

主流大模型综合能力对比报告

一、评估背景

为反映最新技术发展动态，本报告选取当前最具影响力的 5 款大模型进行多维度对比，聚焦 中文能力、技术开发、长文本处理、多模态、成本生态 等核心维度。

二、对比模型清单

模型名称	开发公司	代表版本	国家
ChatGPT	OpenAI	GPT-4o	美国
通义千问 (Qwen)	阿里巴巴	Qwen-Max	中国
DeepSeek-R1	深度求索	R1	中国
智谱AI (GLM)	清华智谱	ChatGLM4	中国
Google Gemini	Google DeepMind	Gemini 1.5 Pro	美国

三、核心维度对比总表

✔️ 强项 | ⚠️ 中性 | ❌ 弱势

维度	GPT-4O	通义千问	DEEPSEEK	智谱GLM	GEMINI
综合推理	✔️✔️✔️	✔️✔️	✔️✔️	✔️✔️	✔️✔️✔️
中文理解生成	✔️✔️	✔️✔️✔️	✔️✔️✔️	✔️✔️✔️	✔️
英文能力	✔️✔️✔️	✔️✔️	✔️✔️	✔️	✔️✔️✔️
长文本处理	✔️✔️	✔️✔️	✔️✔️✔️	✔️✔️	✔️✔️✔️
多模态能力	✔️✔️✔️	✔️✔️	❌	✔️	✔️✔️✔️
编程能力	✔️✔️✔️	✔️✔️	✔️✔️✔️	✔️✔️	✔️✔️✔️
响应速度	✔️✔️✔️	✔️	✔️✔️✔️	✔️✔️	✔️✔️
使用成本	❌	✔️	✔️✔️✔️	✔️	⚠️

维度	GPT-4O	通义千问	DEEPSEEK	智谱GLM	GEMINI
开源程度	✗	✓✓✓	✓✓	✓✓	✗
技术严谨性	✓✓	✓✓	✓✓	✓✓✓	✓✓

上下文窗口对比

GPT-4o: 128K / 通义千问: 128K+ / DeepSeek: 128K / 智谱GLM: 128K / Gemini: 1M

四、分维度详细分析

1. 中文场景能力（国内重点关注）

模型	优势	不足
通义千问	<ul style="list-style-type: none">中文创作天花板（诗词/小说）文化隐喻理解精准	技术文档严谨性稍弱
DeepSeek	<ul style="list-style-type: none">长文本解析能力突出（论文/合同）全免费+API响应快	不支持图文多模态
智谱GLM	<ul style="list-style-type: none">政府/金融场景最安全技术术语表述严谨	创意表达灵活性不足

✅ 场景化推荐：文学创作 → 通义千问 / 技术文档 → 智谱GLM / 长文本分析 → DeepSeek

2. 技术开发能力

模型	优势	典型应用场景
ChatGPT	全栈开发支持 + 插件生态完善	跨平台应用开发
智谱GLM	科研代码生成/论文辅助最强	学术研究/金融量化
DeepSeek	中文注释清晰 + 免费API调用	教学/快速原型验证
Gemini	多模态编程（图转代码）	视觉识别+自动化脚本

3. 成本与生态

模型	成本策略	生态优势
DeepSeek	💰 全免费（含 128K API）	🔒 社区活跃 + 教程丰富
通义千问	💰 免费版 + 阶梯式 API	🌱 开源模型数量最多
智谱 GLM	💰 企业定制 + 学术免费	📋 政府 / 金融准入认证完善
GPT-4o	💰 (20/月 +)10 / 百万 token	🌐 全球开发者生态最成熟
ChatGPT	💰 免费基础版Gpt3.5 + \$20 / 月 Plus 版	🔌 插件生态丰富，与多平台集成便捷

五、选型建议

需求场景	首选模型	次选模型
中文创作（零成本）	DeepSeek-R1	通义千问
金融/科研文档	智谱GLM4	Claude 3
跨模态开发	Gemini 1.5 Pro	GPT-4o
企业私有化部署	通义千问	智谱GLM4
实时交互应用	GPT-4o	DeepSeek-R1