90	0.39	2555	2500	音乐生成; 音频生成; 创作者原型工具
91	google-deepmind/alphafold	医疗健康	14188	14100	88	0.62	2537	2500	蛋白结构预测/生物医学研究
92	guoyww/AnimateDiff	内容媒体 / 文娱 / 广告	11985	11900	85	0.71	1038	1000	文生视频; 动图生成; 镜头运动控制
93	karpathy/makemore	教育与培训	3585	3500	85	2.43	897	883	生成模型教学 / 字符级LM练习
94	fastai/fastai	通用	27783	27700	83	0.3	7682	7700	高层训练库; 快速实验
95	languagetool-org/languagetool	内容媒体 / 文娱 / 广告	13974	13900	74	0.53	1477	1500	写作纠错; 语法检查; 风格建议
96	yzhao062/pyod	制造 / 工业 / 能源	9671	9600	71	0.74	1463	1500	预测性维护; 异常检测; 设备健康监测
97	kermitt2/grobid	政务 / 法务	4576	4500	70	1.56	529	524	PDF结构化 (标题/段落/引用 用于法律/政策文档处理)
98	explosion/spaCy	通用	33069	33000	69	0.21	4636	4600	NLP流水线; 分词/NER/文本分类
99	karpathy/build-a-gpt	教育与培训	4669	4600	69	1.5	735	723	从零实现GPT课程 / 讲解+代码
100	Beckerschen/TransUNet	医疗健康	3065	3000	65	2.17	571	566	医学影像分割(Transformer+UNet)

AIGCLink介绍

AIGCLINK是一个面向AI开发者和从业者的社区组织，致力于推动人工智能技术在各行业的应用、普及和生态建

核心使命与目标

01 推动 AI 解决方案规模化落地

过去三年，AIGCLINK 围绕不同行业真实需求，系统性梳理与输出了近 千个可落地的 AI 解决方案，覆盖内容创作、营销、法律、医疗、制造、教育、金融等多个细分领域，持续推动 AI 从“能力展示”走向“业务实践”。

02 赋能「AI 超级个体」与一人公司（OPC）

AIGCLINK 率先跟进并持续研究“AI 超级个体”与*One-Person Company（OPC，一人公司）的发展趋势，聚焦 AI 对个人生产力、商业化能力与组织形态的重构。

未来计划通过工具支持、案例研究与生态资源对接，帮助 1,000 个 OPC 实现可持续商业化。

主办行业旗舰大会

AIGCLINK 主办年度旗舰峰会——AIGC 开发者大会（ACDC），持续聚焦前沿技术、产业落地与开发者生态建设，已成为 AI 与 AIGC 领域具有代表性的行业交流与趋势观察平台之一。

未来计划通过工具支持、案例研究与生态资源对接，帮助 1,000 个 OPC 实现可持续商业化。

构建开放生态技术

AIGCLINK 长期致力于 AI 技术生态与开发者体系建设，联合发起多项行业倡议，包括：

- MCP 中国开发者联盟：推动 AI Agent（智能体）相关协议与标准的本土化探索与实践
- 芯片适配联盟：促进 AI 应用、模型与国产及异构算力平台的协同适配
- AIGC 开发者“双百扶持计划”：面向开发者与初创团队，提供资源、曝光与成长支持

同时，AIGCLINK 计划投入 5,000 万人民币 设立开发者基金，用于支持早期 AI 项目、工具型产品与开发者生态的长期发展。

AIGCLINK AI 排行榜体系

AIGCLINK 构建并持续迭代多维度的 AI 排行榜体系，通过公开数据、社区反馈与研究模型，对 AI 项目、企业与产品进行长期追踪与评估。

- 中国 AIGC / 产业 AI 影响力榜
- AI Agent、AI 应用、开源项目专项榜
- 行业分榜（医疗、法律、金融、教育、制造等）
- 月度 / 年度趋势榜与专题榜。

排行榜不仅用于展示结果，更服务于研究、投资、产业判断与历史对照。