

DeepSeek-V4：开启国产大模型新时代

技术突破、市场格局与未来展望



北京研精毕智信息咨询有限公司



研精毕智信息咨询

目录

市场规模与预测

竞品对比分析

商业化路径

01

02

03

04

05

06

战略发布：双版本
矩阵

用户生态体系

核心应用场景

战略发布：双版本矩阵

DeepSeek-V4 系列以“百万上下文、MoE 架构、极致效率、国产适配”为核心，通过 Pro 与 Flash 双版本，实现高性能与普惠化的完美平衡，全面覆盖企业级与个人开发者需求。



旗舰版：DeepSeek-V4-Pro

对标国际顶级闭源模型，填补国产高端场景空白



1

核心定位

旗舰级高性能模型，总参数 1.6 万亿，激活参数 490 亿，预训练数据量 33 万亿 tokens，旨在对标 GPT-5.4、Claude Opus 4.7 等国际顶级模型。

2

目标场景

面向企业研发、科研机构、复杂代码开发、智能体 (Agent) 等高要求场景，满足金融风控、工业级代码重构、学术前沿研究等复杂任务需求。

3

性能优势

复杂推理响应速度较 V3.2 提升 40%，百万级代码库处理准确率达 96%，具备替代国际顶级模型的能力。

轻量版：DeepSeek-V4-Flash

聚焦轻量高效，极致优化
算力成本

01

市场定位

轻量高效、经济普惠，总参数 2840 亿，激活参数 130 亿，预训练数据量 32 万亿 tokens，主打低成本与易部署。

02

核心优势

支持在普通 GPU 甚至边缘设备上快速运行，部署成本仅为同类模型的 1/10，极大降低了中小企业的 AI 应用门槛。

03

应用场景

面向个人开发者、中小企业、高频轻量任务（如文本生成、简单问答、基础编程）等场景，实现 AI 技术的普及化。

双轮驱动：市场爆发

“高端引领 + 大众普及”的双轮驱动格局，助力市场份额快速扩张

01

市场策略

Pro 版本聚焦技术制高点，树立行业性能标杆；Flash 版本主打普惠化，快速抢占大众市场，形成差异化竞争优势。

02

增长数据

截至 2026 年 4 月，V4-Pro 企业客户签约量突破 1.2 万家，V4-Flash 开发者下载量超 500 万次，双版本均实现爆发式增长。

市场规模与预测

依托高速增长의 千亿级市场，DeepSeek 凭借技术与成本双重优势，有望快速抢占市场份额成为国产开源大模型的领军者。



千亿市场蓝海

中国 AI 大模型市场高速增长，开源赛道成为核心引擎

01

市场规模

2025 年中国 AI 大模型市场规模达 456.8 亿元，预计 2026 年将突破 700 亿元，年复合增长率达 52%，是全球增长最快的 AI 细分市场之一。

02

开源赛道

开源大模型市场规模增速尤为显著，将从 2024 年的 58.8 亿元增长至 2026 年的 210 亿元，占整体市场的 30%，成为驱动行业增长的核心引擎。

份额快速扩张

凭借综合优势，DeepSeek 有望占据国内开源市场 1/4 强

01

市场地位

在开源大模型细分赛道中，DeepSeek 凭借“长文本 + 价格 + 国产适配”的三重核心优势，市场份额有望快速提升至 25% 以上。

02

营收预测

预计 2026 年 DeepSeek 营收规模有望突破 50 亿元，成为国内营收最高的开源大模型企业。

03

全球布局

到 2028 年，随着海外市场的拓展与生态的进一步成熟，其全球市场份额有望跻身前三，与 OpenAI、Google 形成“两强一极”的全球大模型市场格局。

核心场景规模

代码开发、长文档处理、智能体三大场景市场规模超 390 亿

代码开发领域

2026 年市场规模将达 180 亿元，DeepSeek 凭借 83.7% 的 SWE-Bench 得分，有望占据 30% 以上的市场份额。

01

长文档处理领域

2026 年市场规模将达 120 亿元，其 97% 的长文本检索准确率将支撑其获得 40% 以上的市场份额。

02

智能体与企业自动化

年复合增长率达 78%，2026 年市场规模将达 90 亿元，DeepSeek 的 Agent 适配优势将推动其快速抢占市场。

03

用户生态体系

DeepSeek 已构建起覆盖开发者、企业、政企、C 端用户的全维度生态体系，生态规模与活跃度均处于行业领先水平。



开发者生态



开源社区活跃，开发者留存率与满意度双高

01 生态规模

在 HuggingFace、ModelScope 等平台的模型下载量破千万次，注册开发者超 50 万人，其中企业级开发者占比 30%。

02 社区活跃度

开源社区贡献者超 1.2 万人，较 2025 年增长 80%，形成了活跃的二次开发生态，衍生出 2000+ 基于 DeepSeek 的第三方应用。

03 开发者粘性

开发者留存率达 78%，远超行业平均水平（52%），满意度评分达 8.6 分（满分 10 分），核心得益于其 MIT 宽松开源协议与完善的技术文档。

企业与政企用户

行业覆盖广泛，国产化适配优势显著

01

企业用户

30万+企业用户已接入 DeepSeek API 服务，其中包括 100 家以上世界 500 强企业、十大国有银行、三大运营商。

02

政企用户

已在 23 个省份的政务服务平台落地，覆盖社保、税务、公安等多个场景，用户规模持续扩大。

03

核心优势

国产适配能力与合规性是赢得政企市场的核心原因，其政企用户增速达 120%，远高于行业平均水平。

C 端用户爆发

官方 App 下载量破亿，用户活跃度行业第一

01

增长数据

DeepSeek 官方 App 上线一个月下载量破 1 亿次，周活跃用户最高近 9700 万，无需大规模营销投入即实现爆发式增长。

02

用户粘性

用户日均使用时长达 38 分钟，7 日留存率达 65%，均位列同类产品第一。

03

核心优势

凭借 V4-Flash 的低成本与高性能，快速获得用户认可，主要应用于内容创作、学习辅助、日常问答等场景。

竞品对比分析

DeepSeek-V4 在与国际闭源模型和国内开源模型的竞争中均展现出显著优势，形成难以复制的核心壁垒。



对标国际顶级模型

性能全面超越或持平，成本具备绝对优势

01 上下文长度

1M tokens 远超 GPT-5.4 (128K tokens) 和 Claude Opus 4.7 (200K tokens)，仅与 Gemini-3.1-Pro 持平。

02 编程能力

83.7% 的 SWE-Bench 得分领先于所有国际竞品。

03 价格成本

API 定价仅为国际竞品的 1/5-1/40，具备绝对优势。

04 国产化适配

拥有开源属性与国产芯片深度适配能力，而国际闭源模型普遍仅支持 CUDA 生态，在政企国产化场景中存在天然劣势。

领跑国内开源模型

综合竞争力显著，市场份额有望快速提升

上下文长度

1M tokens 远超其他竞品 (30K-256K tokens) ，
形成绝对的长文本处理壁垒。

01

编程与效率

编程能力与算力效率处于行业顶尖水平，较国内竞
品平均领先 15%-20%。

02

国产芯片适配

覆盖度达 17 家，远超其他竞品的 1-7 家。

03

开源协议

采用 MIT 宽松开源协议，相较于部分国内模型的部
分开源策略，对开发者更友好，生态活跃度更高。

04

核心应用场景

DeepSeek-V4 系列的应用场景已实现全行业覆盖，从个人开发者的轻量任务到企业级的复杂需求，均展现出强大的适配能力。



代码开发与软件工程

研发效率提升超 50% ， 重构周期缩短 75%

全栈开发助手

可实现代码生成、审查、Bug 修复、重构、测试用例生成的全流程自动化，企业研发效率平均提升 50%+，测试用例生成时间缩短 70%。

01

大型代码库管理

支持百万行跨文件理解、架构分析、技术债务清理，某头部科技企业反馈，其核心系统代码优化周期从 3 个月缩短至 2 周。

02

低代码平台赋能

可实现 UI 转代码、业务逻辑自动生成，降低低代码平台的使用门槛，某低代码厂商接入后，平台用户增长 40%。

03

长文档处理与知识管理

分析效率提升 80% ，合同审查时间缩短 85%

01.

行业研报分析

可一次性处理百万字研报，自动提取核心观点、数据、风险点，某券商反馈其研报分析效率提升 80%，分析师人均产出增长 3 倍。

02.

法律金融文档审查

可实现合同比对、条款提取、招股书分析、风险预警，某大型律所使用后，合同审查时间从平均 2 天缩短至 3 小时。

03.

学术科研辅助

支持论文阅读、文献综述、数据关联分析，帮助科研人员快速梳理研究脉络，某高校实验室反馈，文献综述撰写时间缩短 60%。

智能体与企业自动化

业务流程自动化率提升 45% ， 数据分析周期缩短 85%

01

企业级智能助手

可实现跨系统数据整合、业务流程自动化，某制造企业接入后，其供应链管理流程自动化率从 30% 提升至 75%。

02

数据分析 Agent

可自动对接数据库、生成报表、挖掘业务洞察，某零售企业使用后，数据分析周期从 1 周缩短至 1 天。

03

客户服务自动化

智能应答、工单处理、个性化推荐能力显著提升客户体验，某银行接入后，客户咨询满意度提升 30% ，客服人员成本降低 40% 。

国产化与政企场景

政务响应时间缩短 97% ，风控识别准确率提升 12%



01

政务服务

可实现政策解读、公文处理、群众咨询自动化，某省政务服务平台接入后，咨询响应时间从平均 5 分钟缩短至 10 秒。



02

金融科技

应用于风控模型训练、交易数据分析、客户精准营销，某国有银行借助其风控模型训练能力，不良贷款识别准确率提升 12%。



03

智能制造

工业代码生成、设备故障诊断、生产流程优化能力显著提升生产效率，某汽车厂商接入后，生产线故障诊断时间缩短 70%。

个人与普惠场景

开发效率提升 30% ，内容产出周期缩短 83%

个人开发者编程

01

支持 338 种编程语言的基础开发需求，提供实时语法纠错、代码优化建议，某平台个人开发者反馈其开发效率提升 30%。

内容创作

02

可实现长文写作、文案生成、创意策划，某自媒体作者使用后，内容产出周期从 1 天缩短至 2 小时。

学习助手

03

知识问答、论文辅导、技能培训能力帮助用户高效学习，某在线教育平台接入后，用户学习完成率提升 35%。

商业化路径

DeepSeek 采用“多元化变现 + 生态共赢”的商业化模式，通过 API 服务、私有化部署、生态合作、行业解决方案四大核心路径，实现持续增长。



API 服务收费

按调用量阶梯定价，预计 2026 年收入达 30 亿元

01

定价策略

V4-Pro 面向企业客户收取较高费用，根据调用量分为基础版、企业版、旗舰版三个套餐；V4-Flash 针对个人开发者与中小企业推出低价套餐（1 元 / 百万 token 起）。

02

收入预测

预计 2026 年其 API 业务收入将达 30 亿元，占总营收的 60%，其中企业客户贡献 80% 的 API 收入。

私有化部署服务

定制化方案，高毛利且收入稳定

01

服务对象

主要针对金融、政务、军工等强监管、高敏感行业，提供定制化私有化部署方案，包含模型训练、部署优化、运维支持等一站式服务。

02

收入预测

2025 年已签约超 50 个私有化项目，收入超 10 亿元，2026 年预计签约项目数突破 150 个，收入达 15 亿元。

03

核心优势

数据安全可控与定制化适配，毛利率高达 75%，客户续约率达 90%，是公司利润的核心贡献者。

生态合作分成

联合云厂商与芯片厂商，扩大收入来源

01

合作模式

与华为、青云科技等云厂商合作，在其云平台提供 DeepSeek 模型服务，按收入分成；与华为昇腾、沐曦等芯片厂商联合推出“模型+芯片”打包解决方案。

02

收入预测

截至 2026 年 4 月，已与 8 家云厂商、17 家芯片厂商建立合作关系，2026 年生态合作分成收入预计达 3 亿元，未来三年有望保持 100% 的年增长率。

行业解决方案

聚焦垂直领域，提升产品附加值

01 业务模式

聚焦金融科技、智能制造、教育科研等垂直领域，开发专用模型与行业应用，如金融风控模型、工业故障诊断系统等，采用“基础模型 + 行业数据 + 定制开发”的模式。

02 收入预测

单个垂直领域解决方案定价超千万元，2026年预计落地20个行业解决方案项目，收入达2亿元。

03 核心价值

不仅能带来高额收入，更能沉淀行业数据与知识，形成行业壁垒，推动基础模型在垂直领域的性能提升。

分析师声明

负责本研究报告的分析师在本报告中所采用的数据均来自合规渠道，报告的观点、逻辑和论据均为分析师本人研究成果，力求独立、客观和公正，结论不受任何第三方的授意或影响，特此声明。

公司声明

本报告的著作权归北京精毕智信息咨询有限公司（简称为“研精毕智”）所有。本报告是研精毕智研究与统计成果，所载的观点、结论

和建议仅代表行业基本状况，仅为市场及客户提供基本参考。
本报告调研方法主要是桌面研究、行业访谈等，结合公司内部逻辑算法，通过定量和定性分析分析，客观阐述行业的现状，科学预测

行业未来的发展趋势。
我们力求报告内容客观、公正，但受到调研方法及调查资料收集范围的局限，本报告所述的观点、数据并不一定完全准确。

本报告版权仅为本公司所有，未经书面许可，任何机构和个人不得以任何形式篡改、复制和发布。如引用、转载需注明出处，且不得

对本报告进行有悖原意的引用和修改。
本研究报告仅供北京研精毕智信息咨询有限公司客户和经本公司授权机构的客户使用，未经授权私自刊载的机构以及其阅读和使用

应慎重使用报告，本公司不承担由此所产生的相关风险和责任。