

北京大学
PEKING UNIVERSITY

DeepSeek内部研讨系列

AI工具深度测评与选型指南 v1.0

5大类别 - 39个AI工具 - 92个实例测评

AI肖睿团队

（韩露、吴寒、顾跃、王春辉、杜静华、相洪波、李娜）

20250504@北京

- 北大青鸟人工智能研究院
- 北大计算机学院
- 北大教育学院学习科学实验室

讲座内容介绍

一、本次讲座为企业决策者、运营团队以及个人使用者系统地分析主流AI工具的技术特性、应用场景与选型逻辑，提供**深度洞察和经验分享**，帮助大家在AI浪潮中更轻松地选择和使用AI工具。

二、本次讲座的内容主要涵盖4个部分：

1. 导论与分类：剖析AI工具对业务效率与创新的战略意义，提出基于功能与应用场景的分类框架，并提供工具-场景适用性矩阵概览，**快速建立认知地图。**
2. 多维评估体系构建：详细阐述用于AI工具选型的系统性评估框架与方法论，涵盖核心功能、性能指标（如精度、效率、可扩展性）、成本效益、易用性、集成性、行业适配度及安全合规等维度，帮助大家自主**构建符合自身需求的评估标准。**
3. 核心领域工具深度测评与横向对比：聚焦自然语言处理（文本）、计算机视觉（图像/视频）、代码生成、大模型管理应用5个关键领域，选取了39个代表性AI工具（DeepSeek、Gemini、Midjourney、Cursor、可灵等），并进行深度实测与横向性能比较。结合实战案例拆解，**直观展示各工具的优劣势与最佳实践。**
4. 选型策略与未来展望：提炼“需求-场景-工具”的精准匹配框架，规避常见选型误区。提供构建高效AI工具组合的战略性建议，探讨如何在技术迭代中平衡创新投入与业务回报，并展望AI工具的集成化、智能化发展趋势。

三、大家可以参考《人工智能通识教程（微课版）》这本系统全面的入门教材，结合B站“思睿观通”栏目的配套视频进行学习。欢迎关注“AI肖睿团队”的视频号和微信号（ABZ2829），加入ai.kgc.cn社区，我们将提供更丰富的AI工具集合和落地经验，以及交流和分享机会。

本讲座使用的39个AI工具一览

北京大学 PEKING UNIVERSITY

文本生成与处理

DeepSeek　Gemini　通义千问　豆包　Kimi

数字人

硅语　Heygen　智课　奇妙元　闪剪

图像生成与编辑

即梦AI　可灵AI　Midjourney　GPT-4o　Gemini　百度AI图片助手

SnapEdit　通义万相　豆包　腾讯元宝　阶跃星辰

AI编程

Cursor　Trae　Cline　GitHub Copilot　通义灵码　CodeGeeX

MGX　heyBossai

音频生成

海螺AI　MINIMAX　Noiz.ai

视频生成

即梦AI　可灵AI　Vidu　通义万相

大模型管理与应用

CherryStudio　LMStudio　Chatbox

本讲座使用的39个AI工具导航

北京大学 PEKING UNIVERSITY

01 **文本生成与处理** DeepSeek、Gemini、通义千问、豆包、Kimi

02 **图像生成与编辑** 即梦AI、可灵AI、Midjourney、GPT-4o、Gemini
百度AI图片助手、SnapEdit、通义万相、豆包、腾讯元宝、阶跃星辰

03 **音频生成** 海螺AI、MINIMAX、Noiz.ai

04 **视频生成** 即梦AI、可灵AI、Vidu、通义万相

05 **数字人** 硅语、 Heygen、智课、奇妙元、闪剪

06 **AI编程** Cursor、Trae、Cline、GitHub Copilot、通义灵码、CodeGeeX、MGX、heyBossAI

07 **大模型管理与应用** CherryStudio、LMStudio、Chatbox

内容目录

一、从AI模型到AI工具 P6

1. DeepSeek与大模型理论概述 P8
2. AI工具浪潮：为何现在关注AI工具？ P15
3. AI 工具如何赋能千行百业？ P19
4. 系统性认识AI工具 P20
5. AI工具分类 P21
6. AI工具测评的核心价值 P22

二、AI工具测评框架与方法论 P24

1. 12个测评维度定义 P27
2. 6个测评方法详解 P38
3. 根据自身需求选择测评维度权重 P45
4. 不同场景下AI工具测评维度的权重示例 P47
5. 不同场景下测评维度的权重设置参考 P54
6. 总结 P55

三、AI工具测评 P57

1. AI工具测评一览表 P59
2. 文本生成与处理类AI工具测评 P60
3. 图像生成与编辑类AI工具测评 P108
4. 音视频生成与处理类AI工具测评 P171
5. 代码生成与辅助类AI工具测评 P245
6. 大模型管理与应用类AI工具测评 P289

四、AI工具选型指南与未来展望 P310

1. 构建个性化AI工具箱选型决策框架 P312
2. 构建你的AI工具箱：个人AI工具战略选型流程 P313
3. 构建你的AI工具箱：团队AI工具战略选型考量 P314
4. 从工具到工作流，1+1 > 2 的组合力量 P315
5. AI工具的发展趋势 P316
6. 总结 P318

一、从AI模型到AI工具

北京大学 PEKING UNIVERSITY

- 当下，我们正站在AI重塑生产力的关键节点，从ChatGPT到DeepSeek的快速迭代标志着生成式AI已进入实际应用爆发期，AI工具的广泛应用已成为不可忽视的趋势。其作为效率革命引擎对重塑工作方式与未来竞争力起着关键作用。
- 为理解这些AI工具的技术基础，我们以通识视角揭秘了人工智能的核心概念，特别聚焦于驱动当前工具爆发的生成式AI与大语言模型。
- 为了系统性认识庞大的AI工具生态，我们提出了按核心功能（文本、图像、音视频、代码等）与按主要应用领域（办公、创作、研发等）的双重分类框架，并通过直观的“工具x场景”矩阵，初步展示了包括DeepSeek在内的代表性工具的适用范围。通过建立清晰的分类框架，帮助您快速掌握工具生态，为后续的深入评估提供基础认知。

学习交流可加AI肖睿团队助理微信号（ABZ2829）

一、从AI模型到AI工具

北京大学 PEKING UNIVERSITY

1. DeepSeek与大模型理论概述
   - 1.1 现象：DeepSeek的快速出圈
   - 1.2 到底谁是DeepSeek
   - 1.3 大模型的三层能力
   - 1.4 生成模型与推理模型适用场景
   - 1.5 AI工具与AI模型的核心区别
2. AI工具浪潮：为何现在关注AI工具？
3. AI 工具如何赋能千行百业
4. 系统性认识AI工具
5. AI工具分类
6. AI工具评测的核心价值

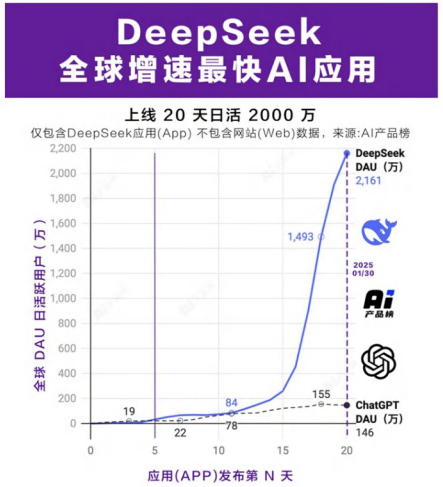
学习交流可加AI肖睿团队助理微信号（ABZ2829）

1.1 现象：DeepSeek快速出圈

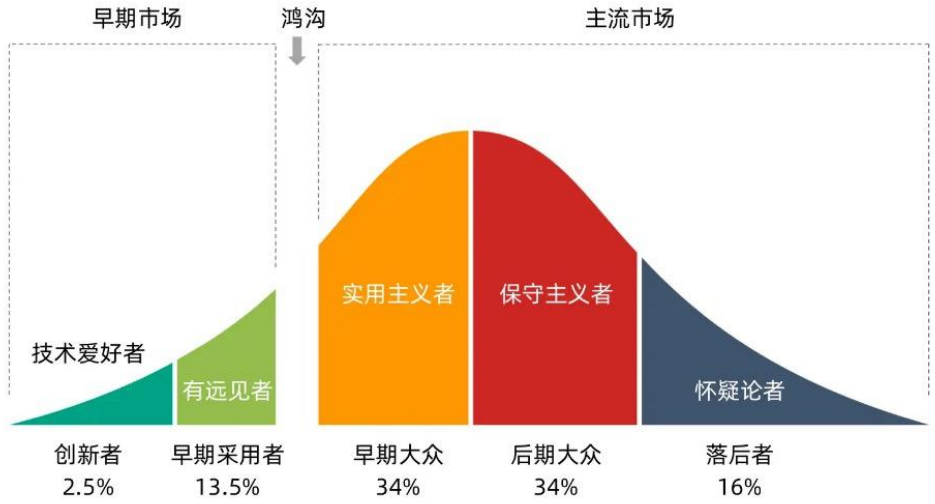
1. 2024年12月26日，DeepSeek推出对标OpenAI GPT-4o的语言模型DeepSeek V3，随后在美国AI行业内部引起轰动。
2. 2025年1月20日，DeepSeek发布对标OpenAI o1的语言模型DeepSeek R1，并于1月24日引起美国投资界KOL关注。
3. 2025年1月26日，关于DeepSeek颠覆了大模型的商业模式的讨论，引发英伟达股价大跌，DeepSeek首先在美国出圈，引发社会讨论。
4. 春节前后，DeepSeek在中国出圈，并上升到中美竞争高度，同时纷纷接入DeepSeek，DeepSeek成为AI和大模型的代名词。
5. **DeepSeek让AI跨越了鸿沟。**



2025年1月20日下午，中共中央政治局常委、国务院总理李强主持召开专家、企业家和教科文卫体等领域代表座谈会，听取对《政府工作报告（征求意见稿）》的意见建议。DeepSeek公司创始人梁文峰作为企业家代表之一参加了此次座谈会。





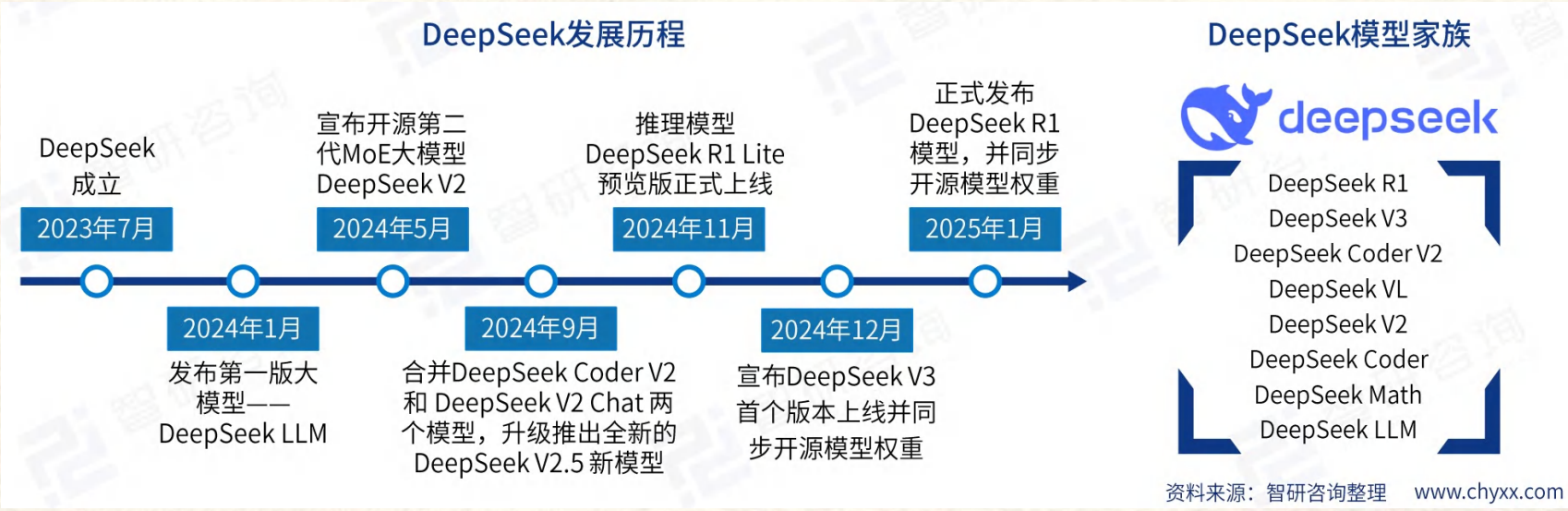




1.2 到底谁是DeepSeek？公司、模型、产品

公司简介

1. DeepSeek于2023年5月成立北京公司，2023年7月成立杭州公司，是幻方量化孵化出的一家大模型研究机构，160人的团队分布在中国杭州和北京，是中国AI2.0（大模型）时代的七小龙之一。
   - 注：AI1.0（深度学习）时代的四小龙：商汤、旷世、云从、依图
2. 除了DeepSeek之外，其它六家大模型公司也被投资界称为中国大模型企业六小虎（智谱AI、百川智能、月之暗面、阶跃星辰、MiniMax、零一万物）。



Model	Base Model	
DeepSeek-R1-Distill-Qwen-1.5B	Qwen2.5-Math-1.5B	**蒸馏模型，能力稍弱** ➢ 实际上是增加了推理能力的Qwen模型和Llama模型，不能称为DeepSeek模型 ➢ 市场上有误解，厂商有误导，Ollama工具的模型选项中也有误导
DeepSeek-R1-Distill-Qwen-7B	Qwen2.5-Math-7B	
DeepSeek-R1-Distill-Llama-8B	Llama-3.1-8B	
DeepSeek-R1-Distill-Qwen-14B	Qwen2.5-14B	
DeepSeek-R1-Distill-Qwen-32B	Qwen2.5-32B	
DeepSeek-R1-Distill-Llama-70B	Llama-3.3-70B-Instruct	
DeepSeek-R1-671B	DeepSeek-V3-Base	满血版，能力最强

1.2 到底谁是DeepSeek？公司、模型、产品

北京大学 PEKING UNIVERSITY

AI系统的四层架构

SaaS
应用服务（网页、APP、桌面软件、设备软件）

MaaS
模型服务（数据工程、推理加速、训练框架、API调用）

PaaS
平台服务（存储、计算、数据、安全）

IaaS
硬件服务（设备/电脑/服务器/GPU、网络）

AI系统的实现方案

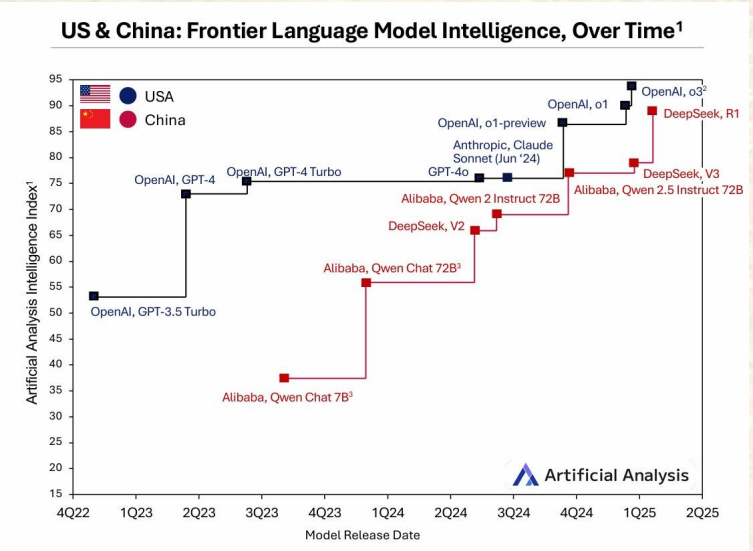
- 生产办公系统：目前多数为定制开发（RAG、LangChain、Agent等）
- 对话机器人：目前多数为标准产品，使用云服务
  - 免费：**DeepSeek**、豆包、Kimi、元宝、Monica、秘塔搜索、问小白、Molly R1等
  - 收费（月费；充值）：文心一言、通义千问、Liblib、ChatGPT等
- 开源模型：**DeepSeek**、Qwen、Llama、SD、微调模型等
  - 向上提供：私有化部署，公有云服务（**DeepSeek**官网、硅基流动、火山引擎等）
- 闭源模型：OpenAI、Claude、百度等
  - 向上提供：公有云服务
- 私有化部署
  - 本地：个人设备或电脑、企业服务器
  - 数据中心（IDC）：企业服务器、服务器集群（私有云）
- 公有云服务：阿里云、火山云、腾讯云、金山云、华为云等；AWS、Azure等

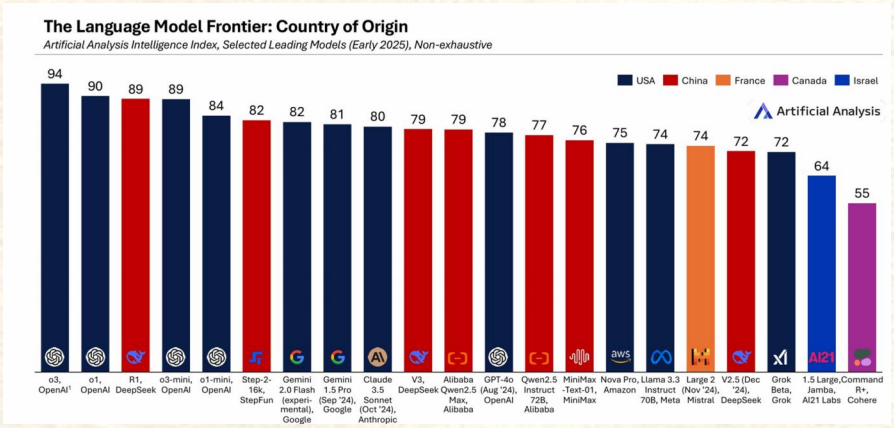
1.2 出圈的是模型：DeepSeek R1模型

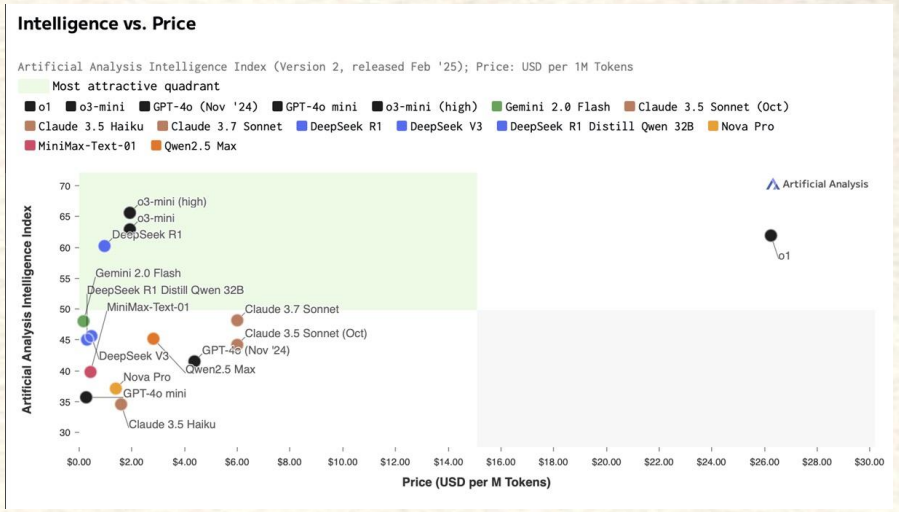
出圈的核心原因

1. 能力突破：推理模型的全球第一梯队
2. 加分项：开源、低成本、国产自主

- 对于大模型研发企业， 更加重视infra工程的价值了。
- 对于大模型应用企业，有了更多高效低成本解决方案。
- 对于社会大众，认识到AI是一个趋势，不是昙花一现。
- 对于市场，用户开始主动引入AI，不再怀疑了。
- 对于国家，大幅缩小了中美的核心技术差距。
- 对于全人类，技术平权，造福一方。







混合专家 MOE | 多头潜注意力 MLA | 多Token预测 MTP | 强化学习 GRPO | 测试时计算 TTC | 混合精度训练 FP8 | 通讯优化 DualPipe | 并行训练框架 HAI | 直接硬件编程 PTX

1.3 大模型的三层能力

语言能力：理解和生成

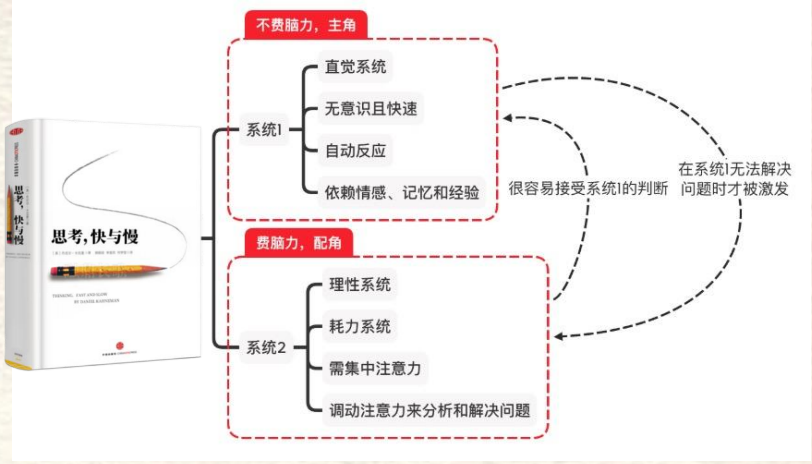
知识能力

幻觉（生成不符合事实的内容）

知识库限制（公开、私有、即时）

上下文窗口限制（记忆、成本）

推理能力







1.4 生成模型与推理模型的适用场景

比较项	生成模型（GPT-4.1、DeepSeek-V3）	推理模型（GPT-o3、DeepSeek-R1）
模型定位	专注于通用自然语言处理，多模态能力突出，适合日常对话、内容生成、翻译以及图文、音频、视频等信息处理。	侧重于复杂推理与逻辑能力，擅长数学、编程和自然语言推理任务，适合高难度问题求解和专业领域应用。
推理能力	在日常语言任务中表现均衡，但在复杂逻辑推理（如数学题求解）上准确率较低。	在复杂推理任务表现卓越，尤其擅长数学、代码推理任务。
多模态支持	支持文本、图像、音频乃至视频输入，可处理多种模态信息。	当前主要支持文本输入，不具备图像处理等多模态能力；未来可能通过社区贡献扩展相关功能。
应用场景	适合广泛通用任务，如对话、内容生成、多模态信息处理以及跨语言交流；面向大众市场和商业应用。	更适合需要高精度推理和逻辑分析的专业任务，如数学竞赛、编程问题和科学研究；在思路清晰度要求高的场景具有明显优势，比如采访大纲、方案
用户交互体验	提供流畅的实时对话体验，支持多种输入模态；用户界面友好，适合大众使用。	可展示部分链式思考过程，便于用户理解推理过程；界面和使用体验具有较高的定制性，但整体交互节奏较慢。

- 生成模型是玩知识和文字的，推理大模型是玩逻辑和推理的，至于计算问题，还是找计算器吧
- 推理模型也不是万能的，其幻觉通常比生成模型大，很多不需要强推理的场合还是传统的生成模型比较适合

1.5 AI工具 vs AI模型：核心区别

AI工具与AI模型的核心关系

- **AI 工具**是基于 AI 模型构建的、面向用户的应用程序或服务；
- **AI 模型**则是提供核心智能能力的底层技术引擎。可以将 AI 模型想象成汽车的发动机，而 AI 工具则是可以直接驾驶的整车。

对比维度	AI工具 (AI Tool)	AI模型 (AI Model)
本质与定义	封装了AI 模型的、**用户友好型应用程序或服务**，通常经过产品化，**开箱即用**，旨在解决特定任务或场景需求。	工具或应用背后的**核心算法与计算框架**，是提供特定 AI 能力（如语言理解、图像生成）的**技术引擎**。
关键特点	• **产品化、用户导向**：注重用户体验和易用性。 • **场景化解决方案**：为特定应用场景设计。 • **多模型集成 (可能)**：有时会整合多个模型或技术。	• **技术/研究者导向**：更关注底层技术实现和能力边界。 • **单一核心能力**：通常聚焦于某项具体的 AI 技术。 • **开源部署**：常以接口或代码形式提供。
目标用户	**终端用户** (所用用户)	**技术开发者**、**AI 研究人员**、需要将 AI 能力集成到自己产品中的团队。
使用方式	• **使用门槛：低**（直接使用） • 通过**图形用户界面 (GUI)**、聊天窗口、浏览器插件等**直接交互使用**，通常**无需编程**。	• **使用门槛：高**（需编程调用） • 需要通过**编程调用 (API、SDK)**，或进行**模型部署、集成**到其他软件或服务中。
形态	成熟的**软件产品**、**在线服务**或**功能模块**	**API 接口**、**软件库 (SDK)**、**预训练模型文件** (如开源模型)。
举例	Midjourney、Cursor、DeepSeek（网页版/App）等	Stable Diffusion、Llama、DeepSeek-R1-671B等
测评重点	易用性、功能满足度、特定场景效果、响应速度、集成度、成本效益等。	性能指标 (准确率、速度等)、技术先进性、可扩展性、部署灵活性、API 稳定性和文档、训练/推理成本等。

2.AI工具浪潮：为何现在关注AI工具？

行业趋势：席卷全球的AI变革

- 麦肯锡全球调查显示，AI采用率正在快速上升，特别是在生成式AI领域。与仅10个月前相比，企业使用生成式AI的比例几乎翻了一倍，达到了65%。根据最新数据：
- 72%的受访企业已经开始采用AI技术，相比往年50%左右的水平显著提升
- 大约50%的企业在两个以上的业务功能中部署了AI技术
- 67%的受访者预计其组织在未来三年将在AI上投入更多资金
- 营销销售和产品开发是最常见的AI应用领域

企业采用

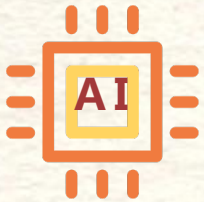
78%

组织报告在2024年使用AI

收入增长

87%

企业预计AI将在3年内带来收入增长



投资回报

97%

投资AI的领导者报告获得积极回报

战略整合

49%

技术领导者称AI已完全融入业务战略

https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

人工智能与算法的协作将催生未来几十年难以想象的科学进步- Demis Hassabis, Google DeepMind联合创始人

2.AI工具浪潮：为何现在关注AI工具？

效率革命：AI驱动的生产力飞跃

■ 创意激发与内容生产加速

创作速度提升

从设计到制作的时间缩短90%

创作成本降低

高质量内容的制作成本大幅降低

创作民主化

降低技术门槛
人人可称为创作者

创作可能性扩展

打破认知局限
探索新创意方向

■ AI内容创作的爆发性增长

3400万
每日AI生成图像

71%
社交媒体图像来自AI

2000+
AI图像生成工具诞生

■ AI内容能力进化

2022
文本生成

2023
文本+图像

2024
多模态

2025
创作Agent

现在，正是拥抱AI工具，提升个人和组织竞争力的关键时刻！

2.AI工具浪潮：为何现在关注AI工具？

效率革命：AI驱动的生产力飞跃

■ 决策效率提升

减轻认知负担

AI处理复杂数据集，帮助管理者聚焦关键决策点，降低认知压力

发现隐藏模式

AI能识别人类难以察觉的数据关联和趋势，挖掘业务洞察

实时决策支持

AI提供即时洞察和建议，支持快速响应变化的市场条件

减少决策偏差

经过适当训练的AI系统可帮助减少人类决策中的认知偏见

■ AI辅助决策流程

数据收集
跨越多种数据源

→

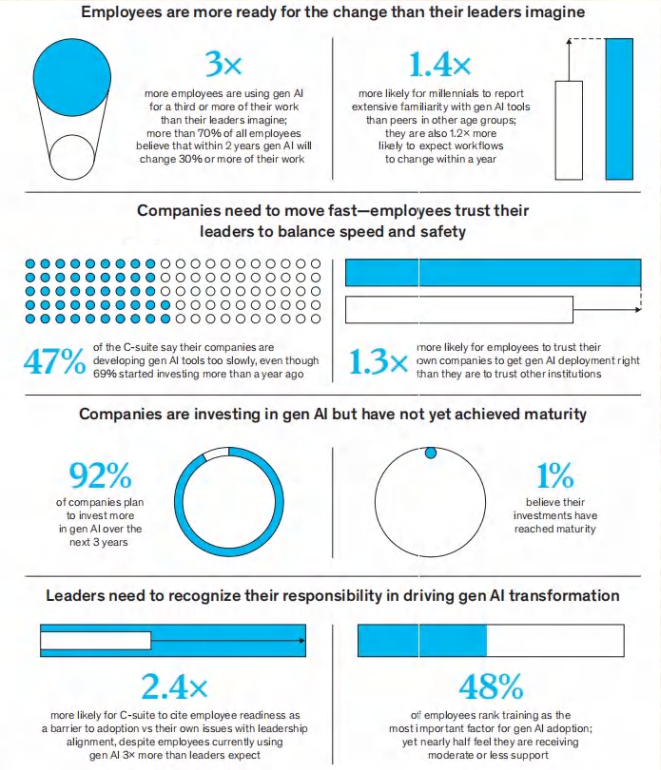
分析与推理
识别模式和洞察

→

预测与建议
生成决策选项

→

决策执行
人机协作



麦肯锡研究调查结果：

员工使用生成式人工智能的比例是领导者想象的3倍，且1.4倍的千禧一代比其他年龄段的人更熟悉相关工具，但47%的C级高管认为其公司开发人工智能工具的速度过慢，仅有1%的公司认为其人工智能投资已成熟，而92%的公司计划在未来三年内投资，领导者需要认识到其驱动变革的责任，因为在采用人工智能方面，员工的准备度是领导者预期的2.4倍，且48%的员工认为培训是采用人工智能的最重要因素。

现在，正是拥抱AI工具，提升个人和组织竞争力的关键时刻！

2.AI工具浪潮：为何现在关注AI工具？

效率革命：AI驱动的生产力飞跃

■ 创新能力增强

- 赋能新产品、新服务、新商业模式的探索

■ 行业变革实例

- **材料科学加速创新**

AI加速新材料发现和制造工艺优化，将传统10年研发周期缩短至1-2年

- **可再生能源优化**

预测性AI在可再生能源领域应用，提高能源产出15-30%，降低运营成本25%

- **个性化教育革新**

AI辅导系统根据学习者进度、风格和表现实时调整教学内容，学习效率提升40%

■ 产业变革

- **商业模式变革：**AI驱动的商业模式使用AI技术以创新方式创造、传递和获取价值，Copilots、agents和AI驱动的服务成为新商业模式的代表
- **医疗行业的巨大机会：**医疗保健AI市场价值达387亿美元，自2023年以来翻了一番
- **设备智能化升级：**30%的新智能手机将配备设备内生成式AI，带来更快、更个性化和更安全的用户体验
- **巨额技术投资：**科技巨头在2025年投资3200亿美元用于AI开发，全球AI服务市场今年将达到2430亿美元

现在，正是拥抱AI工具，提升个人和组织竞争力的关键时刻！

3.AI工具如何赋能千行百业？

AI在不同领域的效率提升与创意激发潜力

✍️ **内容创作提速** (Content Creation Acceleration):
- AI 写作助手、智能文案生成、多语言快速翻译、内容风格转换...
- 相关技术: NLP / LLM

🎨 **设计创意迸发** (Design & Creativity Spark):
- AI 绘画与图像生成、AI 视频/动画制作、个性化设计推荐...
- 相关技术: Generative AI / Diffusion Models

💻 **编程开发增效** (Programming Development Efficiency Boost):
- AI 代码自动补全与生成、智能 Bug 检测与修复、自动化测试用例生成...
- 相关技术: LLM / Code Models

⚙️ **办公自动化升级** (Office Automation Upgrade):
- AI 会议纪要与总结、智能文档信息提取与摘要、自动化业务流程连接...
- 相关技术: NLP / Automation Platforms

🔍 **信息获取革新** (Information Access Revolution):
- AI 驱动的智能搜索、精准知识问答系统、个性化研究助理...
- 相关技术: LLM / Search Technologies / RAG

🌐 **... 更多场景** (More Scenarios):
- 深入教育、医疗、制造、金融、法律等垂直领域的专业级定制化应用

4.系统性认识AI工具

将众多AI工具按核心功能划分

💼 **办公效率** (Office Productivity)
- 场景: 文档处理、会议纪要、邮件撰写、日程管理...

✍️ **内容创作** (Content Creation)
- 场景: 文章写作、营销文案、脚本构思、社交媒体帖子...

📣 **营销推广** (Marketing & Promotion)
- 场景: 广告创意、用户画像分析、市场趋势洞察、SEO 优化...

💻 **软件研发** (Software Development)
- 场景: 代码编写与补全、自动化测试、Bug 修复、项目文档...

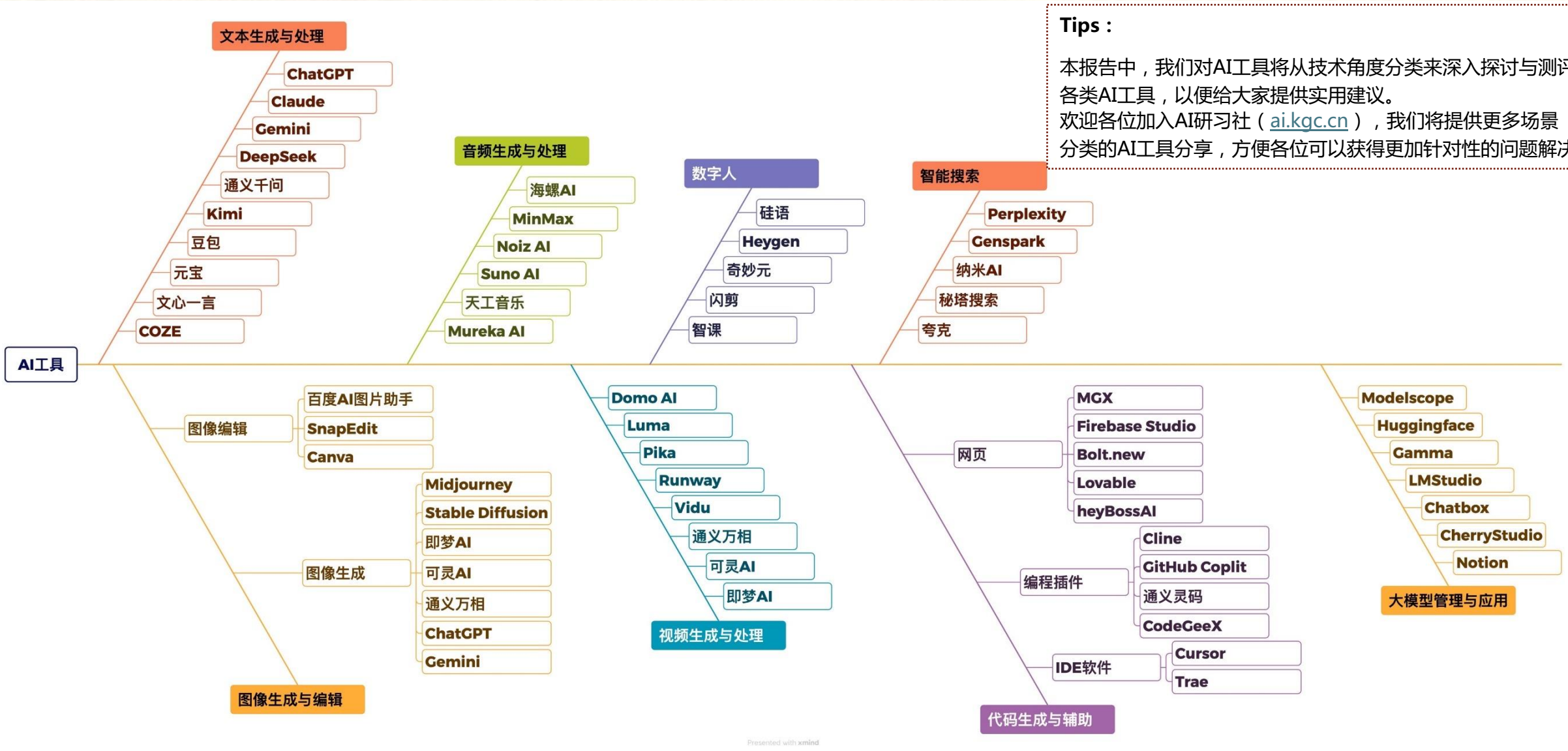
✨ **设计创意** (Design & Creativity)
- 场景: 视觉设计、UI/UX 原型、插画生成、艺术创作辅助...

🎓 **教育培训** (Education & Training)
- 场景: 个性化学习计划、智能辅导、课件生成、知识问答...

📰 **新媒体运营** (New Media Operations)
- 场景: 热点追踪、内容分发、数据分析、用户互动管理...

🏭 **其他关键场景**
- 如:建筑设计、工业制造、金融分析、政企服务等...

5.AI工具分类



Tips：

本报告中，我们对AI工具将从技术角度分类来深入探讨与测评各类AI工具，以便给大家提供实用建议。
欢迎各位加入AI研习社（ai.kgc.cn），我们将提供更多场景分类的AI工具分享，方便各位可以获得更加针对性的问题解决。

6.AI工具测评的核心价值

1. **消除信息不对称**：在繁杂的 AI 工具市场中，为读者提供一个清晰、结构化的认知框架，帮助他们理解不同工具的定位和特点。
2. **实用性与指导性**：不仅停留在理论层面，更侧重于提供可操作的步骤、清晰的评估标准和针对不同场景的建议，帮助大家真正解决选型难题。
3. **客观性与中立性**：尽量避免对特定工具的过度推销，而是提供一个客观的评估框架，引导受众根据自身需求进行选择。
4. **节省时间与精力**：通过系统性梳理和关键信息提炼，大大减少大家自行摸索和调研所需的时间与精力。

内容目录

一、从AI模型到AI工具 P6

1. DeepSeek与大模型理论概述 P8
2. AI工具浪潮：为何现在关注AI工具？ P15
3. AI 工具如何赋能千行百业？ P19
4. 系统性认识AI工具 P20
5. AI工具分类 P21
6. AI工具测评的核心价值 P22

二、AI工具测评框架与方法论 P24

1. 12个测评维度定义 P27
2. 6个测评方法详解 P38
3. 根据自身需求选择测评维度权重 P45
4. 不同场景下AI工具测评维度的权重示例 P47
5. 不同场景下测评维度的权重设置参考 P54
6. 总结 P55

三、AI工具测评 P57

1. AI工具测评一览表 P59
2. 文本生成与处理类AI工具测评 P60
3. 图像生成与编辑类AI工具测评 P108
4. 音视频生成与处理类AI工具测评 P171
5. 代码生成与辅助类AI工具测评 P245
6. 大模型管理与应用类AI工具测评 P289

四、AI工具选型指南与未来展望 P310

1. 构建个性化AI工具箱选型决策框架 P312
2. 构建你的AI工具箱：个人AI工具战略选型流程 P313
3. 构建你的AI工具箱：团队AI工具战略选型考量 P314
4. 从工具到工作流，1+1 > 2 的组合力量 P315
5. AI工具的发展趋势 P316
6. 总结 P318

二、AI工具测评框架与方法论

- 在AI技术爆发式增长的今天，市场上涌现出数以千计的AI工具，它们的功能、性能、适用场景和隐性成本差异显著。若缺乏系统化的评估体系，企业或个人用户极易陷入“技术迷雾”，或被过度营销误导，或因试错成本高昂而错失真正有效的解决方案。



- 本部分将构建一套科学的**AI工具测评方法论**，通过定义核心维度（如效果质量、易用性、成本效益等）、拆解实操测评方法，并指导如何根据业务需求动态调整评估权重，最终形成可量化的决策依据。结合AI工具的测评结果，方便大家快速选取适合自身场景的AI工具，使**AI选型从"经验驱动"转向"数据驱动"**，为数字化转型提供精准的技术支撑。

学习交流可加AI肖睿团队助理微信号（ABZ2829）

二、AI工具测评框架与方法论

北京大学 PEKING UNIVERSITY

1. 12个测评维度定义
   - 1.1 为何需要系统评估AI工具？
   - 1.2 如何全面评估一个AI工具？12个评测维度概览
   - 1.3 核心测评维度定义（1-12个详解）
   - 1.4 AI工具测评维度定义与关键考量（总结）
2. 6个测评方法详解
   - 2.1 常见的6个测评方法
   - 2.2 测评方法详解（1-6个）
3. 根据自身需求选择测评维度权重
   - 3.1 如何根据自身需求选择测评维度权重
   - 3.2 权重应用：构建你的决策矩阵
4. 不同场景下AI工具测评维度的权重示例
   - 4.1 不同场景下AI工具测评维度的权重示例-1
   - 4.2 不同场景下AI工具测评维度的权重示例-2
   - 4.3 不同场景下AI工具测评维度的权重示例-3
   - 4.4 不同场景下AI工具测评维度的权重示例-4
5. 不同场景下测评维度的权重设置参考
6. 总结

1.1 为何需要系统评估AI工具？



1. AI工具爆发带来的四大困境

- 数量爆炸：2024年全球新增AI工具呈现爆炸式增长
- 功能重叠：同类工具差异度不足30%，工具之间功能差异性较低
- 效果存疑：AI工具开发周期太短，很多用户反馈AI工具未达预期效果
- 成本黑洞：无规划采购导致浪费技术预算

2. 系统评估的三大核心价值

- 风险控制：避免敏感数据泄露，防范供应商绑定风险。
- 资源优化：缩短选型周期，精准匹配需求可节省采购成本。
- 效能保障：功能达标率提升，提升员工工作效率。

3. 当前评估的典型障碍

- 标准缺失、主观偏好主导、方法陈旧、维度单一 、短期测试误导、生态兼容性忽视。

目标：构建一套全面且实用的AI工具测评框架与方法论，为后续的工具深度测评奠定基础。

1.2 如何全面评估一个AI工具？12个评测维度概览

核心维度

1. 核心功能 (Core Functionality): 它能做什么？是否满足核心需求？
2. 效果质量 (Effect Quality): 它做得好不好？输出结果是否可靠、优质？
3. 易用性 (Usability): 它好不好用？用户体验是否流畅、直观？
4. 成本效益 (Cost-Effectiveness): 它值不值？投入与回报是否匹配？
5. 集成性 (Integration): 它能否融入现有流程？与其他工具的兼容性如何？
6. 安全合规性 (Security): 它是否安全可靠？数据隐私是否有保障？是否符合相关法规？
7. 社区支持 (Community Support): 遇到问题时，是否有足够的支持资源？
8. 创新前瞻性 (Innovation): 它是否代表了前沿技术？未来潜力如何？

可选维度（根据具体应用场景和需求，可考虑）

1. 偏见 (Bias)：评估工具在处理不同群体或数据时是否表现出不公平或歧视性的行为。
2. 透明度 (Transparency)：评估工具的决策过程是否清晰可解释，用户是否能够理解其工作原理。
3. 伦理考量 (Ethical Considerations)：评估工具的使用是否符合伦理规范，是否可能带来潜在的负面影响。
4. 环境影响 (Environmental Impact)：评估工具的能耗和对环境的影响。

- 面对层出不穷的AI工具，如何系统、客观地进行选择？
- 一个全面的评测框架至关重要。通过以下维度，能够更精准地评估AI工具的价值与适用性。

1.3 核心测评维度定义 - 核心功能 (Core Functionality)

1.定义

AI工具设计用于执行的主要任务和提供的核心特性。

2.关键考量

- 工具是否能够有效地解决用户的实际问题，满足其最主要的需求?
- 工具提供的功能是否完善且性能稳定可靠？
- 重要性： 核心功能是衡量工具价值的基础。即使功能再多，如果不能高效完成核心任务，价值将大打折扣。
- **注意：对于文本生成工具，核心功能可能包括文章撰写、摘要生成、方案起草等。评估时需考察这些功能的质量和效率。**

1.3 核心测评维度定义 - 效果质量 (Effect Quality)

1.定义

AI工具输出结果的优劣程度，涵盖准确性、创造性、真实性、流畅性以及一致性等多个方面 。

2.关键考量

- 文本生成： 逻辑性、可读性、语法规范
- 图像生成： 清晰度、艺术性、是否准确反映用户意图
- 音视频生成： 流畅度、真实感、定制性
- 评估方法： 结合定量指标（如BLEU评分用于翻译）和定性评估（人工评估创意性）。
- **注意： 不同类型的AI应用对效果质量的要求侧重点不同，需根据具体场景选择评估标准。**

1.3 核心测评维度定义 - 易用性 (Usability)

1.定义

AI工具的用户友好程度，包括界面是否直观、操作是否便捷、学习曲线是否平缓，以及是否提供多语言支持和辅助功能等。

2.关键考量

- 用户能否轻松上手并高效地使用工具，而无需过多培训或专业知识?
- 界面设计是否直观易懂？导航是否便捷？错误处理机制是否完善？用户帮助文档是否清晰？
- 重要性： 易于使用的AI工具能显著提高用户采用率和工作效率。
- 用户体验： 用户对AI应用的易用性和用户体验的满意度直接影响其使用意愿。

1.3 核心测评维度定义 - 成本效益 (Cost-Effectiveness)

1.定义

AI工具的成本投入与其所带来的价值或回报之间的平衡，包括订阅费用、API调用费用、可能需要的硬件要求以及使用工具所节省的时间成本等。

2.关键考量

- 在预算范围内，该工具是否能够提供足够的价值，例如提高工作效率或降低运营成本？
- 需要考虑直接经济成本和潜在的长期收益，例如提升竞争力或实现业务增长。
- 定价模式： 了解不同的定价模式（按使用量付费、订阅制、一次性购买等），并根据自身情况选择。
- 开源工具： 通常被认为是具有成本效益的选择，因为没有许可费用。

1.3 核心测评维度定义 - 集成性(Integration)

北京大学 PEKING UNIVERSITY

1.定义

AI工具与其他常用软件、平台或API的兼容性和连接能力。

2.关键考量

- 工具是否提供开放的API接口？
- 是否支持与其他常用应用的连接，例如CRM、ERP或ITSM工具?
- 是否能无缝地融入用户现有的工作流程，减少数据孤岛和重复操作，从而提高整体效率?
- 重要性： 良好的集成性能够减少用户在使用不同工具之间切换的成本，提升工作效率。

1.3 核心测评维度定义 - 安全合规性 (Security)

1.定义

评估AI工具时不可忽视的关键维度，主要考察工具在数据隐私保护、信息安全以及防范恶意攻击等方面的能力。

2.关键考量

- 用户需要确保其数据在使用AI工具时得到妥善的保护。
- 工具是否符合相关的法律法规和安全标准，例如欧盟的GDPR，国内的《生成式人工智能服务管理暂行办法》和《中华人民共和国网络安全法》。
- 关注数据加密措施、用户身份验证机制、隐私政策的透明度以及是否符合行业安全标准。
- 风险评估：包括对AI生成内容潜在风险的评估，例如是否会生成有害或不当的内容。

1.3核心测评维度定义 - 社区支持(Community Support)

1.定义

AI工具是否拥有活跃的用户社区、完善的文档、教程以及官方的技术支持。

2.关键考量

- 当用户在使用过程中遇到问题时，能否及时获得帮助和解决方案?
- 社区是否活跃？文档和教程是否完善？技术支持是否及时有效？
- 重要性： 活跃的社区能够提供丰富的用户经验分享和问题解答，完善的文档和教程能够帮助用户更好地理解和使用工具。
- 开源工具： 活跃的社区支持尤其重要，因为这往往是获取帮助的主要途径。

1.3 核心测评维度定义 - 创新前瞻性(Innovation)

1.定义

评估的是AI工具是否采用了最新的AI技术，是否具备独特的功能或特点，以及未来的发展潜力。

2.关键考量

- 用户需要关注工具是否能够引领行业趋势，为其提供持续的竞争优势。
- 考察工具所采用的算法、模型是否先进，是否具备其他同类工具所没有的独特功能。
- 关注开发团队是否持续投入研发并推出新的功能和改进。
- 重要性： 选择具有创新性的AI工具，有助于用户在快速发展的技术领域保持领先地位。

1.3 核心测评维度定义 - (可选)其他重要维度

其他维度（可选）

- 偏见 (Bias)： 评估工具在处理不同群体或数据时是否表现出不公平或歧视性的行为。
- 透明度 (Transparency)： 评估工具的决策过程是否清晰可解释，用户是否能够理解其工作原理。
- 伦理考量 (Ethical Considerations)： 评估工具的使用是否符合伦理规范，是否可能带来潜在的负面影响。
- 环境影响 (Environmental Impact)： 评估工具的能耗和对环境的影响。

1.4 AI工具测评维度定义与关键考量

测评维度	定义	关键考量
核心功能	工具设计旨在执行的主要任务和提供的核心特性	工具是否有效解决用户痛点，满足核心需求？功能是否完善，性能是否稳定可靠？
效果质量	工具执行核心功能时的表现水平，输出结果的准确性、可靠性、相关性	准确率、精确度、鲁棒性（对异常输入的处理）、生成内容的创造性、逻辑性、流畅性、稳定性。比如：文本生成是否逻辑清晰、可读性强？图像生成是否清晰、具有艺术性？音视频生成是否流畅、真实？
易用性	工具的用户友好程度，包括界面直观性、操作便捷性、学习曲线平缓度等	用户是否能轻松上手并高效使用？界面是否直观易懂？操作流程是否便捷？
成本效益	工具的成本投入与其带来的价值或回报之间的平衡	定价模式（订阅、按量付费、买断）、免费试用（额度）、总体拥有成本（含隐性成本如学习、维护）、与同类工具的性价比、预期投资回报率（ROI）。如：工具是否在预算范围内提供足够价值？是否能提高效率或降低成本？长期收益如何？
集成性	工具与其他常用软件、平台或API的兼容性和连接能力	工具是否能无缝融入现有工作流程？是否提供开放API？是否支持常用应用连接？
安全合规性	工具在数据隐私保护、信息安全、防范恶意攻击及符合相关法规的能力	用户数据是否得到妥善保护？是否符合相关法律法规和安全标准？
社区支持	获取帮助、学习资源和与其他用户交流的便利性	官方文档、技术支持响应速度与质量？社区是否活跃？文档和教程是否完善？开发者生态？
创新前瞻性	工具是否采用了最新的AI技术，是否具备独特的功能或特点，以及未来的发展潜力	技术领先性（模型新颖度）、更新频率、路线图（Roadmap）清晰度、对新趋势的适应能力、背后的研发实力如何？
(可选)偏见	工具在处理不同群体或数据时是否表现出不公平或歧视性的行为	工具是否会产生偏见或歧视性输出？
(可选)透明度	工具的决策过程是否清晰可解释，用户是否能够理解其工作原理	用户是否能理解工具的工作原理和决策过程？
(可选)伦理考量	工具的使用是否符合伦理规范，是否可能带来潜在的负面影响	工具的使用是否符合伦理规范？是否存在潜在负面影响？
(可选)环境影响	工具的能耗和对环境的影响	工具的能耗如何？是否对环境造成影响？

2.1 常见的6个测评方法

北京大学 PEKING UNIVERSITY

1

实测案例分析
(Practical Testing and Benchmarking)

2

用户反馈收集与分析
(User Feedback Collection and Analysis)

3

建立结构化评分体系
(Establishing a Structured Scoring System)

4

(可选)启发式评估与专家评审
(Heuristic Evaluation and Expert Reviews)

5

(可选) A/B测试与对比分析
(A/B Testing and Comparative Analysis)

6

(可选)真实场景数据集测评
(Real-scenario data evaluation)

学习交流可加AI肖睿团队助理微信号（ABZ2829）

2.2 测评方法1：实测案例分析

- **方法：**通过实际操作来作为评估AI工具的性能和效果
- **核心理念:** “Talk is cheap, show me the code (or results).” **实践是检验真理的唯一标准。**也是最直观、最贴合实际需求的评估方式，能发现AI工具官方文档中未提及的问题。
- **关键步骤：**
  - 定义场景: 设计贴近自身业务需求、具有代表性的具体测试任务/案例（Use Cases），覆盖工具的核心功能和常见应用场景。
    - ✓ 例如：测试文本生成工具在不同类型和长度的文章、摘要、翻译等任务上的表现。
    - ✓ 例如：测试图像生成工具在生成特定风格和内容的图像，以及进行图像编辑等方面的能力。
  - 基准测试： 用于对比不同工具在相同任务上的表现。
  - 使用标准化的数据集和评估指标，尽可能量化关键指标（如：生成报告耗时、识别准确率、代码生成通过率）来客观比较性能差异。
  - 基于任务的评估框架： 根据工具执行用户相关特定功能的能力进行评估。
  - 尽早确定“良好”输出的标准： 对于评估过程至关重要.
  - 对比分析: 与预期结果、人工操作或其他工具进行对比。

2.2 测评方法2：用户反馈收集与分析

- **重要性：**用户的真实体验对于评估AI工具的易用性和实用性至关重要
- **测评优势：**
  - 获取大规模、多视角的实际使用体验，弥补单点测试的不足。
  - 了解工具在不同场景下的表现和用户满意度。
- **常用方法：**
  - 查阅专业评测网站、技术博客、论坛。
  - 分析应用商店评分、用户评论。
  - 关注并分析社交媒体上的讨论和口碑。
  - 如果条件允许，进行小范围用户调研或访谈，包括内部员工调查
- **分析过程：**对收集到的用户反馈进行分类、整理和分析，提取有价值的见解，如用户满意度、遇到的问题、改进建议。
- **分析要点：**
  - ①关注共性问题和高频赞扬点 ②区分不同用户群体(如开发者/普通用户/业务人员)的反馈侧重 ③警惕水军或极端评价，注重信息交叉验证。
  - 用户评论和证词： 提供关于AI工具在实际使用中的可用性、满意度以及潜在问题的宝贵信息。
  - 情感分析： 可用于评估大量用户评论数据。

2.2 测评方法3：结构化评分体系

■ **方法：为每个测评维度设计明确的评分标准和等级**

测评实现

- 为每个测评维度设定评分标准（如1-10分，或优/良/中/差或1-5星 ★★★☆☆）。
- 根据实测、反馈和研读结果，对每个工具在各维度上进行打分。
- 结合后续的权重进行加权计算总分。（可选）

优势

- 结构化的评分体系能够对不同AI工具在预定义的评估维度上进行更客观和系统的比较，促进数据驱动的决策。

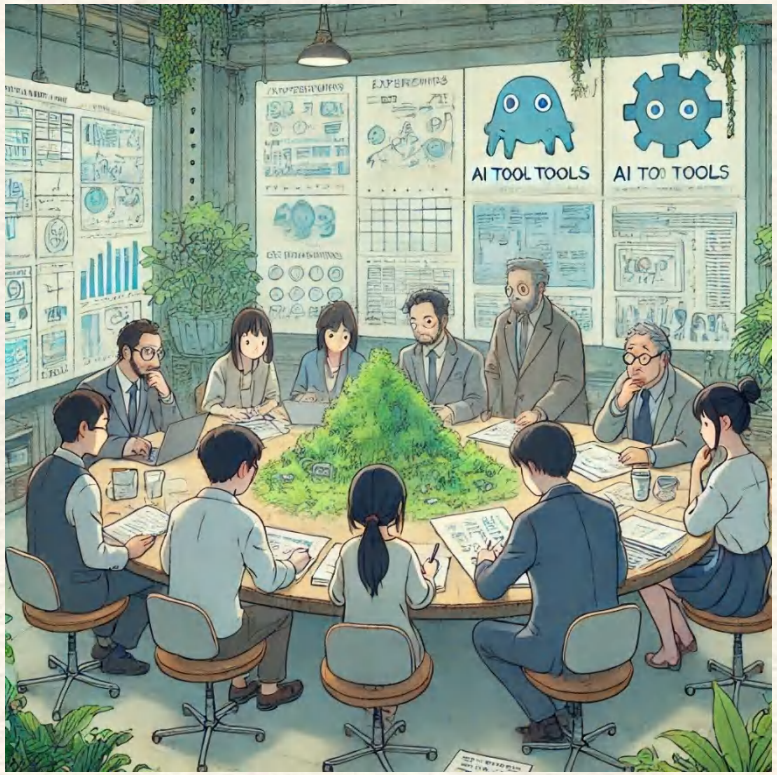
关键要求

- 评分体系应确保客观性、一致性和可操作性，减少主观偏差，保证评估结果的可靠性。

评分表示例

	核心功能效果	易用性	成本	...
Tool A	5	8	5	
Tool B	6	4	7	
Tool C	9	3	9	
....				

2.2 测评方法4：(可选)启发式评估与专家评审



- **启发式评估**
  - 由可用性专家根据预定义的可用性原则（启发式原则）评估工具界面和交互设计
  - 能有效识别评估过程早期潜在的设计和可用性问题
  - 通常聘用五到八名评估人员
- **专家评审**
  - 邀请相关领域的专家对工具的功能、效果和潜在风险进行评估
  - 专家可利用其经验和知识识别潜在问题

2.2 测评方法5：(可选) A/B测试与对比分析



- **A/B测试**
  - 直接比较不同AI工具在相同任务上的表现
  - 直观展示它们在特定任务上的性能差异
- **对比分析**
  - 基于预定义的评估维度，对不同的工具进行系统的比较，突出它们之间的差异和优劣
  - 目的： 为确定哪种AI工具在特定条件或特定任务下表现更好提供实证依据

2.2 测评方法6：(可选) 真实场景数据集测评



- **真实场景数据集测评**
  - 为了得到不同AI工具在某个具体场景下的能力对比，可以使用该场景下的专业测试数据集。
  - 批量获取该场景下专业问题的回复，设计该场景的专业测评维度，针对这些维度进行人工打分，并最终得到每个AI工具在该专业场景下的评测总分。
- **评测步骤**
  1. 由该领域的专业人员给出大批量的测评问题。
  2. 批量获取不同AI工具的回复。
  3. 设定该专业场景下AI回复的测评维度及权重设定。
  4. 该领域的专业人员针对预定义的测评维度进行人工打分。
  5. 获取专业人员对该AI工具的能力评价，并计算每个AI工具在该领域的测评总分，得到最终测评结果。
- **教育场景数据测评参考：**https://arxiv.org/abs/2402.07913

3.1 如何根据自身需求选择测评维度权重

- **核心理念:**
  - 不同用户、不同场景，对AI工具的需求侧重点不同。因此，各测评维度的重要性（权重）也应不同。个性化权重可以使评估结果更贴合实际需求，避免被“平均分”误导。

- **如何确定权重? –” 自我提问 “（根据自身的需求和应用场景，对不同的测评维度赋予不同的权重）**
  - 我的核心目标是什么？(是提高效率、降低成本、提升创作质量，还是探索新技术？)
  - 谁是使用者？(个人、小团队、大企业？技术背景如何？)
  - 预算限制是多少？(对免费/低成本敏感，还是愿意为高性能付费？)
  - 是否需要与现有系统集成？(集成性要求有多高？)
  - 数据敏感性如何？(安全性是否是首要考虑？)
  - 对易用性要求高吗？(需要快速上手，还是愿意投入时间学习？)

- **工具：**加权评分模型和决策矩阵

3.2 权重应用：构建你的决策矩阵

实现步骤：

1. **明确评估目标**：清晰定义使用AI工具要解决的问题或实现的目标
2. **列出候选工具**：Tool A, Tool B, Tool C...
3. **识别关键维度**：根据评估目标，进行“自我提问”，确定最重要的5-8个维度
4. **分配维度权重**：为每个关键维度分配一个表示其重要程度的数值权重（总和可以为1或100%）。例如：核心功能40%，易用性20%，成本20%，集成性10%，安全性10%
5. **评分与计算**：基于之前的测评方法对每个工具在各个维度上进行打分，如1-10分
6. **计算加权得分**：每个工具的加权得分 = Σ (维度得分 × 维度权重)
7. **结果分析与决策**：比较不同工具的总加权得分，选择最符合自身需求的工具

4.1 不同场景下AI工具测评维度的权重示例-1

■ **个人内容创作者：追求效果与效率，灵活运用AI激发创意**

- 独立工作者，如博主、自由撰稿人、设计师、视频UP主等。目标是快速、高质量地生成吸引人的内容，提升个人品牌影响力或满足客户需求。

维度 (权重)	原因分析
效果质量 (高权重)	内容是创作者的生命线。无论是文案的吸引力、图片的创意性，还是视频的流畅度，直接决定了作品的传播力和价值。AI输出的质量必须足够高，才能真正帮助创作者提升作品水平，而非制造“数字垃圾”。
易用性 (较高权重)	创作者通常独立工作，没有专门的技术支持。工具必须直观易懂，学习曲线平缓，能快速上手并融入创作流程，减少在工具摸索上花费的时间，专注于内容本身。
成本效益 (中等权重)	个人预算有限，需要考虑投入产出比。愿意为能显著提升效率或质量的工具付费，但对价格敏感，会倾向于选择性价比高的方案（如免费版、按需付费或价格合理的订阅）。
集成性、安全合规性等（较低权重）	主要关注账号安全和个人作品不被泄露，对企业级的复杂安全合规要求不高。
小结	对个人创作者而言，AI工具是提升内容质量和创作效率的“超级助手”，效果和易用性是核心考量。

4.1 不同场景下AI工具测评维度的权重示例-1

- **自媒体博主，希望提升内容生产效率与创意**
  - 步骤 1 & 2 (关键需求与限制):
    - **核心痛点**: 写作瓶颈，需要快速生成多样化的初稿和社交媒体帖子。
    - **期望结果**: 每周内容产出翻倍，减少构思时间。
    - **关键限制**: 个人预算有限，倾向免费或低成本方案；技术非专长，需要工具易于上手。无敏感数据处理需求。
  - 步骤 3 (映射与权重分配 - 思考过程):
    - 因为 “快速生成多样化初稿” 是核心，所以 【核心功能】(如生成速度、风格多样性) 和 【效果质量】(内容流畅度、创意性) 权重高。
    - 因为 “预算有限” ，所以 【成本效益】权重高。
    - 因为 “技术非专长，需易上手” ，所以 【易用性】权重高。
    - 因为 “无敏感数据” ，所以 【安全性】权重可以较低。
    - 因为 “独立使用，无需协作” ，所以 【集成性】(与其他系统对接) 权重较低。
- **结论 (权重画像)**: 对这位博主而言，选型时应高度关注：效果质量、易用性、成本效益。

4.2 不同场景下AI工具测评维度的权重示例-2

■ **小型企业：聚焦降本增效，AI赋能核心业务流程**

➢ 规模小，资源相对有限，追求快速增长和市场适应能力。希望利用AI解决具体业务问题，提升效率，降低运营成本，增强竞争力。

维度 (权重)	原因分析
成本效益 (高权重)	预算严格，每一笔投入都需要看到明确的回报（ROI）。AI工具必须能带来可量化的价值，如节省人力成本、提高销售转化率、优化营销支出等
核心功能 (较高权重)	需要AI工具能够精准解决业务痛点，如自动化客服、智能营销文案生成、数据分析、简化内部管理等。功能必须稳定可靠，满足业务场景的实际需求。
易用性 (中等权重)	员工可能缺乏专业AI知识，培训资源有限。工具需要相对容易上手，以便团队快速采纳并融入日常工作，避免因使用复杂而降低效率。
集成性（中等权重）	可能已经在使用一些核心业务软件（如CRM、项目管理工具、数据中台）。AI工具若能与这些系统集成，可以打通数据，提升整体工作流效率。
创新前瞻性（适中关注）	对于小型企业而言，AI工具的首要价值在于解决实际业务问题、提高效率和控制成本，因此其对创新性的关注会适中，需确保技术成熟稳定、易于部署且能带来明确回报，而非单纯追求前沿性。
小结	小型企业视AI为提升核心竞争力的杠杆，优先考虑能带来直接经济效益、解决实际问题且易于部署和使用的工具。

4.2 不同场景下AI工具测评维度的权重示例-2

- **某创业公司电商团队（约3-5人），需要制作营销图片和文案**
  - 步骤 1 & 2 (关键需求与限制):
    - **核心痛点:** 缺乏专业设计/文案人员，产品推广素材制作慢、成本高。
    - **期望结果**: 快速生成符合品牌调性的商品图、广告文案，提升点击率。
    - **关键限制**: 团队成员技能不一，需要协作；预算有限但愿意为效果付费；素材需符合平台规范；涉及一定的用户数据（非核心）。
  - 步骤 3 (映射与权重分配 - 思考过程):
    - 因为 “制作营销图片和文案” 是核心，所以 【核心功能】(图像生成、文案撰写能力) 权重最高。
    - 因为 “符合品牌调性” 、 “提升点击率” 要求高，所以 【效果质量】(图像真实感/美观度、文案吸引力、转化导向) 权重高。
    - 因为 “团队协作，技能不一” ，所以 【易用性】权重高。
    - 因为 “预算有限但愿为效果付费” ，所以 【成本效益】权重中高 (需要评估投入产出比)。
    - 因为 “可能涉及用户数据，需合规” ，所以 【安全性】权重中等。
    - 因为 “可能需要接入电商后台或广告平台” ，所以 【集成性】权重中等。
- **结论 (权重画像)**: 该团队应优先考虑：核心功能、效果质量、易用性，并重点评估成本效益和安全性。

4.3 不同场景下AI工具测评维度的权重示例-3

- **大型企业：重视安全合规与系统集成，AI驱动规模化创新**
  - 拥有庞大的组织架构、复杂的业务流程和海量数据。关注风险控制、数据安全、合规性，并需要将AI无缝整合到现有IT基础设施中，实现规模化应用。

维度 (权重)	原因分析
安全合规性（高权重）	涉及大量敏感数据（客户、员工、商业机密），任何安全漏洞或数据泄露都可能导致巨大的经济损失和声誉危机。必须满足严格的行业法规和内部安全标准（如GDPR, ISO 27001）
集成性 (较高权重)	AI工具需要融入复杂的现有IT生态系统，支持单点登录（SSO），确保数据流通顺畅，避免形成新的信息孤岛。
成本效益 (中等权重)	虽然预算相对充足，但大规模部署成本高昂，仍需评估总体拥有成本（TCO）和长期投资回报。 关注规模化应用下的成本控制和效益最大化。
社区/供应商支持（中等权重）	大规模部署和复杂集成需要可靠的技术支持和服务保障。 供应商的响应速度、服务水平、培训资源和问题解决能力至关重要。
核心功能&效果质量	基础要求。工具必须能稳定、高效地完成核心任务，并在大规模应用中保持高质量输出， 满足业务部门的专业标准。
小结	大型企业在引入AI时，将安全、合规与集成置于首位，确保新技术能安全、稳定地融入现有体系， 并获得可靠的供应商支持。

4.3 不同场景下AI工具测评维度的权重示例-3

■ **某大型公司的企业法务部门，需要快速审阅合同、提取关键条款**

- 步骤 1 & 2 (关键需求与限制):
  - **核心痛点**: 合同审阅量大、耗时长，易遗漏风险点。
  - **期望结果**: 提高审阅效率，确保关键条款（如赔偿、管辖权）提取的准确性。
  - **关键限制**: 处理高度敏感的商业机密，数据安全和隐私是绝对红线；需要与内部文档管理系统集成；预算充足但需论证价值；使用者为专业人士，可接受一定学习曲线。
- 步骤 3 (映射与权重分配 - 思考过程):
  - 因为 “处理高度敏感信息”，所以 【安全性】权重最高 (Veto 项)。
  - 因为 “准确提取关键条款”是核心要求，所以 【效果质量】(精准度、召回率) 权重极高。
  - 因为 “审阅合同、提取条款”是核心功能，所以 【核心功能】(法律文本理解能力) 权重极高。
  - 因为 “需要与内部系统集成”，所以 【集成性】权重高。
  - 因为 “使用者为专业人士”，所以 【易用性】权重适中 (功能强大优先于极致简单)。
  - 因为 “预算充足但需论证价值”，所以 【成本效益】权重中等 (更看重长期价值和风险规避)。

■ **结论 (权重画像)**: 对这位博主而言，选型时应 高度关注：效果质量、易用性、成本效益。

4.4 不同场景下AI工具测评维度的权重示例-4

■ 科研机构：追求前沿与精准，AI助力探索未知边界

- 大学、实验室、研究中心等。目标是利用AI进行数据分析、模拟预测、模式识别等，推动科学发现和技术创新。对结果的准确性、方法的新颖性要求极高。

维度 (权重)	原因分析
效果质量 (高权重)	科研成果的准确性和可靠性是基石。AI工具输出的精度、鲁棒性、可复现性直接影响研究结论的有效性。
创新前瞻性 (较高权重)	科研本身就是探索前沿。研究人员倾向于采用最新的AI模型、算法或具有独特功能的工具，以解决现有方法难以处理的问题，或开辟新的研究方向。
透明度/可解释性 (中等权重)	理解AI的决策过程对于验证研究结果、撰写论文、以及确保研究符合伦理规范非常重要。需要了解模型的工作原理，而不仅仅是接受“黑箱”输出。
集成性（中等权重）	对于已在使用一些核心业务软件（如CRM、项目管理工具、邮件系统）。AI工具若能与这些系统集成，可以打通数据，提升整体工作流效率。
成本效益（较低关注）	科研机构评估AI工具时，首要关注其能否支持尖端研究与实现突破，因此性能和功能的重要性远超成本考量，对成本效益的敏感度可能相对较低。
小结	科研场景下，AI工具的核心价值在于其推动知识发现的能力，因此效果的精准性、技术的先进性和方法的透明度是评估的关键。

5. 不同场景下测评维度的权重设置参考

测评维度	个人内容创作者	小型企业	大型企业	科研机构
核心功能	中	高	高	高
效果质量	高	较高	高	高
易用性	较高	中	中	中
成本效益	中	高	中	低
集成性	低	中	较高	中
安全合规性	低	中	高	中
社区支持	低	中	较高	中
创新前瞻性	中	中	中	较高

6. 总结

本节详细介绍了AI工具测评的关键维度和方法，并阐述了如何根据自身需求选择测评维度的权重。建立完善的测评框架对于用户在琳琅满目的AI工具中做出明智的选择至关重要。下一部分将开始应用本节介绍的测评框架和方法，对不同类别的AI工具进行深入测评与探讨。

内容目录

一、从AI模型到AI工具 P6

1. DeepSeek与大模型理论概述 P8
2. AI工具浪潮：为何现在关注AI工具？ P15
3. AI 工具如何赋能千行百业？ P19
4. 系统性认识AI工具 P20
5. AI工具分类 P21
6. AI工具测评的核心价值 P22

二、AI工具测评框架与方法论 P24

1. 12个测评维度定义 P27
2. 6个测评方法详解 P38
3. 根据自身需求选择测评维度权重 P45
4. 不同场景下AI工具测评维度的权重示例 P47
5. 不同场景下测评维度的权重设置参考 P54
6. 总结 P55

三、AI工具测评 P57

1. AI工具测评一览表 P59
2. 文本生成与处理类AI工具测评 P60
3. 图像生成与编辑类AI工具测评 P108
4. 音视频生成与处理类AI工具测评 P171
5. 代码生成与辅助类AI工具测评 P245
6. 大模型管理与应用类AI工具测评 P289

四、AI工具选型指南与未来展望 P310

1. 构建个性化AI工具箱选型决策框架 P312
2. 构建你的AI工具箱：个人AI工具战略选型流程 P313
3. 构建你的AI工具箱：团队AI工具战略选型考量 P314
4. 从工具到工作流，1+1 > 2 的组合力量 P315
5. AI工具的发展趋势 P316
6. 总结 P318

三、AI工具测评

- 针对市面上已经出现的种类繁多的AI工具，本部分对目前主流的人工智能工具进行了全面的分类评估，涵盖了**文本生成与处理**、**图像生成与编辑**、**音视频生成与处理**、**代码生成与辅助类**以及**大模型管理应用**等重要类别。
- 通过对各类工具的核心功能、性能表现、易用性、定价以及适用场景的详细分析和横向比较，我们旨在为读者提供一个清晰、深入的AI工具概览。这些评估结果强调了人工智能技术在各个领域的巨大潜力和快速发展，同时也指出了不同工具在特定任务中的优势和不足。
- 随着技术的不断进步，我们可以预见，人工智能工具将在未来的工作和生活中扮演越来越重要的角色，为各行各业带来效率提升和创新机遇。

学习交流可加AI肖睿团队助理微信号（ABZ2829）

三、AI工具测评

1. AI工具测评一览
2. 文本生成与处理类AI工具测评
   - 2.1 AI文本工具深度测评总述
   - 2.2 核心AI工具测评（DeepSeek、Gemini、通义千问 、豆包、Kimi ）
   - 2.3 文本类AI工具横向对比
   - 2.4 选型指南（构建决策矩阵、常见场景工具推荐、教学场景案例）
3. 图像生成与编辑类AI工具测评
   - 3.1 AI图像工具深度测评总述
   - 3.2 图像生成类AI工具测评（Midjourney 、即梦AI、可灵AI、 GPT-4o、Gemini ）+实测总结
   - 3.3 图像编辑类AI工具测评（百度AI图片助手、SnapEdit ）+ 实测总结
   - 3.4 图像生成+编辑类AI工具测评（通义万相、豆包、腾讯元宝、阶跃星辰 ）+ 实测总结
4. 音频生成与处理类AI工具测评
   - 4.1 AI音频工具深度测评总述
   - 4.2 音频生成与处理类AI工具测评（海螺AI、MINIMAX、Noiz.ai ）+实测总结
   - 4.3 AI视频工具深度测评总述
   - 4.4 视频生成与处理类AI工具测评（即梦AI 、可灵AI 、Vidu、通义万相 ）+实测总结
   - 4.5 AI数字人工具深度测评总述
   - 4.6 数字人AI工具测评（Heygen、硅语、智课、奇妙元、闪剪 ）+实测总结
5. 代码生成与辅助类AI工具测评
   - 5.1 AI编程工具深度测评总述（含工具统计）
   - 5.2 IDE软件测评（Cursor 、Trae ）
   - 5.3 编程插件测评（Cline 、GitHub Copilot、通义灵码、 CodeGeeX ）
   - 5.4 编程网站测评（MGX、heyBossAI ）
   - 5.5 AI编程工具-总结与思考
6. 大模型管理与应用类AI工具测评
   - 6.1 大模型管理与应用类工具深度测评总述（含工具统计）
   - 6.2 大模型管理与应用类工具深度测评（CherryStudio、LMStudio、Chatbox ）

1. AI工具测评一览



Tips：

- 面对当前纷繁复杂的AI工具市场，为了帮助大家有效选择能解决实际问题的工具，我们本次的评测选型经过了全面考量。我们重点关注工具的实际效用和场景表现，并参考了市场热度，力求推荐实用性强、表现突出的代表性AI工具。当然，‘使用体验’具有主观性，且评测范围有限，无法面面俱到。我们真诚欢迎您分享不同见解，共同交流探讨。
- 测评日期：2024.04



北京大学
PEKING UNIVERSITY

01

文本生成与处理类AI工具测评

02

图像生成与编辑类AI工具测评

03

音视频生成与处理类AI工具测评

04

代码生成与辅助类AI工具测评

05

大模型管理与应用类AI工具测评

2.1 文本AI基础：大语言模型 (LLM) 与核心应用

■ 大语言模型基础

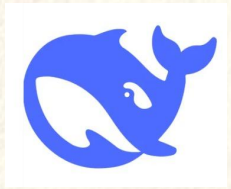
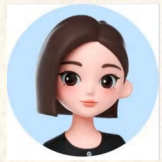
- 基于海量文本数据（如书籍、网页、论文等）训练的深度神经网络模型，通过自监督学习掌握语言规律。
- 核心特性：
  - 理解能力：解析语义、情感、意图（如区分反问与疑问）
  - 生成能力：输出连贯文本（从一句话到长篇文档）
  - 泛化能力：处理未见过的任务（通过少量示例学习新技能）

■ 常见应用场景

应用场景	举例
内容创作	博客文章、营销文案、邮件、小说、剧本等。
信息摘要	长篇文章、报告、会议纪要的关键信息提取。
机器翻译	多语言互译，保持语境和流畅度。
智能问答	回答事实性问题、提供解释、进行知识检索。
代码辅助	代码生成、解释、调试、文档撰写。
创意激发	头脑风暴、生成点子、提供不同视角。

2.1 主流文本AI工具深度测评

- **测评目的:**
  - 全面测评普通用户在真实场景下最易用的工具，展示其效果对比，为用户提供选型依据。
- **测评维度:**
  - 核心功能与效果 (重点)
  - 易用性与交互体验
  - 定价与性价比
  - 集成与扩展能力 (API)






- **测评工具选型：**DeepSeek、Gemini、通义千问、豆包、Kimi
- **测评工具对比:**
  - **DeepSeek：**能力强，六边形战士，尤其是代码及逻辑推理能力，输出内容风格符合国人表达习惯。
  - **Gemini：**代码能力及统筹规划能力超强，长文本理解能力强且有深度调研，超适合复杂推理工作。
  - **通义千问 ：**Qwen3推理速度快，并可对推理进行控制，工具API性价比高，模型更新迭代速度快。
  - **豆包：**文本理解及生成能力弱于DeepSeek，但其工具集成性较好。
  - **Kimi：**kimi1.5长思考能力较好、语言风格符合国人表达习惯，较为风趣，具有长文本理解能力。

2.2 AI工具深度测评 - DeepSeek

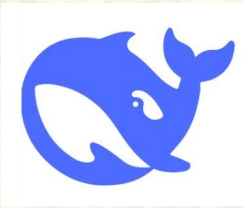
DeepSeek 简介与核心技术

- **开发背景:** 由「深度求索 (DeepSeek AI)」公司开发，专注于AI前沿研究与应用。
- **核心定位:** 在代码生成和中文理解方面具有优势，提供开源模型和商业服务。
- **核心技术：**
  - DeepSeek V3 & R1模型
  - 高效训练方法、对特定领域的优化（代码、数学、推理等）
  - 开源策略，吸引大量开发者和研究者
- **主要产品形态:** 网页、APP、API服务。
  - 链接：https://www.deepseek.com/



混合专家 MOE | 多头潜注意力 MLA | 多Token预测 MTP | 强化学习 GRPO | 测试时计算 TTC | 混合精度训练 FP8 | 通讯优化 DualPipe | 并行训练框架 HAI | 直接硬件编程 PTX

2.2 DeepSeek实测总结



- **DeepSeek总结：核心优势与主要限制:**
  - DeepSeek是一款在代码处理和中文能力上表现卓越，兼具开放性和性价比的AI文本工具。
- **核心优势:**
  - V3的代码能力强（代码编写、补全等，R1的代码能力不如V3，使用时建议不要勾选R1及联网搜索功能）
  - 中文处理能力非常优秀，输出内容符合国人书写习惯和国内文字梗
  - 开源模型能力跻身世界大模型第一梯队
  - DeepSeek是国内大模型API平台中最具性价比的服务提供商
- **主要限制:**
  - 深度思考模式容易陷入逻辑思考循环且推理时间太长非常影响用户体验
  - 网页版联网搜索时需要注意来源的准确性及时效性

2.2 DeepSeek实测：核心功能 - 网站生成

提示词：

给一家AI创业公司做一个首页，公司名：向量智能，公司介绍：向量智能一家专注于人工智能技术研发的科技企业，致力于将科研成果转化为实际市场应用，主要提供AI人才培养和AI技术赋能。为政府、企业和高校提供全方位的解决方案，推动智能化转型。输出html 。

■ **DeepSeek V3生成页面效果:**



一键运行生成的html代码



■ **实践评价:**

- 输出的代码格式工整，代码规范优秀。
- 页面效果生动、UI风格，页面动效符合提示词描述。
- 公司首页的全部网页内容完善，包含常见功能内容。

2.2 DeepSeek实测：核心功能 - UI设计

提示词：
用HTML设计一个拟态风格的睡眠白噪音的APP，要求适配iOS/Android屏幕尺寸，配色用莫兰迪淡雅色系。

- **DeepSeek V3生成页面效果:**





2.2 DeepSeek实测：核心功能 - UI设计

提示词：

1.设计一个小清新拟态风格的睡眠白噪音的手机app界面，简洁清新，输出html

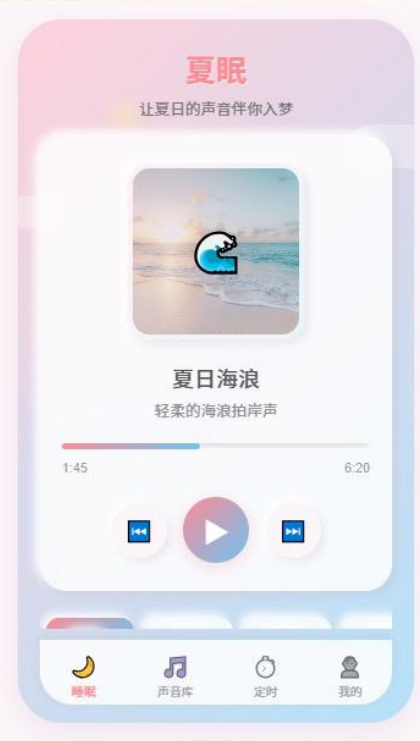
2.再设计一版小清新又可爱的，画幅是9:16，符合手机app尺寸

3.页面显示的按钮不完整，请调整一下，再注意要用简洁可爱的拟态风格，按钮可以点击

■ **DeepSeek V3生成页面效果:**





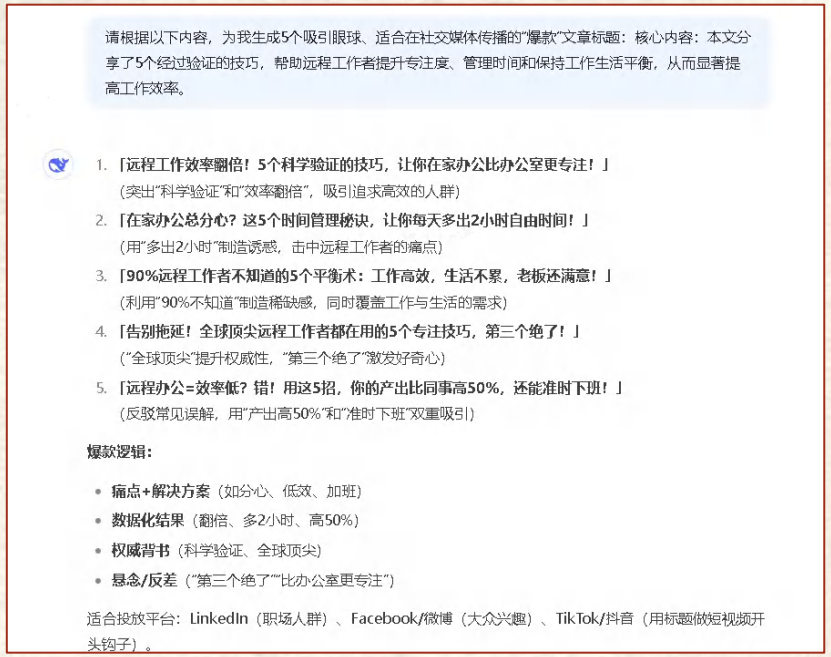


2.2 DeepSeek实测：核心功能 - 爆款标题生成

- **测试任务描述：** 输入同一段核心内容描述（例如：“一篇关于提高远程工作效率的5个实用技巧的文章”），要求DeepSeek生成5个吸引人的“爆款”标题。

> **提示词**：
> 请根据以下内容，为我生成5个吸引眼球、适合在社交媒体传播的“爆款”文章标题：核心内容：本文分享了5个经过验证的技巧，帮助远程工作者提升专注度、管理时间和保持工作生活平衡，从而显著提高工作效率。

- **DeepSeek V3输出结果：**



请根据以下内容，为我生成5个吸引眼球、适合在社交媒体传播的"爆款"文章标题：核心内容：本文分享了5个经过验证的技巧，帮助远程工作者提升专注度、管理时间和保持工作生活平衡，从而显著提高工作效率。

1. **「远程工作效率翻倍！5个科学验证的技巧，让你在家办公比办公室更专注！」**
   （突出"科学验证"和"效率翻倍"，吸引追求高效的人群）
2. **「在家办公总分心？这5个时间管理秘诀，让你每天多出2小时自由时间！」**
   （用"多出2小时"制造诱惑，击中远程工作者的痛点）
3. **「90%远程工作者不知道的5个平衡术：工作高效，生活不累，老板还满意！」**
   （利用"90%不知道"制造稀缺感，同时覆盖工作与生活的需求）
4. **「告别拖延！全球顶尖远程工作者都在用的5个专注技巧，第三个绝了！」**
   （"全球顶尖"提升权威性，"第三个绝了"激发好奇心）
5. **「远程办公=效率低？错！用这5招，你的产出比同事高50%，还能准时下班！」**
   （反驳常见误解，用"产出高50%"和"准时下班"双重吸引）

爆款逻辑：

- **痛点+解决方案**（如分心、低效、加班）
- **数据化结果**（翻倍、多2小时、高50%）
- **权威背书**（科学验证、全球顶尖）
- **悬念/反差**（"第三个绝了""比办公室更专注"）

适合投放平台：LinkedIn（职场人群）、Facebook/微博（大众兴趣）、TikTok/抖音（用标题做短视频开头钩子）。

- **实践评价：**
  - 每个标题都精准融合数字、痛点、承诺和悬念（如：“多出2小时”“第三个绝了”），符合社交媒体传播规律。
  - 角度多样（效率、时间、平衡、权威背书），且通过反问、对比、数据等手法避免同质化，创意突出。
  - 标题紧密围绕“远程工作效率提升”的核心，5个技巧、专注度、时间管理等关键词均得到准确覆盖，无偏离。

2.2 DeepSeek爆款标题生成实际案例

- **实际任务描述：**编写一篇关于人工智能通识教育洞见 + 宣讲活动 + 教材推介的推文。DeepSeek中文理解能力较强，输出标题更符合国内自媒体习惯，因此使用DeepSeek做标题优化。

- **输入 Prompt:**

提示词：

生成公众号推文的标题：
#角色：你是个人工智能专家，主要分享人工智能教育的专业洞见
#目标受众：各大高等职业教育从业者
#内容方向：人工智能通识教育解决方案，加教材推介
#标题需求：运营“人群+痛点输出+解决方案”的方法，增加效果承诺；
#适配平台：公众号
#风险避免：避免使用被算法标记的诱导词

- **DeepSeek V3输出结果:**

1. 《高职教师AI教学难落地？一套通识教育方案+权威教材，快速构建教学体系》
2. 《职业院校AI课程建设遇瓶颈？详解通识教育解决方案，3步实现高质量授课》
3. 《高职教师如何教好AI通识课？从课程设计到教材选择，一站式解决方案来了》
4. 《职业院校AI教学资源匮乏？这套方案+配套教材，助力轻松开课》
5. 《高职院校如何高效开展AI教育？通识课建设指南+优质教材推荐，提升教学效果》
6. 《职业院校AI通识教育无从下手？专家方案+精选教材，让教学更轻松》
7. 《高职教师必看：AI通识课教学难题如何破解？这套方案让课堂效果翻倍》

最终选题：

《人工智能通识课难题如何破解？这套方案让课堂效果翻倍》

2.2 DeepSeek实测：核心功能 - 长文摘要

- **测试任务描述：** 输入一篇中等长度的文章（约1000-1500字，选择一个有时效性或专业性的主题，例如“近期AI伦理法规进展”），要求DeepSeek生成一段150字左右的核心内容摘要。

> **提示词**：
> 请将以下文章内容 [粘贴文章前几段或全文，或提供文章链接让其自行抓取总结 - 如果支持] 总结成一段约150字的核心摘要，要求准确反映文章主旨和关键论点。[此处放置文章节选或说明文章来源]。

- **DeepSeek V3输出结果：**

符合一定条件的会员一年有两次平价购买资格，之后慢慢就转变为需抽签购买。

通过平价飞天茅台引流背后，山姆无疑已是商超巨头沃尔玛的重要增长动力。

沃尔玛CEO道格·麦克米伦曾公开表示，2024年沃尔玛中国市场表现相当强劲、保持两位数增长，尤其是山姆超市和其电商业务表现出色。其中，会员费亦是山姆业务重要的构成，山姆此前就曾披露，截至2023年山姆会员店中国会员数已突破500万；而今年中国春节销售旺季期间，山姆会员费收入增长35%。

过去两年，山姆在中国快速扩店，包括进入泉州晋江、张家港等县级市下沉市场。近期沃尔玛中国区总裁朱晓静称，预计今年将有8家山姆店年销破5亿美元，而2023年这一数据为两家。

【核心摘要】
山姆会员店近期在广州、深圳、珠海三地向卓越会员投放2.4万瓶1499元飞天茅台，今年累计放量达12万瓶，频次和规模显著高于往年。此举通过严苛抽签门槛（如年费680元的卓越会员资格、消费频次要求等）吸引中产消费者，既提升会员活跃度，也助力茅台扩大“茅粉”圈层。贵州茅台通过商超渠道（如山姆、盒马）加大平价茅台投放，以稳定市场价格、缓解经销商压力，同时应对行业动销困境。对消费者而言，中签转售单瓶可获利超600元，但参与门槛逐年提高。山姆借此策略不仅刺激会员消费和办卡，还推动业绩增长，2024年春节会员费收入增长35%。双方合作形成共赢：茅台触达高净值人群，山姆强化市场竞争力，加速在华扩张。

- **实践评价**
  - 摘要包含了原文的核心信息，涵盖山姆放量茅台、会员门槛、茅台渠道策略及双方合作动机等关键内容。
  - 语言流畅简洁，但长度不符合要求，逻辑清晰，无冗余表述。
  - 专业术语准确，如“茅粉”“动销”“批价”等均贴合原文语境，解释恰当。

2.2 DeepSeek实测：核心功能 - 专业问答

- **测试任务描述：** 提出一个具有一定专业性的问题（例如，技术类：“解释一下Transformer模型中的自注意力机制原理”，或法律类：“简述GDPR规定的数据主体权利”）。

> **提示词**：
> 请用简洁明了的语言解释一下Transformer模型中自注意力机制 (Self-Attention Mechanism) 的核心原理和作用。

- **DeepSeek V3输出结果：**



请用简洁明了的语言解释一下Transformer模型中自注意力机制 (Self-Attention Mechanism) 的核心原理和作用。

自注意力机制（Self-Attention）是Transformer的核心组件，其核心原理和作用可简明概括如下：

核心原理

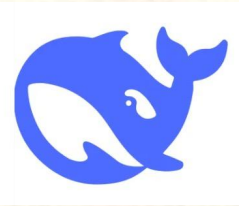
1. **动态权重分配**
   对输入序列的每个词，计算它与其他所有词的关联度（注意力分数），根据相关性动态分配权重，而非固定权重。
2. **三步计算**
   - **查询（Q）、键（K）、值（V）**：每个词通过线性变换生成这三组向量。
   - **注意力分数**：通过Q与K的点积计算词间相关性（后经Softmax归一化）。
   - **加权聚合**：用分数对V加权求和，得到当前词的输出（聚焦重要信息）。
3. **并行处理**
   所有词对的注意力计算同时进行，直接捕获全局依赖关系（无论距离远近）。

核心作用

1. **捕捉长程依赖**
   直接建模序列中任意两个词的关系，解决RNN的长距离遗忘问题。
2. **上下文感知**
   每个词的输出融合了全局信息（如“苹果”能根据上下文关注“水果”或“公司”）。

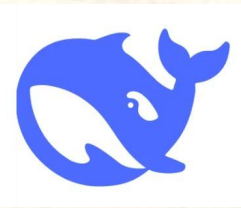
- **实践评价：**
  - 回答准确涵盖自注意力的核心概念（Q/K/V、动态权重、长程依赖）。
  - 通过步骤拆解和类比降低理解难度，逻辑由原理到作用层层递进。
  - 内容基于Transformer原始论文（2017）及主流理解，未涉及最新变体（如2023后的稀疏注意力），但核心原理仍适用。

2.2 DeepSeek实测：效果评估



- **综合优点:**
  - [+] 代码能力突出: 在代码生成、解释、补全方面表现优异 (基于实测结果)。
  - [+] 中文理解与生成: 对中文语境、表达习惯的把握较好。
  - [+] 开放性: 提供强大的开源模型，利于开发者定制和研究。
- **潜在缺点:**
  - [-] 知识范围/时效性: 联网搜索时，知识抽取能力略有不足 。
  - [-] 英文或其他语言: 虽然支持多语言，但最佳性能集中在中文和代码。
  - [-] 响应及推理速度: 实际使用中的响应速度较慢，R1推理时长较长，体验较差（不涉及复杂逻辑推理的问题建议不使用R1推理模型）。
- **特定场景下的表现:**
  - 极力推荐: 编程辅助、代码学习、中文内容生成、技术文档撰写。
  - 可以胜任: 通用问答、文本摘要、信息提取。
  - 需谨慎: API调用时无网络搜索能力，时效性强的问题解答能力较弱，且R1较V3模型幻觉严重。

2.2 DeepSeek实测：易用性、成本效益、集成性



- **易用性与交互体验:**
  - 交互流程: 支持多轮对话记忆，有历史记录、参数调整等便捷功能。
  - 交互体验: DeepSeek R1响应速度较慢、工具整体稳定性一般。生成网页时可以一键运行显示，提高了工具的易用性。
- **定价与免费额度:**
  - 免费政策: AI工具免费使用、无token限制、无使用频率限制。
  - API定价: 生成模型8元/M tokens,推理模型16元/M tokens。在优惠时间段，生成模型仅4元/M tokens,推理模型4元/M tokens。与ChatGPT相比，有很高的性价比。
- **集成与API能力:**
  - 开发平台链接：https://platform.deepseek.com/
  - API可用性: 输出稳定、说明文档完善。
  - 集成生态: API兼容openai调用，生态良好
  - 开发者友好度: 文档质量高、社区支持度好、API功能强大（如流式输出、函数调用）。

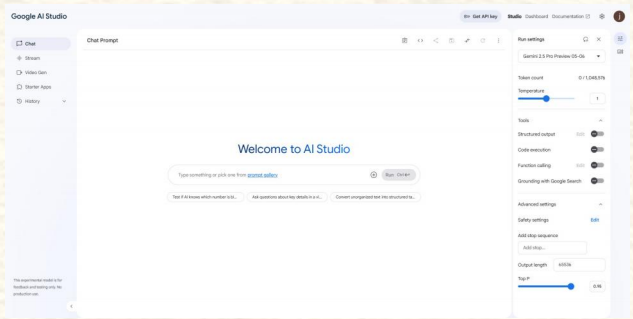
2.2 AI工具深度测评 - Gemini

Gemini 简介与核心技术

- **开发背景:** 由 Google DeepMind（谷歌大脑与 DeepMind 合并后的团队）开发，旨在打造下一代多模态大模型，对标 OpenAI 的 GPT-4o，并推动 AI 在搜索、广告、云计算等领域的应用
- **核心定位:**
  - 模型：Gemini 2.5 Pro Preview 05-06（2025.5.6发布）、Gemini 2.5 Flash Preview04-17、Deep Research和Canvas功能(目前仅限网页版使用)
  - 原生多模态：直接支持文本、图像、视频、音频多模态输入
  - 多语言支持：覆盖多种语言
- **核心技术：**
  - 原生多模态训练：统一学习不同模态数据，而非拼接，提升复杂推理能力
  - 高效训练方法：基于 Google 自研 TPU，优化计算效率
- **主要产品形态:** 网页、Google AI Studio、API服务。
  - 网页：gemini.google.com
  - Google AI Studio: aistudio.google.com



网页：gemini.google.com



Google AI Studio: aistudio.google.com

2.2 Gemini实测总结

- **Gemini工具总结：核心优势与不同版本模型核心特点**
  - Gemini是一款在多模态理解和生成能力上领先，并具备强大推理和整合能力的AI模型工具。
- **核心优势:**
  - 强大的推理和理解能力：最新的Gemini 2.5 Pro Preview 05-06代码编写能力比DeepSeek、GPT、Claude都要强劲。
  - 工具自带联网搜索能力，且基于联网搜索的知识整合能力比DeepSeek要强，特别是Deep Research功能，可以一键生成一份质量极高的研究报告。
  - Google公司团队研发，模型更新迭代速度快，能力强劲。
- **不同版本模型核心特点：**

	Gemini 2.5 Pro(05-06)	Gemini 2.5 Flash(04-17)	Deep Research
核心特点	网页版免费好用、性价比高	网页版免费好用、性价比最高	好用且有深度调研，网页版限量使用
	处理复杂任务、多模态理解最好	分析复杂问题效果较好	获取某领域的研究报告
	代码及编程能力最强	生成及输出速度快	系统性的统筹规划
	自带联网功能且效果不错	自带联网功能且效果不错	自带联网功能且效果不错

2.2 Gemini实测：核心功能 - 代码解释

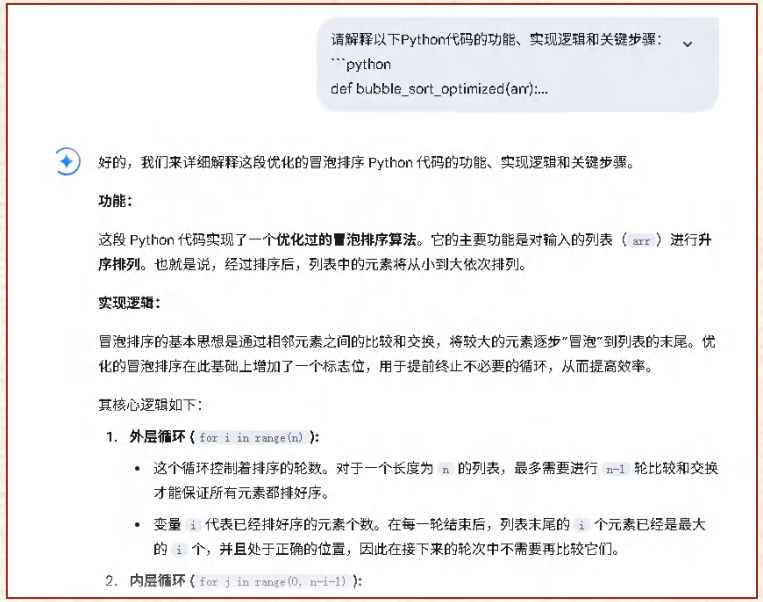
- **测试任务描述：**提供一段中等复杂度的代码片段（例如，Python实现的快速排序，或一段JavaScript的异步操作代码），要求Gemini解释其功能和逻辑。

提示词：

请解释以下Python代码的功能、实现逻辑和关键步骤：

```python
[在此处粘贴代码片段]
```

- **Gemini 2.5 Pro 输出结果：**



请解释以下Python代码的功能、实现逻辑和关键步骤：
```python
def bubble_sort_optimized(arr):...
```

好的，我们来详细解释这段优化的冒泡排序 Python 代码的功能、实现逻辑和关键步骤。

功能：

这段 Python 代码实现了一个**优化过的冒泡排序算法**。它的主要功能是对输入的列表（`arr`）进行**升序排列**。也就是说，经过排序后，列表中的元素将从小到大依次排列。

实现逻辑：

冒泡排序的基本思想是通过相邻元素之间的比较和交换，将较大的元素逐步"冒泡"到列表的末尾。优化的冒泡排序在此基础上增加了一个标志位，用于提前终止不必要的循环，从而提高效率。

其核心逻辑如下：

1. **外层循环**（`for i in range(n)`）：
   - 这个循环控制着排序的轮数。对于一个长度为 `n` 的列表，最多需要进行 `n-1` 轮比较和交换才能保证所有元素都排好序。
   - 变量 `i` 代表已经排好序的元素个数。在每一轮结束后，列表末尾的 `i` 个元素已经是最大的 `i` 个，并且处于正确的位置，因此在接下来的轮次中不需要再比较它们。
2. **内层循环**（`for j in range(0, n-i-1)`）：

- **实践评价：**
  - 功能描述准确，逻辑步骤的拆解清晰。
  - 对冒泡排序的算法思想和关键的比较、交换操作解释也比较好。
  - 虽然没有直接提及潜在问题或优化点，但提到了优化的效果（最好情况下的时间复杂度），可以视为间接提及。

2.2 Gemini实测：核心功能 - 代码生成

- **测试任务描述:** 根据视频内容，复刻app，输出html。

提示词：

根据视频，复刻这个app，输出html，用到的图标可以CDN引用在线的库

- **参考视频:**



- **Gemini 2.5 Pro 输出结果:**



- **实践评价:**
  - 已理解视频中APP的使用过程，并根据视频复刻出APP中的功能完整。
  - 页面内容基本完整，且美观大方，但与原APP还有差距，手机屏幕兼容略有不足，例如：底部菜单栏没有固定、页面布局没有完全一致等。

2.2 Gemini实测：核心功能 - 爆款标题生成

- **测试任务描述：**输入同一段核心内容描述（例如：“一篇关于提高远程工作效率的5个实用技巧的文章”），要求Gemini生成5个吸引人的“爆款”标题。

> **提示词**：
> 请根据以下内容，为我生成5个吸引眼球、适合在社交媒体传播的“爆款”文章标题：核心内容：本文分享了5个经过验证的技巧，帮助远程工作者提升专注度、管理时间和保持工作生活平衡，从而显著提高工作效率。

- **Gemini 2.5 Flash 输出结果：**



- **实践评价：**
  - 每个标题都巧妙地融入了数字、悬念、痛点和明确的价值承诺，具备成为“爆款”的潜质。
  - 提供的标题在表达方式和侧重点上有所不同，展现了一定的创意性和多样性。
  - 所有标题都紧密围绕远程工作者的专注度、时间管理和工作生活平衡的核心内容展开。

2.2 Gemini实测：核心功能 - 文案生成

- **测试任务描述:** 编写一篇关于人工智能通识教育洞见 + 宣讲活动 + 教材推介的推文。Gemini在教育领域具有优势，正文选择Gemini编写。

- **输入 Prompt:**

提示词：

写一篇微信公众号的文章：
#文章主题：高职教师必看：AI通识课教学难题如何破解？这套方案让课堂效果翻倍
#目标读者：高等院校教师
#文章目的：信息传递、引发讨论、增加粉丝
#期望的风格和语气：专业解读、分享的语气
#关键信息或要点：人工智能通识教育的痛点，我们的AI时代的技术观和人才观，高校通识教育的新挑战，AI通识课：目标与设计理念，AI通识课的内容，人工智能教学赋能解决方案，教材推介
#大致的篇幅要求：1000字左右

- **Gemini 2.5 Flash 输出内容节选:**

标题：高职教师必看：AI通识课教学难题如何破解？这套方案让课堂效果翻倍

（建议配图：一张富有科技感和教育元素的图片，如图书与电路板结合）

各位奋战在高等教育一线的老师们，大家好！

人工智能（AI）的浪潮正以前所未有的速度席卷各行各业，教育领域首当其冲。开设AI通识课，培养学生的AI素养，已成为高校教育不可回避的重要议题。然而，理想丰满，现实骨感。在实际教学中，我们常常面临诸多挑战：

- **课程内容难把握**：AI技术日新月异，如何设计既前沿又体系化的教学内容？如何从“AI 1.0”的传统科普，升级到适应“AI 2.0”生成式、推理式AI时代的思维与应用？[68]
- **教学资源难获取**：缺乏合适的实操平台、鲜活的教学案例和系统的课程资源，空有理论难以落地。[71]
- **教学方法难创新**：如何避免照本宣科？如何引导学生从“会用工具”走向“善用AI解决问题”？如何评估AI时代的学习成果？[50]
- **师资能力难提升**：面对全新的AI领域，教师自身也需要快速学习和适应，如何获得有效支持？[77]

这些痛点，是否也让你在AI通识课的教学路上感到困惑与焦虑？别担心，今天我们为您带来一套旨在破解这些难题、让AI通识课堂效果翻倍的解决方案。

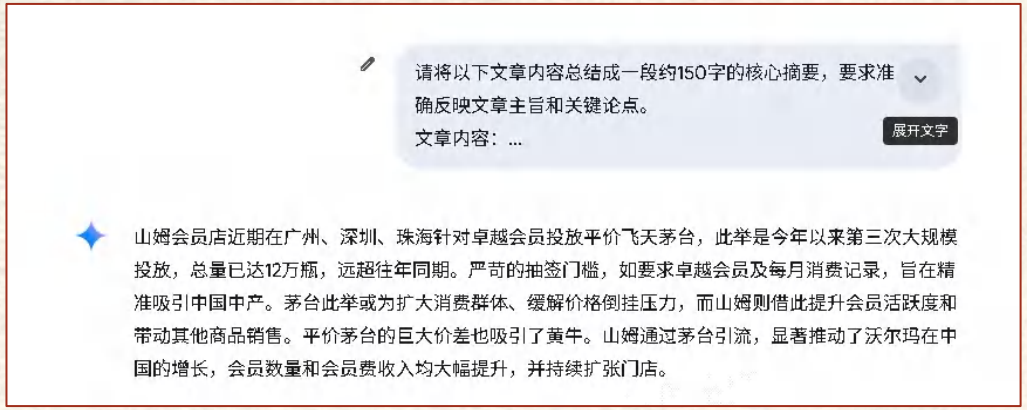
（建议配图：一张描绘教师面对AI教学挑战场景的插画）

2.2 Gemini实测：核心功能 - 长文摘要

- **测试任务描述:** 输入一篇中等长度的文章（约1000-1500字，选择一个有时效性或专业性的主题，例如“近期AI伦理法规进展”），要求Gemini生成一段150字左右的核心内容摘要。

> **提示词**：
> 请将以下文章内容 [粘贴文章前几段或全文，或提供文章链接让其自行抓取总结 - 如果支持] 总结成一段约150字的核心摘要，要求准确反映文章主旨和关键论点。[此处放置文章节选或说明文章来源] 。

- **Gemini 2.5 Flash 输出结果：**



- **实践评价:**
  - 摘要总结了山姆投放平价茅台、吸引中产和茅台自身策略的核心信息。并且，语言表达也流畅简洁，总结长度符合提示词要求。

2.2 Gemini实测：效果评估

- **综合优点:**
  - [+] 代码能力突出: Gemini 2.5 Pro Preview 05-06，在代码生成、解释、补全方面表现优异，整体优于其他同类模型，例如DeepSeek、GPT、Claude。
  - [+] 知识范围/时效性: Gemini全系列内置联网搜索，知识时效性较好、覆盖面较广，知识总结能力优异 。
  - [+] 创意写作/多模态: 具有较好的多模态理解和创意能力。
- **潜在缺点:**
  - [-] 开放性: 未提供开源模型，不利于开发者定制和研究。
  - [-] 响应速度: 需要科学上网，实际使用响应速度与网络情况强关联。
- **特定场景下的表现:**
  - 推荐: 编程辅助、代码学习、内容生成、统筹规划、方案规划。
  - 可以胜任: 通用问答、文本摘要、信息提取。

2.2 Gemini实测：易用性、成本效益、集成性

- **易用性与交互体验:**
  - 交互流程: 输入输出流畅，支持多轮对话记忆，有历史记录、角色选择（Gem）等便捷功能。
  - 交互体验: 工具平台响应速度较慢，生成时流畅、稳定性一般。
- **定价与免费额度:**
  - 免费政策: AI工具免费使用、无token限制、无使用频率限制。
  - API定价: 用户可申请$300的试用额度（有3个月左右的使用限制，在申请的3个月后失效）。
- **集成与API能力:**
  - 开发平台链接：https://ai.google.dev/gemini-api/docs/quickstart
  - API可用性: 输出稳定程度与网络情况及上网工具相关联、说明文档完善。
  - 集成生态: 兼容openai调用，生态良好。
  - 开发者友好度: 文档质量高、API功能强大（如流式输出、函数调用等）。

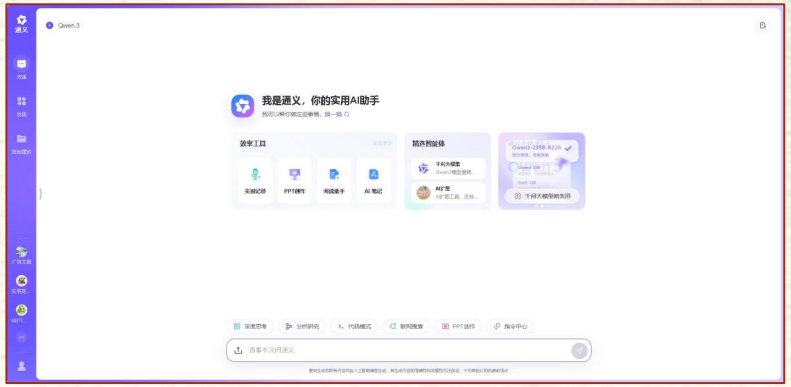
2.2 AI工具深度测评 - 通义千问

通义

通义千问 简介与核心技术

- 2025.4.19凌晨，阿里发布Qwen3系列模型，暴击OpenAI o1和DeepSeek R1，横扫各大基准，登顶全球开源模型王座。在推理、指令遵循、工具调用、多语言能力等方面均大幅增强。

- **Qwen3核心特点:**
  - Qwen3模型引入了一种混合问题解决方式，它们支持两种模式-思考模式和非思考模式。
  - 支持119种语言和方言，意味着Qwen3有极大潜力创建风靡全球的国际应用。
  - 阿里提高了Qwen3模型的编码和智能体能力，并且还加强了对MCP的支持。



- **主要产品形态:** 网页、API服务、APP助手。
  - 网页链接（均支持Qwen3满血模型）：

https://tongyi.aliyun.com/qianwen/（模型不可选，默认：Qwen3)

https://chat.qwen.ai/（模型可选，且实现多模型同时回复，更适合开发者使用）

2.2 通义千问实测总结

通义千问工具总结：

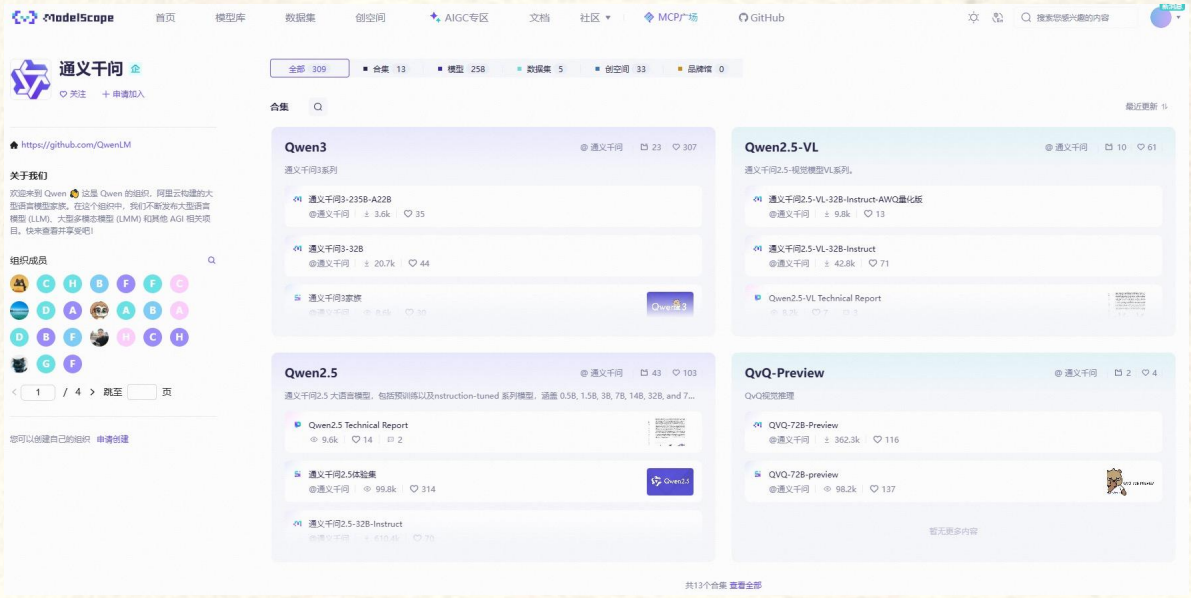
- Qwen3在混合推理，多语言理解和智能体能力上具有极大优势。

■ **Qwen3核心优势:**

- 混合推理模式极大提高推理能力与推理速度。
- 支持119种语言和方言的多语言能力让人惊艳。
- 编码和智能体能力极强，同时增强了MCP的支持。
- 同等性能效果下，模型体量更小，推理成本更低。

■ **主要限制:**

- 对网络搜索的知识总结能力一般 。



大量、多场景的开源模型

2.2 通义千问实测：核心功能 - 代码生成

- **测试任务描述：** 提供一段页面的描述，要求通义千问输出HTML代码。

> **提示词**：
> 帮我生成一个AI工具登陆页
> "暗黑底色+霓虹蓝紫光效，全息悬浮的输入框，粒子背景随鼠标移动流动，玻璃拟态按钮赛博朋克字体，有太空电梯纵向视觉元素"。

- **通义千问（代码模式）输出结果：**



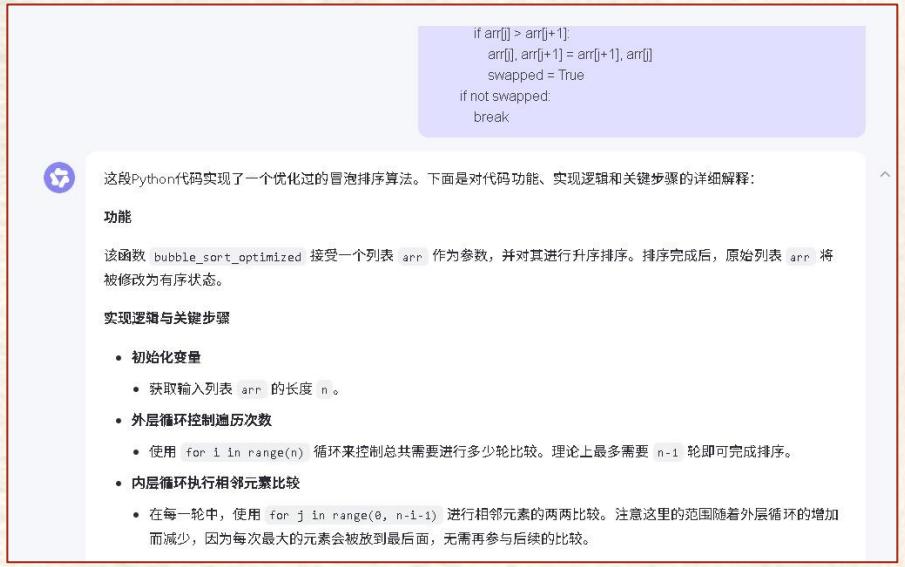
- **实践评价：**
  - 代码模式下，页面功能集成度较高，一键预览代码效果，易用性很好。
  - 提示词遵从度较高，完美实现提示词描述效果。
  - 代码生成速度很快，且代码规范较好。

2.2 通义千问实测：核心功能 - 代码解释

- **测试任务描述：** 提供一段中等复杂度的代码片段（例如，Python实现的快速排序，或一段JavaScript的异步操作代码），要求通义千问解释其功能和逻辑。

> **提示词**：
>
> 请解释以下Python代码的功能、实现逻辑和关键步骤：
>
> ```python
> [在此处粘贴代码片段]。
> ```

- **通义千问（代码模式）输出结果：**



```
            if arr[j] > arr[j+1]:
                arr[j], arr[j+1] = arr[j+1], arr[j]
                swapped = True
        if not swapped:
            break
```

这段Python代码实现了一个优化过的冒泡排序算法。下面是对代码功能、实现逻辑和关键步骤的详细解释：

功能

该函数 `bubble_sort_optimized` 接受一个列表 `arr` 作为参数，并对其进行升序排序。排序完成后，原始列表 `arr` 将被修改为有序状态。

实现逻辑与关键步骤

- **初始化变量**
  - 获取输入列表 `arr` 的长度 `n`。
- **外层循环控制遍历次数**
  - 使用 `for i in range(n)` 循环来控制总共需要进行多少轮比较。理论上最多需要 `n-1` 轮即可完成排序。
- **内层循环执行相邻元素比较**
  - 在每一轮中，使用 `for j in range(0, n-i-1)` 进行相邻元素的两两比较。注意这里的范围随着外层循环的增加而减少，因为每次最大的元素会被放到最后面，无需再参与后续的比较。

- **实践评价：**
  - 准确地描述了对输入列表进行升序排序的功能。
  - 代码被分解为五个部分逐一解读。
  - 对冒泡排序的算法思想和比较、交换操作解释到位。
  - 提到了在处理接近有序的数据集时，该优化版本的冒泡排序可以显著提高效率，这是一项额外的优化点说明。

2.2 通义千问实测：核心功能 - 分析研究

- **测试任务描述：** 输入一篇中等长度的文章（约1000-1500字，选择一个有时效性或专业性的主题，例如“近期AI伦理法规进展”），要求通义千问生成一段150字左右的核心内容摘要。

> **提示词**：
> 五一准备从北京出发去苏州旅行，帮我做一下规划

- **通义千问（分析研究模式）输出结果：**



- **实践评价：**
  - 分析研究模式非常适合处理极度复杂的任务，工具会对任务进行逐步拆解并调用多种工具解决任务，并生成一份详细的研究报告。
  - 由于是复杂任务，会调用多个工具，因此耗时明显较长，但任务处理效果令人满意。

2.2 通义千问实测：效果评估

通义

- **综合优点:**
  - [+] 响应速度: 实际使用响应及回复输出速度较快，推理模式下的回复速度较DeepSeek R1快一倍。
  - [+] 混合推理模式：能够在思考模式（用于复杂的逻辑推理、数学和编码）和非思考模式（用于高效的通用聊天）之间无缝切换，从而确保在各种场景中实现最佳性能
  - [+] 代码能力突出: 开启代码模式后，在代码生成、解释、补全方面表现优异。
  - [+] 智能体能力显著增强: 支持在思考和非思考模式下与外部工具的精确集成。
  - [+] 开放性: 提供了一系列开源模型，并持续推出最新模型，利于开发者定制和研究。
- **潜在缺点:**
  - [-] 知识时效性: 开启联网搜索后，对网络搜索的知识总结能力一般 。
- **特定场景下的表现:**
  - 极力推荐: 代码学习、多语言场景、角色扮演。
  - 可以胜任: 通用问答、创意写作、信息提取、统筹规划。

2.2 通义千问实测：易用性、成本效益、集成性

北京大学 PEKING UNIVERSITY

通义

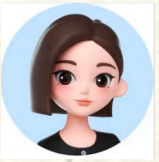
- **易用性与交互体验:**
  - 交互体验: 推理模式下速度很快，生成时流畅、功能稳定性一般。
  - 交互流程: 输入输出流畅，支持多轮对话记忆，有历史记录等便捷功能。
- **定价与免费额度:**
  - 免费政策: AI工具免费使用、无token限制、无使用频率限制。
  - API定价: 新用户开通即享每个模型100万Tokens试用额度。详情见下图。

模型名称	模式	最大回复长度	输入成本	输出成本
			(每千 Token)	
qwen3-235b-a22b	非思考		0.004 元	0.012 元
	思考			0.04 元
qwen3-32b	非思考		0.002 元	0.008 元
	思考			0.02 元

旗舰模型	通义千问-Max 适合复杂任务，能力最强	通义千问-Plus 效果、速度、成本均衡	通义千问-Turbo 适合简单任务，速度快、成本极低	通义千问-Long 适合大规模文本分析，效果与速度均衡、成本较低
最大上下文长度 (Token 数)	131,072	131,072	1,000,000	10,000,000
最低输入价格 (每千 Token)	0.0024 元	0.0008 元	0.0003 元	0.0005 元
最低输出价格 (每千 Token)	0.0096 元	0.002 元	0.0006 元	0.002 元

- **集成与API能力:**
  - 开发平台链接：https://bailian.console.aliyun.com/
  - API可用性: 批量请求输出稳定程度较好、说明文档完善。
  - 集成生态: 兼容openai调用，生态良好。
  - 开发者友好度: 文档质量高、提供模型选型建议、API功能强大（支持联网搜索、流式输出、函数调用等）。

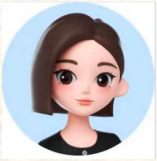
2.2 AI工具深度测评 - 豆包



豆包 简介与核心技术

- **开发背景:** 由字节跳动（全球知名的互联网技术公司）开发，旨在打造功能强大、表现优异、能够广泛服务用户的人工智能，推动 AI 在信息查询、知识问答、文案创作、智能交互等多领域的应用。
- **核心定位:**
  - 多领域知识覆盖：具备丰富的知识储备，能够回答科学、技术、文化、历史等各类领域的问题。
  - 自然语言处理能力：擅长理解自然语言的语义和意图，进行准确且自然流畅的对话交流和内容生成。
  - 多样化任务支持：可以完成知识科普、文案写作（如故事、诗歌、论文等）、代码解释与辅助、数据分析建议等多种任务，满足不同用户的需求。
- **主要产品形态:** 网页、API服务、APP、PC桌面版。
  - 网页链接： https://www.doubao.com/chat/
  - 豆包桌面版：https://www.doubao.com/download/desktop
  - API服务：https://chat.qwen.ai/

2.2 豆包实测总结

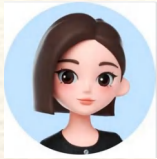


- **豆包工具总结：核心优势与最佳适用场景**
  - 豆包在自然语言处理等方面有着出色的能力，能为用户提供高质量的语言及多模态交互服务。
- **核心优势:**
  - 字节跳动出品，模型能力更新迭代快速
  - 优秀的自然语言处理能力
- **主要限制:**
  - 豆包没有开源模型，无法私有化部署
  - 对复杂专业领域的深度理解存在局限
  - 语言表达风格的多样性相对较差

谁适合用豆包

- **需要知识查询和解答的用户（普通大众）**：无论是学习、工作还是日常生活中遇到各种问题，如：辅导孩子写作业等都可以通过豆包获取相关知识和答案。
- **进行文本创作的用户**：比如写作文章、诗歌、故事等，豆包可以提供创意灵感、思路建议以及对文本的润色和优化。

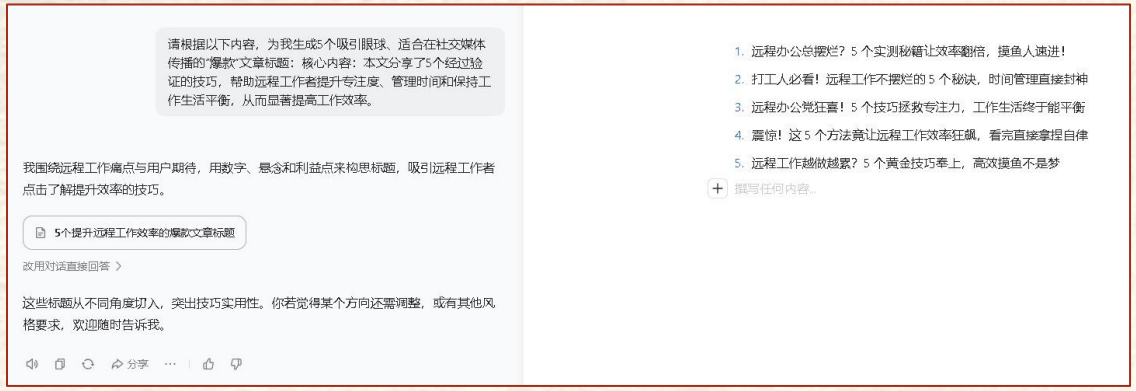
2.2 豆包实测：核心功能 - 爆款标题生成



- **测试任务描述：**输入同一段核心内容描述（例如：“一篇关于提高远程工作效率的5个实用技巧的文章”），要求豆包的”帮我写作“模式生成5个吸引人的“爆款”标题。

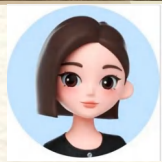
> **提示词**：
> 请根据以下内容，为我生成5个吸引眼球、适合在社交媒体传播的“爆款”文章标题：核心内容：本文分享了5个经过验证的技巧，帮助远程工作者提升专注度、管理时间和保持工作生活平衡，从而显著提高工作效率。

- **豆包（帮我写作）输出结果：**



- **实践评价：**
  - 使用数字明确技巧数量，以 “摆烂” “效率狂飙” 等制造悬念。
  - 结合网络流行语和夸张表述，从不同场景和情绪点切入，风格涵盖轻松幽默、震惊体等，创意新颖且类型多样。
  - 紧扣分享 5 个技巧帮助远程工作者提升效率、平衡生活的核心，清晰传递文章关键信息，契合度高。

2.2 豆包实测：核心功能 - 长文摘要



- **测试任务描述:** 输入一篇中等长度的文章（约1000-1500字，选择一个有时效性或专业性的主题，例如“近期AI伦理法规进展”），要求豆包生成一段150字左右的核心内容摘要。

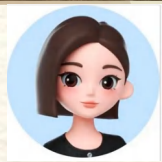
> **提示词**：
> 请将以下文章内容 [粘贴文章前几段或全文，或提供文章链接让其自行抓取总结 - 如果支持] 总结成一段约150字的核心摘要，要求准确反映文章主旨和关键论点。[此处放置文章节选或说明文章来源]。

- **豆包 输出结果：**



- **实践评价:**
  - 摘要涵盖了原文中茅台商超投放、山姆放量情况、茅台调整目的、山姆抽签资格及对会员影响、山姆的地位等核心信息。
  - 摘要遗漏了部分细节如每次放量的具体时间、不同平台活动更详细的资格要求等关键点。
  - 语言表达流畅简洁，符合长度要求。

2.2 豆包实测：核心功能 - 专业问答



- **测试任务描述:** 提出一个具有一定专业性的问题（例如，技术类：“解释一下Transformer模型中的自注意力机制原理”，或法律类：“简述GDPR规定的数据主体权利”）。

> **提示词**：
> 请请用简洁明了的语言解释一下Transformer模型中自注意力机制 (Self-Attention Mechanism) 的核心原理和作用。

- **豆包（AI搜索）输出结果：**



搜索一下：请用简洁明了的语言解释一下Transformer模型中自注意力机制 (Self-Attention Mechanism) 的核心原理和作用。

Transformer 模型中自注意力机制的核心原理和作用如下：

核心原理

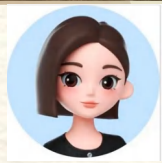
- **生成向量**：将输入序列中的每个元素通过线性变换生成三个向量，即查询（Query）、键（Key）和值（Value）。
- **计算分数**：通过计算查询向量与所有键向量之间的点积，得到一个分数矩阵，该矩阵反映了每个元素与其他元素之间的相关性。为稳定梯度，会用键向量的维度对分数进行缩放。
- **计算权重**：将缩放后的分数通过 Softmax 函数进行归一化，得到每个位置的注意力权重，这些权重表示了每个元素在关注其他元素时的重要程度，且权重之和为 1。
- **生成表示**：将注意力权重应用到值向量上，通过加权求和得到每个位置的输出向量，该向量包含了整个序列的上下文信息。

作用

- **捕捉长距离依赖**：能有效捕捉序列中任意两个位置之间的依赖关系，解决了传统 RNN 在处理长序列时的梯度消失问题，比如在处理长句子时，能让模型理解相距较远的单词之间的语义联系。
- **并行计算**：可以并行处理序列中的所有位置，显著 ⌄ 计算效率，大大加快模型的训练速度。
- **自适应关注**：注意力权重是动态计算的，能够根据输入内容自适应地调整关注的重点，使模型在处理每个元

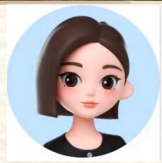
- **实践评价:**
  - 回答准确，包含了自注意力机制的关键概念，如查询、键、值向量的生成，注意力权重计算及作用等。
  - 解释清晰易懂，按照原理和作用的逻辑顺序，逐步阐述了自注意力机制的核心内容，层次分明。
  - 自注意力机制原理描述仍符合当前研究（如Transformer架构），但未提及最新变体（如稀疏注意力），时效性中等。

2.2 豆包实测：效果评估

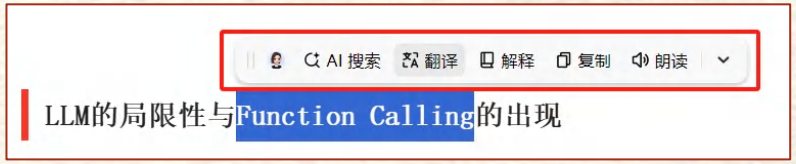


- **综合优点:**
  - [+] 特定任务性能: 深度思考模式下，逻辑推理、数学问题有优势，达到DeepSeek R1水准。
  - [+] 响应速度: 实际使用响应及回复输出速度较快。
  - [+] 代码能力突出: 开启代码模式后，在代码生成、解释、补全方面表现优异，画图、页面代码一键展示。
- **潜在缺点:**
  - [-] 开放性: 开源模型较少，不利于开发者定制和研究。
  - [-] 意图识别：对长文本提示的意图识别能力较弱。
  - [-] 知识时效性: 开启联网搜索后，对网络搜索的知识总结能力一般 。
- **特定场景下的表现:**
  - 极力推荐: 思维导图绘制、PPT制作、知识类学习、内容生成。
  - 可以胜任: 通用问答、文本摘要、信息提取、孩子辅导、生活小助手。
  - 需谨慎: 高度创造性的艺术创作。

2.2 豆包实测：易用性、成本效益、集成性



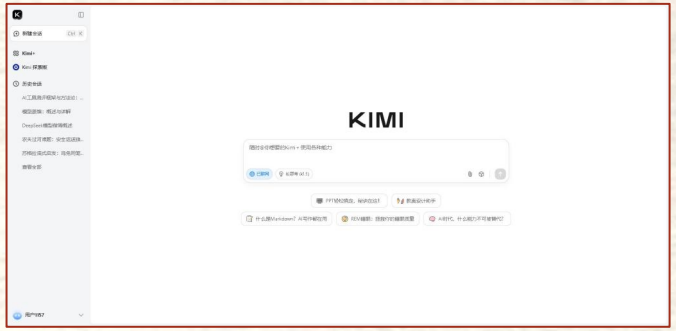
- **易用性与交互体验:**
  - 交互流程: 输出流畅，支持多轮对话记忆，有历史记录，收藏对话，编辑及一键下载回复内容等便捷功能。
  - 交互体验: 响应速度较快，生成时流畅、稳定性较好。桌面PC端选中文字跳出快捷菜单功能。



- **定价与免费额度:**
  - 免费政策: AI工具免费使用、无token限制、无使用频率限制。
  - API定价: 新用户开通即享每个模型50万Tokens试用额度，性价比等同DeepSeekAPI，详情见火山引擎开发平台。
- **集成与API能力:**
  - 开发平台链接：https://console.volcengine.com/ark
  - API可用性: 批量请求输出稳定程度较好、说明文档完善、提供API使用示例。
  - 集成生态: 兼容openai调用，生态良好。
  - 开发者友好度: 文档质量高、提供模型选型建议、API功能强大（支持联网搜索、流式输出、函数调用等）。

Kimi 简介与核心技术

- **开发背景:** 由 Moonshot AI（国内领先的人工智能公司）开发，旨在打造高性能、长文本处理能力强大的大语言模型，推动 AI 在内容创作、教育、企业服务等领域的应用。
- **核心定位:**
  - 超长上下文支持：原生支持超长文本输入与理解，擅长处理复杂文档和深度对话。
  - 中文优化：针对中文场景深度优化，在中文理解和生成任务上表现优异。
  - 多场景适配：覆盖写作、编程、数据分析等多种任务，兼顾通用性与垂直领域能力。



- **核心技术：**
  - 长上下文建模技术：突破传统模型的文本长度限制，实现更连贯的上下文关联与记忆。
- **主要产品形态:** 网页、API服务、APP助手。
  - 链接：https://kimi.moonshot.cn/

2.2 Kimi实测总结

- **Kimi工具最佳适用场景**
  - Kimi产品化能力较强，“PPT助手”可以自动做PPT，且支持PPT流式输出，PPT大纲可借助其他能力强的模型完成。
- **核心特点：**
  - Kimi 1.5的长思考模式能力效果不如DeepSeek
  - Kimi的“PPT助手”实用性强
- **主要限制：**Kimi不提供开源模型，无法私有化部署
- **PPT助手的实践评价：**Kimi的“PPT助手”模板选择较多
  - 支持二次编辑，易用性较高
  - PPT流式输出，用户体验很好



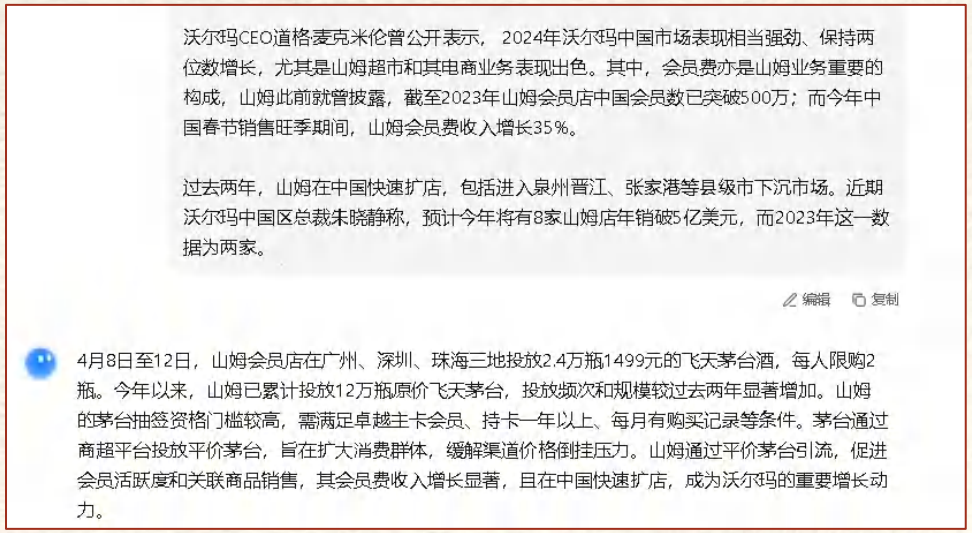
2.2 Kimi实测：核心功能 - 长文摘要

- **测试任务描述：** 输入一篇中等长度的文章（约1000-1500字，选择一个有时效性或专业性的主题，例如“近期AI伦理法规进展”），要求Kimi生成一段150字左右的核心内容摘要。

- **输入 Prompt (示例):**

> **提示词**：
> 请将以下文章内容 [粘贴文章前几段或全文，或提供文章链接让其自行抓取总结 - 如果支持] 总结成一段约150字的核心摘要，要求准确反映文章主旨和关键论点。[此处放置文章节选或说明文章来源]。

- **Kimi 输出结果：**



- **实践评价：**
  - 摘要涵盖了山姆会员店茅台投放活动的主要内容、目的以及对双方的意义，反映了文章的主旨。
  - 摘要未遗漏投放规模、会员资格要求、茅台与山姆的战略意图等关键点。
  - 语言表达流畅简洁，但超出150字的长度要求。

2.2 Kimi实测：效果评估

- **综合优点:**
  - [+] 长上下文支持：原生支持超长文本输入与理解，擅长处理复杂文档和深度对话。
  - [+] 响应速度: 实际使用响应及回复输出速度较快。
  - [+] 中文理解与生成: 对中文语境、表达习惯的把握较好。
- **潜在缺点:**
  - [-] 开放性: 未提供开源模型，不利于开发者定制和研究。
  - [-] 知识时效性: 开启联网搜索后，对网络搜索的知识总结能力一般 。
- **特定场景下的表现:**
  - 极力推荐: PPT制作、代码学习、内容生成。
  - 可以胜任: 通用问答、文本摘要、信息提取。

2.2 Kimi实测：易用性、成本效益、集成性

- **易用性与交互体验:**
  - 交互流程: 输入输出流畅，支持多轮对话记忆，有历史记录等便捷功能。
  - 交互体验: 响应速度较快，生成时流畅、稳定性一般。
- **定价与免费额度:**
  - 免费政策: AI工具免费使用、无token限制、无使用频率限制。
  - API定价: 新用户有15￥的试用额度。详情见右图。
- **集成与API能力:**
  - 开发平台链接：https://platform.moonshot.cn/
  - API可用性: 批量请求输出稳定程度较好、说明文档完善。
  - 集成生态: 兼容openai调用，生态良好。
  - 开发者友好度: 文档质量高、API功能强大（支持联网搜索、流式输出、函数调用等）。

生成模型 kimi-latest

模型	计费模型	计费单位	输入价格（缓存命中）	输入价格（缓存未命中）	输出价格	模型上下文长度
kimi-latest	kimi-latest-8k	1M tokens	￥1.00	￥2.00	￥10.00	8,192 tokens
	kimi-latest-32k	1M tokens	￥1.00	￥5.00	￥20.00	32,768 tokens
	kimi-latest-128k	1M tokens	￥1.00	￥10.00	￥30.00	131,072 tokens

生成模型 Moonshot-v1

模型	计费单位	输入价格	输出价格	模型上下文长度
moonshot-v1-8k	1M tokens	￥2.00	￥10.00	8,192 tokens
moonshot-v1-32k	1M tokens	￥5.00	￥20.00	32,768 tokens
moonshot-v1-128k	1M tokens	￥10.00	￥30.00	131,072 tokens
moonshot-v1-8k-vision-preview	1M tokens	￥2.00	￥10.00	8,192 tokens
moonshot-v1-32k-vision-preview	1M tokens	￥5.00	￥20.00	32,768 tokens
moonshot-v1-128k-vision-preview	1M tokens	￥10.00	￥30.00	131,072 tokens

2.3 文本AI工具横向对比

■ 关键维度性能矩阵

维度	DeepSeek	Gemini	通义千问	豆包	Kimi
核心功能	★★★★★	★★★★★	★★★★☆	★★★★☆	★★★☆☆
效果质量	★★★★☆	★★★★★	★★★★☆	★★★★☆	★★★☆☆
易用性	★★★★☆	★★★☆☆	★★★★★	★★★★★	★★★★★
成本效益	★★★★☆	★★☆☆☆	★★★★☆	★★★★☆	★★☆☆☆
集成性	★★★★☆	★★☆☆☆	★★★★★	★★★★☆	★★★★☆
安全合规性	★★★★★	★★★★☆	★★★★★	★★★★★	★★★★★
社区支持	★★★★★	★★★★☆	★★★★★	★★★★☆	★★★☆☆
创新前瞻性	★★★★★	★★★★★	★★★★★	★★★★☆	★★★☆☆

2.4 选型指南：构建决策矩阵

- **基于场景构建决策矩阵示例：**独立工作者-up主
  - **不同维度权重占比：**

	核心功能	效果质量	易用性	成本效益	集成性	安全合规性	社区支持	创新前瞻性
权重（%）	10	30	20	10	5	5	5	15

  - **最终计算得分：**将权重占比与上页内容合并计算所得（每颗星20分，满分100分）

	DeepSeek	Gemini	通义千问	豆包	Kimi
最终得分	87	81	88	85	68

  - **最终选择AI工具：**通义千问

2.4 选型指南：常见场景下的工具推荐

- **场景1：内容创作 & 营销人员**
  - 需求: 创意写作、文案生成、标题优化、内容摘要、多语言翻译。
  - 推荐工具: DeepSeek (综合/创意), Gemini (长文本/创意), Kimi (长文本/中文)。
  - 考虑因素: 创造力、文本长度支持、易用性。
- **场景2：研究人员 & 学生**
  - 需求:文献摘要、专业问答、研究思路启发、数据分析辅助 (部分工具)。
  - 推荐工具: Gemini Deep Research (研究整合能力), DeepSeek (长文档处理), 通义千问 分析研究模式(综合知识)。
  - 考虑因素: 知识准确性、信息来源追溯 (部分支持)、长文本处理能力。

2.4 教学场景下不同AI工具能力测评

- **测评方法**：真实场景数据模型测评
- **测评数据集**：北大青鸟教学问答测评数据集（文本类）
- **测评方式**：基于测评人员的个体差异，采用交叉测评方式，以保证测评数据的公正，客观

参考论文：
https://arxiv.org/abs/2402.07913

- **测评说明**：
  - 五个知识方向总共有474条问题数据，其中python 83条 智能体 180条 提示词 60条 大模型应用开发 50条 大模型微调 101条。测评时，从每个知识方向的不同模块随机选取，每个知识方向共30-60条数据进行评测。
  - 针对模型回复评分时，主要从4个维度进行打分（百分制）：① 技术表述正确性&严谨性（含代码示例）- **权重占比60%** ② 语言组织逻辑性、生动性- **权重占比5%** ③ 回答的内容是否解决问题- **权重占比20%** ④ 教学角度考虑：是否切近用户（根据问题的用户标签来进行判断确定）- **权重占比15%**
  - 在计算总分时，根据维度权重占，进行总分计算。
  - 注：测评数据为调用API获取，temperature参数统一设置为0.7

QA评价					
技术表述正确性&严谨性（含代码示例）	语言组织逻辑性、生动性	回答的内容是否解决问题	教学角度考虑：是否切近用户（根据问题的用户标签来进行判断确定）	最终评价分数	代码正确性 <空：无代码； 1：正确； 2：错误>
60%	5%	20%	15%		
90	80	80	70	84.5	

2.4 教学场景下不同AI工具能力测评

AI工具排名（教学场景）

- 第一梯队：Gemini、DeepSeek、GPT
- 第二梯队：通义千问（QwenMax、QwQ）
- 第三梯队：Kimi、文心一言

- **Gemini测评结论（概要）：**
  - **优势**：表达自然生动、案例教学能力强、用户体验最佳，尤其在Python、提示词工程和大模型应用开发知识方向中表现突出，适合互动式辅导和轻松风格的教学场景。
  - **不足**：部分场景回复偏冗长，需优化简洁性；技术类知识块需人工验证新知识准确性。
  - **推荐场景**：优先用于入门教学、案例演示、互动辅导，适合注重趣味性和易懂性的用户。
- **DeepSeek测评结论（概要）：**
  - **优势**：技术解析扎实、逻辑性强，适合系统性讲解（如大模型开发、微调知识块），综合能力均衡。
  - **不足**：理论表达不够简洁，偶有冗长或术语堆砌；对新框架或前沿知识可能存在幻觉。
  - **推荐场景**：优先用于技术深度教学、原理分析，需搭配人工审核补充最新知识。
- **GPT测评结论（概要）：**
  - **优势**：中文表达流畅，问题理解能力较好，教学逻辑清晰，适合Python基础、大模型微调等课程的理论讲解和快速知识梳理。
  - **不足**：部分领域（如Agent开发、向量数据库）存在技术幻觉，案例较少且本土化不足，对最新平台（如Dify/FastGPT）的实操了解较弱，表达冗长。
  - **推荐场景**：适合理论教学和知识梳理，对Python基础、大模型原理知识表现良好。适合提示词设计基础教学，但需补充案例。在智能体等实操内容上，仅适用于概念讲解不适用于新知识或强实操的教学需求。
- **通义千问（特定场景适用）测评结论（概要）：**
  - **QwQ**：输出简洁清晰，适合偏好高效风格的用户，但需约束造词倾向，适合快速解答或时间有限的场景。
  - **QwenMax**：逻辑性较强但细节不足，适合快速概览或入门指导，实操建议需补充。。
- **Kimi测评结论（概要）：**
  - 风格轻松但多语言混杂，技术错误较多，仅适合非严谨的辅助答疑。
- **文心测评结论（概要）：**
  - 表达生硬、本土化案例不足，理解偏差频发，适用性较低，仅建议用于基础规范内容。

2.4 教学场景下不同AI工具能力测评

■ **测评结果**（🌟百分制，优：90~100；良：80~90；中：60~70；差：60以下）

Python方向

LLMs	最终得分	思考过程得分
gemini-2.5-pro	97.54	-
DeepSeek V3	96.36	-
DeeepSeek R1	93.23	92.8
GPT4.1	92.35	-
kimi-latest	92.33	-
文心X1	91.37	86.5
文心4.5	89.68	-
QwQ-Plus	89.27	84
QwenMax	88.1	-

提示词工程方向

LLMs	最终得分	思考过程得分
DeepSeek V3	84.6	-
GPT4.1	82.9	-
gemini-2.5-pro	82.3	-
QwenMax	81.2	-
DeeepSeek R1	80.7	84.8
文心4.5	80.6	-
QwQ-Plus	79.2	84.5
文心X1	78.6	80.6
kimi-latest	75.4	-

智能体方向

LLMs	最终得分	思考过程得分
GPT4.1	80.9	-
gemini-2.5-pro	80.9	-
DeeepSeek R1	80.1	80.6
DeepSeek V3	79.2	-
文心4.5	76	-
QwQ-Plus	75.9	75
kimi-latest	75.2	-
文心X1	74.1	77
QwenMax	69.9	-

大模型应用开发方向

LLMs	最终得分	思考过程得分
GPT4.1	81.9	-
gemini-2.5-pro	79.0	-
DeepSeek V3	78.7	-
DeeepSeek R1	74.2	88.1
QwenMax	72.7	-
QwQ-Plus	71.9	77.7
文心4.5	70.1	-
kimi-latest	64.0	-
文心X1	61.3	71.8

大模型微调方向

LLMs	最终得分	思考过程得分
gemini-2.5-pro	84.3	-
GPT4.1	80.6	-
DeeepSeek R1	74.9	81.1
DeepSeek V3	74.7	-
文心X1	74.0	81.4
文心4.5	73.8	-
QwQ-Plus	73.2	84.2
QwenMax	70.3	-
kimi-latest	69.3	-



北京大学
PEKING UNIVERSITY

01 文本生成与处理类AI工具测评

02 图像生成与编辑类AI工具测评

03 音视频生成与处理类AI工具测评

04 代码生成与辅助类AI工具测评

05 大模型管理与应用类AI工具测评

3.1 AI图像生成工具测评总述

本次测评聚焦于多款主流图像类AI工具，涵盖以下三大类别，适用于插画、设计、自媒体、办公等众多场景：

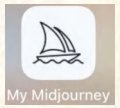
- **硬核效果类**：Midjourney、即梦AI、可灵AI
- **综合功能类**：ChatGPT、Gemini、豆包、元宝、阶跃星辰

工具	类型	核心能力	效果评估	易用性与Prompt技巧	优势	适用场景	支持API
Midjourney	图像生成	图像生成	顶级艺术表现力，细节丰富	易用性较强，需要一定Prompt技巧	社区活跃，风格多样，可控性强	插画、设计、广告、短视频、电商、影视、教育、社交媒体	否
即梦AI	图像/视频生成	图像生成、视频生成	亚洲审美出色，色彩鲜明	易用性强，中文Prompt友好，可用DeepSeek优化Prompt，包容度较高，有参数预设	免费试用，速度快，成本可控	插画、设计、广告、短视频、电商、教育、社交媒体	是
可灵AI	图像/视频生成	图像生成、视频生成	亚洲审美出色，光影效果逼真	易用性强，中文Prompt友好，可用DeepSeek优化Prompt，包容度较高，有参数预设	免费试用，本土化强，影视级质量，视频能力强	插画、设计、广告、短视频、电商、影视、教育、社交媒体	是
海螺AI	图像/视频生成	图像生成、音频生成、视频生成	语义理解能力较好，风格化能力强，卡通风格较强	易用性强，中文Prompt友好，包容度较高，有参数预设	免费试用，速度快，功能集成性较强	插画、设计、广告、短视频、电商、教育、社交媒体	是
ChatGPT	功能集成	多模态图像生成与理解	部分风格化能力较强（例如吉卜力等）	易用性强，语义理解强，中文Prompt友好，包容度较高	速度快， 与文本生成无缝结合	插画、设计、广告、教育、社交媒体	是
Gemini	功能集成	多模态图像生成与理解	表现力较好	易用性强，语义理解强，中文Prompt友好，包容度较高，多模态交互能力强	免费，功能集成性强， 谷歌生态整合	插画、设计、广告、教育、社交媒体	是
豆包	功能集成	多模态图像生成与理解	风格较稳定，写实能力较强	易用性极强，中文Prompt友好，包容度较高，有参数预设	免费，功能集成性强	插画、设计、广告、短视频、教育、社交媒体	是
阶跃星辰	功能集成	图像生成	风格迁移能力强，图片局部修改效果好，生图准确度较差	易用性强，自然语言出图	免费，速度快	日常办公	是
元宝	功能集成	多模态图像生成与理解	擅长卡通风格	易用性强，中文Prompt友好，包容度较高，有参数预设	免费，功能集成性强	日常办公	是
通义万相	图像/视频生成	图像生成、视频生成	风格稳定，写实能力较强	易用性强，中文Prompt友好，包容度较高，有参数预设	免费试用，阿里云生态支持	日常办公	是
百度AI图片助手	图像修改	图像修改	功能齐全，整体效果基本可用	易用性强，中文友好	细节处理能力较弱 AI味重	日常办公，快速修图	否
SnapEdit	图像修改	图像修改	细节处理出色，可商用	易用性较强	需要付费	设计场景快速修图	是

图像生成类AI工具测评

3.2 图像生成类AI工具深度测评

- **测评目的: 全面评估不同AI工具在具体业务场景下的表现，为用户提供选型依据。**
- **测评维度:**
  - 核心功能与效果 (重点)
  - 易用性与交互体验
  - 定价与性价比
  - 集成与扩展能力 (API)











- **测评工具选型:**
  - Midjourney v7、GPT-4o、Gemini 2.0 Flash 、即梦AI 3.0、可灵AI 2.0
- **测评结果对比：**
  - **Midjourney：**能力最强且发挥稳定，各种细节处理和质感都是最强的AI绘画工具，但目前对于中文提示词的支持效果较差，适合专业设计场景，学习成本较高。（收费）
  - **GPT-4o：**风格化效果出色，出图质感及氛围感很棒，仅次于Midjourney（收费）
  - **Gemini：**出图速度快，语义理解能力强，出图效果可做设计创意，满足基本商用效果（免费不限量）
  - **即梦AI：**中文提示词友好，“国风”效果出色，（中文）文字准确度超高，社区支持优秀，学习成本低，但复杂场景的控图较差（收费）
  - **可灵AI：**对标即梦AI，国风效果惊艳，但文字生成能力弱（无论中英文，出错概率高）（收费）

3.2 AI工具深度测评 - Midjourney

- **Midjourney 简介与核心技术**
  - Midjourney V7 （2025.4.3更新） 拥有全新改进的扩散模型架构，不仅在图像生成质量上取得显著提升，还在生成速度、输出分辨率、提示词控制和风格多样性方面实现了突破性进步。
  - 作为AI绘画领域的领军者，Midjourney V7 以其更高的图像质量、更快的生成速度和更强的个性化功能，再次刷新了AI绘画的边界。
- **核心定位:**
  - 致力于探索思维新媒介并扩展人类想象力的独立研究实验室，通过AI技术将文字描述转化为艺术化视觉创作
- **核心技术突破（Midjourney V7）：**
  - 新版本显著提升了文本提示理解能力，在图像质量上实现质的飞跃。在材质质感、人体结构、手部细节以及物体呈现等方面。相较过往版本（以及其他AI绘图工具），V7生成的图像已达到难辨真伪的水准—成功解决了长期以来困扰AI绘画的"手部畸形"和复杂形体塑造难题。
- **主要产品形态:** 网页、Discord APP
  - Midjourney：https://www.midjourney.com/
  - Discord APP：https://discord.com/channels/

3.2 AI工具深度测评 - Midjourney

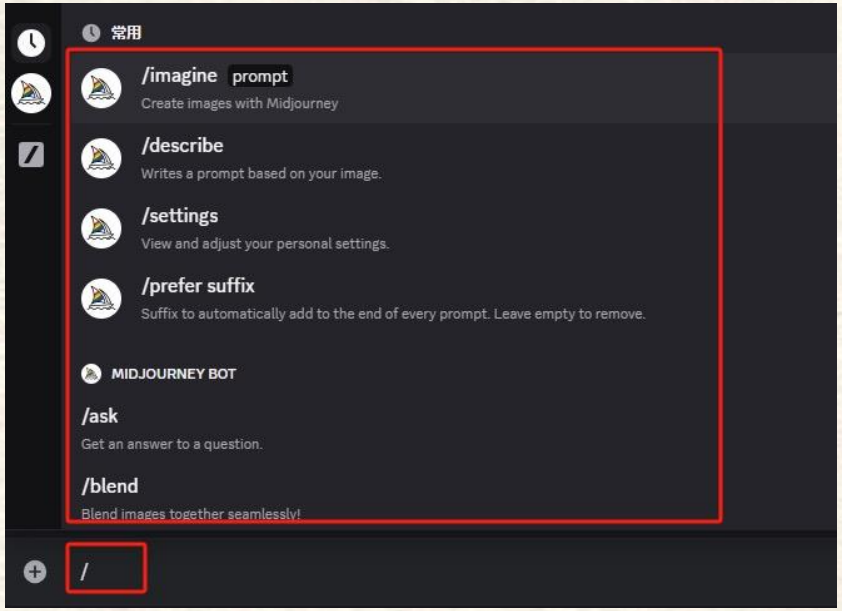
Midjourney 核心功能

- **图片生成:**
  - 文生图：根据用户输入的文字描述，创造出全新的、富有想象力的图像；
  - 图生图：将用户上传的一张或多张图片作为灵感，结合文字提示生成新的图像；
  - 图片混合：允许用户将多张图片融合成一张全新的、具有各图元素的合成图像。
- **图片描述：**分析用户上传的图片，并生成多个 Midjourney 风格的文字提示，帮助用户理解AI如何“看”图或用于反向生成类似风格图片
- **社区联动：**Midjourney主要在Discord平台上运行，用户通过在社区服务器中输入指令与Bot交互，并能方便地与其他用户交流、分享和获取灵感。

3.2 AI工具深度测评 - Midjourney

Midjourney 易用性与交互体验-Discord

Midjourney可以在Discord和Midjourney控制台中使用，以对话形式发送提示词（中文支持较弱）即可，Discord可以参考社区作品，Midjourney控制台编辑能力较强。

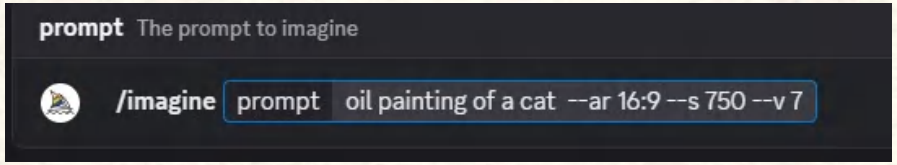


命令唤醒：

英文输入法下在输入框输入反斜杠/，系统就会自动提示常见的命令以及全部可用的命令

Discord常用命令

命令	含义	使用频率
imagine	最常用绘图命令，紧跟着提示词	★★★★★
blend	也即是常说的融图，上传多张（2-5张）图片，系统会将其融合为一张	★★★★
describe	图片生成提示词，上传一张图片，系统会生成四段提示词供你参考	★★★★
settings	Midjourney生成图片相关的设置	★★★★
prefer suffix	设置提示词后缀	★★
info	查看账户信息	★★
subscribe	订阅Midjourney会员	★★
fast	切换到快速模式，订阅会员后可在开启，或在setting开启	/
ask	向系统咨询问题，系统会弹出答案和提示	/
help	向系统求助，系统会弹出新人快速指南	/
invite	获得一个Midjourney Discord服务器的邀请链接	/
prefer option	设置自定义选项，比如设置自定义后缀参数	/
prefer auto_dm	设置完成的任务会自动发送到私信	/
prefer remix	开启/关闭混音模式	/
private	切换到隐身模式（仅专业版会员）	/
public	切换到公共模式	/
relax	切换到放松模式	/
show	通过人物ID恢复指定图片	/
stealth	切换到隐身模式（仅专业版会员）	/



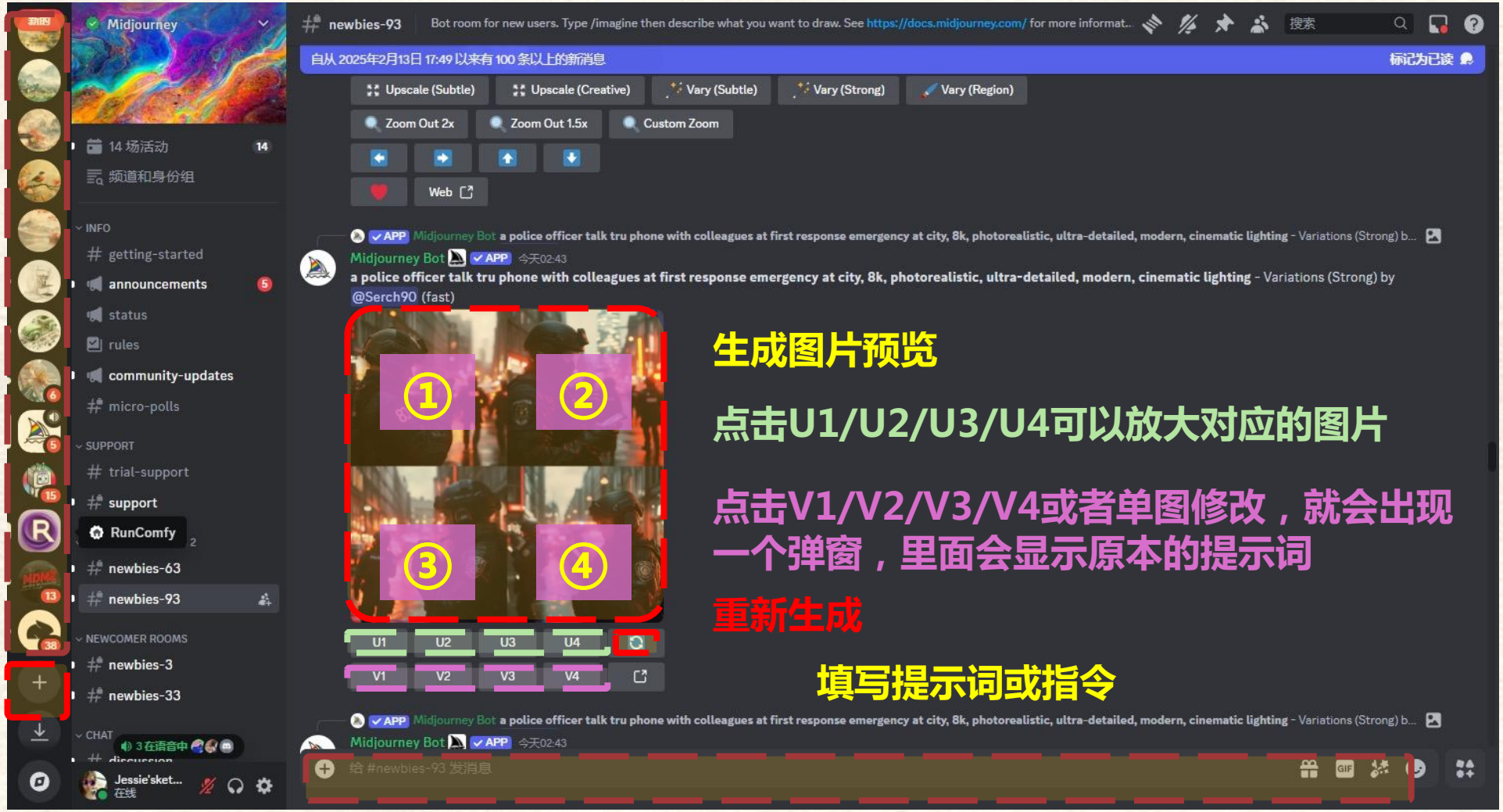
输入提示词即可出图

3.2 AI工具深度测评 - Midjourney

基本操作

选择服务器 →

创建服务器 →



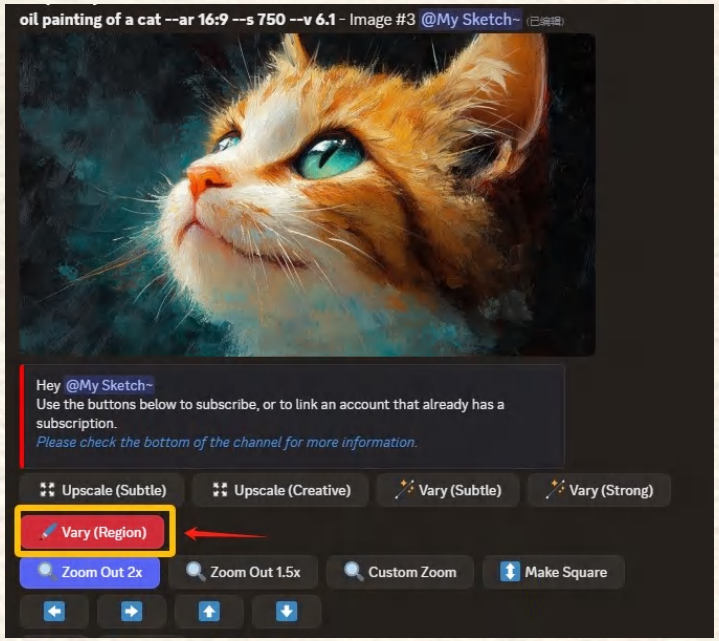
3.2 AI工具深度测评 - Midjourney

Midjourney 易用性与交互体验-Discord

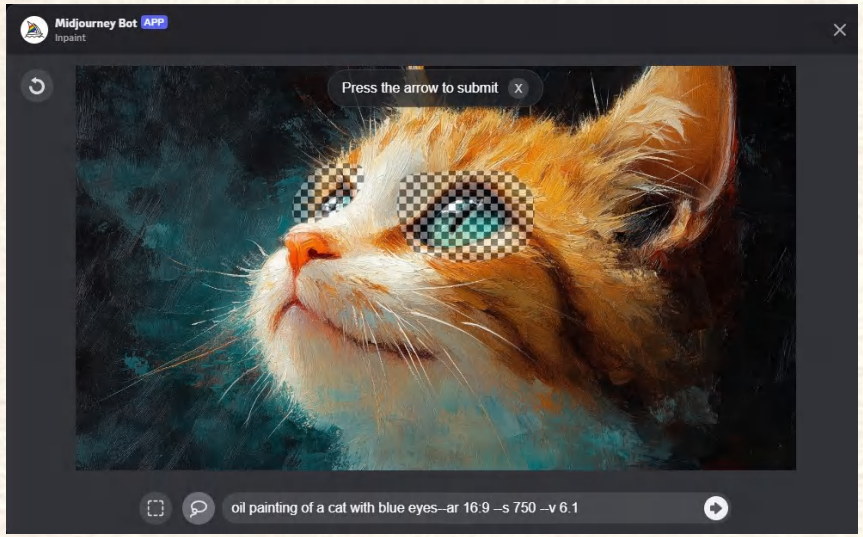
■ 图片编辑

在Discord中，生成图片后，点击Vary Region，可以通过选区或套索工具配合提示词，进行图像编辑

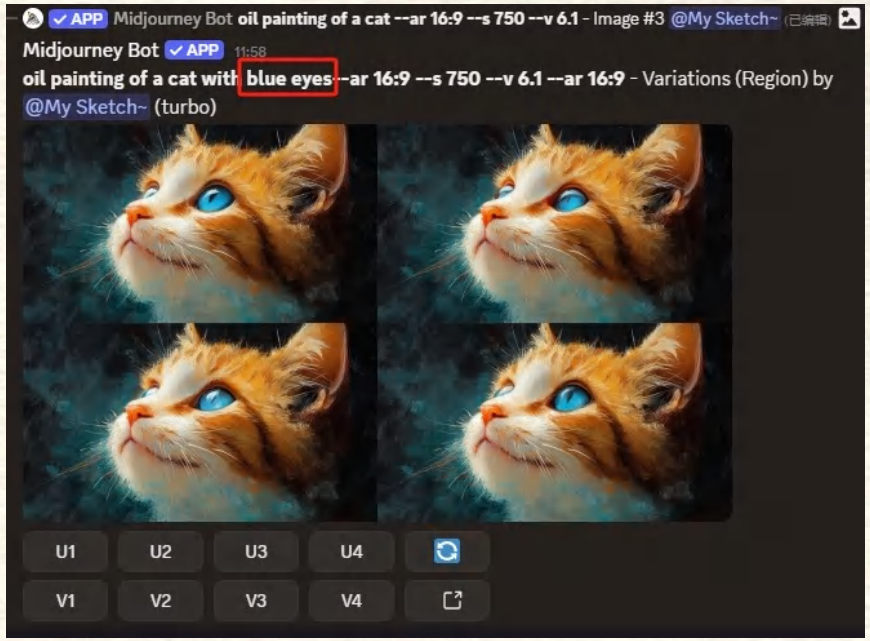
例如将图中猫眼睛从绿色改成蓝色。



除了修改之外，还可以通过Zoom扩图



选择修改区域之后，在对话框编辑提示词，尤其是要修改的部分，例如“blue eyes”



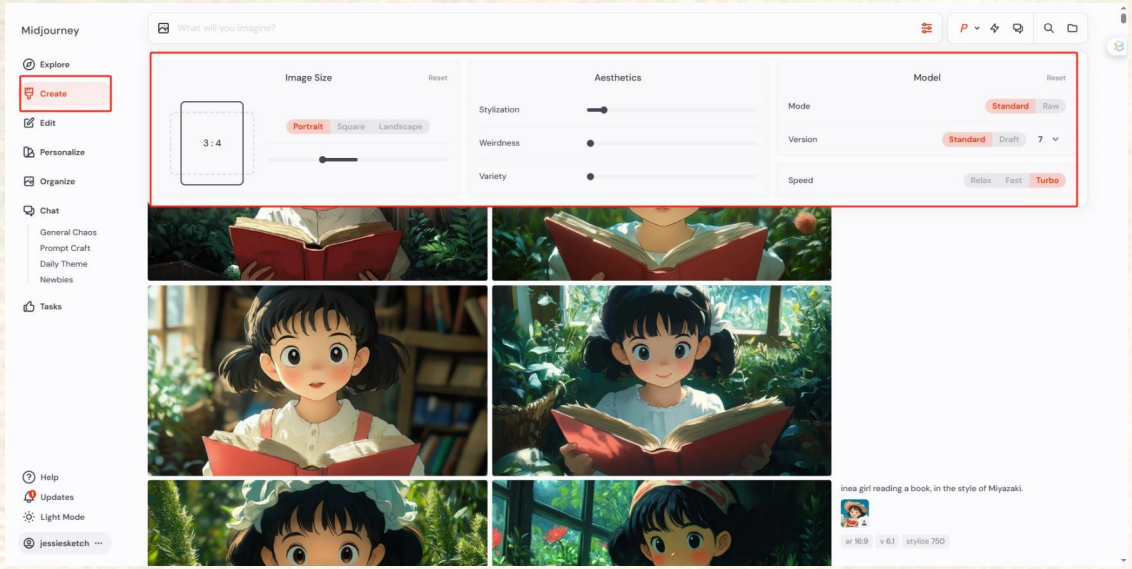
修改之后，重新出图即可获得修改后的图片

3.2 AI工具深度测评 - Midjourney

Midjourney 易用性与交互体验-Midjourney官网

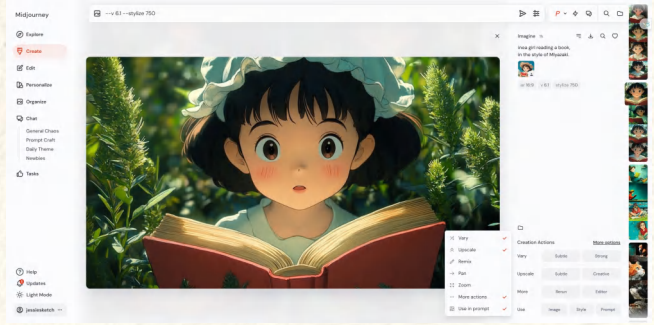
■ 出图设置

Midjourney 官网可以通过按钮操作出图参数设置，包括画幅比例、风格化程度、模型版本、出图速度

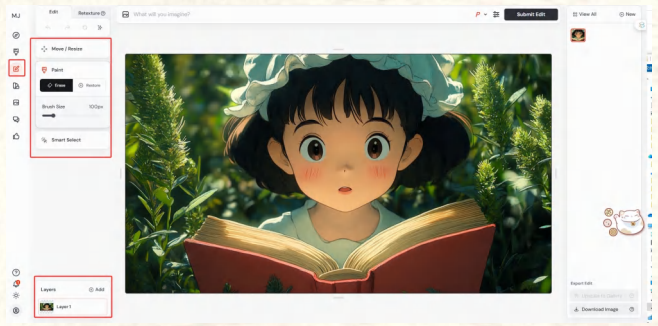


■ 图片编辑

修改图片相比Discord，更为灵活。



可以在生成轻微调整



也可以在Edit界面进一步调整

3.2 AI工具深度测评 - Midjourney

Midjourney V7 效果实测（提示词遵从程度-1）

- Midjourney的提示词规则明确，但有一定的学习成本
- 生成图像对提示词遵从程度较高（首次生图，四张均为不同风格的油画画风）

1. 写实主义与印象派融合：猫的形象比较写实，但背景采用了略微粗糙、抽象的笔触和高饱和度的颜色，像是印象派中的光影与色彩处理方式。整体画面有种复古工业感。
2. 表现主义：猫的细节虽然清晰，但色彩夸张，尤其是毛发中橙红与绿色背景的强烈对比，用色情绪化，表达情感强烈，符合表现主义风格的特征。
3. 写实主义：画面的细节密度极高，是典型的超写实主义风格。
4. 超写实主义：细节表现非常到位，尤其是毛发的层次、眼睛的光泽、皮肤质感等，都达到了“照片级”的精度



以油画风格猫咪为例，效果出色√
测试提示词：oil painting of a cat --ar 16:9 --s 750 --v 7.0

Midjourney V7 效果实测（提示词遵从程度-2）

- Midjourney尝试多风格出图，例如卡通黏土风

测试提示词：
Claymation-style scene of a single plasticine baker joyfully decorating an oversized cake at a small table in a pastel outdoor setting --ar 16:9 --s 750 --v 7.0

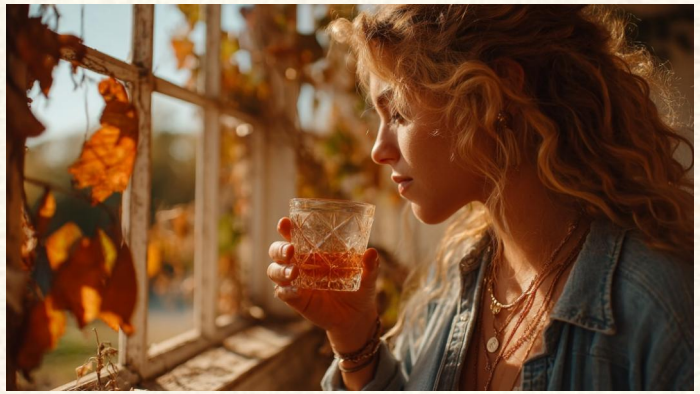




评价：风格一致，细节丰富√

Midjourney V7 效果实测（手部细节测试）

- 手部问题是所有图像类大模型的最大难点，Midjourney 从V5版本之后，有了明显改善，本次V7版本在复杂动作中表现更佳，例如拿玻璃杯、织毛衣、做陶艺，用手比心等动作，都能做到动作协调，手指正确





3.2 AI工具深度测评 - Midjourney

Midjourney V7 效果实测（风格化）

- Midjourney V7在中国画和水彩画的表现，相比V6.1版本略差一点

Midjourney V7





Midjourney V6.1





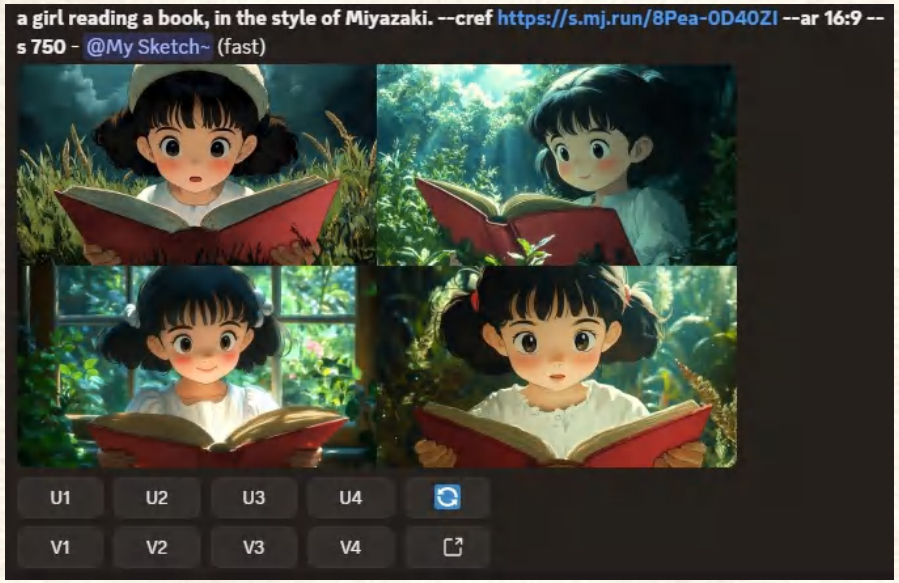
3.2 AI工具深度测评 - Midjourney

Midjourney V7 效果实测（角色一致性）

- 角色一致性是AI出图的难点之一，Midjourney V7的角色一致性还没开放，此处用V6.1来测试，使用cref参照命令（测评时间：2025年4月24日）
- 任务：将原图的角色通过提示词替换动作，从吃西瓜变成看书



原图



四张图都抓住了角色的主要特征√



角色基本接近√

Midjourney V7 效果实测（风格一致性）

- 风格一致性、风格参照/迁移，是AI绘画、设计等场景中最重要的应用之一，也是展示模型对画面风格捕捉的能力



原图



四张图在配色上，与原图基本一致√



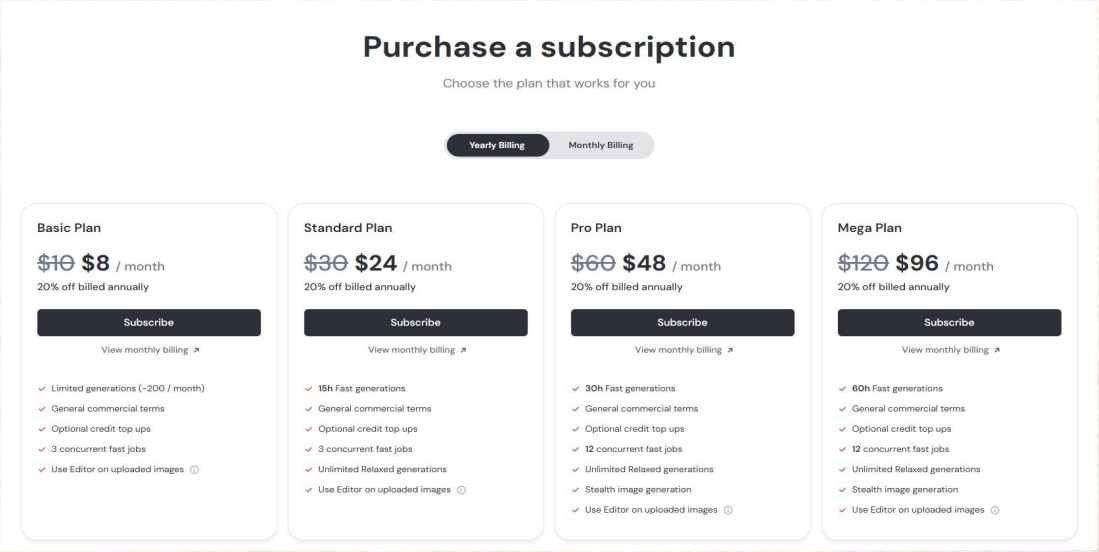
风格基本接近，且细节更清晰√

3.2 AI工具深度测评 - Midjourney

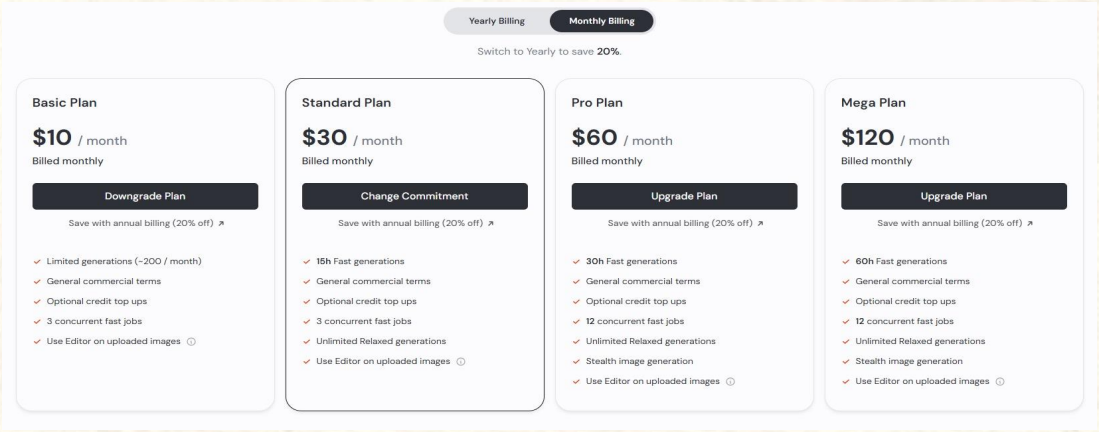
Midjourney 定价与性价比

- Midjourney是一款付费AI生图工具，用户可以根据自己的使用量，订阅不同级别的付费方案。
  - 初步尝试、长期使用，可以按年付费选择Basic Plan
  - 如果用量始终，且长期使用，推荐Standard Plan（每月24美金）
  - 如果短期尝试，可以选择月付Basic Plan（每月10美金）

评价：Midjourney出图效果惊艳，可用于专业设计与创作，相比人工制作，性价比非常高



按年付费



按月付费

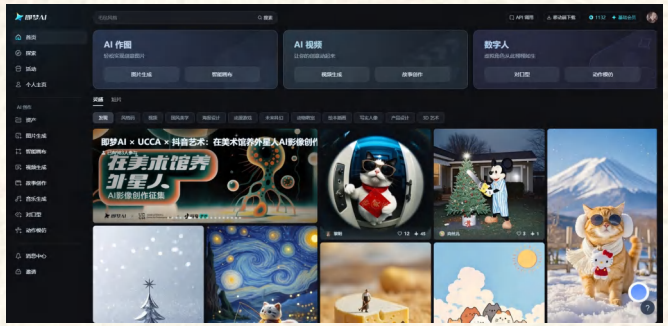
3.2 AI工具深度测评 - Midjourney

Midjourney整体总结

- **优点**
  - 👍 出图整体效果好：Midjourney是目前整体效果最好的AI绘画工具，无论是场景氛围感、人物稳定性还是工业产品质感等都有出色表现
  - 👍 提示词遵从性强：通过Midjourney提示词规则描述，即可相对准确控图
  - 👍 易用性高：提示词出图的形式，相比传统设计工具Photoshop、3ds Max等等学习成本大幅降低
  - 👍 二次编辑：可以通过涂抹或套索工具结合提示词，对图片进行简单修改
- **缺点**
  - 👎 语言：提示词通常需要使用英文，对英文基础较弱的用户不太友好
  - 👎 成本高：相较于国内AI绘画工具，成本较高
  - 👎 API：Midjourney 的服务条款不允许任何形式的自动化，因此不提供API调用
- **使用场景**
  - ➢ **平面设计、UI设计、工业设计、自媒体等追求高质量出图的场景，适合专业设计类需求的用户使用**

3.2 AI工具深度测评 - 即梦AI

- **即梦AI 简介与核心技术**
  - 即梦AI 3.0（2025.4.3更新）凭借中文生成能力、影视级画质和高清输出等核心升级，成为当前中文AI绘图领域的标杆。
- **核心定位:**
  - 提供简单易用的文本到图像、视频生成工具，强调电影级视觉质量、精准排版（尤其是英文）以及多模态创作能力，旨在降低创作门槛并赋能数字艺术、商业营销等领域
- **核心技术突破（即梦AI 3.0）：**
  - 将图像渲染质量提升至影视级水准，具备2K高清直出能力，实现了对中文文字的精准控制和灵活调整，生成速度提升30%，且模型对用户意图的理解更准确，并针对中英文混合排版和中文设计需求进行了深度优化。
- **主要产品形态:** 网页、APP
  - 网页：https://jimeng.jianying.com/
  - APP：https://ugurl.cn/JHnhP





3.2 AI工具深度测评 - 即梦AI

即梦AI 核心功能（图像部分）

- 即梦AI在图像方面操作简洁易用、效果好，提供DeepSeek进行提示词优化，及基于社区作品“做同款”大幅降低AI创作的门槛，在二次编辑方面功能简单明确，出图之后也可丝滑切换到视频制作，为视频制作提供极大便利，同时相比同类产品性价比较高，且出图速度快

- **AI作图:**
  - 文生图：根据用户输入的文字描述，创造出全新的、富有想象力的图像
  - 图生图：将用户上传的一张或多张图片作为灵感，结合文字提示生成新的图像
  - 图片混合：允许用户将多张图片融合成一张全新的、具有各图元素的合成图像

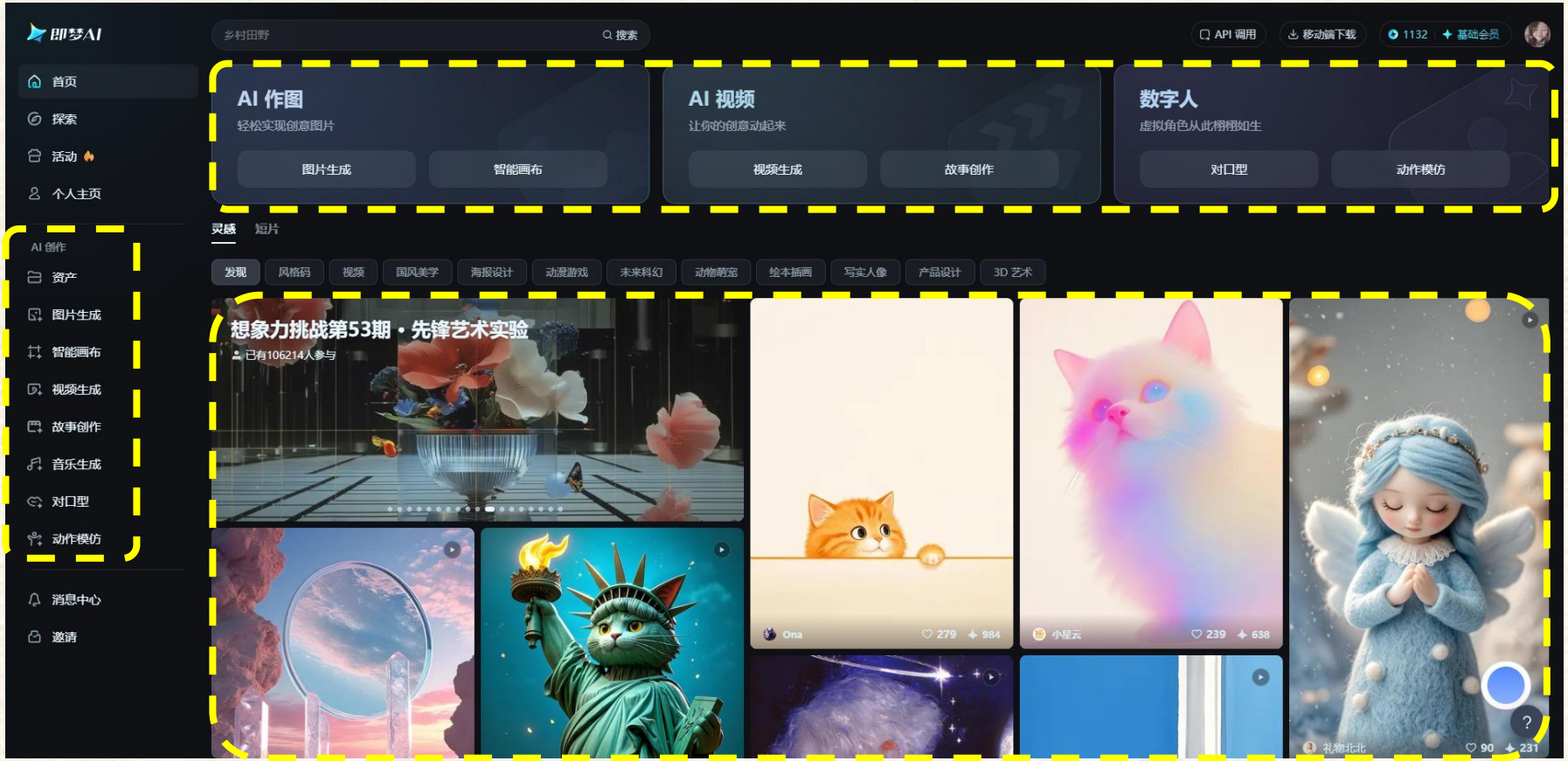
- **智能画布:**
  - 局部重绘: 涂抹图像的某一部分，结合文字描述，让AI重新绘制该区域
  - 扩图 : 扩展现有图像的边界，智能填充延展部分的内容

- **社区联动：**提供官方或用户共享的特定风格模型，通过“做同款”方便用户快速生成特定风格的图像

3.2 AI工具深度测评 - 即梦AI

即梦AI 易用性与交互体验

即梦AI界面简洁直观，且符合国内使用习惯，易于上手、交互体验流畅。图片与视频制作丝滑切换。



全部功能 ⟶

核心功能 ⟵

社区活动及作品 ⟵

3.2 AI工具深度测评 - 即梦AI

即梦AI 易用性与交互体验

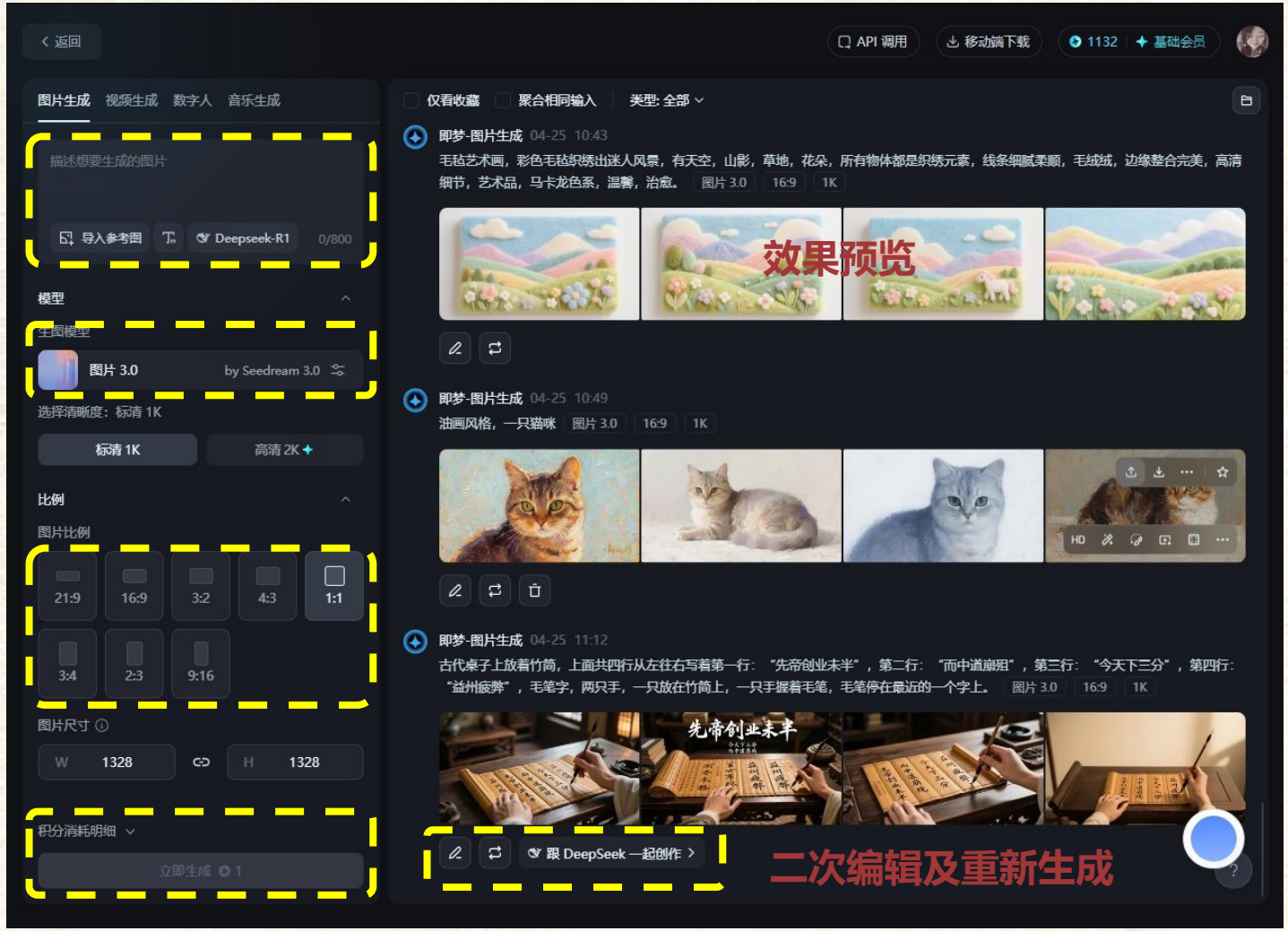
选择功能 →

输入提示词
(可通过DeepSeek优化，也可上传参考图实现角色和风格一致性) →

选择模型
（3.0提供2K高清出图） →

选择图片比例 →

点击即可出图
（预览积分消耗，有助于用户控制成本） →





手机端

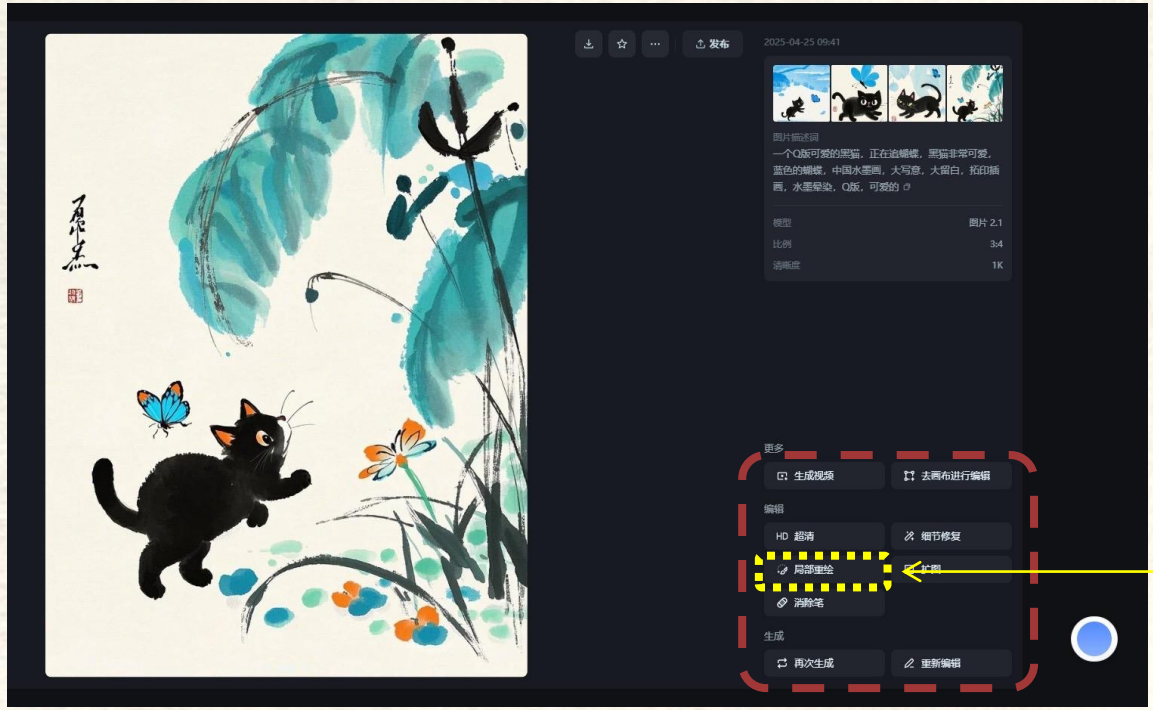
3.2 AI工具深度测评 - 即梦AI

即梦AI 易用性与交互体验

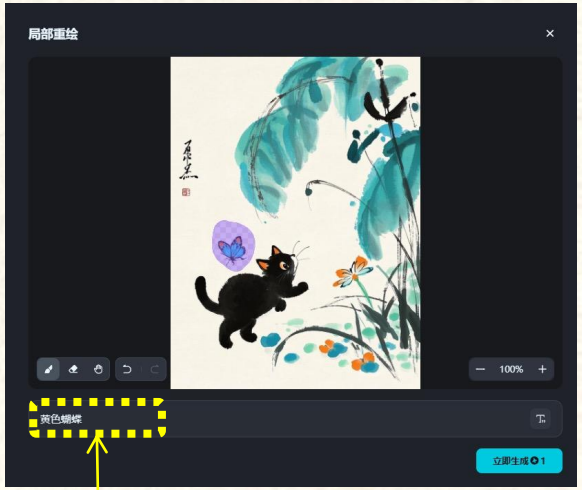
- 图片编辑

在即梦中，生成图片后，提供多种常用的二次编辑功能，例如：超清、局部重绘、消除笔、细节修复、扩图等.

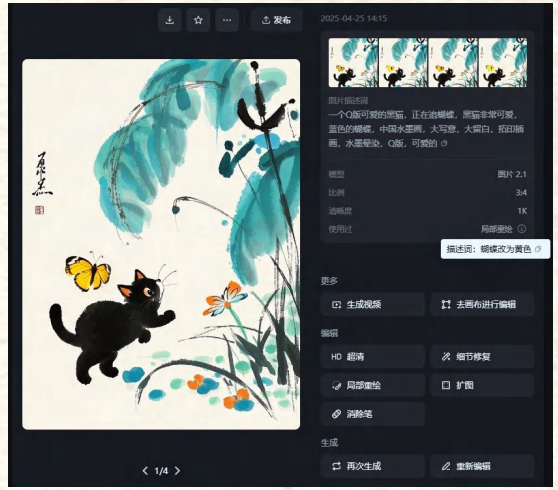
也可以直接点击“生成视频”跳转到“图生视频”界面



二次编辑功能区



涂抹修改区域，输入提示词
例如“黄色蝴蝶



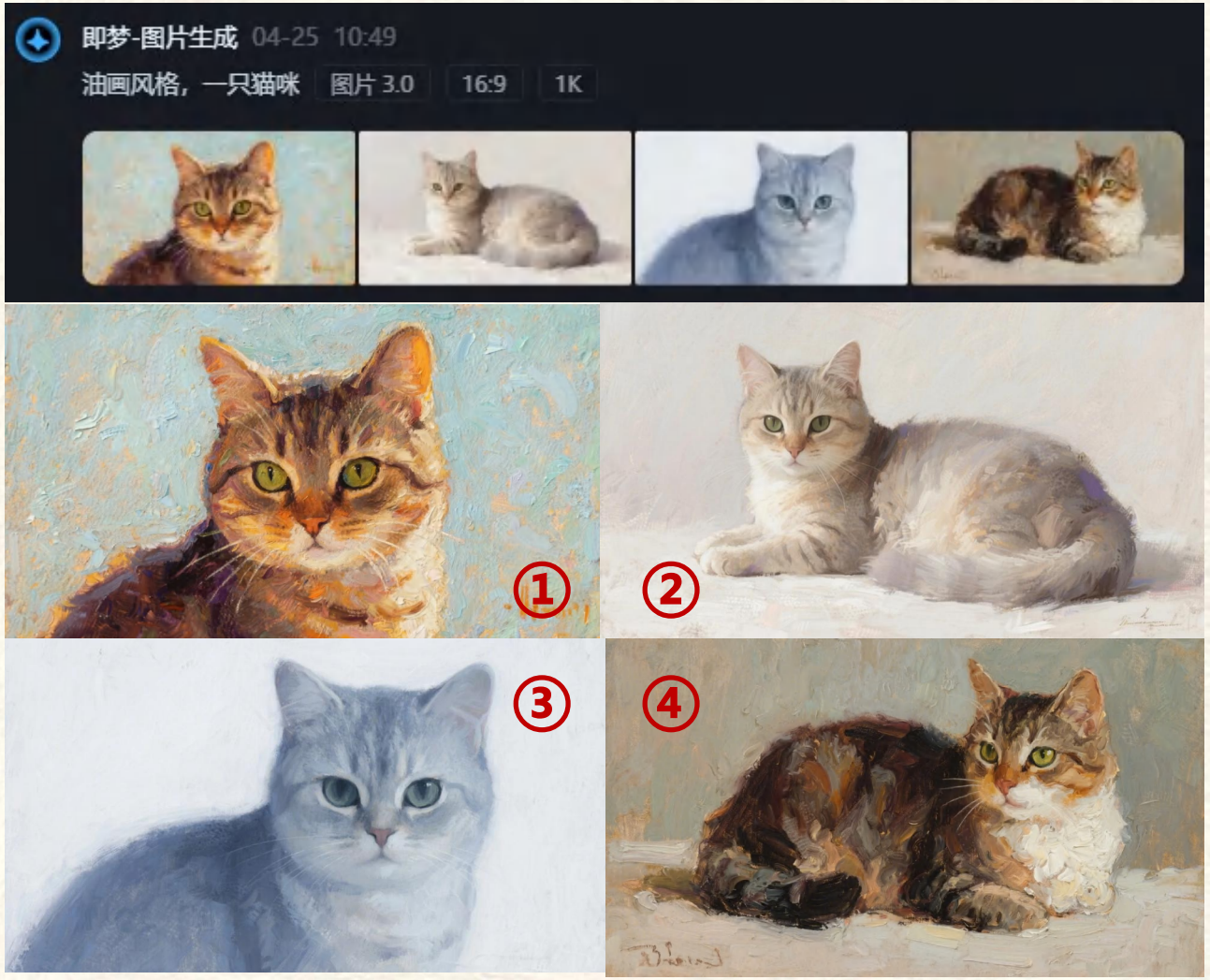
修改后可在新图的“局部重绘”看到修改的信息

3.2 AI工具深度测评 - 即梦AI

即梦AI 3.0 效果实测（提示词遵从程度-1）

- 即梦AI 对中文用户极度友好，且可以用DeepSeek进行提示词优化，学习成本极低
- 生成图像对提示词遵从程度较高（首次生图，四张均为不同风格的油画画风）

1. 印象派写实风格：用笔松散，色彩丰富，明显可以看到笔触，尤其在背景和猫咪毛发上。整体色彩明快，有一定光影对比但不过度强调细节，强调的是“印象”而非精细刻画
2. 新古典风：接近传统的古典油画风格。用笔细腻，光影过渡柔和，结构严谨。猫咪的姿态、毛发、眼神都处理得非常精细，非常强调真实感和三维感
3. 现代写实风格：虽然是写实的，但整体色调冷静，几乎没有明显的笔触，光影呈现出一种非常理性和平面的感觉
4. 表现主义：保留写实基础，但用笔更加奔放，有一定的情绪表达。背景的笔触更粗犷，颜色较为浓重，带有一定的个人表现风格

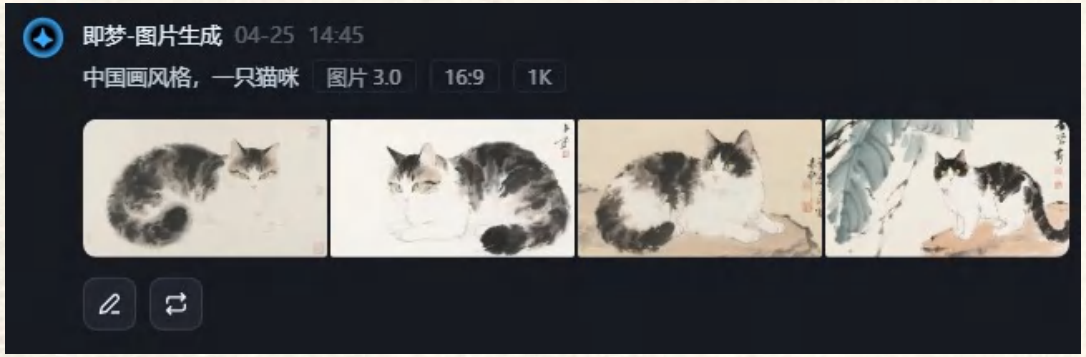


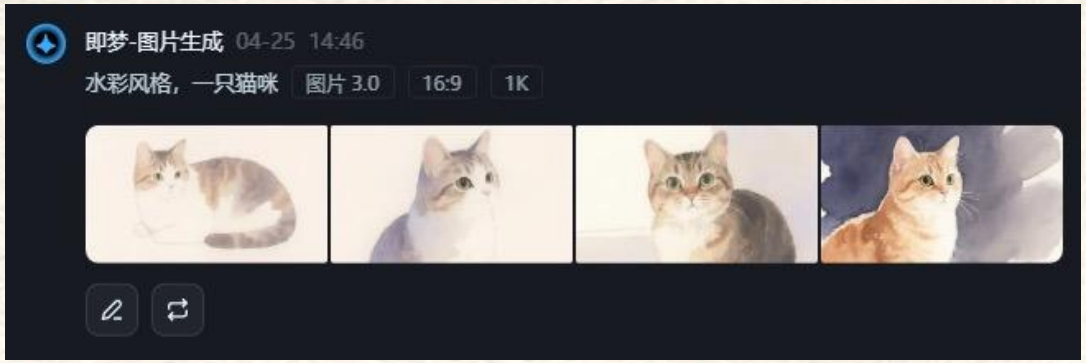
3.2 AI工具深度测评 - 即梦AI

即梦AI 3.0 效果实测（提示词遵从程度-2）

- 即梦AI 的中国画风格出图效果非常惊艳

实测示例：尝试中国画风格、水彩风格







中国画风格 √



水彩风格 √

3.2 AI工具深度测评 - 即梦AI

即梦AI 3.0 效果实测（风格化）

- 即梦AI 提供强大的社区，有众多优秀作品可作为参考，将提示词的主体内容替换，即可生成同款风格的图片



写实人像

（3.0更新后，影视质感是一大亮点）



Jellycat风格



国风美学



海报设计

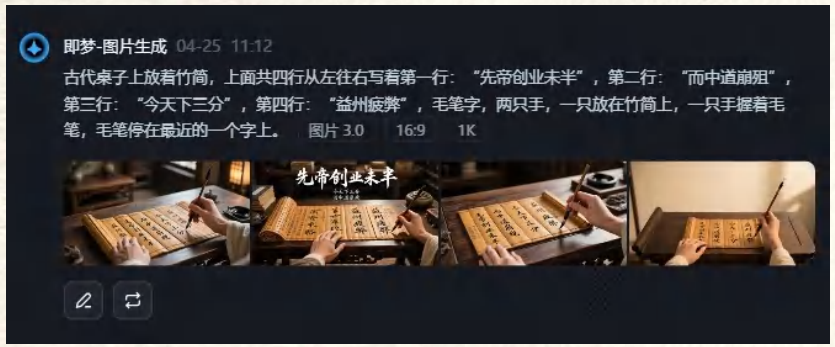
3.2 AI工具深度测评 - 即梦AI

即梦AI 3.0 效果实测（文字准确性几乎可达100%）

文字生成是所有AI生图的最大难点之一，本次3.0更新，对文字做了精准处理。尝试用“书简写字”测试文字的准确度，同时也可测试到手部处理的精细化表达。

测试提示词：古代桌子上放着竹简，上面共四行从左往右写着第一行：“先帝创业未半”，第二行：“而中道崩殂”，第三行：“今天下三分”，第四行：“益州疲弊”，毛笔字，两只手，一只放在竹简上，一只手握着毛笔，毛笔停在最近的一个字上。





虽然四张图中，有文字顺序错误和布局错误

但有两张是完全正确的，且手指处理也完全正确

√



文字准确，手部正确且环境、光影协调√

3.2 AI工具深度测评 - 即梦AI

即梦AI 3.0 效果实测（角色一致性）

- 角色一致性是AI出图的难点之一，即梦3.0的角色一致性还没开放，此处用V2.1进行测试（测评时间：2025年4月25日）
- 任务：将原图的角色通过提示词替换动作，从吃西瓜变成看书





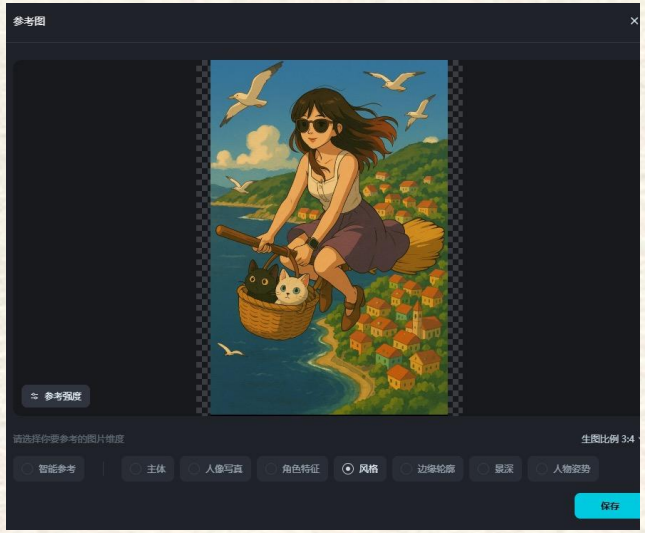
上传参考图，
卡通人物选“主体”

四张图角色确实一致，但提示词失效，角色没有看书

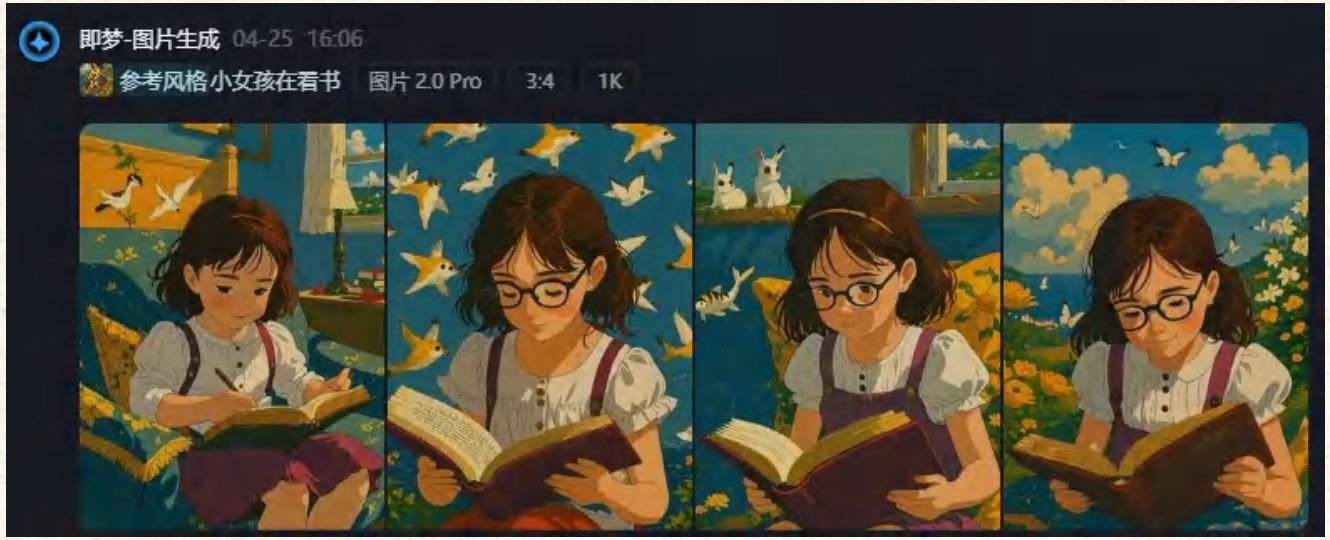
角色完全一致

即梦AI 3.0 效果实测（风格一致性）

- 即梦AI的风格参照和角色参照的操作类似，但风格参照，仅用于2.0 Pro和之前版本



上传参考图，
选择“风格”
（Tips：可以上传多张参考图）



四张图在配色上，与原图基本一致



风格一致
（手指有错误）

3.2 AI工具深度测评 - 即梦AI

即梦AI 定价与性价比

- 即梦AI是一款付费AI图像、视频综合工具，效果基本可商用，相比人工和国外AI工具，性价比非常高。每日免费60积分，可供体验
- 用户可以根据用量选择订阅计划，或购买积分（适合少量使用）



积分购买



连续包年



连续包月



单月购买

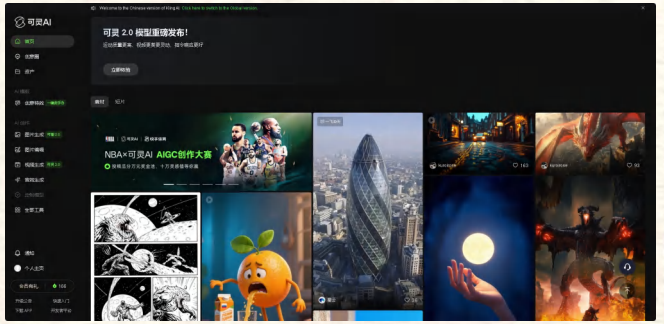
3.2 即梦AI实测总结

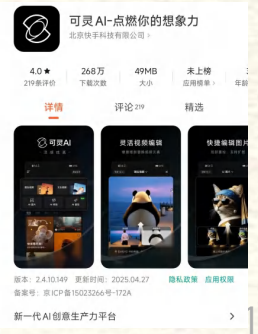
即梦AI 整体总结

- **优点**
  - 👍 中文提示词友好：作为国产AI工具的最大优势是通过中文提示词自然语言出图
  - 👍 角色、风格一致性较好：少量“抽卡”即可获得优秀的角色、风格一致性图片，最新“风格码”功能可稳定输出制定风格图片
  - 👍 **文字准确性：文字几乎无需“抽卡”即可生成包含正确文字的图片（目前所有AI生图工具中最强）**
  - 👍 国风效果惊艳：对于中国风的理解非常，无论是中国画还是中国风氛围感场景都有出色表现
  - 👍 易用性高：操作简单几乎零门槛，还可以通过社区优秀作品“做同款”，大幅降低学习成本
  - 👍 二次编辑：几乎涵盖主要图片修改功能，进一步降低设计门槛
  - 👍 成本低：相比Midjourney，即梦AI提供更多档位的付费选择空间，同时提供每天积分赠送，新手可放心试用
  - 👍 API：提供API调用，可灵活接入自己的程序，且与COZE联动友好
- **缺点**
  - 👎 稳定性：复杂场景仍需“抽卡”，且多次修改易出现物体变形的情况
  - 👎 审美：首次出图审美相比Midjourney较弱一点
- **使用场景**
  - ➢ **平面设计、UI设计、工业设计、自媒体等追求高质量出图的场景，以及对美学要求不太高的场景**

3.2 AI工具深度测评 - 可灵AI

- **可灵AI 简介与核心技术**
  - 可灵AI（KLING）是快手自主研发的全球领先视频生成大模型，可图 2.0 基座模型2025年4月上线，大幅度提升提示词遵循能力，能响应近百种风格画面更具电影质感。
- **核心定位**
  - 旨在降低视频创作门槛，赋能个人创作者与企业用户，推动AI生成内容（AIGC）的普及。提供高效、低成本的视频生成工具，激发用户创意表达。可通过API服务赋能企业，优化广告、电商、短剧等内容生产效率
- **核心技术突破（可图 2.0）**
  - 可图2.0则在静态图像生成方面取得重大突破，对复杂指令的遵循能力显著提升，能够准确响应多种元素、颜色位置等细节要求。支持超过60种艺术风格，从特殊材质到数字艺术再到各类绘画技法，使普通用户也能轻松创作出具有高级电影质感和独特美学调性的图像。
- **主要产品形态:** 网页、APP
  - https://app.klingai.com/cn/





可灵AI 核心功能（图像部分）

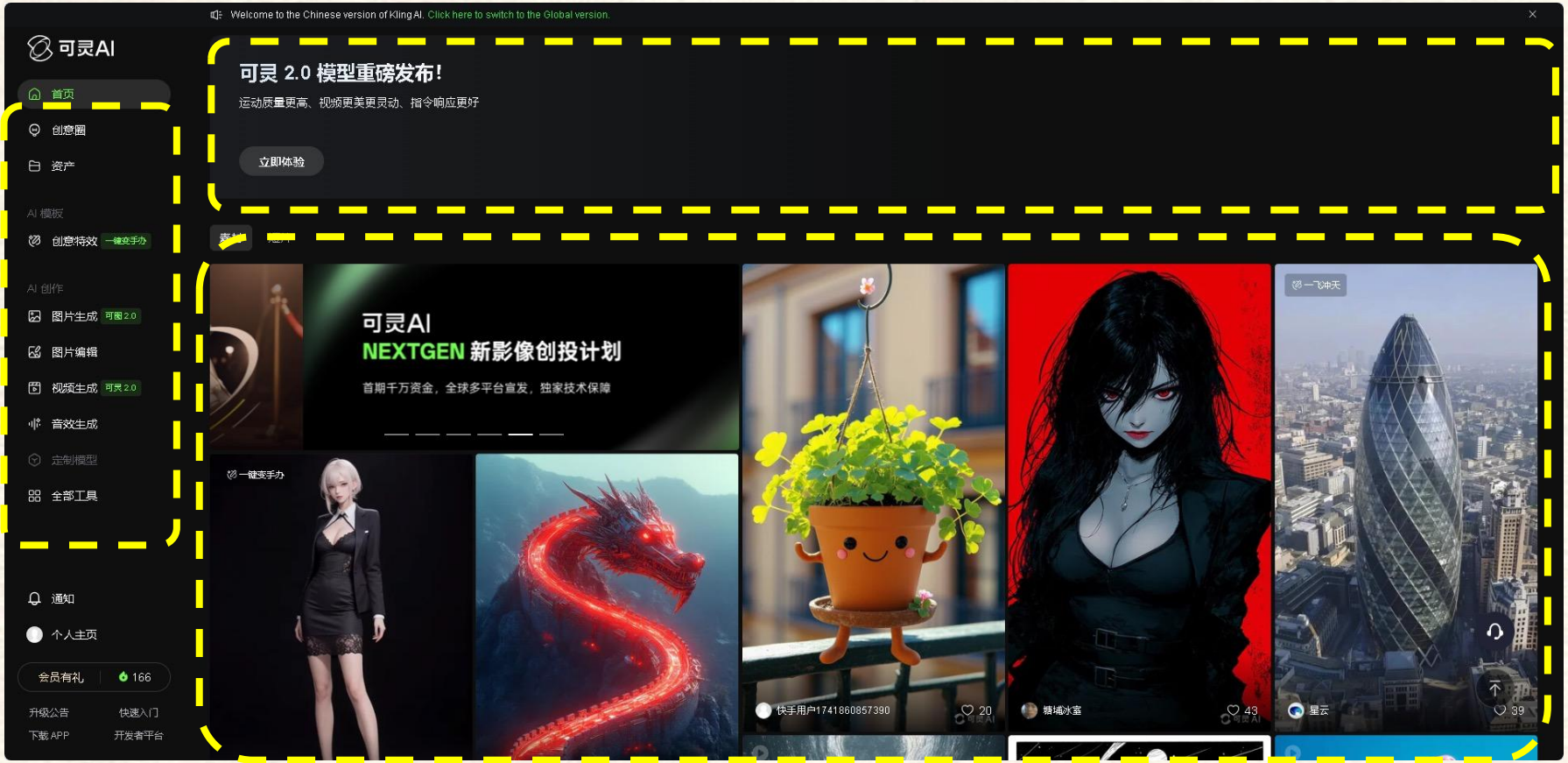
- 可灵AI在图像方面操作简洁易用、效果好，提供DeepSeek进行提示词优化，及基于社区作品“一键同款”大幅降低AI创作的门槛，在二次编辑方面功能简单明确，出图后可一键生成视频、局部重绘、扩图、画质增强、一键作为垫图，同时相比同类产品性价比较高，且出图速度快。

- **AI作图:**
  - 文生图：根据用户输入的文字描述，创造出全新的、富有想象力的图像
  - 图生图：将用户上传的一张图片作为参考图，选择提取特征选项（角色特征、人物长相、风格转绘、通用垫图）结合文字提示生成新的图像
- **图片编辑:**
  - 局部重绘: 涂抹图像的某一部分，结合文字描述，让AI重新绘制该区域
  - 扩图：扩展现有图像的边界，智能填充延展部分的内容
- **社区联动：**提供官方或用户共享的特定风格模型，通过“做同款”方便用户快速生成特定风格的图像

3.2 AI工具深度测评 - 可灵AI

可灵AI 易用性与交互体验

可灵AI界面简洁直观，符合国内使用习惯，易于上手、交互体验流畅。图片与视频制作丝滑切换。



全部功能 ⟶

⟵ 最新发布

⟵ 社区活动及作品

3.2 AI工具深度测评 - 可灵AI

可灵AI 易用性与交互体验

选择模型
（2.0提供2K高清出图）

输入提示词
(可通过DeepSeek优化，也可上传参考图实现角色和风格一致性)

选择功能
（图片生成、视频生成、音效生成、AI试衣、对口型、视频延长）

选择图片比例及数量

点击即可出图
（预览积分消耗，有助于用户控制成本）

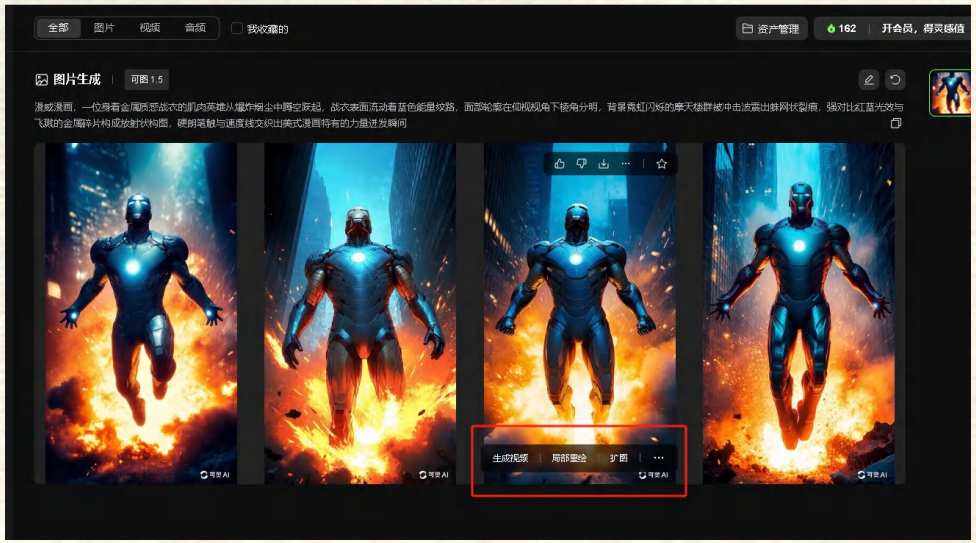


Tips：建议使用网页端进行生图操作，手机端操作易用性较差一些

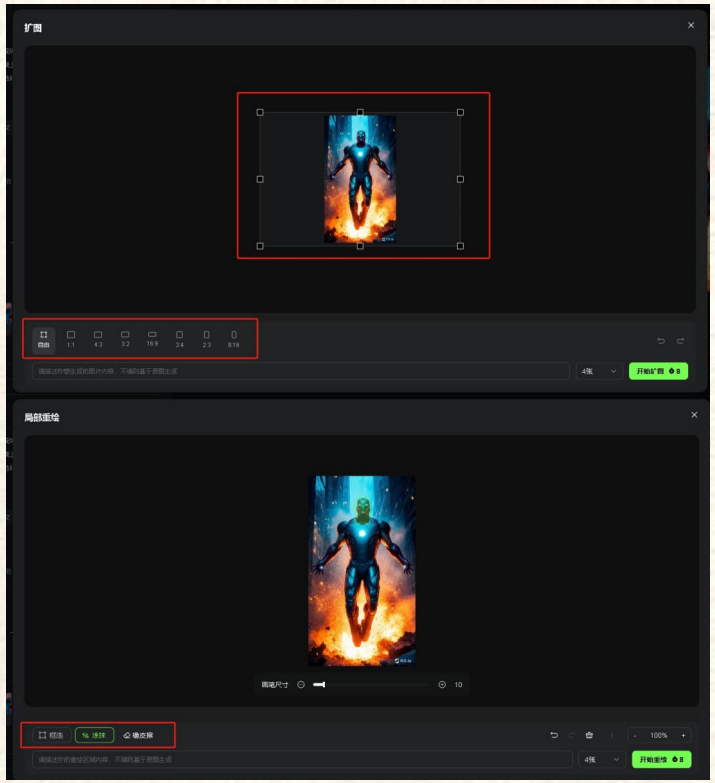
可灵AI 易用性与交互体验

■ 图片编辑

可灵AI生成图片后，提供多种常用的二次编辑功能，例如：生成视频、局部重绘、扩图、作为垫图、画质增强等。（画质增强功能需要会员权限）



二次编辑功能区（或者点击图片展开功能区）



扩图：在生成图像基础上，选择不同尺寸，输入提示词进行扩图操作，扩图后可继续进行编辑

局部重绘：该功能下可对图片中内容进行擦除，局部重绘支持框选、涂抹并根据输入提词进行局部重绘

3.2 AI工具深度测评 - 可灵AI

可灵AI 2.0效果实测（提示词遵从程度-1）

- 可灵AI 对中文用户极度友好，且可以用DeepSeek进行提示词优化，学习成本极低

测试提示词：

“复古风格的麦当劳美食海报，采用丝网印刷技术，色彩浓郁饱满。画面中央是一份金黄酥脆的大麦克汉堡，两片芝麻面包之间夹着牛肉饼、芝士、生菜和特制酱汁，诱人垂涎。汉堡旁摆放着热腾腾的薯条和一杯冰凉的可口可乐。背景是50年代美国街头的繁华景象，霓虹灯闪烁，汽车穿梭。海报上方用大胆的无衬线字体写着“Taste the Tradition”，下方是麦当劳标志性的金色拱门Logo。整体构图平衡，色彩对比强烈，充满怀旧气息。”





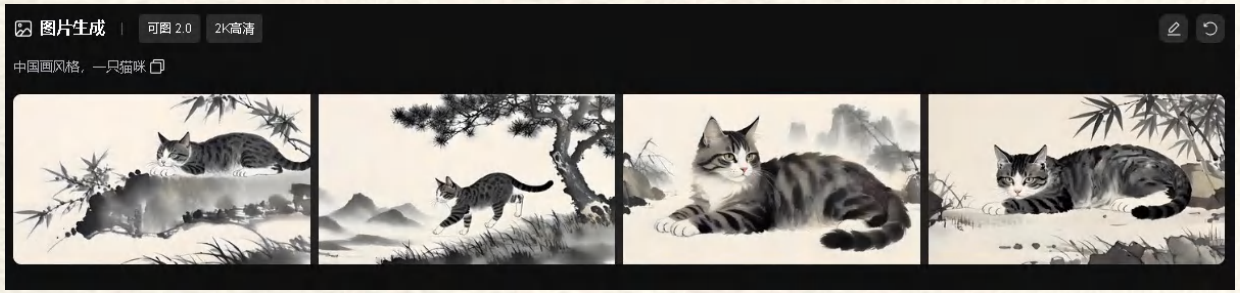
- **测试评价**：
  - 复杂提示词测试下，四张生成图像风格、物品主体和空间位置与提示词中大部分描述内容一致。且设计极富艺术感。
  - 但生成的四张图中“Tradition”这个单词全部错误。

3.2 AI工具深度测评 - 可灵AI

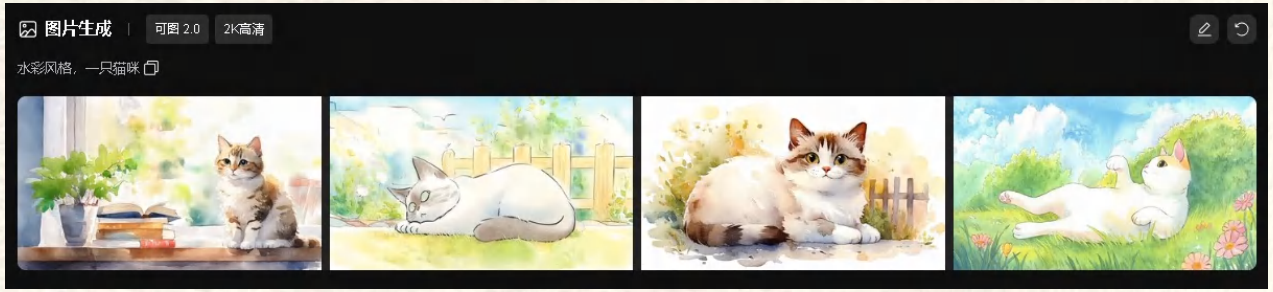
可灵AI 2.0效果实测（提示词遵从程度-2）

- 即梦AI 的中国画风格出图效果非常惊艳

评测示例：尝试中国画风格、水彩风格









3.2 AI工具深度测评 - 可灵AI

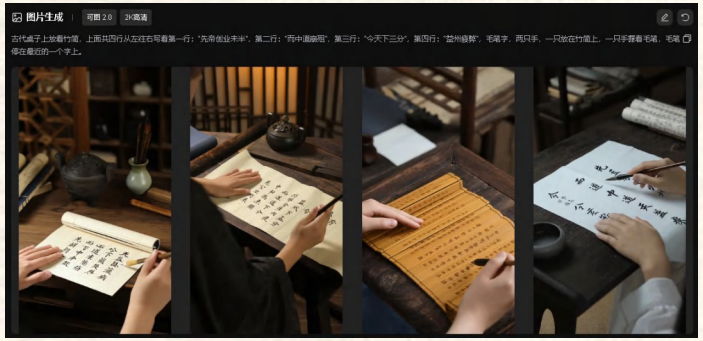
可灵AI 2.0效果实测（文字准确性）

- 文字生成是所有AI生图的最大难点之一。尝试用“书简写字”测试文字的准确度，同时也可测试到手部处理的精细化表达。

测评结果：表现不佳，文字内容全部错误，但是手部处理正确。弱于即梦AI3.0

测试提示词：古代桌子上放着竹简，上面共四行从左往右写着第一行：“先帝创业未半”，第二行：“而中道崩殂”，第三行：“今天下三分”，第四行：“益州疲弊”，毛笔字，两只手，一只放在竹简上，一只手握着毛笔，毛笔停在最近的一个字上。





四张图中，文字内容全部错误

但手指处理完全正确



手部正确且环境、光影协调

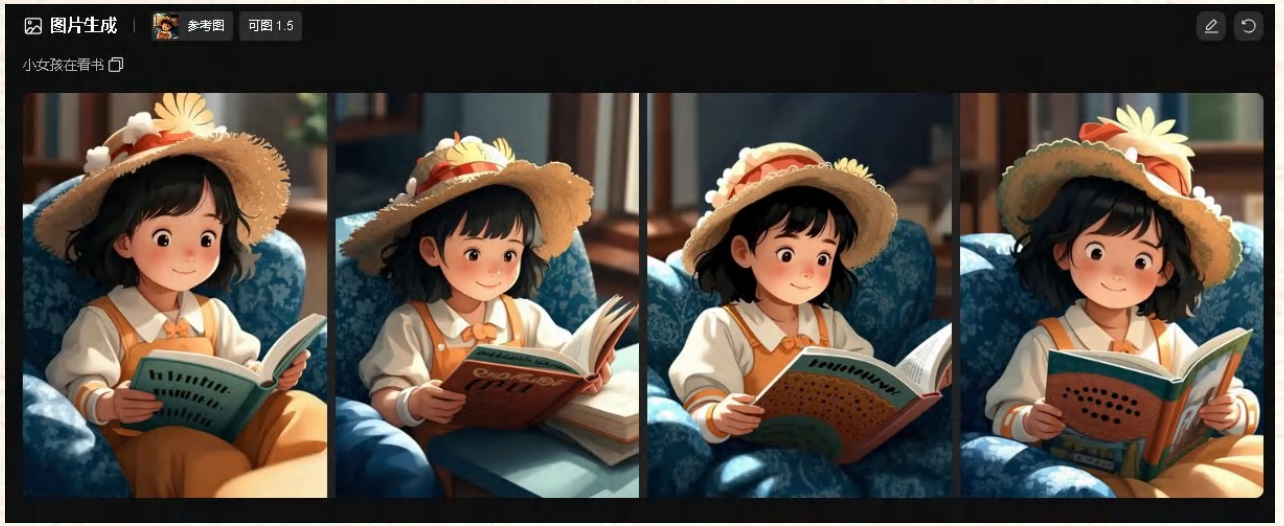
3.2 AI工具深度测评 - 可灵AI

可灵AI 2.0效果实测（角色一致性）

- 角色一致性是AI出图的难点之一，既要考虑角色主体与参考图一致又要考虑与提示词所给描述内容一致。目前（测评日期：2025.4.28）V2.0的角色一致性不支持，此处用V1.5来测试
- 任务：将原图的角色通过提示词替换动作，从吃西瓜变成看书



上传参考图，
设置脸部参考及主题参考值



四张图角色确实在看书，但与所给图片中人物及画面风格稍有区别（脸部参考及主体参考设置为90）



动作与提示词一致

3.2 AI工具深度测评 - 可灵AI

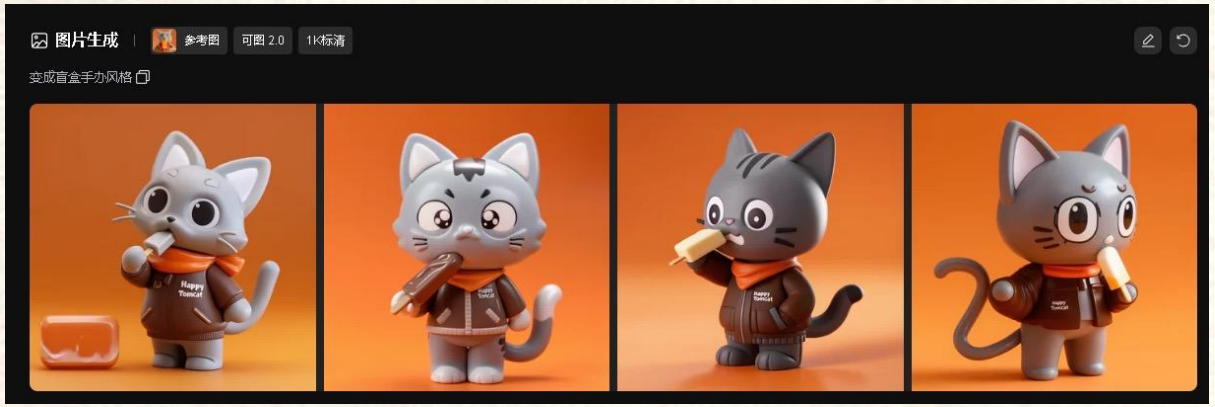
可灵AI 2.0效果实测（风格转绘）

- 可灵AI的风格转绘主要维持画面的主要内容，包括主体、构图、姿势等不变，通过prompt描述改变画面风格。

测试提示词："变成盲盒手办风格"



上传参考图，
选择“风格转绘”



四张图在图片风格上与提示词描述一致



风格与提示词描述一致，
但人物的姿势构图与原图有区别

3.2 AI工具深度测评 - 可灵AI

可灵AI 2.0 效果实测（人物长相）

- **可灵AI的人物长相功能**：维持人物的人脸长相不变，通过文本描述(prompt)即可随意改变人物的服装、发型、配饰、场景等元素，且可保持人物长相与参考图高度相似，轻易实现同一人物在多场景的创作需求。注：v1.5支持，v2.0目前不支持（测评日期：2025.4.28）

测试提示词：“黑白风格，银色窄框眼镜，黑白照片，黑白质感人像摄影, 简约灰色背景，基础款白色棉质T恤, 搭配黑色高级定制西装外套, 黑色高腰西裤, 黑色短发自然蓬松, 眉眼深邃有神, 东亚男性特征, 姿态慵懒随性, 侧身45度站立，背景里有一些光晕。”

上传参考图，选择“人物长相”，选择图片中人物并设置参考强度





四张图在图片背景、人物细节上与提示词描述一致，但其中三张图片的人物头部细节与提示词描述不同



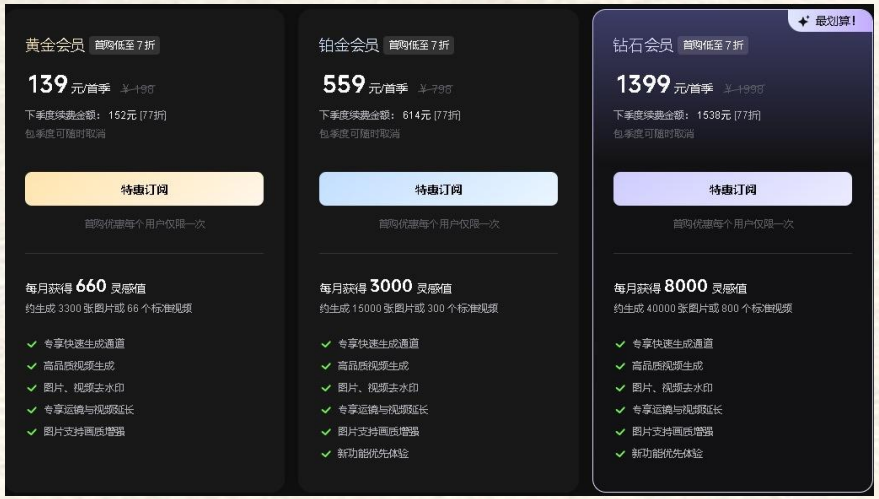
3.2 AI工具深度测评 - 可灵AI

可灵AI 定价与性价比

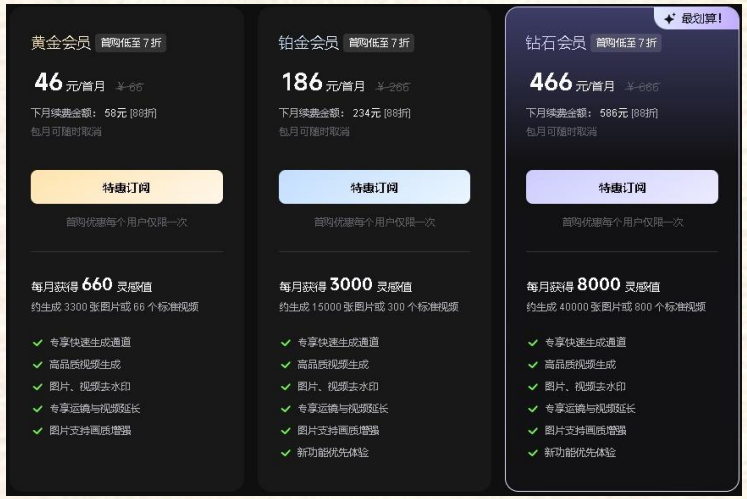
- 可灵AI是一款付费AI图像、视频综合工具，效果基本可商用，相比人工和国外AI工具，性价比非常高。每月免费166灵感值，可供体验
- 用户可以根据用量选择订阅计划，或购买积分（适合少量使用）



灵感值购买



连续包年



连续包月



按月购买（3个月9折、12个月7折）

3.2 可灵AI实测总结

可灵AI整体总结

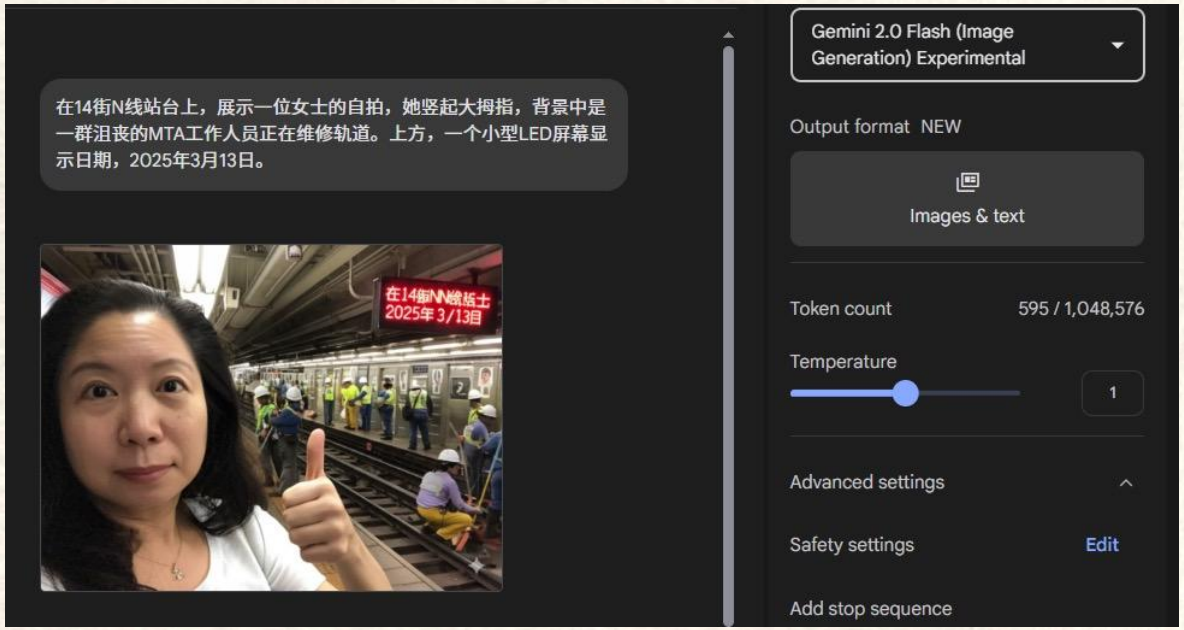
- **优势：**
  - 👍出图整体效果好: 国内目前整体效果较好的AI绘画工具，无论是场景氛围感、人物稳定性还是工业产品质感等都有出色表现
  - 👍提示词遵从性强: 对于复杂提示词内容理解较好
  - 👍易用性好: 功能使用设计上良好，符合国人使用习惯，提高设计人员及其它用户的工作效率
  - 👍二次编辑:可以通过涂抹或套索工具结合提示词，对图片进行简单修改
- **不足：**
  - 👎文字生成: 当需要在生成图片中添加文字时，不管是中英文，表现总体很差
  - 👎成本高: 相较于国内其它AI绘画工具，成本较高
- **推荐场景**：平面设计、UI设计、工业设计、自媒体等追求高质量出图的场景。

3.2 AI工具深度测评 - Gemini

Gemini 图像方面

https://aistudio.google.com/

Gemini 2.0 Flash (Image Generation) Experimental 具有强大的图片生成及修改能力，虽然尚未达到完全商用标准，但日常办公，生活场景、设计推敲完全满足，而且免费在谷歌的AI Studio中使用，方便快捷。



尝试生成一张带有动作和文字的图，提示词：
在14街N线站台上，展示一位女士的自拍，她竖起大拇指，背景中是一群沮丧的MTA工作人员正在维修轨道。上方，一个小型LED屏幕显示日期，2025年3月13日。



修改文字基本准确，但需要“抽卡”，且文字有变形（图中是抽卡2次获得的正确效果）

3.2 AI工具深度测评 - Gemini

Gemini 图像方面（ Gemini 2.0 Flash ）

Gemini 出图速度非常快，可用于创意或头脑风暴阶段，通过自然语言设计并生成地中海风格客厅效果图只需5秒。

简单的局部修改可以省去Photoshop手动修图的工作量，且操作零门槛。



提示词：
设计一个地中海风格的客厅，并生成效果图



将沙发全部改成白色



将女孩衣服改为黄色衬衫



改后效果：女孩发型和电脑细节略有变化，背景和环境保持稳定，适合对图片要求不高的日常办公或生活使用

3.2 AI工具深度测评 - GPT-4o

GPT-4o 核心功能

https://chatgpt.com/

GPT-4o 的核心是理解能力，你可以不断地反馈提出自己的要求，并在连续修改的过程中保持图片一致性可达95%。但单论一次性生图，审美上、细节上基本可用，但依然不如Midjourney。



将图片改为
吉卜力风格





Q版建筑卡通风格图片生成



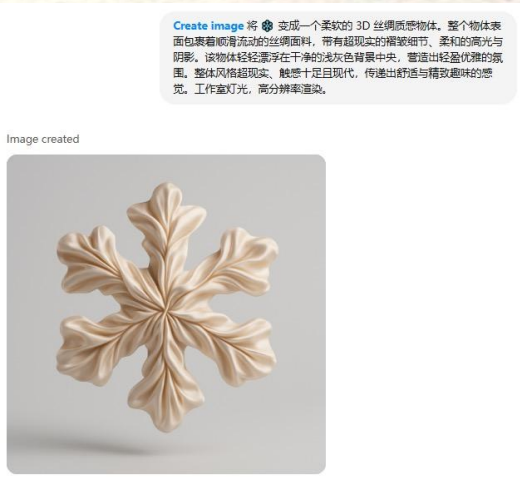
LOGO改色成功



给每个人物戴帽子成功，
但路灯也有帽子了->错误



海报生成
（风格、文字正确）



将Emoji表情生成丝绸质感

3.2 Gemini vs GPT-4o 实测总结

Gemini、GPT-4o 整体总结

工具	优点	缺点	适用场景
Gemini	1.语义理解能力强，自然语言指令友好 2.速度快，网络正常情况下，5秒左右即可完成出图 3.免费不限量	1.图片修改稳定性相比GPT-4o稍弱一点 2.生成文字准确性稍弱一点	设计创意推敲、快速出图、日常办公等
GPT-4o	1.风格化效果十分出色 2.出图质感及氛围感很棒，仅次于Midjourney 3.图片修改稳定性强 4.文字生成稳定性较强	1.生图速度慢，且免费用户有限额 2.对提示词质量有一定要求	设计创意推敲、产品材质更改、日常办公，以及已升级plus的用户

分享讨论环节：

作者认为Gemini和4o的能力不相上下，Gemini语义理解强，4o的风格化出色，都可以适合专业设计师使用（当然也有设计师自身偏好的影响啦），但从国内用户的易用性和成本考量上来讲，Gemini更胜一筹。各位读者有什么使用感受欢迎交流沟通～～

3.2 图像生成类AI工具 - 实测总结

工具	优点	缺点	适用场景
Midjourney	1.能力最强+发挥稳定 2.细节处理及质感最强 3.提示词遵从性强，可相对准确控图	1.中文提示词支持效果较差，对英文基础较弱的用户不太友好 2.不提供API调用	追求高质量出图的场景，适合专业设计类需求的用户使用
即梦AI	1.中文提示词友好，角色风格一致性较好，国风效果出色 2.（中文）文字准确度超高（文字生成几乎无需抽卡-目前AI生图工具最强） 3.易用性高，社区支持优秀，学习成本低 4.二次编辑强，几乎涵盖主要图片修改功能，降低设计门槛 5.提供API调用，可灵活接入自己的程序，且与COZE联动友好	1. 复杂场景空图较差，仍需“抽卡”，且多次修改易出现物体变形的情况 2.审美方面：首次出图审美相比Midjourney较弱一点	平面设计、UI设计、工业设计、自媒体等追求高质量出图的场景，以及对美学要求不太高的场景（国内用户首选）
可灵AI	1.出图整体效果好，国风效果惊艳 2.提示词遵从性强: 对于复杂提示词内容理解较好 3.易用性好: 功能使用设计符合国人使用习惯 4.二次编辑:可以通过涂抹或套索工具结合提示词，对图片进行简单修改	1.文字生成能力弱（无论中英文，出错概率高） 2.相较于国内其它AI绘画工具，成本较高	平面设计、UI设计、工业设计、自媒体等追求高质量出图的场景，以及对美学要求不太高的场景（国内用户首选）
Gemini	1.语义理解能力强，自然语言指令友好 2.速度快，网络正常情况下，5秒左右即可完成出图 3.免费不限量	1.图片修改稳定性相比GPT-4o稍弱一点 2.生成文字准确性稍弱一点	设计创意推敲、快速出图、日常办公等
GPT-4o	1.风格化效果十分出色 2.出图质感及氛围感很棒，仅次于Midjourney 3.图片修改稳定性强 4.文字生成稳定性较强	1.生图速度慢，且免费用户有限额 2.对提示词质量有一定要求	设计创意推敲、产品材质更改、日常办公，以及已升级plus的用户

图像编辑类AI工具测评

3.3 图像编辑类AI工具深度测评

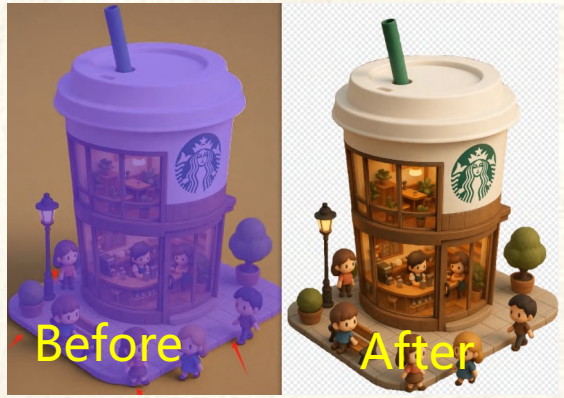
- **测评目的:**
  - 在图像创作领域，仅靠AI生成可能难以商用，需要配合类似Photoshop的图像编辑的工具完成特定任务。本次选取常用修图功能对工具进行测评
- **测评维度:**
  - 核心功能与效果 (重点)
  - 易用性与交互体验
  - 定价与性价比
- **测评工具选型：**
  - 百度AI图片助手（免费） 百度AI图片助手 https://image.baidu.com/
  - SnapEdit（限量免费） SnapEdit https://snapedit.app/create
- **测评结果对比：**
  - **百度AI图片助手：**大众赛道，适合普通用户使用，完全免费，功能全面，便捷且效果不错。
  - **SnapEdit ：**专业赛道，更适合设计工作者，要求较高的用户，专供3个核心功能（便清晰、消除涂抹、去背景），部分付费，效果出色。

3.3 AI工具深度测评 - 百度AI图片助手

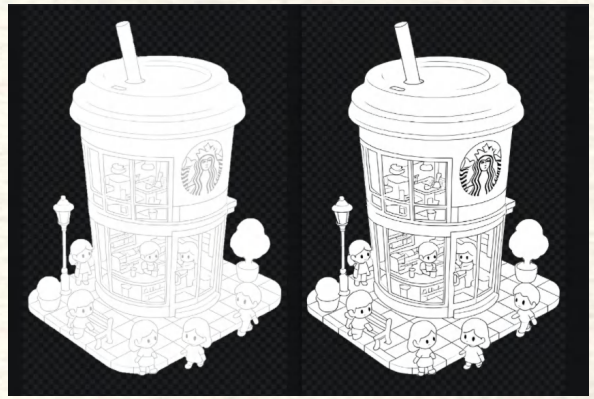
百度AI图片助手核心功能及实测效果



功能1：变清晰（质量较高√）



功能2：智能抠图
（基本可用，边缘需要细化）



功能3：提取线稿
（线稿准确且提供两种明度的线稿）



功能4：去水印（易用零门槛，细节略有丢失）



功能5：涂抹消除（有印记）



功能6：背景替换（基本可用）

3.3 AI工具深度测评 - SnapEdit

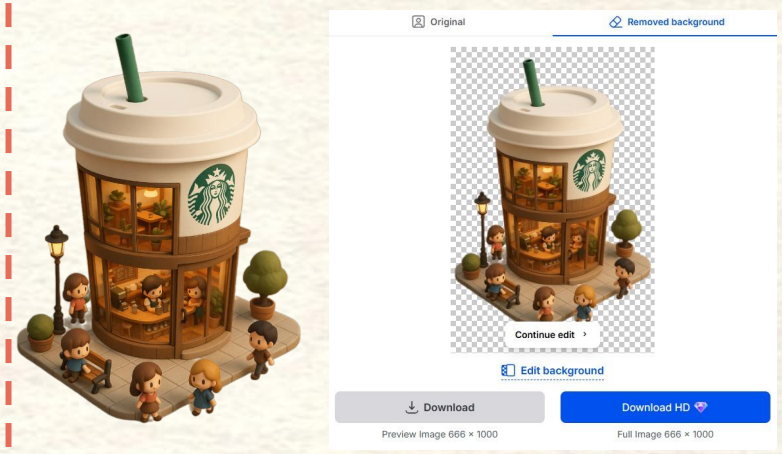
SnapEdit 核心功能实测效果及使用成本



功能1：变清晰
（质量较高√，有2K、4K付费可选）

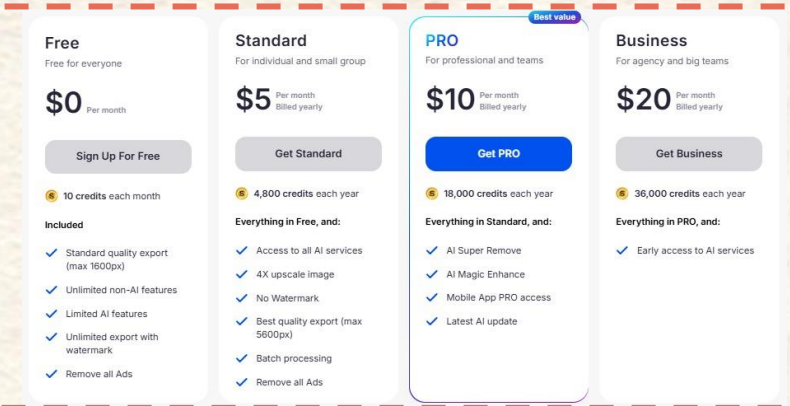


功能2：消除涂抹
（质量非常高√）

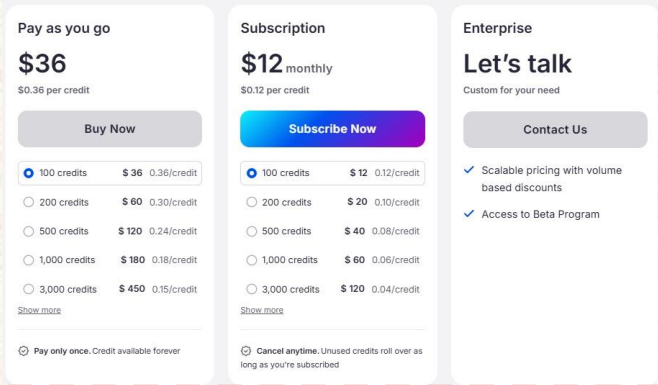


功能3：去除背景
（免费的质量非常高√还可付费下载高清版）

付费价格，可以按需订阅



提供API付费调用



3.3 图像编辑类AI工具 - 实测总结

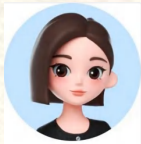
百度AI图片助手、SnapEdit 整体总结

工具	优点	缺点	适用场景
百度AI图片助手	1.功能齐全，涵盖主要修图场景 2.图片变清晰质量较高，普通办公场景基本可以用 3.免费不限量	1.细节处理能力较弱 2.风格化AI味较重	适合用户：大众使用 适合快速修图及日常办公场景
SnapEdit	1.细节处理出色，可商用 2.仅包含3个核心修图功能，但每个都十分出色 3.图片修改稳定性强	1.需要付费 2.对网络有一定要求	适合用户：设计者 适合快速修图及设计图辅助修改，专业度更高

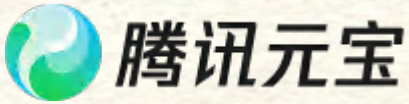
图像生成+编辑类AI工具测评

3.4 图像生成+编辑类AI工具深度测评

- **简介: 本部分统一测评免费AI工具通义万相、豆包、腾讯元宝、阶跃星辰的生图、编辑功能。**
- **测评工具介绍**
  - **通义万相：**阿里推出的生图工具（https://tongyi.aliyun.com/wanxiang/creation）
  - **豆包：**字节跳动推出的AI聊天工具，带有生图功能（https://www.doubao.com/chat/）
  - **腾讯元宝：**腾讯推出的AI聊天工具，带有生图功能（https://yuanbao.tencent.com/chat/)
  - **阶跃星辰：**阶跃星辰的AI聊天工具，带有生图功能（stepfun.com）
- **测评结果对比：**
  - **豆包：**升级3.0后，文字排版能力强，扩图效果好，但AI味儿较重，易用性最强堪称0门槛。
  - **腾讯元宝：**仅能通过对话形式生成图片，易用性较差，没有提示词优化等小白友好型功能。
  - **阶跃星辰：**速度最快，风格迁移能力相对较强，出图的整体审美能力一般般。
  - **通义万相：**能力最弱，生图效果出错率高，需要多次抽卡，人物动作错误明显。









3.4 AI工具深度测评 - 通义万相

核心功能及实测效果展示（图像部分）

- 易用性和交互体验（附加功能）
  - 提示词智能扩写：根据输入的提示词文本，智能生成描述更详细的提示词。小白友好型工具。
  - 灵感模式：增加创意灵感，提升图片丰富度与表现力，可能会与输入提示词有差异。
  - 咒语书：提供多种图像方面专业提示词、方便用户选择、提高工具易用性。



文生图功能页面



提供多个创作模型



咒语书

3.4 AI工具深度测评 - 通义万相

北京大学 PEKING UNIVERSITY

■ 实测效果

➢ 输入 Prompt

提示词：（复杂的肢体运动、大幅度动作、镜头的移动）

微型立体场景呈现，运用移轴摄影的技法，呈现出Q版【孙悟空三打白骨精】场景，其中孙悟空手持金箍棒要打白骨精，猪八戒沙和尚唐僧白龙马在一旁围观。

➢ 评价：专业版和极速版效果都较差，无法理解和准确生成西游记中的经典人物形象。（可对比GPT4o生图效果）



万相2.1专业版输出



万相2.1极速版输出

Vs



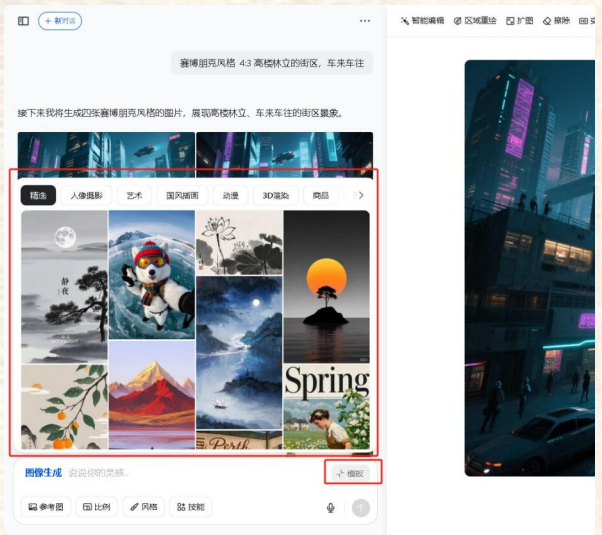
GPT4o输出

3.4 AI工具深度测评 - 豆包

核心功能及实测效果展示（图像部分）

- 核心功能
  - 豆包3.0图像生成的文字排版能力为此次更新的一大亮点，文字效果直逼付费工具即梦3.0
  - 区域重绘、扩图、擦除、变清晰
  - 参考图：根据所给提示词及参考图生成需要的图片
  - 根据生图一键生成视频

不同的图像风格模板非常丰富，使用便捷





文生图功能页面

3.4 AI工具深度测评 - 豆包

实测效果





文字生成及排版效果非常好(质量非常高√)



消除涂抹（质量非常高√）



区域重绘（质量较好，但需“抽卡”）





扩图（质量很高√扩图连接自然、整体画面风格融洽）



变清晰（质量较高√树木纹理、人物均变清晰）

3.4 AI工具深度测评 - 腾讯元宝

核心功能及实测效果展示（图像部分）

- 易用性和交互体验：腾讯元宝图片生成功能单一，通过对话形式生成图片，且易用性较差，没有提示词优化等小白友好型功能，仅提供了图片风格及比例快捷选项。
- 实测效果

3D Q版迷你风格，一个充满奇趣的迷你星巴克咖啡馆，外观就像一个巨大的外带咖啡杯，还有盖子和吸管。建筑共两层，大大的玻璃窗清晰地展示出内部温馨而精致的设计：木质的家具、温暖的灯光以及忙碌的咖啡师们。街道上有可爱的小人偶漫步或坐着，四周布置着长凳、街灯和植物盆栽，营造出迷人的城市一角。整体采用城市微缩景观风格，细节丰富、逼真，画面光线柔和、呈现出午后的惬意感受。



3D Q版风格理解较好、图片中内容与提示词一致性较好

生成海报图片：复古风格的麦当劳美食海报，采用丝网印刷技术，色彩浓郁饱满。画面中央是一份金黄酥脆的大麦克汉堡，两片芝麻面包之间夹着牛肉饼、芝士、生菜和特制酱汁，诱人垂涎。汉堡旁摆放着热腾腾的薯条和一杯冰凉的可口可乐。背景是50年代美国街头的繁华景象，霓虹灯闪烁，汽车穿梭。海报上方用大胆的无衬线字体写着"Taste the Tradition"，下方是麦当劳标志性的金色拱门Logo。整体构图平衡，色彩对比强烈，充满怀旧气息。



四张图文字全部错误但布局正确画面风格与提示词一致性良好

古代桌子上放着竹简，上面共四行从左往右写着第一行："先帝创业未半"，第二行："而中道崩殂"，第三行："今天下三分"，第四行："益州疲弊"，毛笔字，两只手，一只放在竹简上，一只手握着毛笔，毛笔停在最近的一个字上。



文字全部错误，且三张手指处理错误但布局正确画面风格与提示词一致性良好

3.4 AI工具深度测评 - 阶跃星辰

核心功能及实测效果展示（图像部分）

- 阶跃星辰2025.4.27宣布开源图像编辑大模型Step1X-Edit，性能达到开源SOTA。该模型总参数量为19B，具备语义精准解析、身份一致性保持、高精度区域级控制三项关键能力；支持11类高频图像编辑任务类型，如文字替换、风格迁移、材质变换、人物修图等。
- 实测效果展示



将图片转成吉卜力风格



风格迁移成功

将“2025”改为“2046”



文字修改失败

3D Q版迷你风格，一个充满奇趣的迷你星巴克咖啡馆，外观就像一个巨大的外带咖啡杯，还有盖子和吸管。建筑共两层，大大的玻璃窗清晰地展示出内部温馨而精致的设计:木质的家具、温暖的灯光以及忙碌的咖啡师们。街道上有可爱的小人偶漫步或坐着，四周布置着长凳、街灯和植物盆栽，营造出迷人的城市一角。整体采用城市微缩景观风格，细节丰富、真，画面光线柔和、呈现出午后的惬意感受



3D Q版风格理解较好、审美较好，但LOGO多一个，且有错误



将LOGO改成红色

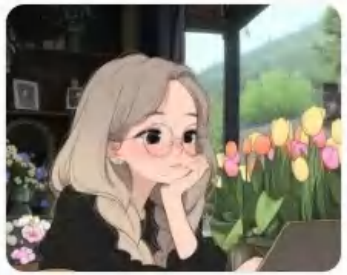


LOGO改色失败，只识别了红色，但没有改色，只执行的填充

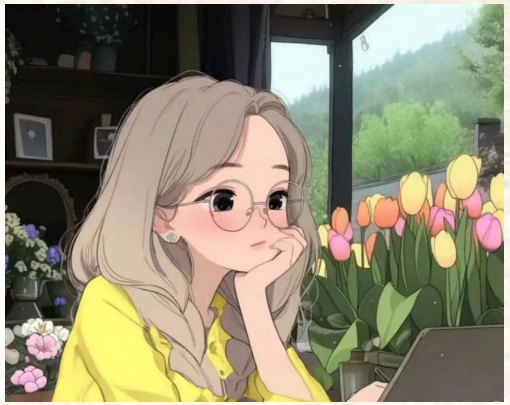
地点是“英国伦敦”，生成一张夏季的彩色矢量艺术海报，顶部有大的“LONDON”标题，下方有较小的“UNITED KINGDOM”标题



海报生成
（风格、文字正确）



将女孩衣服改为黄色衬衫



将女孩衣服改成黄色衬衫，改后效果：完全正确，且人物、背景和环境保持稳定

3.4 图像生成+编辑类AI工具 - 实测总结

通义万相、豆包、腾讯元宝、阶跃星辰 整体总结

工具	优点	缺点	适用场景
通义万相	1.功能齐全，涵盖图像、视频生成 2.免费不限量	1.稳定性差（出错概率高，需多次抽卡） 2.角色肢体动作处理能力差	适合快速修图及日常办公场景—低配要求使用 （推荐指数：🌟）
豆包	1.消除涂抹效果好 2.扩图能力较强 3.图片清晰度提升质量高 4.免费不限量 5.易用性强，有提示词优化功能，小白友好	图像生成的“AI味”较重	适合快速修图及设计图辅助修改 （推荐指数：🌟🌟🌟🌟🌟）
腾讯元宝	1.提示词理解能力较好 2.免费不限量	1.文字不准确 2.手部不准确	适合快速出图，对图像要求不高的场景 （推荐指数：🌟）
阶跃星辰	1.图像生成速度快 2.风格迁移效果好 3.语义理解能力强 4.免费不限量	图像生成及修改的准确度较低	适合需要快速出图的办公场景及要求质量不高的设计图风格切换 （推荐指数：🌟🌟）



北京大学
PEKING UNIVERSITY

01 文本生成与处理类AI工具测评

02 图像生成与编辑类AI工具测评

03 音视频生成与处理类AI工具测评

04 代码生成与辅助类AI工具测评

05 大模型管理与应用类AI工具测评

音频生成与处理类AI工具测评

4.1 音频类AI工具深度测评总述

- **测评目的：**通过系统化、多维度的评估，帮助用户了解不同工具的功能特点、使用效果和性价比，以选择最合适的音频处理工具
- **测评维度:**
  - **核心功能与效果 (重点)**
  - **易用性与交互体验**
  - **定价与性价比**
  - **集成与扩展能力 (API)**
- **测评工具选型：**
  - **海螺AI**--https://hailuoai.com
  - **MINIMAX**--https://www.minimax.io/audio
  - **Noiz.ai**--https://noiz.ai
- **测评结果对比：**
  - **海螺AI：**速度快，可根据预设音色生成语音，灵活参数设置满足不同情绪的语音生成，效果出色，但不支持声音克隆。（限时免费）
  - **MINIMAX：**支持声音克隆，支持情绪、音色效果等设置，整体效果出色，对网络要求高，可达到高品质语音合成，适合专业用户。（收费）
  - **Noiz.ai：**支持声音克隆，效果基本可用，没有情绪、音色调节等功能，支持语种少，多语言切换能力弱。（免费）

4.2 音频类AI工具深度测评 - 海螺AI

海螺AI 核心功能

- **多语种&多音色：**
  - 海螺AI支持中文、英文、日语、阿拉伯语、西班牙语、荷兰语等多个国家语言
  - 海螺AI原始语音库中包含多种音色，如专业女主持、活泼男声、憨憨萌兽等
- **语音调节：**
  - 基础调节：在右侧设置栏中可调节语速快慢、音量大小、声调高低
  - 情绪调节：在右侧设置栏中可以选择输出情绪，如自动、开心、生气、中性等
  - 音色效果调节：如右图，可以对已选择的音色进行调整



- **社区联动：**海螺AI提供Discord平台，方便用户在社区服务器中输入指令与Bot交互，并能与其他用户交流、分享和获取灵感。

海螺AI 效果实测（情绪调节）

- 首先选择音色库中的**港普女声**
- **输出情绪**：自动

> **测试提示词**：
> 哎哟，这个idea很可以喔~ 就像我的新歌前奏，你以为是钢琴solo？nonono，后面直接给你加段二胡。

- 在右侧设置栏中更改输出情绪：开心 ； 难过

实测评价：

- 两种情绪效果均ok，难过情绪相较于开心情绪的生成结果，语速更加缓慢，声调更加柔和
- 对于一般需求，直接使用默认的【自动情绪】模式即可满足大多数场景，效果自然又省心

4.2 音频类AI工具深度测评 - 海螺AI

海螺AI 效果实测（音色调节）

海螺AI**不支持自定义语音模型**，只能使用原始音色库中的声音

但海螺AI提供**音色效果调节功能**，可以根据需求，在原始音色的基础上进行调节

测试提示词：

哎哟，这个idea很可以喔~ 就像我的新歌前奏，你以为是钢琴solo？nonono，后面直接给你加段二胡。

- 原音频（参数默认）：
- 音色更改：

实测评价：

- 参数更改后音色发生较大变化，但同时导致生成语音的连贯性下降，影响生成质量，且生成时间较长

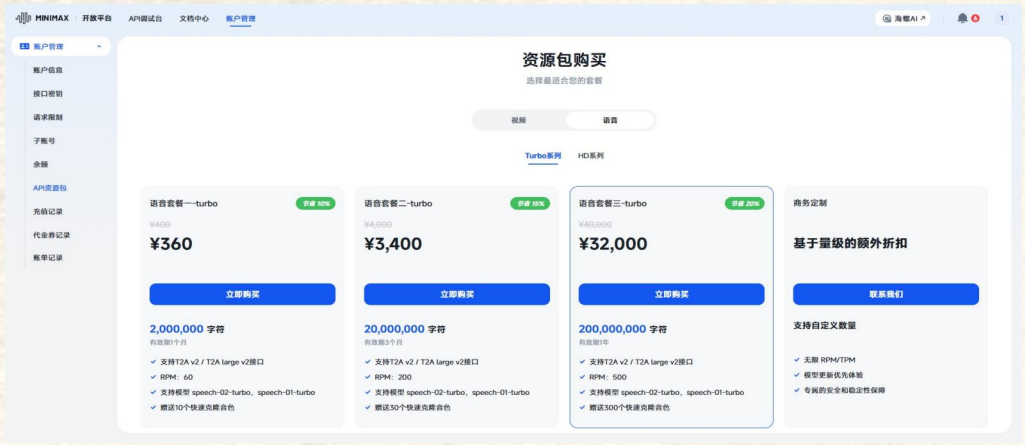
参数设置：

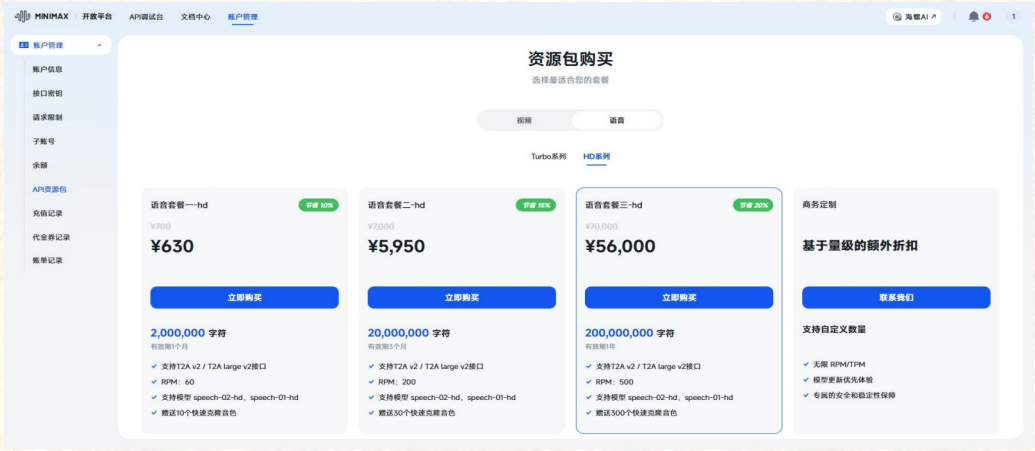


4.2 音频类AI工具深度测评 - 海螺AI

海螺AI 价格与性价比

- 海螺AI 是一款**限时免费**的AI语音生成平台，音色库丰富、多语言支持、高质量的语音输出，可以达到高效便捷的语音合成体验。
- 虽然目前暂**不支持声音克隆**，但其出色的语音质量和灵活的调节功能，仍能满足大多数场景需求，适合需要快速生成高品质语音，且希望直接调用API的用户。
- 语音生成速度快，支持调节情绪、语速、声调等参数，并**开放API**方便开发者集成到各类应用中，用户可根据需求购买资源包。





4.2 音频类AI工具深度测评 - MINIMAX

MINIMAX 核心功能

- **大部分功能与海螺AI一致：**
  - 参数设置：可调节语速、声调、音量、音色效果、输出情绪；
  - 多语种&多音色：支持中文、英文、阿拉伯语、荷兰语等多个国家语言，且音色库包含多种音色。
  - 社区联动：提供Discord平台，方便用户在社区服务器中输入指令与Bot交互，并能与其他用户交流、分享和获取灵感。
- **上传自定义语音模型：**与海螺AI不同在于，MINIMAX可以进行声音克隆

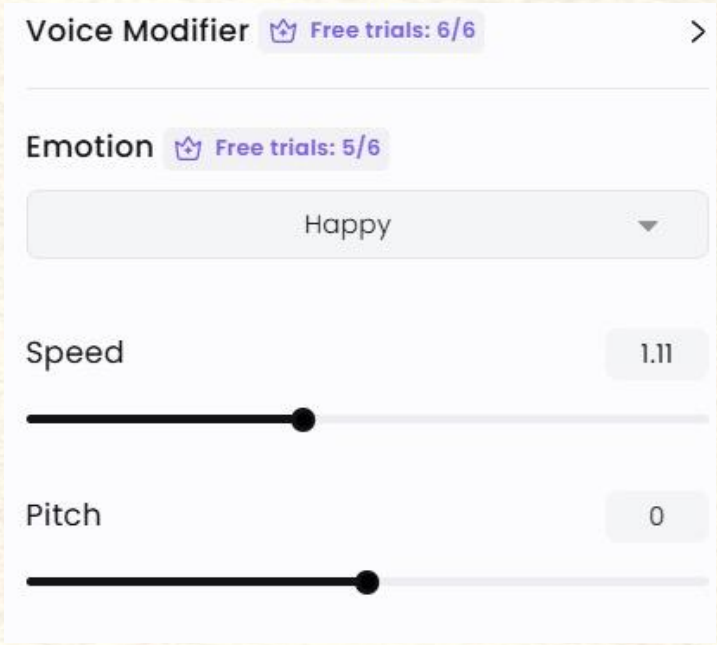
4.2 音频类AI工具深度测评 - MINIMAX

MINIMAX 效果实测

- 参数默认
- 原音频：
- 克隆音色后文本转音频生成

> **测试提示词**：
> 哎哟，这个idea很可以喔~ 就像我的新歌前奏，你以为是钢琴solo？nonono，后面直接给你加段二胡。

- 更改参数后（如右图）：



实测评价：

语速提升至1.11倍速，输出情绪为开心，与默认参数的音频相比，生成效果较默认设置更显活泼生动

4.2 音频类AI工具深度测评 - MINIMAX

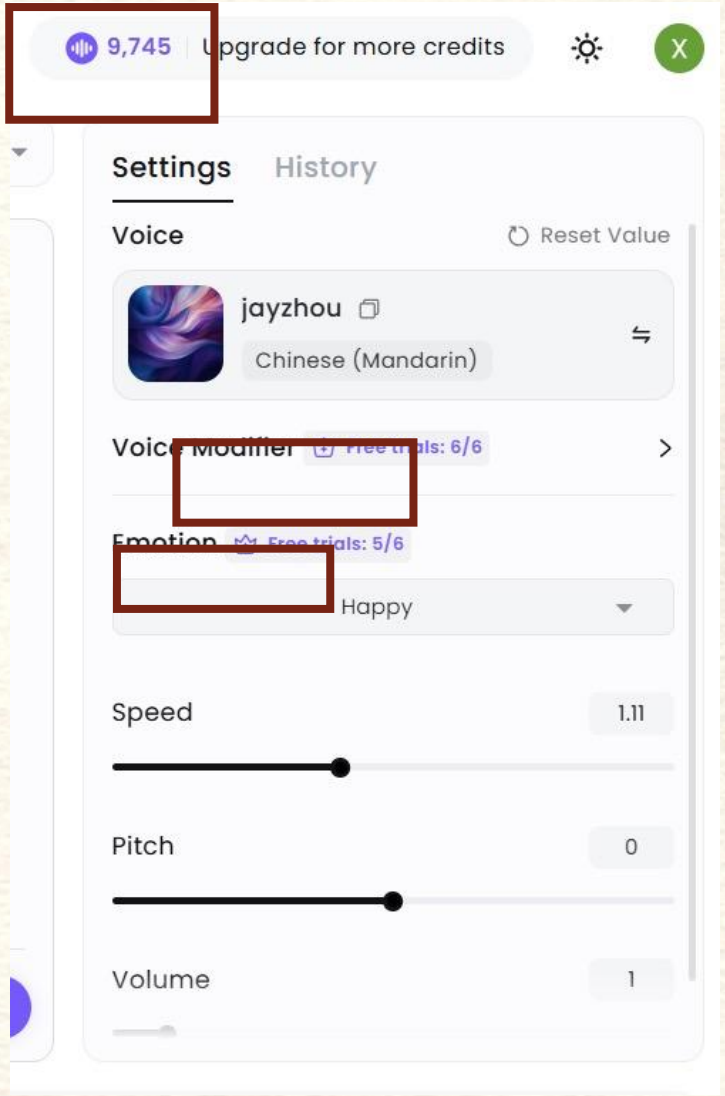
MINIMAX 价格与性价比

- Minimax 提供**每月10000积分免费额度**，音色库丰富、多语言支持和高质量的语音生成，为开发者和企业用户提供了专业级的语音合成解决方案。
- 每生成一次语音消耗**75积分**，且用户仅有**6次**音色和情绪调节试用次数，该平台语音生成速度快，支持自定义语音模型，并可精细调节情绪、语速、声调等参数，同时提供API方便集成到各类应用中。
- 访问平台需使用**网络加速工具**，对部分用户可能不够便捷。

计费说明

计费项	接口说明	单价
T2A v2	支持音量、语调、语速调整和混音功能，支持比特率、采样率相关参数调整特性，支持音频时长、音频大小等返回参数，适用于需要短文本快速得到结果的场景，比如闲聊、对话等场景	3.5元/万字符（speech-02-hd/speech-01-hd） 2元/万字符 （speech-02-turbo/speech-01-turbo / speech-01-240228 / speech-01-turbo-240228）
T2A large v2	支持基于文本到语音的异步生成，单次文本生成传输最大支持1000万字符，生成的完整音频结果支持异步的方式进行检索。	3.5元/万字符（speech-02-hd/speech-01-hd） 2元/万字符 （speech-02-turbo/speech-01-turbo）
Voice Generation（文生音色）	支持基于用户输入的声音性别、声音年龄、声音描述信息，来生成音色（voice_id）；并支持使用该生成的音色（voice_id）在T2A v2接口中进行语音生成。	试听字符收费：2元/万字符； 后续实际使用T2A v2以及T2A Large v2推理时收费。
voice_clone（快速复刻）	基于大语言模型的音色克隆更加精准快速，无需数小时时长的超高质量原音频、无需传统TTS的超长工期，可以在极短时间内完成音色复刻，并通过大语言模型加持，使复刻后的音色与原音色进行高质量还原，从而满足客户需求。	9.9元/音色。调用本接口获得复刻音色时，不会立即收取音色复刻费用。音色的复刻费用将在首次使用此复刻音色进行语音合成时收取。 试听字符收费：2元/万字符。

文档反馈



4.2 音频类AI工具深度测评 - Noiz AI

Noiz AI 效果实测（情绪匹配）

- 在Noiz.ai中利用**标点符号**调节语音情绪，在文本转音频时，标点符号的设置尤为重要
- 任务：克隆周杰伦的声音，根据提示词生成一段新的音频

原音频：

测试提示词：

哎哟，这个idea很可以喔~ 就像我的新歌前奏...你以为是钢琴solo ？？ nonono，后面直接给你加段二胡！

- 改进提示词中的标点符号：

改进测试提示词：

哎哟，这个idea很可以喔~ 就像我的新歌前奏，你以为是钢琴solo ？nonono，后面直接给你加段二胡。

提高语音质量的技巧

1. 标点符号很重要！尽可能合理使用它们。
2. 口音和风格将取决于参考音频和文本。根据需要调整输入，以获得理想的输出效果。

如何丰富声音层次感

3. **感叹号（!）**：提高震惊语气，适合表达人物情感、制造紧张感场景。
4. **分句（,、 。 空格 换行）**：短暂停顿，适合复杂内容中插入呼吸间隙。
5. **省略号（...）**：进一步增加停顿，适合长停顿、犹豫、制造悬念场景。
6. **问号（？）**：增加强疑问语气，多适用于故事转折、互动问答场景。
7. **破折号（——）**：改善吞字问题，被吞的字前面加"——"，例如"你说的——对"。

- **实测评价：**对比两段音频发现，克隆情绪平淡的声音时，标点符号越简单，生成效果越好

Noiz AI 效果实测（情绪匹配）

如果需要比较激进的情绪，该如何调整？

- 更改原音频：

- 依然采用刚刚的提示词：

测试提示词：
哎哟，这个idea很可以喔~ 就像我的新歌前奏，你以为是钢琴solo ？nonono，后面直接给你加段二胡。

- 可以发现，生成后的音频语气较平，需要对提示词的标点符号进行更改

- 更改后的提示词：

更改测试提示词：
哎哟！！！这个idea很可以喔， 就像我的新歌前奏....你以为是钢琴solo ？？？nonono ...后面直接给你加段二胡！

4.2 音频类AI工具深度测评 - Noiz AI

Noiz AI 效果实测（连贯性）

以前面的例子为例

- 提示词1：

测试提示词1：

哎哟，这个idea很可以喔~ 就像我的新歌前奏，你以为是钢琴solo ？nonono，后面直接给你加段二胡。

- 提示词2：

测试提示词2：

哎哟！！！这个idea很可以喔， 就像我的新歌前奏....你以为是钢琴solo ？？？ nonono ... 后面直接给你加段二胡！

实测评价：

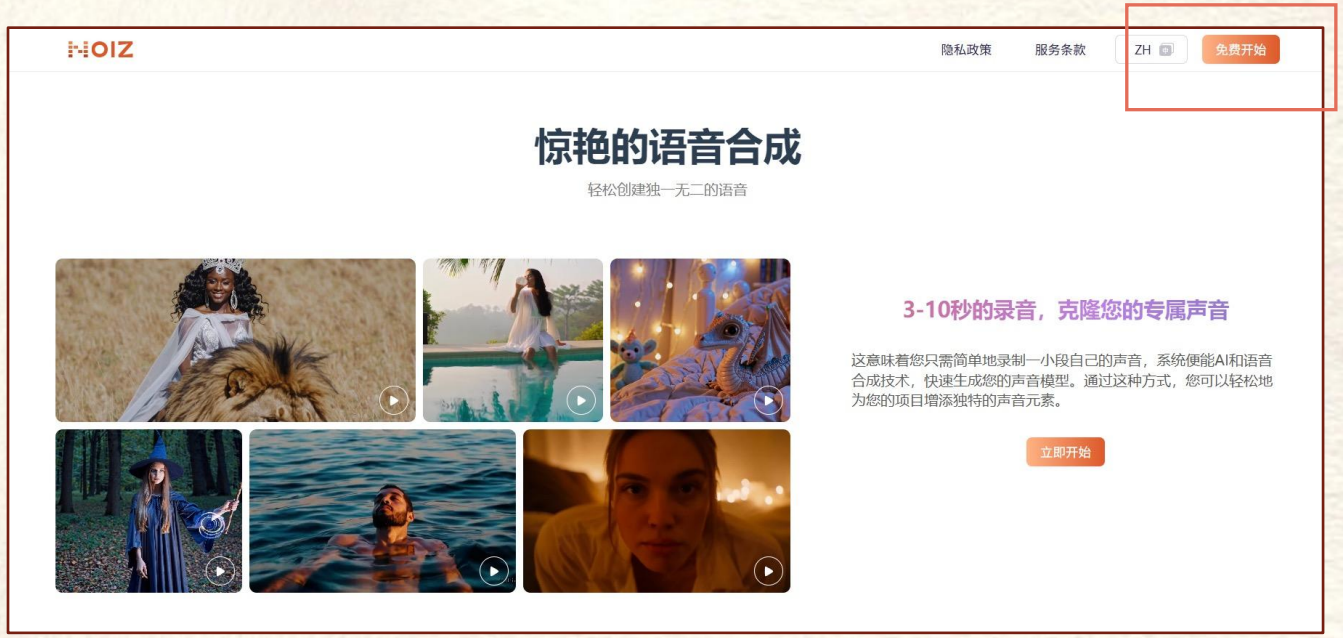
在调整音频情绪的同时，过多使用标点符号（如多个感叹号和问号）可能会影响语音的自然流畅度。所以需要多次调整标点符号的组合方式，对比生成效果，才能找到最自然的表达节奏。当然整体体验感对于用户操作来讲较为繁琐，需要不断调整提示词，才能得到满意效果。

4.2 音频类AI工具深度测评 - Noiz AI

Noiz AI 价格与性价比

Noiz.ai 是一款支持**免费试用**的AI语音工具，凭借其**快速生成**和**自定义音色**功能，为普通用户和轻度需求者提供了高性价比的选择。

用户可免费上传自己的声音样本，生成个性化语音，适合**日常使用或临时项目**需求。但Noiz.ai未提供API，无法集成到其他应用上。



4.2 音频类AI工具 - 实测总结

工具	优点	缺点	适用场景
海螺AI	可以**调用API** 音频生成功能**限时免费** 参数设置：输出情绪、音色效果等，**根据预设音色生成语音，效果出色**	不支持声音克隆	需要**快速**生成高品质语音或**直接调用API**的用户
MINIMAX	可以**调用API** **支持声音克隆，效果出色** **参数设置**：输出情绪、音色效果等	对网络要求较高 积分消耗和功能次数限制	需要**高品质**语音合成，并愿意接受积分消耗和网络要求的**专业**用户
Noiz.ai	**免费**使用 支持**声音克隆，效果基本可用** 可以通过标点符号调节情绪	支持语种相对较少 无法直接调节情绪 无法调节音量、音色效果 不能调用API	适合**日常使用或临时项目**需求

视频生成与处理类AI工具测评

4.3 视频生成类AI工具深度测评总述

- **测评目的:**
  - 国产视频类AI工具飞速发展，本次测评聚焦国内视频类AI工具，帮助用户选择最适合其需求的工具并推动效率提升与创新。
- **测评维度:**
  - 核心功能与效果 (重点)
  - 易用性与交互体验
  - 定价与性价比
  - 集成与扩展能力 (API)

即梦AI　可灵　通义万相　Vidu

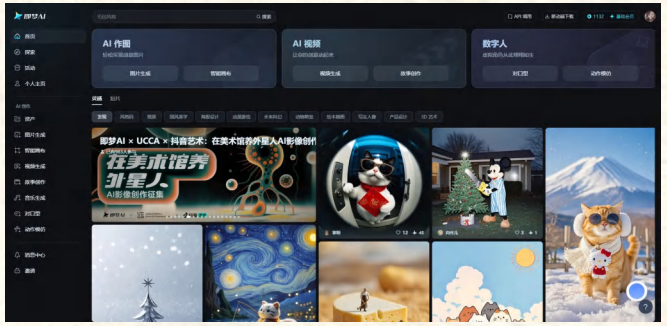
- **测评工具选型:**
  - **即梦AI**：提示词遵从度高、工具集成度高、操作简单、复杂场景制作效率高，性价比最高。
  - **可灵AI**：更适配影视制作和商业项目，提供电影级画质、物理模拟，操作简单，风格多样且生成速度快，效果富有表现力，长视频生成功能。
  - **Vidu**：适合短视频创作者和内容爱好者，更擅长卡通场景。
  - **通义万相**：完全免费，普通场景效果不错，复杂场景能力较弱。

4.4 AI工具深度测评 - 即梦AI

即梦AI 简介与核心技术

即梦AI 3.0（2025.4.3更新）在各方面都是国内一线水准，在人物肢体动作等领域甚至接近顶级。并且价格便宜，速度极快。成功率极高基本不需要”抽卡“并且免费用户也可以使用视频3.0模型。

- **核心定位:** 提供简单易用的文本到图像、视频生成工具，强调电影级视觉质量、精准排版（尤其是英文）以及多模态创作能力，旨在降低创作门槛并赋能数字艺术、商业营销等领域。
- **核心技术突破（即梦AI 3.0）：**
  - 在视频生成方面实现了多项核心技术突破，主要包括影视级画质、2K高清直出、精准语义控制以及多模态生成能力。在多段动作遵循、运镜美学与动幅、情绪表现力与质感等方面都表现出顶级水准。
- **主要产品形态:** 网页、APP、API。
  - 网页：https://jimeng.jianying.com/
  - APP：https://ugurl.cn/JHnhP
  - API：www.volcengine.com/product/jimeng





4.4 AI工具深度测评 - 即梦AI

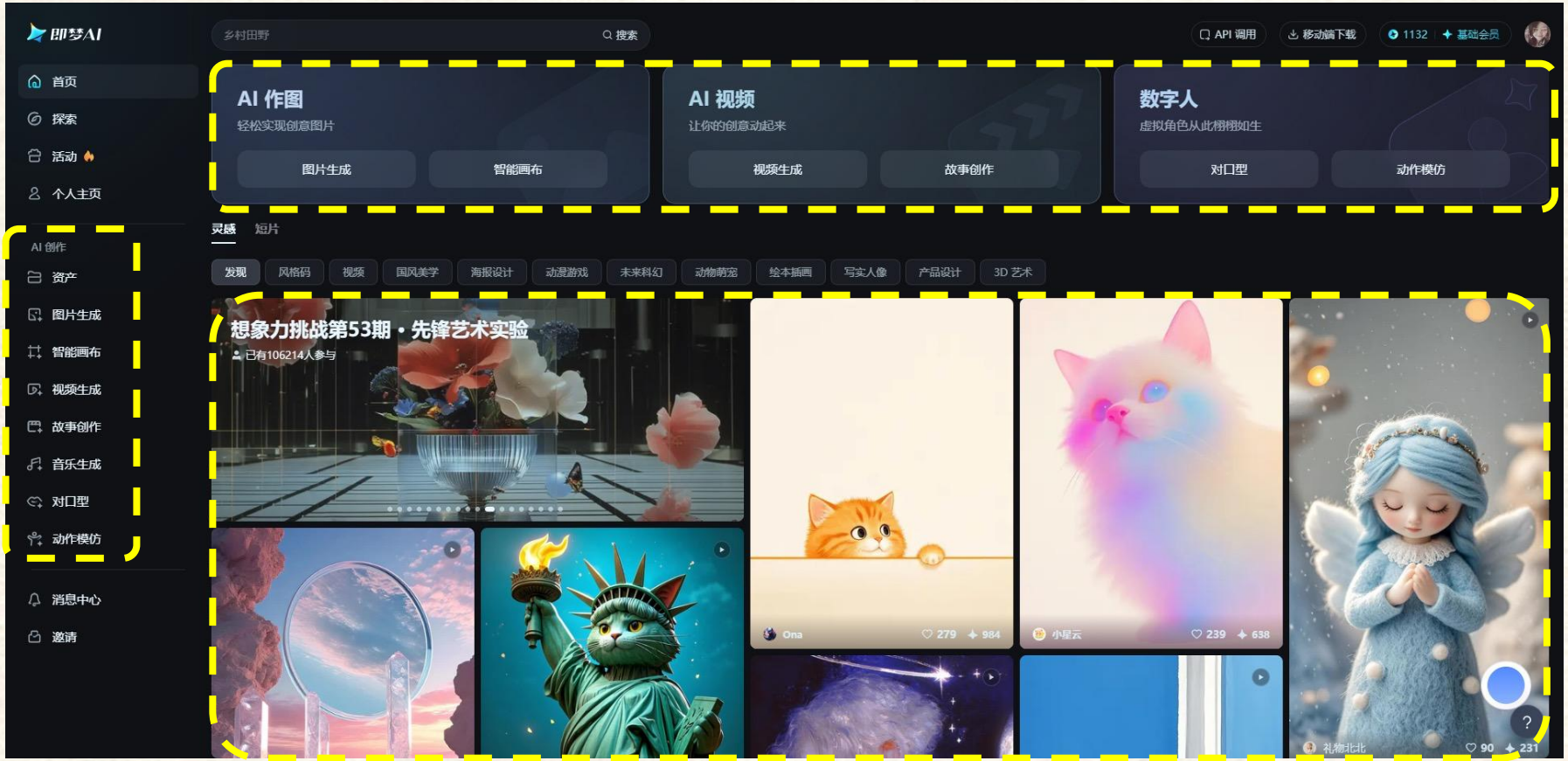
即梦AI 主要功能（视频部分）

- **AI视频:**
  - 文生视频：根据用户输入的文字描述，创造出全新的、富有想象力的视频；
  - 图生视频：将用户上传的一张或多张图片，结合提示词生成视频；
- **视频编辑:**
  - 补帧: 对已生成的视频进行帧率提高操作，让视频更加流畅。
  - 提升分辨率：让生成视频更高清，提升视频观看感受。
  - 视频延长：扩展已生成的视频，智能填充延展部分的视频内容。
- **故事创作：**长视频创作平台，批量导入分镜图片一键生成连续视频，并将文生视频、图生视频集成到该平台，极大提升视频创作效率，易用性极高。
- **社区联动：**提供官方或用户共享的特定风格模型，通过“做同款”方便用户快速生成特定风格的视频。

4.4 AI工具深度测评 - 即梦AI

即梦AI 易用性与交互体验

即梦AI界面简洁直观，且符合国内使用习惯，易于上手、交互体验流畅。图片与视频制作丝滑切换。



核心功能

全部功能

社区活动及作品

4.4 AI工具深度测评 - 即梦AI

即梦AI 视频生成 功能说明

选择功能 →

上传图片输入提示词(文生视频可通过DeepSeek优化) →

选择模型（3.0提供2K高清出图） →

选择生成时长及比例（文生视频可自定义、图生视频根据上传图片自动设置比例） →

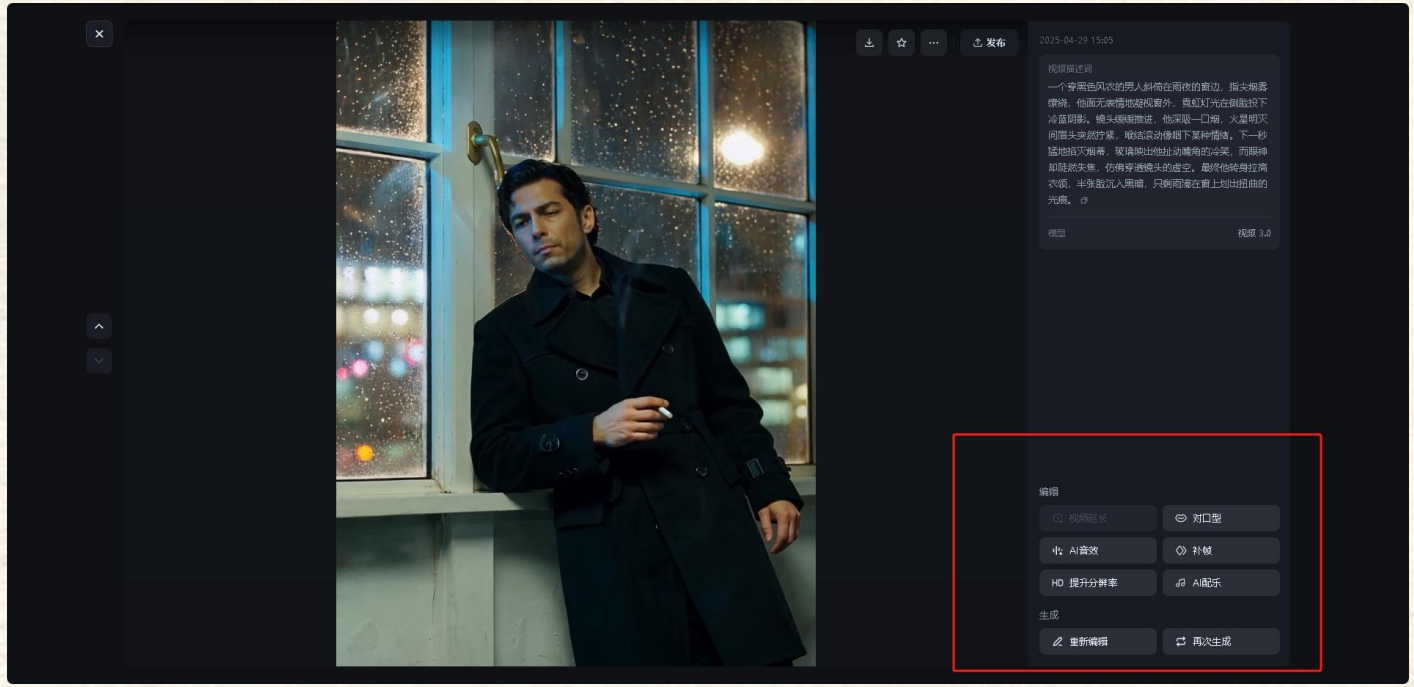
点击即可出图（预览积分消耗，有助于用户控制成本） →



4.4 AI工具深度测评 - 即梦AI

即梦AI 易用性与交互体验

- 视频二次编辑
- 生成视频后，提供多种常用的二次编辑功能，如：视频延长、对口型、AI音效、补帧、提升分辨率、AI配乐等。



二次编辑功能区



AI音效生成：一次生成三种音效可选功能

4.4 AI工具深度测评 - 即梦AI

即梦AI3.0 文生视频 效果实测（提示词遵从程度-动作遵循）

- 即梦AI 对中文用户极度友好，且可以用DeepSeek进行提示词优化，学习成本极低
- 生成视频对提示词遵从程度较高，提示词里如果有很多个镜头、很多个动作，能够在有限的时长里，按提示词的需求去依次生成

【镜头1 特写】一位25岁左右的亚洲女性面对镜头露出温柔的微笑，午后阳光透过窗户在她脸上投下斑驳的光影。她轻轻眨动右眼（时长0.8秒），睫毛在光线下闪烁微光。

【镜头2 中景】她突然以舞蹈般的姿态向右转身（时长1.2秒），亚麻色长发在空中划出优美弧线，米色针织衫随着动作微微飘动。

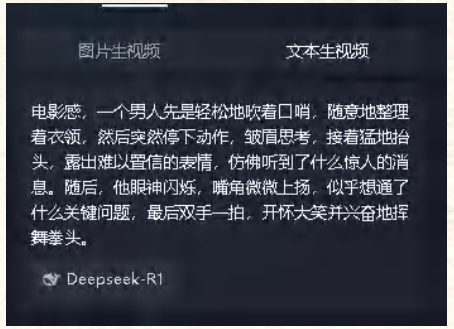
【镜头3 过肩镜头】转回正面时（时长0.5秒），她的瞳孔猛然扩大，倒映着某个看不见的物体。面部肌肉瞬间紧绷，右手无意识地捂住张开的嘴。

【镜头4 微表情特写】震惊表情持续1秒后，眉头突然舒展（时长0.3秒），左眼微微眯起，嘴角扬起带着自嘲意味的弧度，手指轻轻点向太阳穴。

【镜头5 动态广角】她突然放声大笑（时长1.5秒），双手抱住后脑勺时碰散了发髻，几缕发丝垂落在泛红的脸颊旁，背景虚化的书架上有本翻开的日记隐约可见。



对于复杂程度很高的提示词，
动作遵循效果依然很强。

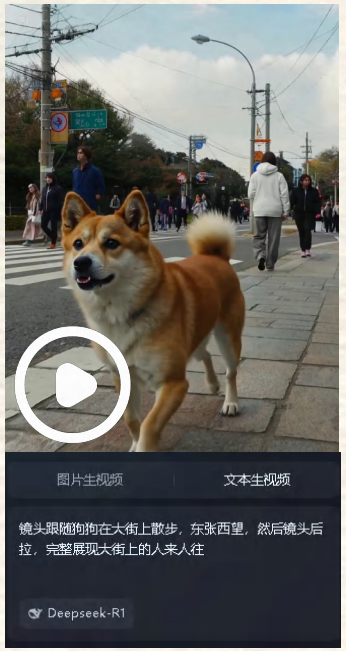


保证复杂提示词动作遵循效果的同时，视频氛围保持的也很好。

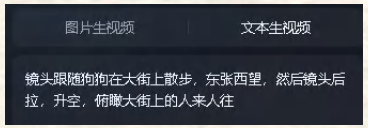


即梦AI3.0文生视频 效果实测（提示词遵从程度-镜头遵循能力）

- AI生成的画面普遍呆板，运镜能弥补“动态不足”。所以，镜头语言是让AI视频“活过来”的关键。镜头遵循能力对于AI视频生成工具来说非常重要。

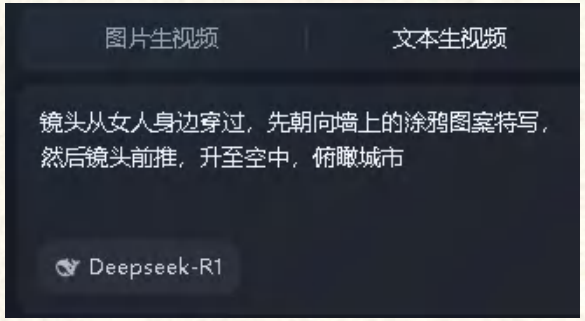




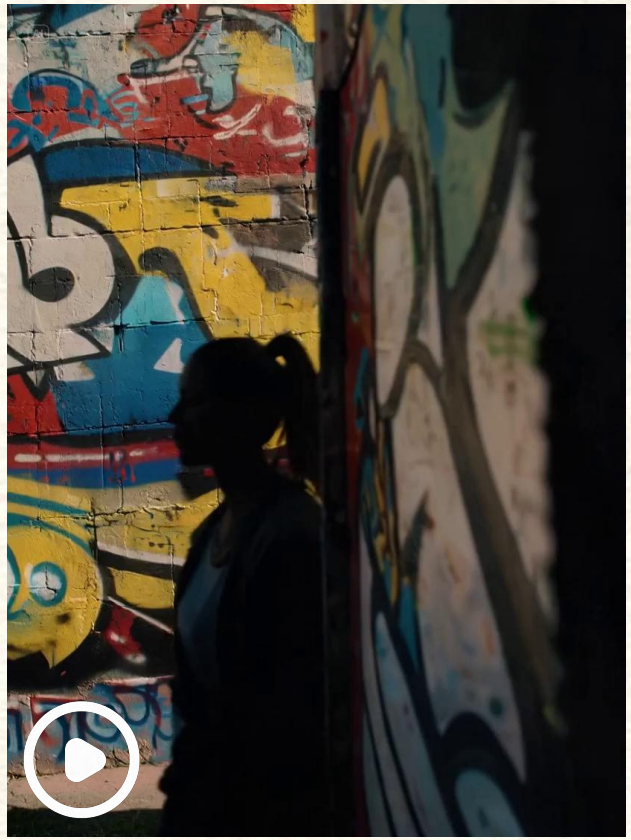


镜头跟随效果非常好，但对提示词要求较高

（第一次提示词没有写好，所以后半段升空的画面没有展现，修改后运镜效果正常。）



虽然画面景别在变，画面主体在变，但这个一镜到底的完成度非常高，唯一的瑕疵是从在涂鸦墙面过度的时候，画面有点崩。



4.4 AI工具深度测评 - 即梦AI

即梦AI3.0 文生视频 效果实测（情绪表现力与质感）

- 控制人物的情绪表现是做影视剧中经常使用的特性。主要看情绪的张力和表演的到位性。



图片生视频 | 文本生视频

太空舱中，男子发现了什么，然后非常恐惧的站了起来

生成的视频不仅涵盖了提示词所给的情绪描述。还增加了由发现到思考再行动的整体过程。



图片生视频 文本生视频

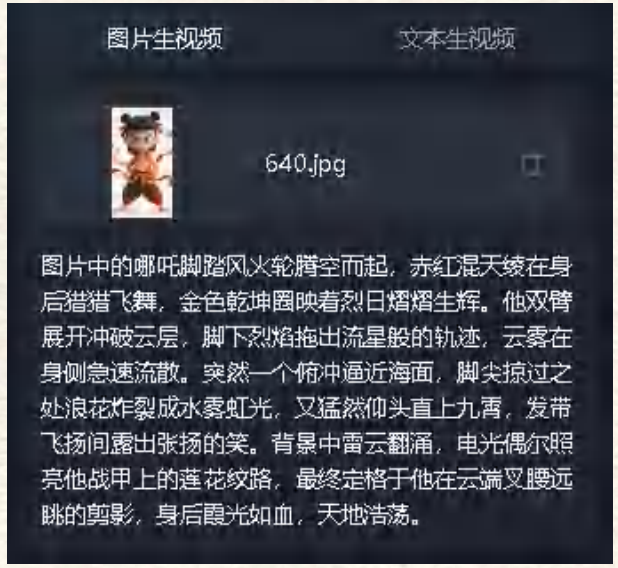
一个穿黑色风衣的男人斜倚在雨夜的窗边，指尖烟雾缭绕，他面无表情地凝视窗外，霓虹灯光在侧脸投下冷蓝阴影。镜头缓缓推进，他深吸一口烟，火星明灭间眉头突然拧紧，喉结滚动像咽下某种情绪。下一秒猛地掐灭烟蒂，玻璃映出他扯动嘴角的冷笑，而眼神却陡然失焦，仿佛穿透镜头的虚空。最终他转身拉高衣领，半张脸沉入黑暗，只剩雨滴在窗上划出扭曲的光痕。

可以看出不管是人物动作还是人物表情张力，即梦3.0都已经不是什么问题了

4.4 AI工具深度测评 - 即梦AI

即梦AI3.0 图生视频 效果实测

- 即梦AI 3.0 图生视频的唯一短板在于缺少首尾帧及多图片参考功能。



生成视频的提示词遵从程度非常好，并且视频中的人物各部位及细节均取自参考图







加入运镜控制后，生成的视频效果也非常好

4.4 AI工具深度测评 - 即梦AI

即梦AI 定价与性价比

- 即梦AI是一款付费AI图像、视频综合工具，效果基本可商用，相比人工和国外AI工具，性价比非常高。每日免费60积分，可供体验。用户可以根据用量选择订阅计划或购买积分（适合少量使用）。

即梦3.0积分消耗：

- 文生视频：5s--10积分--耗时约1分钟
  10s--20积分--耗时约2分钟
- 图生视频：5s--10积分--耗时约2分钟
  10s--20积分--耗时约4分钟



积分购买



连续包年



连续包月



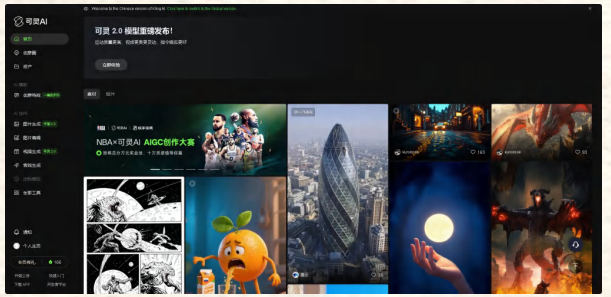
单月购买

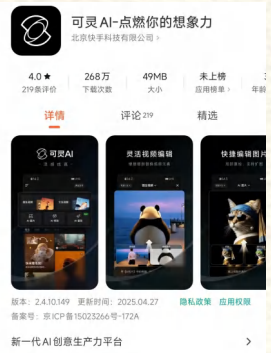
4.4 AI工具深度测评 - 可灵AI

可灵AI 简介与核心技术

可灵AI（KLING）是快手自主研发的全球领先视频生成大模型，自2024年6月发布以来，凭借高频迭代、技术创新及商业化落地能力，迅速成为行业标杆。

- **核心定位:**旨在降低视频创作门槛，赋能个人创作者与企业用户，推动AI生成内容（AIGC）的普及。提供高效、低成本的视频生成工具，激发用户创意表达。可通过API服务赋能企业，优化影视、电商等内容生产效率。
- **核心技术突破（可灵 2.0大师版）**
  - 2025年4月15日，可灵发布2.0大师版在视频生成技术上实现了质的飞跃，其创新的动态扩散模型显著提升了长视频生成的稳定性，动作过渡流畅度达到专业动画水准。系统能够精准模拟头发、布料等柔性体的自然运动轨迹，并实时计算复杂材质的光线反射效果。更突破性的是其首创的语义导演系统，可智能解析剧本内容并自动匹配合适的镜头语言，用户仅需自然语言指令即可调整包括胶片质感和特殊运镜在内的专业参数。
- **主要产品形态:** 网页、APP
  - 网页：https://app.klingai.com/cn/





可灵AI 主要功能（视频部分）

- **文生视频:**
  - 输入文字描述（如“一只小猫在草地上跳舞”），AI自动生成匹配的动态视频。
- **图生视频:**
  - 支持局部动画（仅让图片中的某部分动起来，如人物眨眼）；
  - 上传静态图片（如一张风景照），AI为其添加动态效果（如云朵流动、树叶摇曳）。
- **多模态编辑：**
  - **混合输入：**结合文字、图片、音频甚至手势控制来生成或修改视频；
  - **精细化调整：**替换视频中的物体、修改人物动作、调整光影/色调。
- **智能视频优化：**
  - 补帧增强：提升视频流畅度，支持60/120FPS高帧率输出；
  - 时长扩展：自动延展视频内容并保持剧情逻辑连贯。

4.4 AI工具深度测评 - 可灵AI

可灵AI--可灵2.0大师版 效果实测（文生视频）

提示词：夕阳西下，染红了整片天空，云层如同燃烧的火焰般绚烂。巨大的移动城堡喷吐着蒸汽，齿轮转动发出沉重的金属声响，摇摇晃晃地行走在连绵的绿色山丘上。城堡的烟囱歪歪扭扭，铁皮外壳布满锈迹和补丁，却透着奇异的生命力。近景处，苏菲站在城堡的露台上，风扬起她朴素的蓝色裙摆和棕色的发丝。她望着远方，眼神既带着对未知的忐忑，又有一丝隐约的期待。哈尔的黑鸟形态从她身旁掠过，翅膀划过空气，投下转瞬即逝的阴影。背景里，山脚下的小镇点缀着暖黄的灯火，蜿蜒的河流反射着暮光，像一条流动的金色丝带。整个画面既奇幻又温暖，蒸汽朋克的机械感与自然风光的柔美奇妙地融合在一起，仿佛在诉说：即使是最笨重的钢铁，也能拥有童话般的心跳。

效果评价：

- 画面清晰，连续性与稳定性强
- 文本理解：

对**背景环境**的描写较为**细致**，如天空的晚霞、移动城堡的机械构造、远方的山脉与河流等元素都得到了充分呈现。

但原文中关于**近景特写**的细节——特别是小女孩被风吹起的发丝和眼神中交织的不安与期待在画面中未被提及。



可灵AI--可灵2.0大师版 效果实测（图生视频）

- 设置首帧图，生成5s视频



无图片描述



有图片描述

效果评价：

- 右侧视频在细节呈现上更为丰富，不仅捕捉到山风**掀起女孩裙摆**的细微动态，还展现了**黑鸟振翅掠过**的轨迹。
- 左侧视频在**结尾**处存在瑕疵——女孩的动作出现了**不自然的卡顿变形**

可灵AI--可灵2.0大师版 效果实测（图生视频）

- 无论是否添加详细的图片描述，前页的两个版本都缺乏对近景人物神态的刻画（如眼神等细节）
- 将生成视频时长增加到10秒



有图片描述

- 图片描述：夕阳西下，染红了整片天空，云层如同燃烧的火焰般绚烂。巨大的移动城堡喷吐着蒸汽，齿轮转动发出沉重的金属声响，摇摇晃晃地行走在连绵的绿色山丘上。城堡的烟囱歪歪扭扭，铁皮外壳布满锈迹和补丁，却透着奇异的生命力。近景处，苏菲站在城堡的露台上，风扬起她朴素的蓝色裙摆和棕色的发丝。她望着远方，眼神既带着对未知的忐忑，又有一丝隐约的期待。哈尔的黑鸟形态从她身旁掠过，翅膀划过空气，投下转瞬即逝的阴影。背景里，山脚下的小镇点缀着暖黄的灯火，蜿蜒的河流反射着暮光，像一条流动的金色丝带。整个画面既奇幻又温暖，蒸汽朋克的机械感与自然风光的柔美奇妙地融合在一起，仿佛在诉说：即使是最笨重的钢铁，也能拥有童话般的心跳。

- 依然没有近景与远景的切换

4.4 AI工具深度测评 - 可灵AI

可灵AI--可灵2.0大师版 效果实测（近景与远景）

提示词：暮色中的油屋灯火通明，朱红色的廊桥横跨在幽蓝的水面上，灯笼的光晕在雾气中晕染开来。远处的神明浴场层层叠叠，屋檐翘角上蹲踞着模糊的石像鬼影。蒸汽从建筑群的缝隙间升腾，与河面飘荡的薄雾交融，将整个异界笼罩在氤氲的暖光里。对岸的树林已沉入黛青色阴影中，唯余几盏引魂灯在水面投下细碎的倒影，一条无人乘坐的扁舟正随波轻轻摇晃。这幅远景完美捕捉了神秘与温馨并存的氛围，那些发光的窗户后，或许正上演着人类看不见的百鬼夜行。



远景

提示词：动画人物：千寻的瞳孔在眼眶里急速收缩，虹膜边缘泛着湿润的光——那是泪水在积聚却尚未滚落的临界状态。眉弓上方渗出细密的汗珠，顺着太阳穴滑到微微抽搐的苹果肌上。鼻翼随着急促呼吸不断张合，在灯光下投出蝴蝶翅膀般颤动的阴影。最精彩的是嘴唇的变化：上唇因恐惧而绷紧，露出一点门牙的弧光；下唇却不受控制地发抖，让涂着淡淡唇膏的表面泛起细纹，像被石子打破的水面。当她突然倒吸冷气时，人中部位会先出现一道短暂的纵向凹陷，随后整张脸的肌肉如同被无形的手揉皱的纸，从额头到下巴泛起涟漪般的颤动。



近景

效果评价：

- 可灵2.0在视频生成方面展现出强大的场景刻画能力，无论是**人物近景**的细腻神态还是**宏大远景**的空间层次都能精准呈现
- 但当文本指令**同时包含近、远景特写**要求时，系统会**优先保证远景质量**，而忽略近景细节的生成

4.4 AI工具深度测评 - 可灵AI

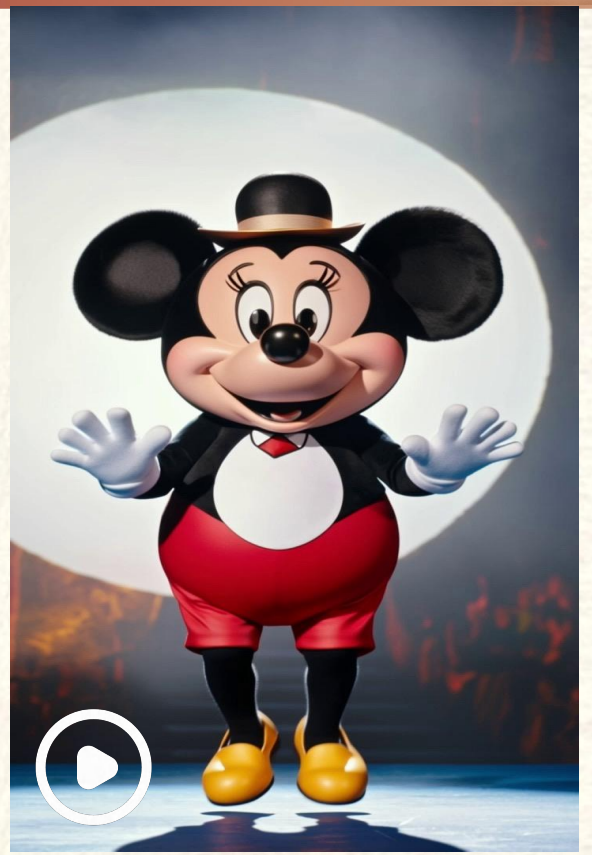
可灵AI--可灵1.6 效果实测（多模态编辑）

- 目前【多模态编辑】仅支持在可灵1.6模型下使用

- 将原视频中的米奇角色
- 替换为小猫形象



替换形象



原视频



替换结果

效果评价：

- **整体表现优秀，画面流畅清晰，动作自然连贯**
- 小猫形象：细节丰富，包括四肢自然的**毛流感**、脚掌的**肉垫形状**等特征，都高度还原了真实小猫的形态
- 动态光影效果：小猫的**瞳孔颜色**会随光线变化而变化，**尾巴**在地面的**投影**随着舞姿摇曳变化，毛发上还能看到灯光的**高光反射**
- **不足：**渲染时遗漏了小猫的眉毛和头部的毛流感细节

4.4 AI工具深度测评 - 可灵AI

可灵AI 价格与性价比

- 可灵AI是一款付费AI图像、视频综合工具，效果基本可商用，相比人工和国外AI工具，性价比非常高。**每月免费166**灵感值，可供体验。

- 可根据用量决定购买方案：
  - 少量使用--灵感值购买
  - 长期使用--购买会员（连续包年、包月）

可灵2.0灵感值消耗：

- 文生视频：5s--100灵感值--耗时约5分钟
- 图生视频：5s--100灵感值--耗时约5分钟
  10s--200灵感值--耗时约10分钟
- 多模态编辑（目前仅支持可灵1.6模型）
  5s--50灵感值--耗时约4分钟



灵感值购买



按月购买（3个月9折、12个月7折）

4.4 AI工具深度测评 - Vidu

Vidu 简介与核心技术

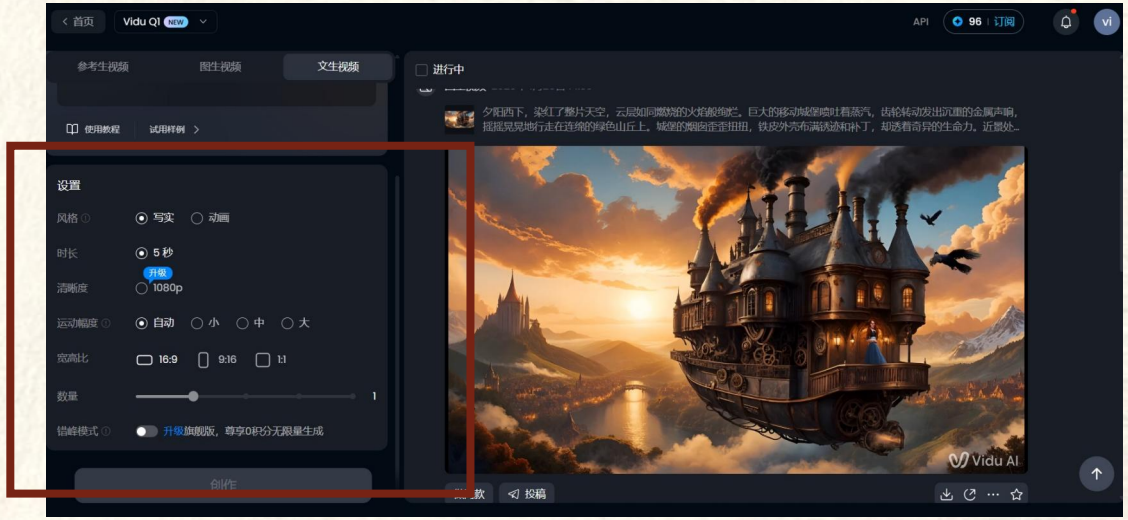
Vidu AI是由生数科技联合清华大学正式发布的中国首个长时长、高一致性、高动态性视频大模型。Vidu 在语义理解、推理速度、动态幅度等方面具备领先优势，并上线了全球首个“多主体参考”功能，突破视频模型一致性生成难题。

- **核心定位:** 致力于打造专业级视频生成解决方案，赋能影视制作与商业内容生产，推动AI视频技术的工业化应用。通过突破性的物理仿真与电影级渲染技术，为创作者与企业提供高效、高质的内容生产工具，重塑影视、广告、游戏等领域的创作流程。可提供定制化API服务，助力企业实现高质量视频内容的规模化生产。
- **核心技术突破（Vidu Q1）：**
  - Vidu Q1版本在视频生成技术领域取得重大突破，可精准还原皮肤纹理、金属反光等微观细节，同时保持大场景的空间连贯性。最具颠覆性的是其智能运镜控制系统，通过深度学习百万部电影镜头语言，能自动生成符合叙事逻辑的专业级镜头调度，用户只需简单描述即可实现包括斯坦尼康稳定拍摄和无人机俯瞰镜头在内的高难度运镜效果。
- **网页：**https://www.vidu.cn/

4.4 AI工具深度测评 - Vidu

Vidu Q1 核心功能（视频部分）

- **文生视频**
  - 风格选择：动画、写实
  - 清晰度：1080p（试用3次后需会员）
  - 运动幅度：自动、小、中、大
- **图生视频**
  - 时长：5秒
  - 清晰度：1080p（试用3次后需会员）
  - 运动幅度：自动、小、中、大
- **参考生视频：**允许用户通过上传视频片段、图片作为创作蓝本，系统会智能提取参考素材的视觉风格、运镜节奏甚至情感基调，结合深度学习算法生成既保留参考特征又符合用户文字描述的全新视频内容。
- **社区联动**
  - 提供官方或用户共享的特定风格模型，通过“做同款”方便用户快速生成特定风格的视频。

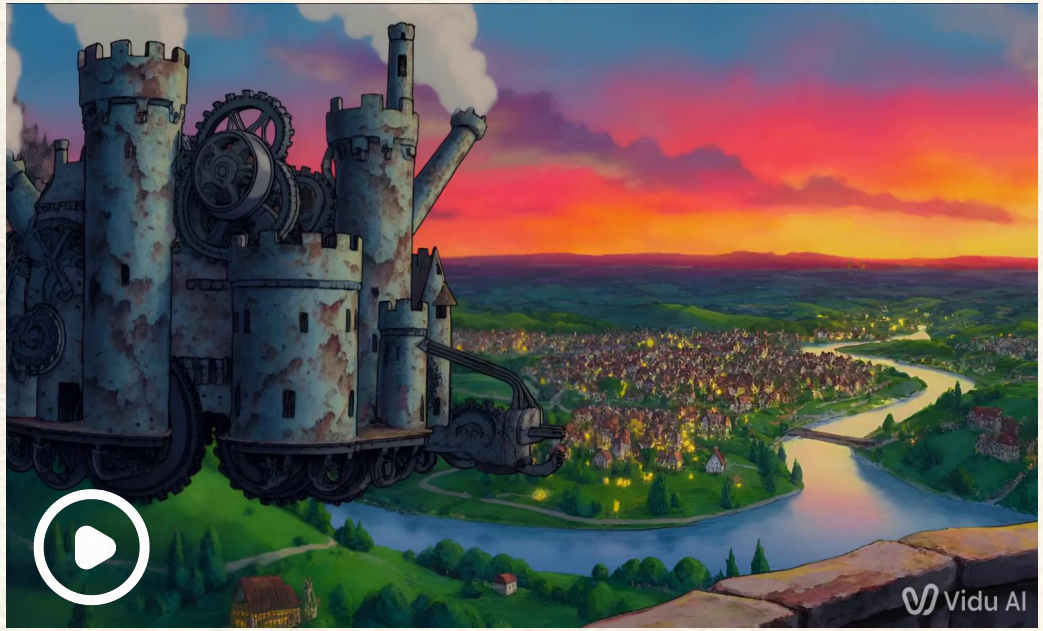




4.4 AI工具深度测评 - Vidu

Vidu--Vidu Q1 效果实测（文生视频）

提示词：夕阳西下，染红了整片天空，云层如同燃烧的火焰般绚烂。巨大的移动城堡喷吐着蒸汽，齿轮转动发出沉重的金属声响，摇摇晃晃地行走在连绵的绿色山丘上。城堡的烟囱歪歪扭扭，铁皮外壳布满锈迹和补丁，却透着奇异的生命力。近景处，苏菲站在城堡的露台上，风扬起她朴素的蓝色裙摆和棕色的发丝。她望着远方，眼神既带着对未知的忐忑，又有一丝隐约的期待。哈尔的黑鸟形态从她身旁掠过，翅膀划过空气，投下转瞬即逝的阴影。背景里，山脚下的小镇点缀着暖黄的灯火，蜿蜒的河流反射着暮光，像一条流动的金色丝带。整个画面既奇幻又温暖，蒸汽朋克的机械感与自然风光的柔美奇妙地融合在一起，仿佛在诉说：即使是最笨重的钢铁，也能拥有童话般的心跳。



动画风格



写实风格

效果评价：

- 文本理解：人物应当站在飞行城堡露台上，与城堡**共同运动**；但生成的视频呈现为人物站在固定城堡中看着飞行城堡移动
- 生成效果：画面清晰连续，细节刻画精准，但**整体视觉**效果略显**传统**

4.4 AI工具深度测评 - Vidu

Vidu--Vidu Q1 效果实测（图生视频）



效果评价：

- 生成效果：画面清晰连续，细节刻画精准（人物飞起的裙摆等），但城堡没有呈现出运动的状态，一直在下沉
- 与可灵2.0**相同，没有呈现出文本中对近景（如人物眼神）的描述，优先保证远景质量**

4.4 AI工具深度测评 - Vidu

Vidu--Vidu2.0 效果实测（参考生视频）

- 上传三张图片，系统智能提取参考素材的视觉风格
- 提取第1张图中古人、第2张图的背景和飞行城堡、第3张图的炼药炉，集合到一个场景中
- 因古人形象和药房搭配，系统智能将两者结合









效果评价：

- **画面清晰流畅，细节刻画精准**
- 稳定性较差，容易出现结合失败的情况

4.4 AI工具深度测评 - Vidu

Vidu 价格与性价比

- Vidu是一款付费AI图像、视频综合工具，效果基本可商用，相比人工和国外AI工具，性价比不错。**每日免费20**积分，可供体验（新人礼包80积分）。
- 可根据用量决定购买方案：
  - 商务定制--积分充值套餐
  - 长期使用--购买会员（连续包年、包月）



按年购买

能力	模型版本	时长	清晰度	风格	积分消耗
图-视频	vidu2.0	4S	720p	通用	4
图-视频	vidu2.0	4S	1080p	通用	10
图-视频	vidu2.0	8S	720p	通用	10
首尾帧	vidu2.0	4S	720p	通用	4
首尾帧	vidu2.0	4S	1080p	通用	10
首尾帧	vidu2.0	8S	720p	通用	10
参考-视频	vidu2.0	4S	720p	通用	8

积分消耗
（Vidu2.0）

金额（RMB）	购得积分数	可用并发数	有效期
100	320 积分	≤ 5	1年
1000	3200 积分	≤ 5	1年
2000	6400 积分	≤ 5	1年
5000	16000 积分	≤ 5	1年
10000	32000 积分	≤ 5	1年

积分充值套餐

4.4 AI工具深度测评 - 通义万相

通义万相 简介与核心技术

- **开发背景:** 通义万相是阿里云通义系列AI绘画创作大模型，由阿里巴巴研发，该模型可辅助人类进行图片、视频创作，于2023年7月7日正式上线。
- **市场定位:** 主要面向 创意专业人士、设计师，以及 普通用户，提供 AI 辅助的视频、图像生成工具，降低创作门槛。
- **核心技术：**
  - Diffusion Transformer (DiT) 架构，结合 自研高效 VAE（变分自编码器），支持 无限长 1080P 视频生成。
- **关键技术特点：**
  - 多模态输入支持：可接受 文本、图像 作为输入，并生成高质量输出。
  - 运动控制技术（类似 Motion Brush）：能精准模拟 复杂肢体动作、物理规律（如滑雪、花样滑冰等）。
  - 中文文字生成：首个支持 中文文字特效 的视频生成模型，适用于广告、短视频。
  - 时空上下文建模：采用时空全注意力机制，增强视频连贯性。

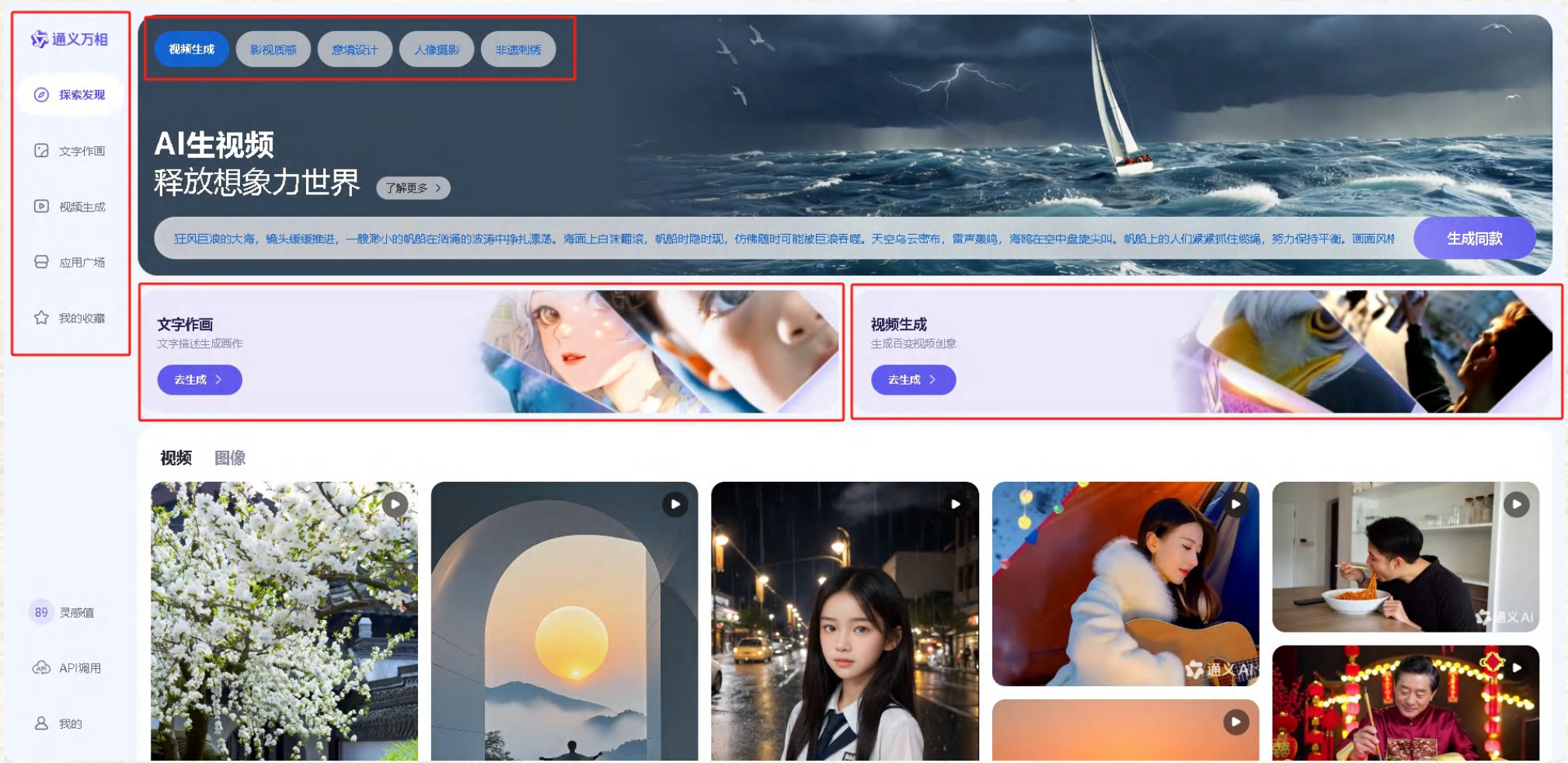
通义万相

- **主要产品形态:** 网页、API服务
  - 链接：https://tongyi.aliyun.com/wanxiang/
- **主要功能**
  - **文生视频**：根据文本描述生成高质量视频（如“滑雪运动员在阿尔卑斯山滑行”）
  - **图生视频**：输入静态图像，根据提示词生成动态视频（如将建筑照片转换为动态花灯效果）
- **附加功能**
  - **提示词智能扩写**：根据输入的提示词文本，智能生成描述更详细的提示词。
  - **灵感模式**：增加创意灵感，提升画面丰富度与表现力，可能会与输入提示词有差异。
  - **视频音效**：为视频内容生成合适的声音效果，若无明确音效则生成背景音乐。
  - **视频比例**：文生视频支持设置生成视频的长宽比。
  - **首尾帧**：根据给出的视频的第一张和最后一张图片以及描述提示词生成视频。

4.4 AI工具深度测评 - 通义万相

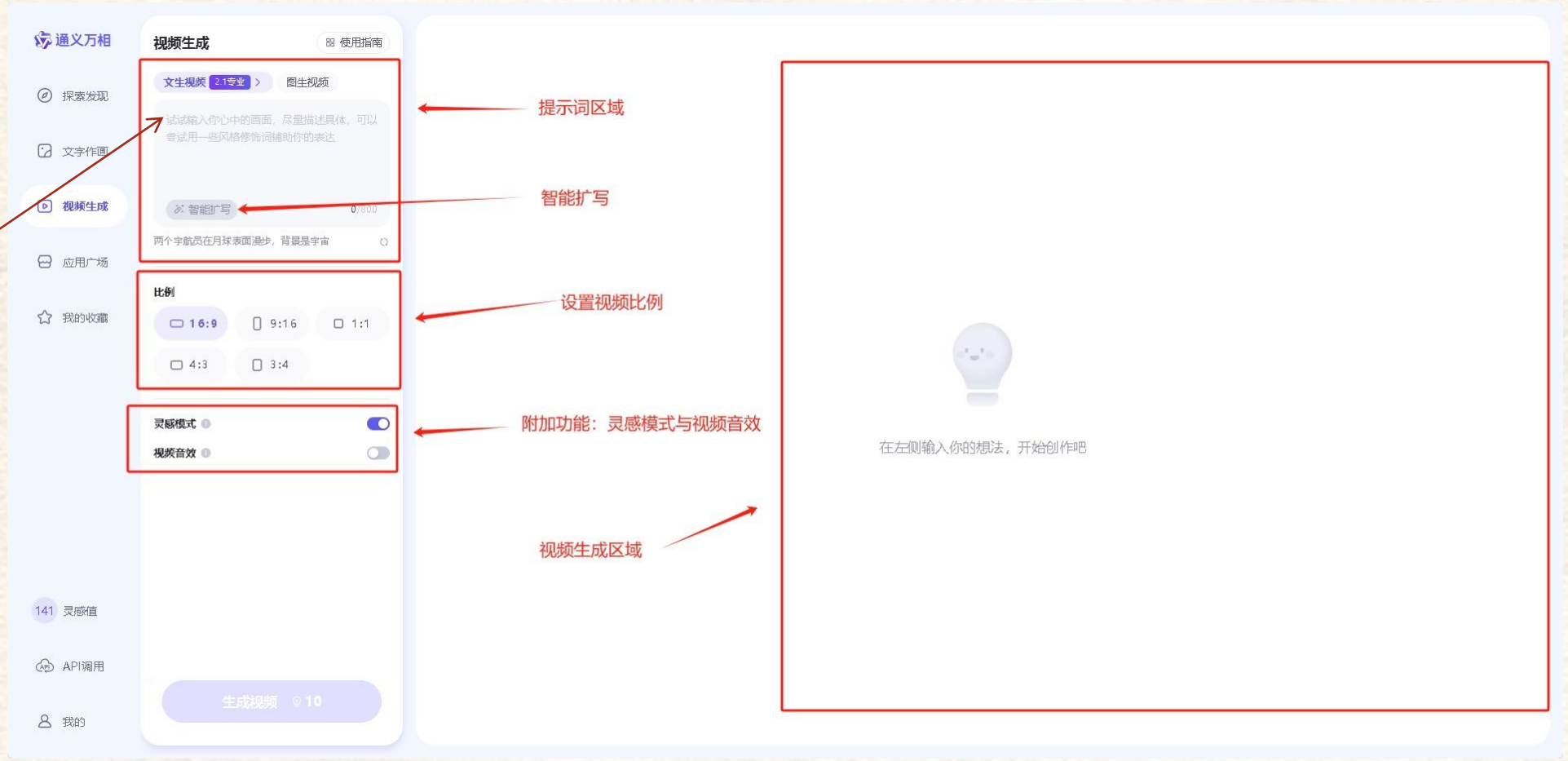
通义万相

功能选择区 →



主页截图

通义万相


模型选择区
创作模型
文生视频 2.1 极速
New 全新升级
生成速度更快，大幅度提升复杂运动能力，真实反馈现实物理规律，视频质感与美观度进一步升级
文生视频 2.1 专业
New 全新升级
生成品质更高，大幅度提升复杂运动能力，真实反馈现实物理规律，视频质感与美观度进一步升级
文生视频 2.0
通用文生视频，丰富的风格表现力，擅长中式元素生成
通义万相
探索发现
文字作画
视频生成
应用广场
我的收藏
141 灵感值
API调用
我的
视频生成
使用指南
文生视频 2.1专业
图生视频
试试输入你心中的画面，尽量描述具体，可以尝试用一些风格修饰词辅助你的表达
智能扩写
0/800
两个宇航员在月球表面漫步，背景是宇宙
比例
16:9
9:16
1:1
4:3
3:4
灵感模式
视频音效
生成视频 10
提示词区域
智能扩写
设置视频比例
附加功能：灵感模式与视频音效
视频生成区域
在左侧输入你的想法，开始创作吧

视频生成页面

4.4 AI工具深度测评 - 通义万相

- **核心功能实测-文生视频**
- **测试任务描述:** 主要测试提示词遵从程度，测评输出视频是否与提示词描述一致。

提示词：（**复杂的肢体运动、大幅度动作、镜头的移动**）
在室内，镜头平拍一个外国男子跳霹雳舞的全景，男子身穿灰色的上衣和绿色的裤子，镜头随着男子的动作而移动，男子在舞台上进行一系列的翻滚和旋转动作，背景中可以看到观众席上的观众和一些模糊的舞台灯光，但焦点始终保持在舞者的动作上。

- **文生视频2.1专业版输出评价：**人物动作及肢体运动变形、不符合物理运动规律、但镜头移动、跟随及画面氛围效果较好。
- **文生视频2.1极速版输出评价：**人物肢体扭曲、动作变形、镜头运动幅度较小。



文生视频2.1极速版输出



文生视频2.1专业版输出

4.4 AI工具深度测评 - 通义万相

- **核心功能实测-文生视频**
- **测试任务描述:** 主要测试提示词遵从程度，测评输出视频是否与提示词描述一致。

> **提示词**：（**真实世界的物理规律，也能逼真还原、雨滴落在伞上，会溅起水花**）
> 对穿着正式晚礼服的夫妇在回家途中遭遇大雨他们撑着黑色雨伞。平拍镜头下，男士穿着黑色西装，女士穿着白色长裙。他们在雨中缓缓行走，雨水沿着伞面滴落。镜头跟随他们的步伐平稳移动，展现出他们在雨中的优雅姿态。

- **文生视频2.1专业版输出评价：**画面、人物动作细腻真实，整体效果非常逼真。
- **文生视频2.1极速版输出评价：**人物肢体稍显变形、但雨滴等物体运动符合真实世界运动规律，表现较好。



文生视频2.1极速版输出



文生视频2.1专业版输出

4.4 AI工具深度测评 - 通义万相

- **核心功能实测-图生视频**
- **测试任务描述:** 主要测试提示词遵从程度，测评输出视频是否与提示词描述一致。
- **输入 Prompt：**

提示词：

夏日午后，吉卜力风格画面，90年代日本动画的视频片段，一位笑容满面的小女孩坐在古朴凉亭内的长椅上，享受着清凉的西瓜。她身着鲜亮的红色短袖与蓝色短裙，头发被细心地编成两条活泼的辫子。一旁，一只好奇的小猫正歪头观察着她的一举一动，似乎也被这甜蜜的场景吸引。阳光透过密集的竹林，投射出斑驳陆离的光影效果，为画面增添了几分生动与活力。凉亭四周环绕着繁茂的绿色植被，竹叶随微风轻轻摇曳，营造出一种宁静而和谐的氛围。近景特写，捕捉到小女孩快乐的表情与小猫灵动的姿态，展现出一幅充满生活情趣的画面。



输入参考图



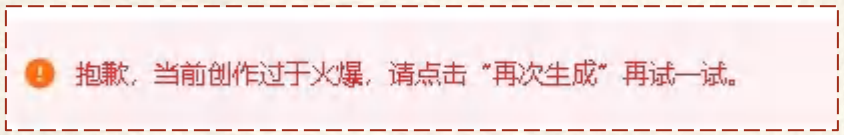
图生视频2.1专业版输出

效果评价：

提示词一致性效果较好，且画面细腻，镜头转换流畅，但有些细节处处理不到位（切近景时小猫头部变形）

4.4 AI工具深度测评 - 通义万相

- **易用性与交互体验:**
  - 界面设计: Web界面简洁、直观、易于上手。
  - 交互流程: 智能扩写功能增加了工具的易用性有历史记录功能。
  - 交互体验:
    1. 由于免费，用户基数庞大，使得工具整体稳定性一般，测试过程中图生视频经常出现下图所示输出。（曾出现10次申请仅1次成功的情况）



    2. 复杂动作场景下，人物肢体细节处理较差。
    3. 镜头遵循和复杂动作遵循能力与即梦3.0相比还有差距

- **集成与API能力:**
  - 开发平台链接：https://help.aliyun.com/zh/model-studio/text-to-video-api-reference
  - API可用性: 输出稳定性一般、说明文档完善
  - 开发者友好度: 文档质量高、社区支持度好

4.4 AI工具深度测评 - 通义万相

北京大学 PEKING UNIVERSITY

AI工具定价与免费额度:

- 免费政策: 当前（2025.4.25）通义万相AI工具使用免费，没有付费使用，但视频生成需要消耗灵感值。每日签到、创作反馈、创作投稿都会免费获取灵感值。
- 生成消耗:
  - ✓ 文生视频 2.1专业版一次生成需要消耗10灵感值。
  - ✓ 文生视频 2.1极速版一次生成消耗5灵感值。
  - ✓ 图生视频一次生成消耗10灵感值



API定价

文生视频

模型名称	模型简介	计费单价	免费额度
wanx2.1-t2v-turbo	生成速度更快，表现均衡。	0.24元/秒	免费额度：各200秒 有效期：阿里云百炼开通后180天内
wanx2.1-t2v-plus	生成细节更丰富，画面更具质感。	0.70元/秒	

图生视频

模型名称	模型简介	计费单价	免费额度
wanx2.1-i2v-turbo	速度更快，耗时仅为plus模型的三分之一，性价比更高。	0.24元/秒	免费额度：各200秒 有效期：阿里云百炼开通后180天内
wanx2.1-i2v-plus	生成细节更丰富，画面更具质感。	0.70元/秒	

4.4 视频类AI工具 - 实测总结

综合对比

工具	时长	单价（元）	API	优点	缺点	适用场景
即梦AI	5s	0.22	√	生成速度较快、准确率较高，基本不需要“抽卡”，工具集成度高，易用性好 性价比最高	缺少首尾帧及多图片参考 功能。高难度动作不够好	复杂运镜动幅、大段提示词语义、体育模拟、风格化内容，适合设计推敲
可灵AI	10s	1	×	高质量**视觉美学** 动态表现力强 文本理解能力较强 提供首尾帧功能	成本高，抽卡次数多	影视级真实感、商业项目（如产品广告） 适用人群：短剧创作者
Vidu	8s	0.35	√	**动画风格**表现力较高 （文生视频中可选择动画或写实风格） 提供首尾帧功能	艺术化风格稍弱 文本理解能力相对较弱 高难度动作处理	高动态表现、风格化内容（如动漫混剪、游戏素材）
通义万相	6s	-	√	工具免费，简单场景生成效果较好	稳定性较差，经常系统繁忙。人物复杂动作生成较差，易出现肢体变形	低成本制作，简单场景生成

注：单价估算不含免费赠送积分

数字人AI工具测评

4.5 AI数字人工具深度测评总述

- **测评目的**
  - 全面评估不同工具在数字人视频生成场景下的表现，为用户提供选型依据。
- **测评维度:**
  - **核心功能与效果 (重点)**
  - **易用性与交互体验**
  - **定价与性价比**
- **测评工具选型**
  - **Heygen、硅语、智课、奇妙元、闪剪**
- **测评结果对比**

1. **硅语**：声音和数字人效果非常好，适合对画面和语音要求极高的场合
2. **Heygen**：声音和数字人效果非常好，适合大多数场景，尤其是出海业务
3. **智课**：声音和数字人效果好，有开源大模型加持，非常适合制作视频课程
4. **奇妙元**：声音操控功能强大，适合对声音要求高的场合，但数字人动作略显僵硬
5. **闪剪**：专业版服务好，功能多，适合长期大量做IP的业务场景，数字人效果一般般

4.6 AI数字人工具测评 - 硅语

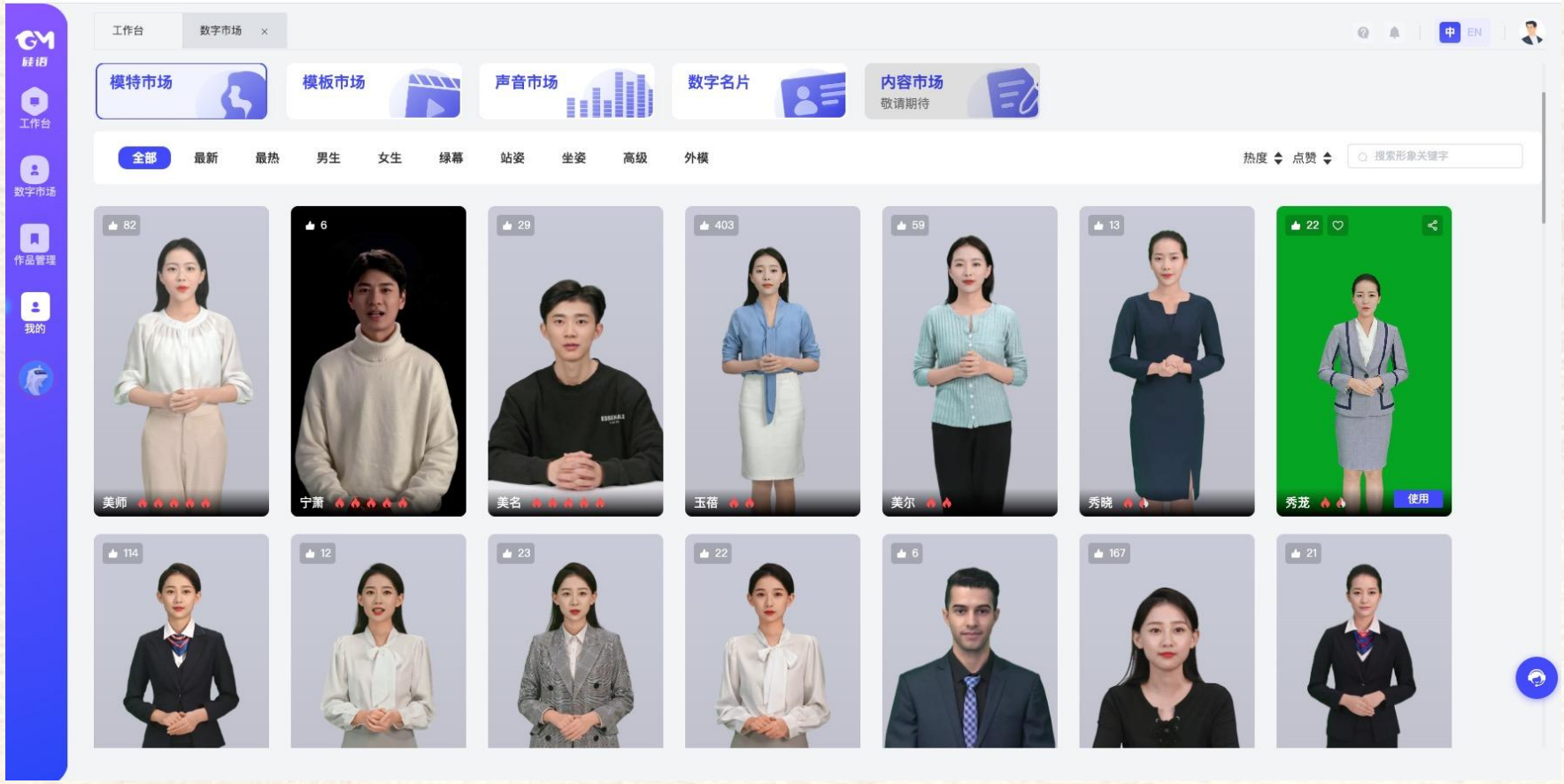
硅语简介：

硅语AI，是硅基智能的AI数字人工具，核心功能涵盖了AI数字人克隆、语音合成、智能提词、精准字幕生成以及美颜滤镜等。

硅语

地址：https://meta.guiji.ai/

扫码定制数字人
体验数字人工具



4.6 AI数字人工具测评 - 硅语

硅语的核心功能与效果：声音效果顶级

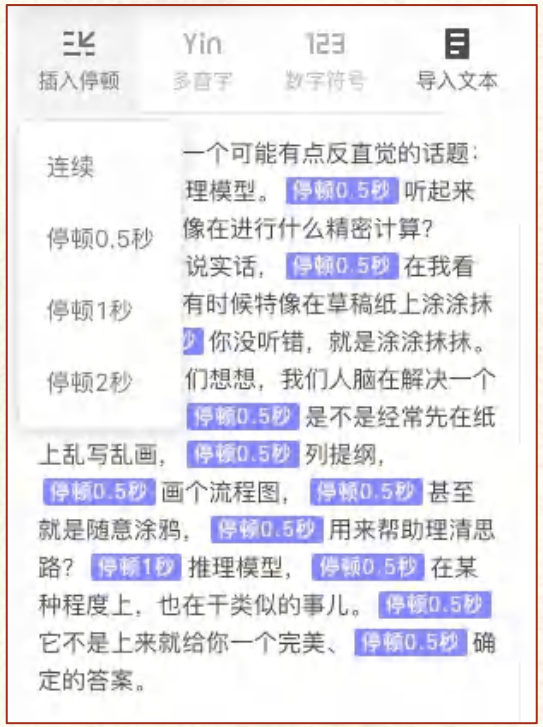
- **声音：**
  - 直出：直出好，可用
  - 声音克隆：支持1:1声音克隆，分不同级别，适合不同用户
- **数字人动作/口型：**
  - 动作、口型真实
  - 无法操控表情/动作
- **AI味儿：**基本没有什么机械感，通过声音细节操控更佳



4.6 AI数字人工具测评 - 硅语

硅语的易用性与交互体验

- 画幅支持：支持16:9、9:16
- 易用性：一般，声音操控功能不太好找
- 协作功能：团队版支持共享工作区
- 操控性：可以调节语调、速度、停顿
- 素材丰富度：1000+公用数字人
- 试听体验：片段试听功能不太好用





4.6 AI数字人工具测评 - 硅语

硅语的定价与性价比

- **价格方案：**
  - 视频制作每分钟5元，音频制作每分钟1元
- **性价比分析：**
  - 收费采用充值方式，较灵活



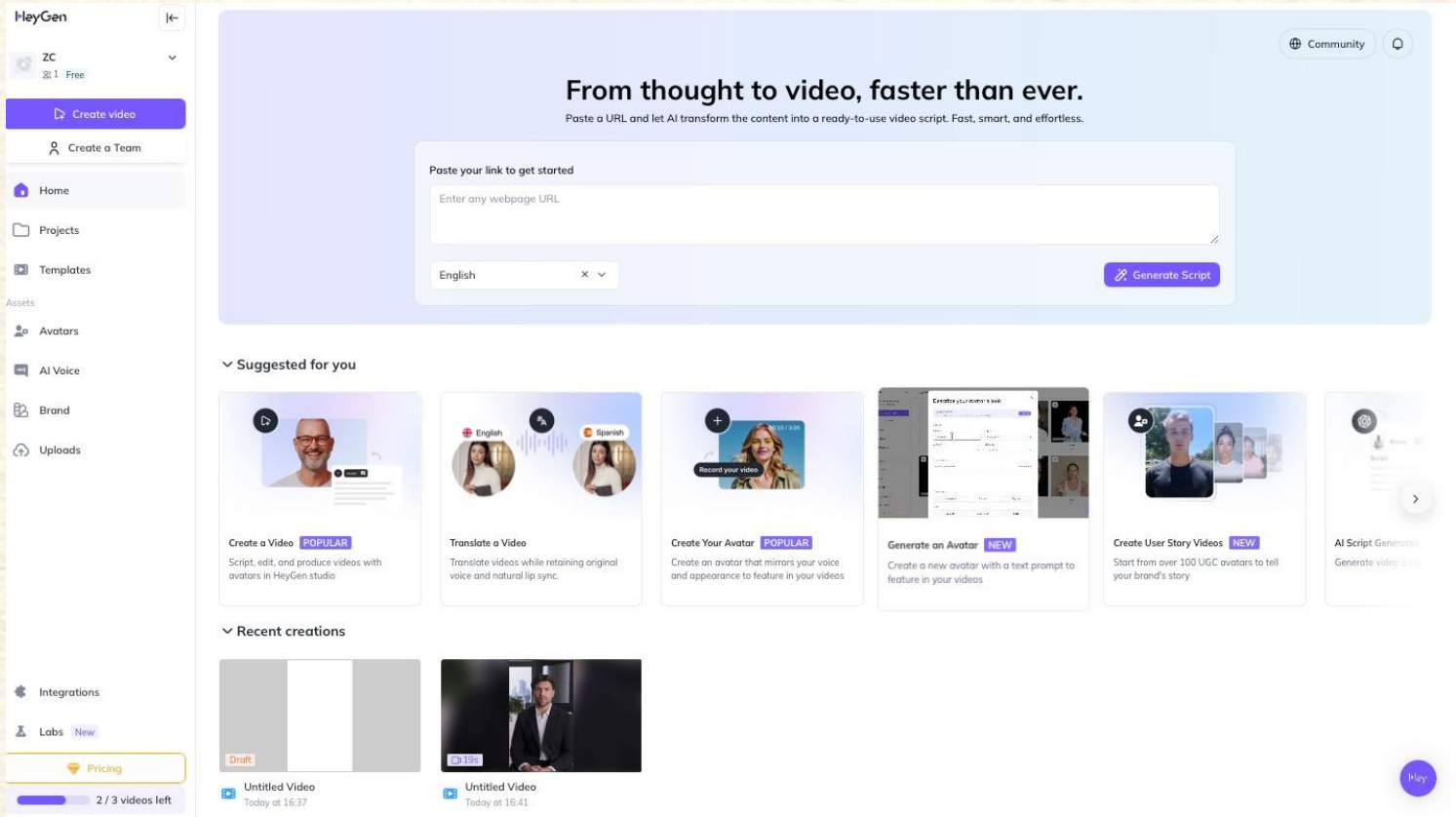
4.6 AI数字人工具测评 - Heygen

- HeyGen是一款热门的AI驱动的数字人视频创作平台，简化视频制作过程，让用户能够迅速制作出具有专业水准的数字人视频。该工具的核心优势在于其先进的人工智能技术，不仅赋予用户对视频中数字人物形象的完全控制权，还提供了一个丰富的素材库，包括多样化的背景、插图和文字模板，以支持用户打造个性化的宣传视频。（需要梯子）

HeyGen

https://app.heygen.com

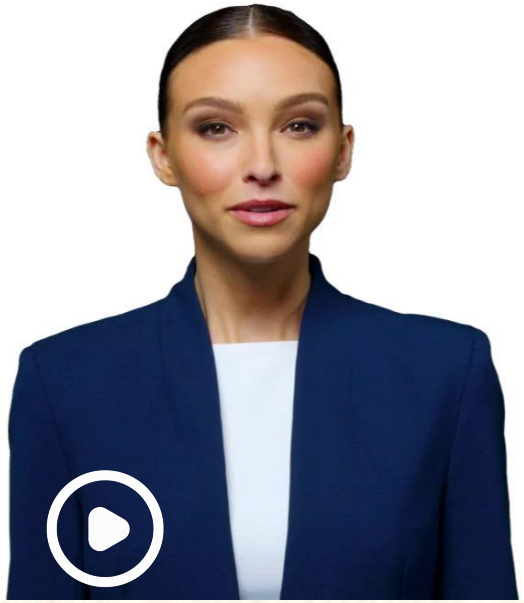
扫码定制数字人
体验数字人工具



4.6 AI数字人工具测评 - Heygen

Heygen的核心功能与效果：适合“出海”等跨境业务

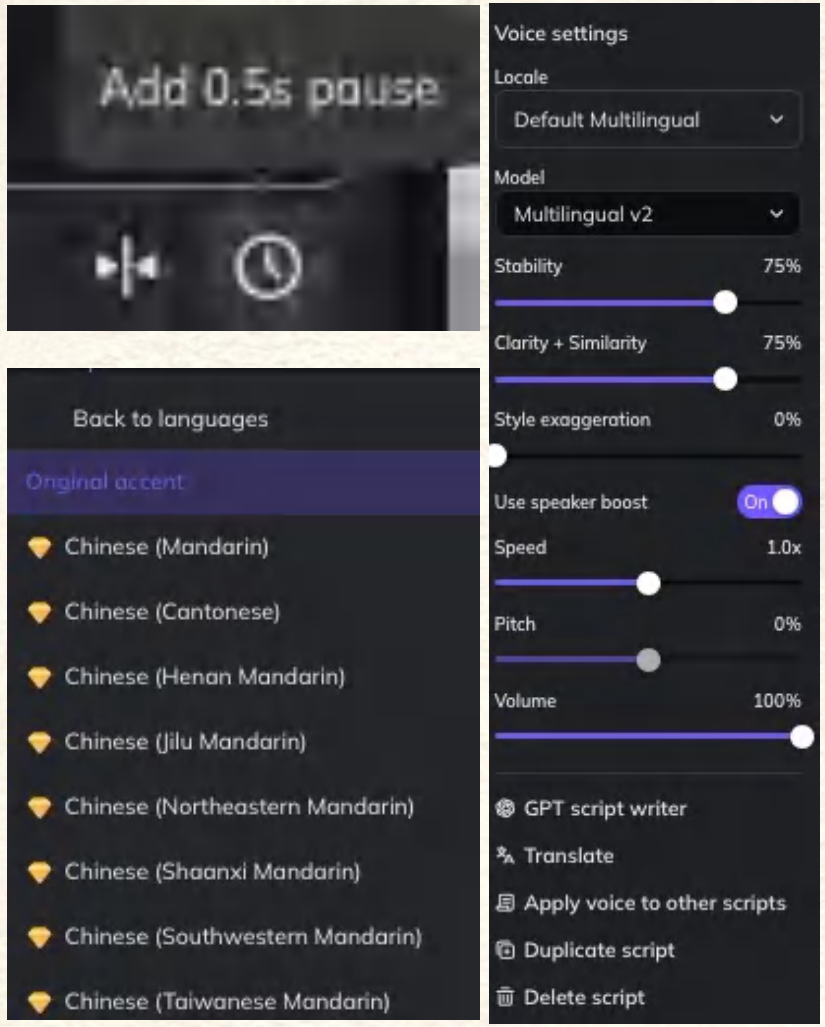
- **声音：**
  - 直出：直出效果正常，外国人说中国话的味道都能出来
  - 声音克隆：支持1:1声音克隆、方言、情感语气。
- **数字人动作/口型：**
  - 动作、口型接近真实
  - 无法操控表情/动作
- **AI味儿**：基本没有什么机械感



4.6 AI数字人工具测评 - Heygen

Heygen的易用性与交互体验

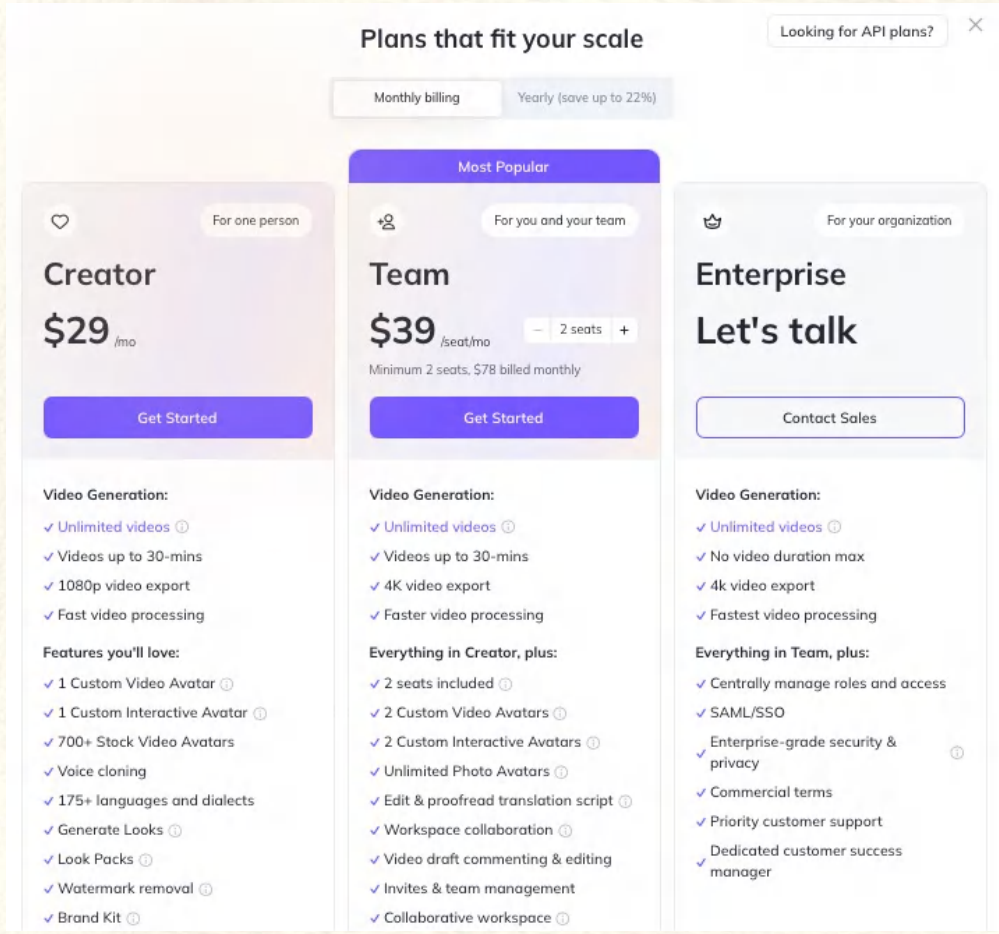
- 画幅支持：多画幅（16:9、9:16、4:3）
- 易用性：英文界面，需科学上网
- 协作功能：团队版支持共享工作区
- 操控性：可以调节音色、速度、停顿、多语言（方言）支持跨语言迁移、情感语气生成，实时声音转换
- 素材丰富度：700+模板，1个自定义头像，资源较丰富
- 试听体验：缺少片段试听功能，只能整体试听



4.6 AI数字人工具测评 - Heygen

Heygen的定价：

- **价格方案**
  - 个人版≈210元/月，每分钟视频约8元
  - 团队版≈280元/月（4K导出+协作工具）
- **性价比分析**
  - 团队协作场景性价比高
  - 单用户成本较高；适合高频次需求



4.6 AI数字人工具测评 - 智课

- 智课是一款基于数字人技术的视频制作工具，支持上传PPT并自动生成文稿，提供真人形象克隆、声音还原（最高达99%音色匹配）及通用数字人模板选择，同时内置多场景素材库。适用于企业培训、在线教学、产品演示等场景。产品通过简化传统视频制作的脚本撰写、团队协作、设备租赁等环节，降低时间与人力成本，并提供线下定制服务以满足个性化需求。

智课

http://zhike.bdqn.cn/

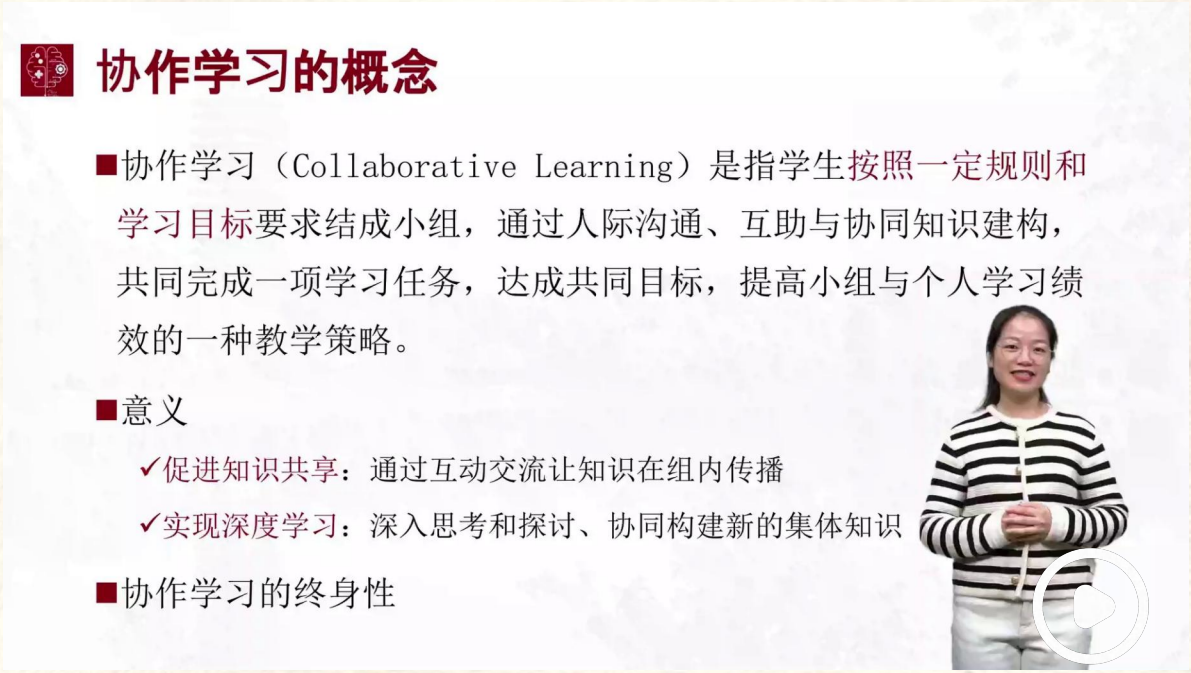
扫码定制数字人
体验数字人工具



4.6 AI数字人工具测评 - 智课

智课的核心功能与效果：最适合做视频课程

- **声音：**
  - 直出：直出效果正常
  - 声音克隆：支持1:1声音克隆、方言、情感语气。
- **数字人动作/口型：**
  - 动作、口型接近真实
  - 无法操控表情/动作
- **AI味儿**：没有机械感



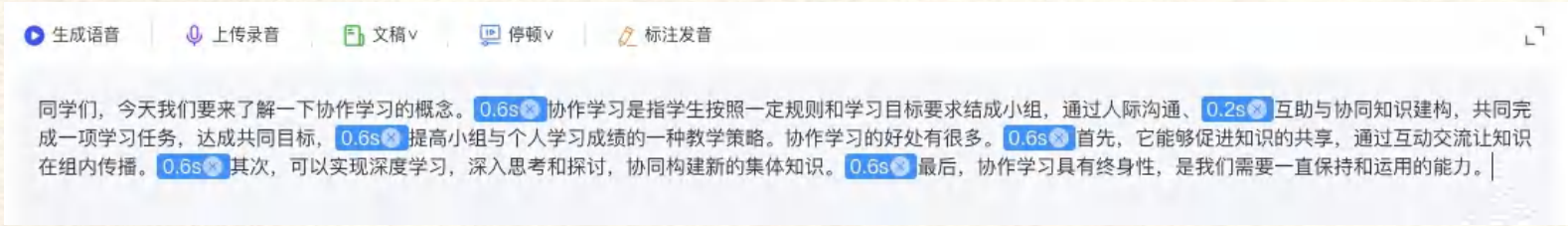
4.6 AI数字人工具测评 - 智课

智课的易用性与交互体验

- 画幅支持：多画幅（16:9、9:16、4:3）
- 易用性：**DeepSeek、Qwen3等大模型加持PPT讲课稿生成，改写。支持私有化部署。**
- 协作功能：团队版支持共享工作区
- 操控性：可以调节音色、速度、停顿、多语言。支持跨语言迁移、情感语气最丰富
- 素材丰富度：较少
- 试听体验：缺少片段试听功能，只能整体试听







4.6 AI数字人工具测评 - 智课

智课的定价：

- **价格方案**
  - 每分钟视频约5元
  - 数字人定制分两个级别，适合不同用户
- **性价比分析**
  - 定制E级数字人性价比高
  - 视频制作成本性价比高



4.6 AI数字人工具测评 - 奇妙元

- 奇妙元是一站式数字人平台，支持图像建模及多种驱动方式，轻松制作数字人视频和直播，广泛应用于各行业。

奇妙元

https://www.weta365.com

扫码定制数字人
体验数字人工具



4.6 AI数字人工具测评 - 奇妙元

奇妙元的核心功能与效果：

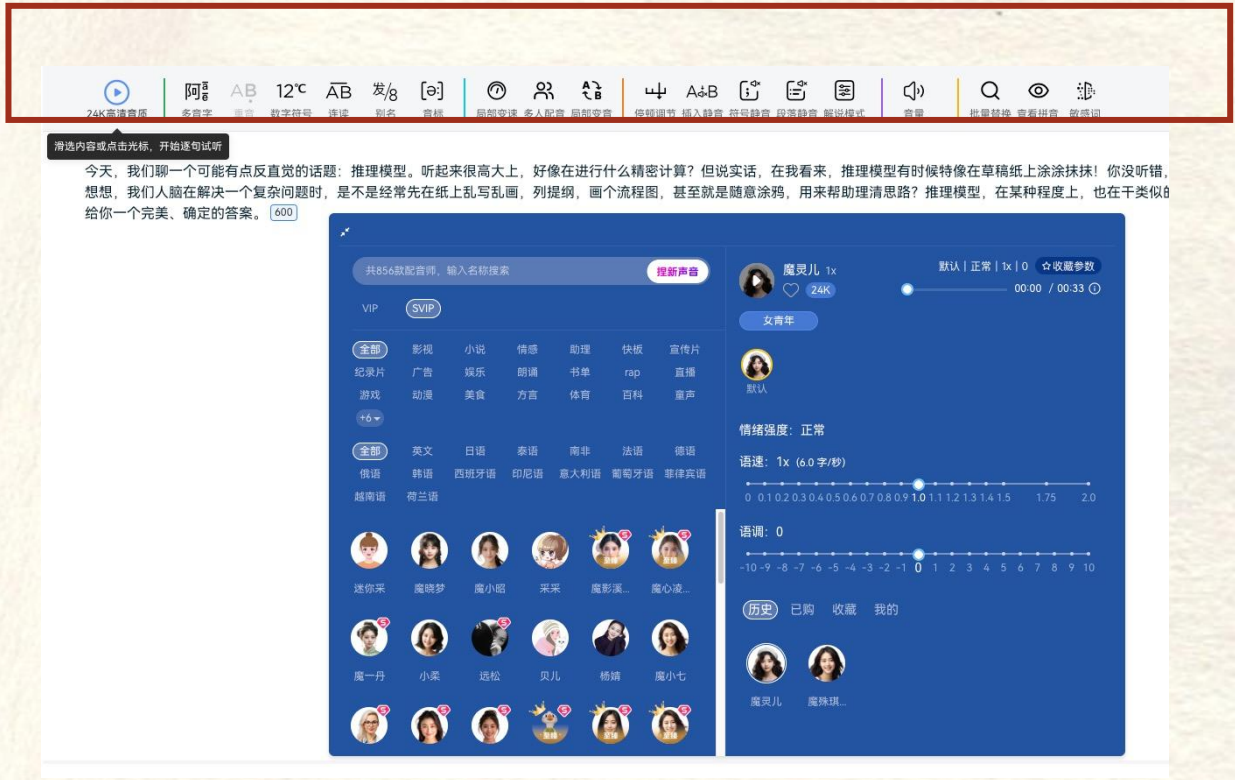
- **声音：**
  - 直出：直出不可用，需要通过增加停顿调节节奏
  - 支持方言、情感语气
- **数字人动作/口型：**
  - 动作有些僵硬、口型接近真实
  - 无法操控表情/动作
- **AI味儿：**有一定的机械感



4.6 AI数字人工具测评 - 奇妙元

奇妙元的易用性与交互体验：声音操控功能强大

- 画幅支持：常用的画幅尺寸都支持
- 易用性：团队协作功能强（共享工作区）
- 操控性：强大的声音编辑器，可以调节音色、速度、停顿、静音、多音字等多种功能，适合制作多场景的内容。
- 素材丰富度：60+数字人就，1000+声音风格，37种语言，15种方言。预制资源充足，适合批量生产
- 试听体验：支持片段试听，速度快



4.6 AI数字人工具测评 - 奇妙元

奇妙元的定价：

- **价格方案**
  - 月卡298元/月（60+数字人+500+声音）
  - 团队版≈280元/月（4K导出+共享工作区），平均每分钟视频7元
- **性价比分析**
  - 预制资源丰富，适合快速批量生产
  - 团队版性价比高



4.6 AI数字人工具测评 - 闪剪

- 闪剪是一款AI智能定制你的专属数字人，形象、声音意见克隆，告别真人出镜，并提供多样模板与配音资源。

闪剪

https://shanjian.tv

扫码定制数字人
体验数字人工具



4.6 AI数字人工具测评 - 闪剪

闪剪的核心功能与效果：

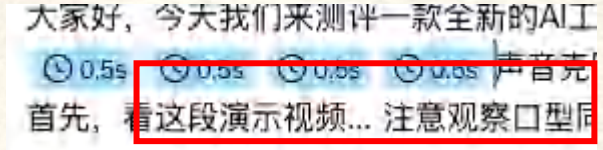
- **声音：**
  - 直出：直出不可用，需要通过增加停顿调节节奏
  - 支持方言、情感语气
- **数字人动作/口型：**
  - 动作较自然、口型接近真实
  - 无法操控表情/动作
- **AI味儿**：有一定的机械感

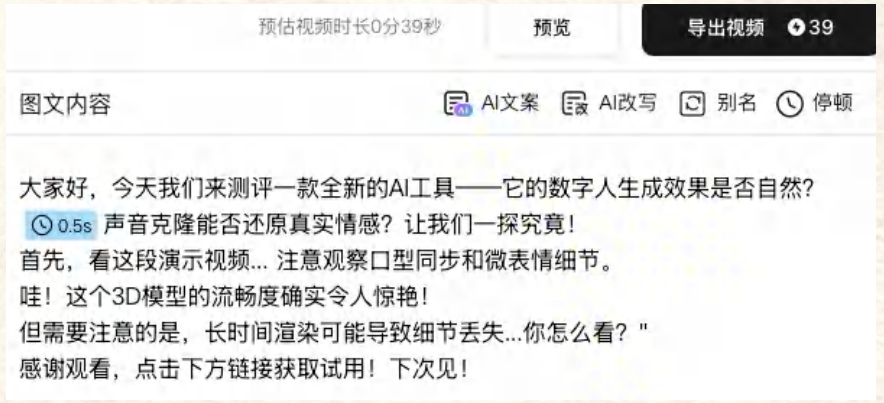


4.6 AI数字人工具测评 - 闪剪

闪剪的易用性与交互体验：只适合做短视频类内容

- 画幅支持：只支持9:16尺寸
- 易用性：接入了DeepSeek，有AI文案功能加持
- 操控性：声音调节功能少，只有停顿调节，别名，AI文稿改写等
- 素材丰富度：基础版60+数字人，专业版无限定制；素材库适中
- 试听体验：不支持片段试听，速度一般

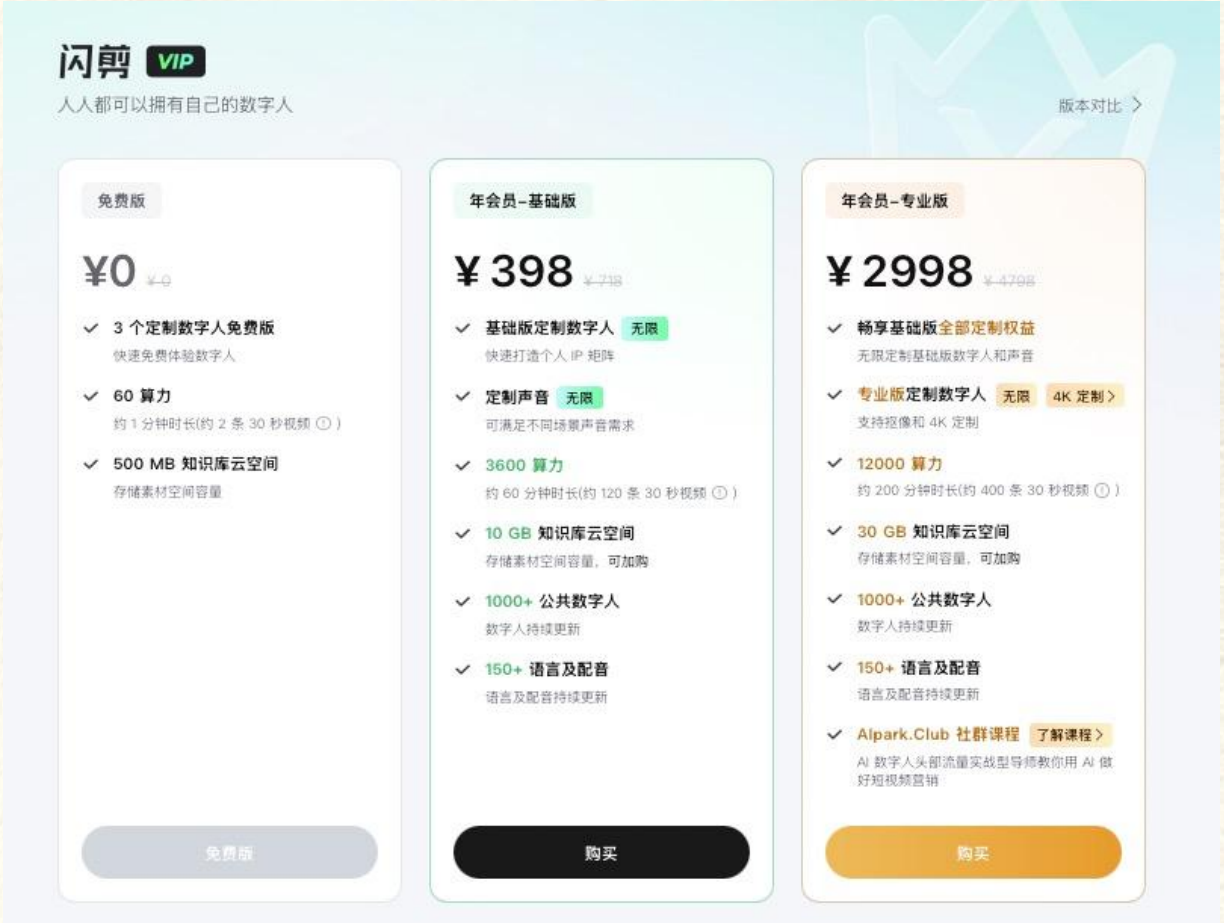




4.6 AI数字人工具测评 - 闪剪

闪剪的定价：

- **价格方案：**
  - 基础版398元/年（无限定制数字人/声音）
  - 专业版2998元/年（4K+200分钟视频+课程）
- **性价比分析：**
  - 年会员提供的服务丰富，专业版无限定制数字人，适合长期IP运营



4.6 AI数字人工具 - 实测总结

工具	优点	缺点	成本	适用场景
硅语	1.声音效果非常好，强烈推荐 2.数字人效果真实	易用性一般，有bug	1. 充值模式 2. 视频5元/分钟	要求较高的画面和声音的短视频、课程等
Heygen	1.外文支持好 2.各类语音效果真实	声音可操控性一般	1.订阅制 2. 视频8元/分钟	适合大多数业务，尤其是出海业务
智课	1.声音情感语气最丰富 2.AI加持授课讲稿生成，最适合做视频课程 3.支持私有化部署	素材数量较少	1.充值模式 2.视频5元/分钟	适合做视频课程，学校私有化部署
奇妙元	1.声音操控功能强大 2.资源丰富	数字人动作较僵硬	1.订阅制 2. 视频7元/分钟	适合对声音操控较高的场合
闪剪	1.接入DeepSeek，有AI能力 2.收费服务好，免费资源丰富	声音可操控性一般	1. 专业版可以无限训练数字人 2. 免费版可以训练数字人 3. 视频6元/分钟	适合长期大量做IP运营

扫码定制数字人
体验数字人工具

01 文本生成与处理类AI工具测评

02 图像生成与编辑类AI工具测评

03 音视频生成与处理类AI工具测评

04 代码生成与辅助类AI工具测评

05 大模型管理与应用类AI工具测评

5.1 AI编程工具深度测评总述

- **测评目的:**
  - 全面评估不同工具（IDE软件、编程插件、编程网站）在编程场景下的表现，为用户提供选型依据
- **测评维度:**
  - 核心功能与效果 (重点)
  - 易用性和体验度（重点）
  - 定价与性价比
  - 不同工具集成和编程语言能力
  - 社区活跃度情况
- **测评插件选型:**
  - IDE软件：Cursor、Trae
  - 编程插件：Cline、GitHub Copilot、通义灵码、CodeGeeX
  - 编程网站：MGX、heyBossAI



说明：所有的编程插件统一基于Visual Studio Code评测

5.1 AI编程工具深度测评总述

工具	类型	版本	核心功能	效果质量	集成性（支持模型）	易用性	成本效益	适用场景
Cursor	IDE软件	0.48.9	可以创建、维护复杂的、大型的项目工程代码，支持MCP	质量较高，全能、模型支持的较多、发布较早，功能稳定	Claude、DeepSeek、gemini、GPT系列及自定义兼容openai格式的模型	安装后，开箱即用	分个人版和企业版，均收费	无论是专业开发人员维护已有项目还是小白想要快速验证想法均适用
Trae	IDE软件	1.3.2	可以创建、维护复杂的、大型的项目工程代码，带有MCP市场	质量较高，高级模型高峰期可能需要排队。中文、免费、生成质量尚可	国际版支持Claude、DeepSeek、gemini、GPT系列的最新主流模型及自定义兼容openai格式的模型	安装后，开箱即用	免费	使用未接触过专业IDE的小白或预算有限想要实现工程化项目的建设和维护的人员
Cline	插件	3.13.2	功能丰富，可自动运行终端命令，自动进行调试等，带有MCP市场	Agent功能好用，生成质量取决于所选择的模型，全自动安装MCP服务，输出效果渲染较好	市面上绝大部分的模型服务商	安装后，需要些许配置	插件免费、模型服务商收费	功能较多，使用范围较广，可以对已有项目维护、也可以对新项目开发，还可以当做一个AI小助手对话使用，核心点在于所选择的模型的能力
GitHub Copilot	插件	1.303.0	支持主流IDE软件，嵌入交互式编程，compose智能模式	编码质量优秀稳定，操作便捷高效，综合体验好	Claude及GPT系列	IDE中安装后登录使用，需使用魔法	收费	专业开发人员，已有中、大型生产项目，对代码质量要求高，可结合传统IDE软件在保证质量前提下极大限度提效
通义灵码	插件	2.3.0	支持主流IDE软件；支持图片多模态交互问答，AI程序员提供工程化能力	生成失败概率较高，内容质量不太稳定	qwen系列、deepseek系列	IDE中安装后登录使用，不用魔法，中文支持更友好	个人版基础版免费 个人专业版及企业版收费	专业开发人员，已有中、大型项目，且对国内模型有特殊要求，需要图片多模态交互的推荐使用
CodeGeeX	插件	2.22.3	联网搜索、本地模型、项目地图（理解项目结构）、语言转换等编程工具箱	长文本输出效果不佳，交互式编程直接应用效果一般	CodeGeeX Lite、CodeGeeX Pro	IDE中安装后登录使用，不用魔法，中文支持更友好	个人用户免费 企业私有化部署收费	专业开发人员，辅助快速理解项目。需要联网搜索或者使用本地模型的推荐使用。
MGX heyBossAI	网站	/	两个工具都是利用多个Agent生成工程化代码，略有区别 • MGX：多个AI虚拟员工，各司其职按流程生成工程化代码 • heyBossAI：多位同事，在讨论组随机发言，形成项目功能点，可视化修改界面，支持模板上传	• MGX：对于简单需求来说生成质量尚可，可以选择高级模型提高生成质量，复杂功能需求需要人工介入 • heyBossAI：生成效果一般（模型不可选，且不公开），交互体验较好，适合非开发人员	• MGX：Claude、GPT、DeepSeek少量模型 • heyBossAI：无法切换模型	• MGX：省去了下载安装繁琐过程，小白可以打开网页即可使用，对小白友好 • heyBossAI：侧重简单模糊提示词生成项目	• MGX：每日月有免费额度，可以根据使用频率按需订阅 • heyBossAI：积分用来创建项目下载源码需要额外付费	适用于小白快速开发前端页面，部署上线（非开发者友好），但对于专业开发来说不是很适用

5.2 AI编程工具深度测评 - Cursor

Cursor简介与核心技术

- **开发背景**：Cursor 由 Anysphere 公司于 2022 年推出，目标是“创造一款能编写世界上所有软件的工具”
- **核心定位**：Cursor 定位于 “AI-first Code Editor” ，核心目标是通过深度集成 AI 技术重构开发流程
- **特别说明：**以下评测基于Cursor【0.48.9】 版本，模型选择Claude-3.7-Sonnet
- **核心技术**：
  - AI 代码生成与重构：自然语言生成代码，代码自动优化，跨语言转换
  - 智能上下文编程：@符号索引，多文件理解，缺陷预判
  - 交互式 AI 协作：行内聊天，差异对比，追问调试
  - 全栈开发支持：前端智能组件，API 衔接，文档同步
  - 工程化增强：智能测试生成，依赖管理，CI/CD 集成
  - 多模态开发：图表代码化，UI 草图转代码，语音编程
- **主要产品形态**：桌面客户端【链接：https://www.cursor.com/】

学习交流可加AI肖睿团队助理微信号（ABZ2829）

5.2 AI编程工具深度测评 - Cursor

Cursor整体总结

- **优点**

  👍 **功能全面**：作为一个AI编程工具，主要功能也是围绕编程展开，相比较其他对标工具（Trae、Windsurf）功能相似

  👍 **代码生成**：可以创建、维护复杂的、大型的项目工程代码，其中前端代码能力较强，小程序代码能力需要严谨的提示词，后端的代码需要严谨的提示词以及规范的Rule才可以达到日常使用要求

  👍 **学习成本**：基础功能（代码补全、简单生成）零门槛上手，但高级功能（Agent 模式、规则配置）需一定编程经验

- **缺点**

  👎 **使用成本**：Pro 版 20 美元/月定价对个人开发者偏高，企业版功能需 40 美元/月/用户

  👎 **MCP服务**：配置支持JOSN格式粘贴，不支持界面可视化配置以及对话中针对性开启个别MCP

5.2 Cursor实测：核心功能 - 网页代码生成

- **测试任务描述：** 开启Agent模式，模型选择Claude-3.7-Sonnet，在Cursor的输入框中输入一段网页开发需求，等待Cursor输出后使用浏览器验证开发效果
- **输入 Prompt (示例)：**使用HTML, CSS, JavaScript设计一款移动端的瑞士卷小游戏, 内容是很多瑞士卷和其他食物, 厨具等等, 物品从上往下掉, 用户点击到6个瑞士卷即可获胜, 一共两关, 第一关超简单, 第二关超难, 速度快而且掉落的内容多而繁杂







实践评价

- 网页界面生成速度快、稳定性高、界面美观
- 工程化项目，多个文件会分开且独立，目录结构清晰
- 网页端不好排错，需要到浏览器控制台复制错误信息回到Cursor提问（可以借助browser-tools-mcp打通Cursor与浏览器）

5.2 Cursor实测：核心功能 - 小程序代码生成

- **测试任务描述：** 开启Agent模式，模型选择Claude-3.7-Sonnet，在Cursor的输入框中输入一段微信小程序开发需求，等待Cursor输出后使用微信开发工具编译并验证开发效果
- **输入 Prompt (示例)：**在当前目录下按照原型图开发一个微信小程序, 小程序方案如下: ... ... (此处提示词过多 省略)

实践评价

- 生成微信小程序代码时，初始代码建议在微信开发工具中自动生成，Cursor生成大概率还需要调试
- 代码生成的准确性需要严谨的提示词，才能达到较好的效果
- 反复迭代，或者多次Restore时，一定概率出现代码修改Bug，无法进行自动Apply，需要手动



5.2 Cursor实测：核心功能 - 后端代码生成

- **测试任务描述：** 开启Agent模式，模型选择Claude-3.7-Sonnet，在Cursor的输入框中输入一段后端开发需求，等待Cursor输出后使用IDEA启动程序并使用接口工具进行验证
- **输入 Prompt (示例)：**backend/recruitment-backend目录是我初始化的一个SpringBoot项目，该项目目前是由idea默认创建，现在请按照这个接口文档, 帮我实现相关接口，目前数据库信息如下: ... ...（此处省略数据库信息）

实践评价

- 在后端代码生成中，Cursor生成的效果较为一般，不如使用代码生成器生成三层基础代码
- 迭代时，如果前后端项目在同一目录，当前端出现问题时，Cursor会索引当前目录所有文件，排错不仅会排查前端异常，也会排查后端代码







5.2 Cursor实测：核心功能 - MCP服务

- **测试任务描述：** 在Cursor的MCP配置中添加高德地图的MCP服务，在对话窗口中输入有关地图路线的问题，等待Cursor调用MCP服务后，配合地图工具验证输出结果是否正确
- **输入 Prompt (示例)：**我现在西二旗地铁站，我要去北京西站，该如何坐地铁？

实践评价

- MCP配置文件为通用文件，在其他客户端的配置文件可以平移到Cursor中
- MCP配置仅支持Json格式配置不支持可视化界面配置
- MCP调用时会出现弹出按钮，待用户确认时才会真正调用，避免危险调用（可关闭）




```
{
  "mcpServers": {
    "amap-amap-sse": {
      "isActive": true,
      "name": "amap-amap-sse",
      "type": "sse",
      "url": "https://mcp.amap.com/sse?key=
    }
  }
}
```

5.2 Cursor实测：效果评估

- **综合优点：**
  - [+] 工程化代码能力：可以创建、维护复杂的、大型的项目工程代码，其中前端代码能力较强，小程序代码能力需要严谨的提示词，后端的代码需要严谨的提示词以及规范的Rule才可以达到日常使用要求
  - [+] 灵活的自定义 AI 配置：支持接入 OpenAI API 等第三方模型，开发者可自由选择最新模型并调整参数，实现个性化代码生成
  - [+] 集成智能测试生成、依赖管理优化、CI/CD 模板等工程化功能，覆盖全流程开发需求
- **潜在缺点：**
  - [-] 对于小程序或后端项目的开发中，需要写复杂的提示词与Rule文件，无疑增加了学习和使用成本

5.2 Cursor实测：易用性、成本效益、集成性

- **易用性与交互体验：**
  - 界面继承与操作优化：基于 VS Code的界面设计，开发者可无缝迁移原有配置、主题和插件，降低学习成本，快捷键深度集成，交互流程自然流畅。
  - AI 功能直观易用：自然语言生成代码（如:输入“创建用户登录界面”）直接输出完整模块，支持行内差异对比和逐行采纳,自动上下文感知（通过 @ 符号引用项目内函数），减少手动查找依赖的繁琐操作
  - 学习曲线分层：基础功能（代码补全、简单生成）零门槛上手，但高级功能（Agent 模式、规则配置）需一定编程经验
- **定价与免费额度**：
  - 免费版限制明显：免费用户仅限 50 次慢速 GPT-4 调用和 2000次基础补全，难以满足高频开发需求
  - 订阅模式灵活性不足：
    - Pro 版 20 美元/月（无限补全 + 500 次快速模型调用），对比同类工具（如:免费的Trae）定价偏高
    - 企业版（40 美元/月）虽含隐私模式和团队管理，但中小团队可能因成本超支转向开源替代方案
- **集成性与生态兼容**：
  - 开发工具链深度兼容
    - ✓ 完全保留 VS Code 插件体系（如GitLens、Prettier），支持主流语言（Python/Java/Go）和框架（React/Spring）
    - ✓ 预置 CI/CD 模板（GitHub Actions/GitLab CI），无缝衔接工程化流程
  - 多模型服务商聚合：支持GPT-4、Claude 3.5等第三方模型接入，企业版可部署本地化模型（如:Llama 3），平衡性能与隐私

5.2 Cursor实测总结

- **核心优势：**
  - 多模态交互创新：支持Figma 草图转前端代码、图表代码化生成，拓宽开发交互边界
  - 灵活可扩展的模型生态：支持 OpenAI、Anthropic 等第三方模型接入，企业版允许私有化部署，平衡性能与数据隐私
  - AI 深度赋能开发流程：通过 GPT-4、Claude 等模型实现自然语言生成代码，支持跨文件级上下文理解与智能重构，减少 50% 以上重复编码
  - 零迁移成本的生态兼容：基于 VS Code 内核构建，完全兼容其插件、主题和快捷键，开发者无需改变原有习惯即可享受 AI 增强
- **主要限制：**
  - 中文语境适配不足：中文提示词生成效果弱于英文；国内用户订阅需跨境支付，流程繁琐
  - 高阶功能学习门槛：Agent 模式（自动化任务流）、Rules 配置（代码规范检查）等需开发者具备较强工程经验
  - 长期使用成本压力：Pro 版 20 美元/月定价对个人开发者偏高，企业版功能（如 SOC2 合规）需 40 美元/月/用户
- **适用场景：**
  - 全栈程序员开发提效
  - 快速原型验证：用语音/草图描述需求，10 分钟内生成可运行的最小可行性产品（MVP）代码框架

5.2 AI编程工具深度测评 - Trae

Trae简介与核心技术

- **开发背景**：Trae是由字节跳动开发的国内首个免费AI原生IDE，目标是通过深度集成人工智能技术提升开发效率
- **核心定位**：定位于"真正的AI工程师（The Real AI Engineer）"，致力于通过AI原生架构重构开发流程。与AI插件模式不同，其核心技术深度嵌入代码编辑器，实现智能开发的全流程自动化。
- **特别说明：以下评测基于Trae【1.3.2】 国际版本**
- **核心技术**：
  - 模型底座：集成了Claude，Gemini，GPT，DeepSeek等最新模型
  - 双模开发系统：包含从零构建的Builder模式和代码对话的Chat模式
  - 智能代码生成：多文件理解、错误预判、跨文件自动补全导入
  - 本土化增强：中文友好界面、国产API适配、企业级规范代码生成
- **主要产品形态**: 桌面客户端、插件【链接：https://www.trae.com.cn/】

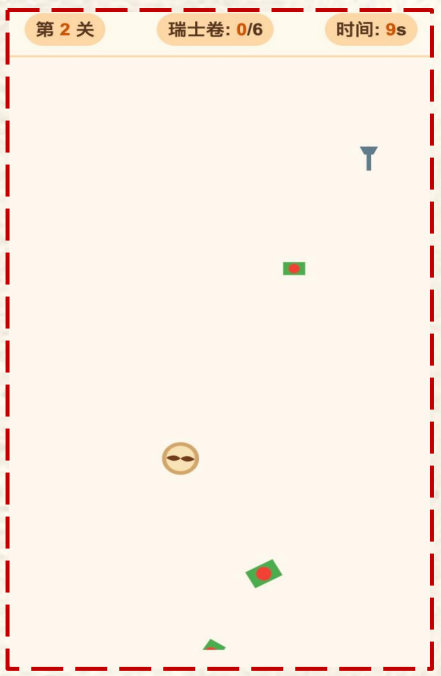
Trae整体总结

- **优点**
  - 👍 本土化优势：国内版完全支持中文界面与国产API调用，提供符合国内企业开发规范的代码生成能力
  - 👍 零成本使用：完全免费策略，集成DeepSeek等国内顶尖模型，无需额外支付模型调用费用
  - 👍 模式创新：Builder+Chat双模式分离设计，前者实现0基础项目构建，后者满足专业开发者调试需求
  - 👍 MCP市场：提供了MCP市场，配置起来也较容易
- **缺点**
  - 👎 复杂任务处理：生成复杂交互需求（如多部件联动、动态效果）时易出现功能缺失
  - 👎 响应效率：同类任务平均耗时比Cursor多2-3倍
  - 👎 功能限制：Builder模式自动纠错机制不完善，需频繁手动干预调试流程

5.2 Trae实测：核心功能 - 代码生成

- **测试任务描述：** 选择Agent模式，模型选择Claude-3.7-Sonnet，在输入框中输入一段网页开发需求，等待Trae输出后使用浏览器验证开发效果
- **输入 Prompt (示例)：**使用HTML, CSS, JavaScript设计一款移动端的瑞士卷小游戏, 内容是很多瑞士卷和其他食物, 厨具等等, 物品从上往下掉, 用户点击到6个瑞士卷即可获胜, 一共两关, 第一关超简单, 第二关超难, 速度快而且掉落的内容多而繁杂







实践评价

- 生成速度与生成效果取决于所选择的模型，调用速度取决于Trae，高峰时期调用速度会慢一些
- 一次性生成的网页，功能正常，效果不错，可以正常玩
- 创建多个文件时，需要每个文件单独点击一次应用，没有全部应用选项，只有全部接受选项

5.2 Trae实测：核心功能 - python代码生成

- **测试任务描述：** 开启Agent模式，模型选择Claude-3.7-Sonnet，在Trae的输入框中输入一段开发需求，等待Trae输出后验证开发效果
- **输入 Prompt (示例)：**创建api_demo目录，包含：

- config/config.py（存储API密钥）
- services/map_service.py（实现百度地图坐标转换）
- main.py（调用服务的示例代码）



实践评价

- 在没有给出百度地图API接口文档的情况下，Trae还是很顺利的写出了几段代码，经验证一次成功
- 只是给出了终端命令，却无法自动执行终端命令，需要自己手动执行

5.2 Trae实测：核心功能 - MCP服务

- **测试任务描述：** 在Trae的MCP配置中添加高德地图的MCP服务，在对话窗口中输入有关地图路线的问题，等待Trae调用MCP服务后，配合地图工具验证输出结果是否正确
- **输入 Prompt (示例)：**我现在在北京, 五一的时候想要去云南玩5天, 请帮我制作一个详细的旅游攻略, 需要给出具体的路线, 与时间安排, 还需要考虑天气情况





实践评价

- MCP配置文件为通用文件，在其他客户端的配置文件可以平移到Trae中
- 提供MCP市场，市场中的MCP服务较少，一些MCP还是需要手动配置
- 配置好MCP Server后，需要将该MCP Server添加到某个Agent助手方可使用
- MCP Server启动维度是助手维度，相比较其他客户端，启动不够灵活

5.2 Trae实测：效果评估

- **综合优点：**
  - [+] 工程化代码能力：可以创建、维护复杂的、大型的项目工程代码，工程效果主要取决于模型能力，高质量的模型如Claude-3.7-Sonnet等高峰期可能需要等待
  - [+] 中文IDE深度定制：各个Agent对于中文理解的效果也比较好，生成的质量会有显著改善
- **潜在缺点：**
  - [-] 模型数量少：目前集成模型服务商较少，很多模型还需要自定义接入服务商
  - [-] MCP市场服务较少：目前在MCP市场中提供了很多的MCP Server，但本土化常用的MCP服务（如高德地图MCP）不在市场中，还需要等待后续添加

5.2 Trae实测：易用性、成本效益、集成性

- **易用性与交互体验：**
  - 全中文界面设计：对小白友好，尤其是没接触过专业IDE人士，可以轻松上手
  - 对话模式：Builder+Chat双模式分离设计，前者实现0基础项目构建，后者满足专业开发者调试需求
- **定价与免费额度**：
  - 免费策略激进：作为完全免费工具，Trae 在预算敏感场景吸引力强，适合学生、独立开发者或低频需求用户，避免订阅制压力
  - 商业化路径模糊：长期可持续性存疑，若未来引入付费功能（如高速 GPT-4 调用），可能破坏现有免费优势，用户迁移成本将陡增
- **集成性**：
  - 模型选择单一化：仅支持少量模型，虽简化操作但牺牲了灵活度，尤其在代码生成场景无法根据任务特性切换模型（如 Claude 长文本处理 vs GPT-4 逻辑推理），专业性弱于 Cursor 的多模型聚合策略

5.2 Trae实测总结

- **核心优势：**
  - 零迁移成本生态：基于VS Code内核深度定制，兼容VS Code插件，支持主题/快捷键无缝迁移
  - 全流程自动化开发：Builder模式实现自然语言驱动的项目构建，自动生成环境配置，提升项目搭建效率
- **主要限制：**
  - 复杂任务处理局限：在交互逻辑复杂的场景存在功能缺失，相比Cursor实现完整度要低一些
  - 响应速度瓶颈：生成同等复杂度的小游戏（如记忆卡牌），平均响应时间比Cursor慢一些
- **适用场景：**
  - 中文环境快速原型开发：通过自然语言描述完成基础业务代码框架搭建，中文错误提示包含语义分析（如“第23行分号冗余”），结合代码作用域检查建议，降低调试门槛
  - 成本较低：适用于创业团队免费调用高级模型，快速完成微信小程序/网页/APP等前端程序验证产品效果

5.3 AI编程工具深度测评 - Cline

Cline简介与核心技术

- **核心定位**： 一个可以使用 **终端** 和 **编辑器** 的 AI 助手
- **核心技术**：
  - 使用任何 API 和模型：Cline 支持 OpenRouter、Anthropic、OpenAI等多家API 提供商
  - 在终端中运行命令：Cline 可以直接在终端中执行命令并接收输出
  - 创建和编辑文件：Cline 可以直接在编辑器中创建和编辑文件，展示更改的差异视图
  - 使用浏览器：Cline 可以启动浏览器，点击元素，输入文本和滚动，在每一步捕获截图和控制台日志
  - 集成MCP服务：Cline 可以创建和自动安装适合的特定工作流程的工具
  - 检查点：比较和恢复：可以使用“比较”按钮查看快照和当前工作区之间的差异，并使用“恢复”按钮回滚到该点
- 主要产品形态: VsCode插件【链接：https://cline.bot/】
- **特别说明：以下评测基于Cline【3.13.2】 版本**

Cline整体总结

- **优点**
  - 👍 **功能全面**：作为一款AI编程插件，功能很丰富，代码生成、问答、MCP市场、终端执行
  - 👍 **代码生成**：两种生成模式，Plan与Act模式，在Plan中进行工程设计、在Act中开发显著提高代码生成的准确性与结构的规范性
  - 👍 **免费开源**、生态庞大，集成多家模型服务商，使用时按量付费，较为合理
  - 👍 **MCP市场**：支持通过对话的形式安装MCP服务，简化了自己配置的繁琐过程
- **缺点**
  - 👎 **学习成本**：安装后需要些许的配置，对小白不利
  - 👎 **Token量消耗**：整体功能依赖于分步骤的Agent，每次模型请求会将当前上下文以及当前环境进行发送，Token量耗费巨大，尤其是MCP调用时，有别于其他MCP客户端基于Function Calling调用，Cline是完全基于提示词调用，会发送大量提示词

5.3 Cline实测：核心功能 - 代码生成

- **测试任务描述：** 选择Act模式，模型选择deepseek-chat，在输入框中输入一段网页开发需求，等待Cline输出后使用浏览器验证开发效果

- **输入 Prompt (示例)：**使用HTML, CSS, JavaScript设计一款移动端的瑞士卷小游戏, 内容是很多瑞士卷和其他食物, 厨具等等, 物品从上往下掉, 用户点击到6个瑞士卷即可获胜, 一共两关, 第一关超简单, 第二关超难, 速度快而且掉落的内容多而繁杂

实践评价

生成速度与生成效果取决于所选择的模型

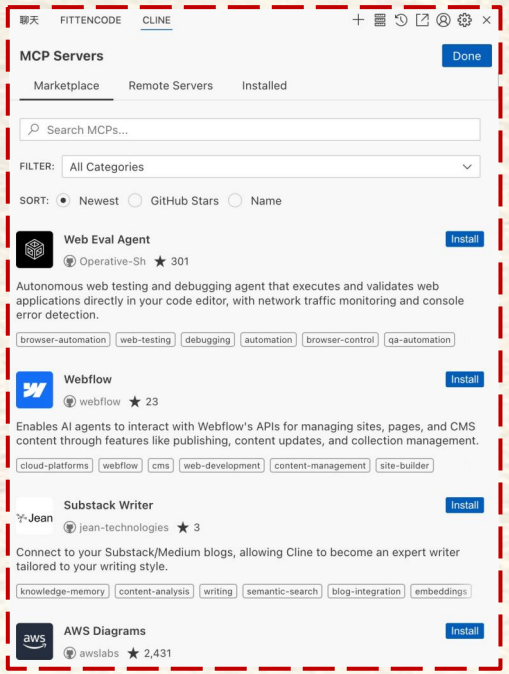
- Cline在代码生成结束后，会自动启动程序，遇到错误会自动进行调试修改并重新启动知道成功后完成
- 生成逻辑是先生成任务列表，再逐步实现每个任务，Token消费量较多



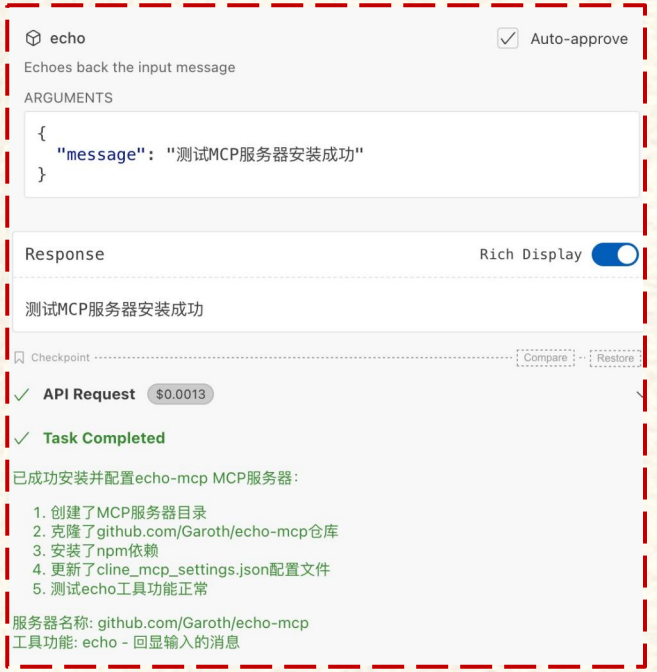


5.3 Cline实测：核心功能 - MCP服务安装

- **测试任务描述：** 在Cline的Marketplace中安装插件
- **找到Cline的MCP市场**
- **找到自己需要的MCP Server**
- **点击Install，根据提示逐步安装**







实践评价

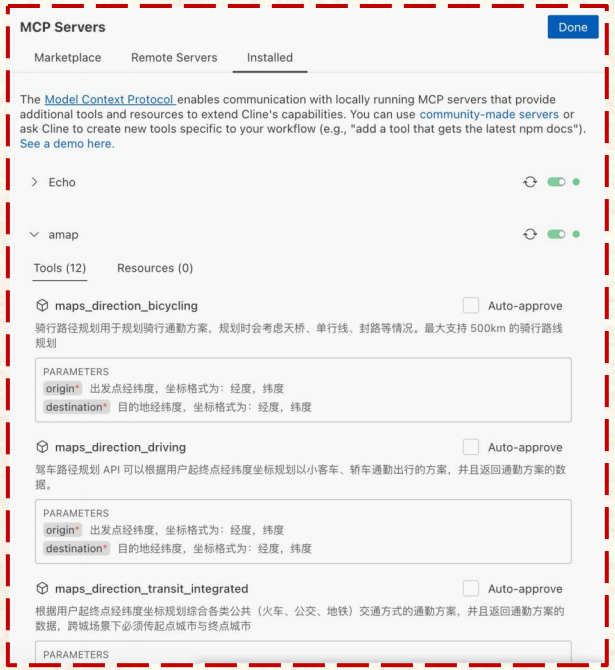
- 对于MCP小白来说，安装很方便，安装过程类似于应用商店安装一个APP
- 对于有经验的人士来说，安装部分有点冗余，时间较长
- MCP市场中的MCP Server目前还在维护中，不包含世面上所有的MCP，需要的MCP Server不在列表中时（如高德地图MCP）还是需要自己手动安装

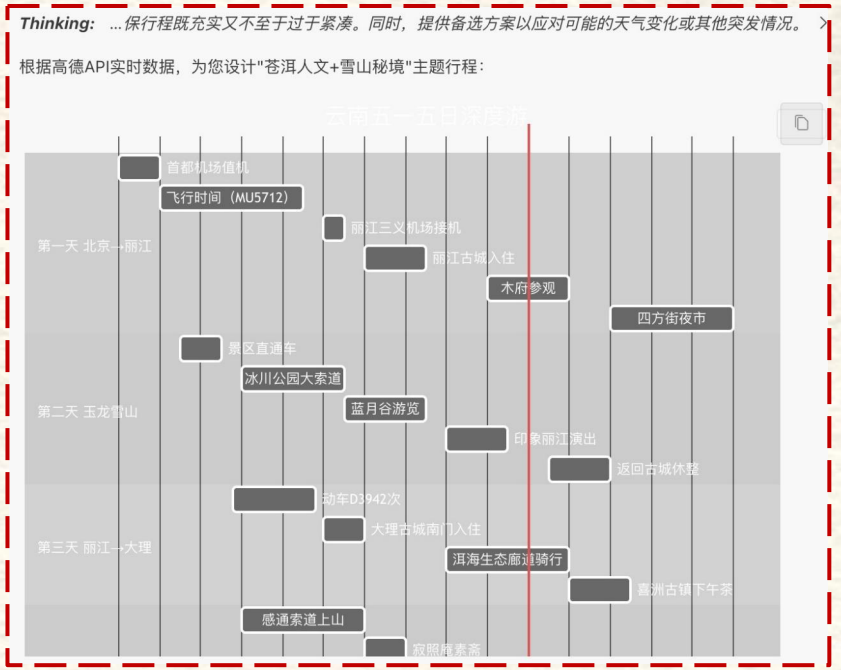
5.3 Cline实测：核心功能 - MCP服务使用

- **测试任务描述：** 在Cline中配置高德MCP服务
- **模型选择**：deepseek-reasoner
- **输入 Prompt (示例)：**使用HTML, CSS, JavaScript设计一款移动端的瑞士卷小游戏, 内容是很多瑞士卷和其他食物, 厨具等等, 物品从上往下掉, 用户点击到6个瑞士卷即可获胜, 一共两关, 第一关超简单, 第二关超难, 速度快而且掉落的内容多而繁杂

实践评价

- 与其他MCP客户端一样，Cline可以正常调用MCP服务，值得一提的是Cline还会根据输出结果的不同渲染出不同的图形，例如饼图、柱状图、甘特图等，结果一目了然
- Cline在MCP的调用上与其他客户端实现逻辑上不同，Cline是将MCP所有的选择甚至MCP的一些demo都会一并发送给大模型，待大模型的输出后再去调用MCP服务





5.3 Cline实测：效果评估

- **综合优点：**
  - [+] 工程化代码能力：可以创建、维护复杂的、大型的项目工程代码，最终生成的效果还是取决于所选择的大模型的能力
  - [+] 根据模型输出的数据不同，会有针对性的图形渲染，展现了Cline输出的强大，以至于可以不局限于当做代码生成器来使用

- **潜在缺点：**
  - [-] 消耗Token量较大，部分服务商可能会限制模型调用频率进而导致输出超时
  - [-] MCP市场目前覆盖面不全，部分好用的MCP Server不在市场中，还是需要自己手动配置

5.3 Cline实测：易用性、成本效益、集成性

- **易用性与交互体验：**
  - 相比较Cursor/Trae等开箱即用的工具而言，Cline还需要配置一些模型秘钥等操作，对小白略微不友好
  - Cline中提供的MCP市场对小白较友好，可以点击Install，进行一键安装
  - 交互逻辑还是相对较好的，划分的Plan与Act模式功能也较为清晰

- **定价与免费额度**：
  - Cline插件免费，模型服务商收费，当使用频率较高时，尤其是像Claude-3.7-Sonnet等专业模型可能会花费较多

- **集成性与生态兼容**：
  - 集成市面上大部分的知名厂商，填入Key即可使用，同时也提供了兼容OpenAI模型的自定义厂商配置

5.3 Cline实测总结

- **核心优势：**
  - 强大的代码理解与生成能力：在处理大型代码库时表现出色，能深入理解复杂项目结构，准确分析代码结构和上下文，生成高质量的代码，且对多仓库项目支持良好
  - 高度可定制：支持选择多种底层 AI 模型，如 Google Gemini、DeepSeek Chat 等，并可混合使用，还能自定义提示词模板和工作流，满足不同开发者对模型性能和功能的个性化需求
  - 丰富的功能支持：具有代码生成、终端命令执行、Web 开发辅助、自动化无头浏览器测试、主动修复错误等多种功能，还可通过 Model Context Protocol (MCP) 连接外部服务
- **主要限制：**
  - 资源消耗较高：其代码上下文编排虽然高效，但资源消耗也较高，可能会导致 token 使用成本增加，进而影响工作流。
  - 可能存在速率限制：使用预置模型 API 服务时会有频率限制，如果频繁调用 API，可能会触发速率限制，从而减慢工作流
- **适用场景：**
  - 代码生成与优化/Web 开发辅助/项目理解与任务执行/自动化无头浏览器测试/生成专业图

5.3 AI编程工具深度测评 - GitHub Copilot

GitHub 与 OpenAI 联合推出的 AI 编程伙伴 ，提供智能代码建议与生成等功能，支持主流编程语言及 IDE

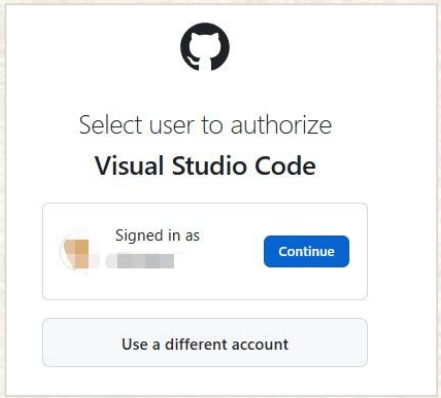
特别说明：以下评测基于GitHub Copilot 【1.303.0】 版本

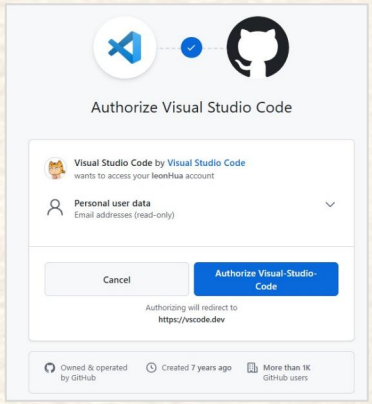
功能	说明
核心功能	代码补全 (从单行到整个函数/类)、代码生成 (根据注释或上下文)、代码解释、代码重构建议、调试辅助、单元测试生成 (辅助)、智能问答 (通过 Copilot Chat 实现)
支持的 IDE	Visual Studio Code, Visual Studio, JetBrains 全家桶、Neovim、Xcode、Azure Data Studio、Eclipse
支持的语言	支持几乎所有主流编程语言，如 Go, Python, C++, C, C#, Java, PHP, Rust, HTML, TypeScript, JavaScript, CSS, Kotlin, Objective-C, Swift, Ruby, Shell, SQL, R 等
是否收费	个人版（免费版和付费版）、企业版（付费版）
支持模型	免费版支持：Claude 3.5 Sonnet、Gemini2.0 Flash、GPT-4.1(Preview)、GPT-40、o3-mini 付费可支持Claude 3.5 Sonnet、o1、GPT-4.5等更好的模型
社区情况	提供官方文档、GitHub Community 论坛进行讨论和获取帮助、用户可以直接在 IDE 插件或网站上提供反馈。作为 GitHub 的产品，拥有庞大的开发者社区基础。

5.3 GitHub Copilot部分实测效果

实践评价

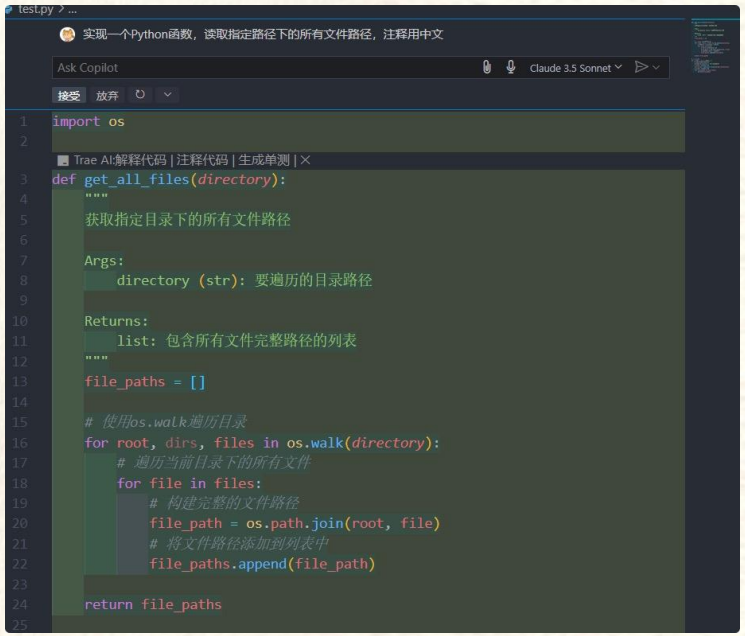
- GitHub Copilot生成内容视选择模型而定，依托于Claude、Gemini等系列模型，代码生成质量不错，生成相应速度较快
- 可提供类似Cursor IDE的体验效果（局部修改、一键替换）
- 登录需要使用Github账号进行登录，需要提前准备一个Github账号





授权认证

内嵌聊天



对话聊天



5.3 GitHub Copilot实测总结

评测维度	评测总结
核心功能	功能全面，覆盖编程开发的各个环节。附加上下文环境最为丰富，除代码段、文件外还支持终端、VS Cose API、当前编辑器等上下文环境
效果质量	生成内容视选择模型而定，依托于Claude、Gemini等系列模型，代码生成质量不错，生成相应速度较快
易用性	使用方式丰富灵活，可聊天式对话，也可在文件中内嵌聊天。内嵌聊天支持局部修改并覆盖，效率很高 支持语音聊天接入（需要安装相应扩展），方便性进一步提升 需要使用GitHub账号登录认证，对外部网络有特殊要求，有一定的门槛。文档、模型反馈默认英文，不如中文便捷
成本效益	免费版有使用限制：每月50次请求，模型限制版本 付费版：个人付费每月10$、39$，企业版每月19$、39$。付费后不限制请求次数，模型可使用最新版本
社区情况	官方提供视频教程（YouTube），推出后关注和使用群体量大，相关教程资源非常丰富 自推出以来，受到了广泛关注和使用，口碑整体较为积极，综合使用评价为同类工具产品中的领先者
使用建议	对代码质量要求非常高的场景优先选择，可使用Claude等代码效果更好的模型 对使用体验要求更高的场景优先选择，可提供类似Cursor IDE的体验效果（局部修改、一键替换） 若不想付费，可结合国产类插件使用，在低要求场景使用国产插件，高要求场景使用其免费功能

5.3 AI编程工具深度测评 - 通义灵码

基于阿里云通义代码大模型打造的智能编码助手，提供代码智能生成、研发智能问答、任务自主执行等能力

特别说明：以下评测基于通义灵码 【2.3.0】 版本

功能	说明
核心功能	基础功能：代码补全 、代码生成 、代码编辑、代码解释 、代码注释生成、单元测试生成、智能修复、智能问答 AI程序员：多文件代码修改（Multi-file Edit）和工具使用（Tool-use）的能力，可以与开发者协同完成编码任务等
支持的 IDE	Visual Studio Code、Visual Studio、JetBrainsIDEs 等主流编程工具
支持的语言	支持Java、Python、Go、JavaScript、TypeScript、C/C++、C#等主流语言
支持的模型	qwen2.5、qwen2.5-max(beta)、qwq-plus、deepseek-r1、deepseek-v3
是否收费	个人基础版、个人专业版、企业标准版和企业专属版，个人版基础版免费，个人专业版限时免费
社区情况	官方提供详细的教程和文档，提供持续的公开课与其它活动，活跃度与知名度都较高

5.3 通义灵码部分实测效果

- **测试任务描述：** 分别让其输出前端页面，python程序代码等，进行高频度使用，测试其结果

评价

- 通过一段时间测试，通义灵码的功能还是比较全面的，AI程序员也支持跨文件创建、修改等，类似Cursor的Composer agent模式、Trae的Builder模式
- 生成不稳定，经常性的出现如有图所示的调用异常错误

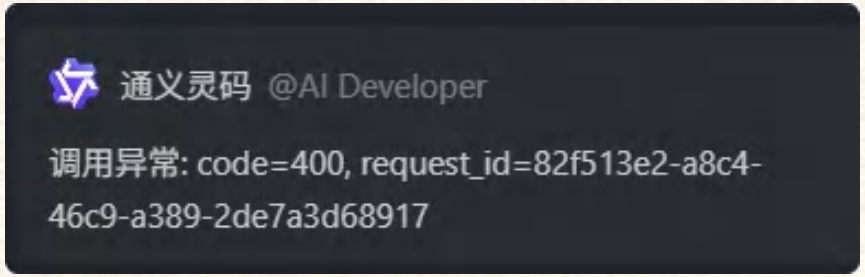
AI程序员交互



普通交互问答



调用出错



5.3 通义灵码实测总结

评测维度	评测总结
核心功能	功能全面，覆盖编程开发的主要环节。AI程序员类似Cursor的Composer agent模式、Trae的Builder模式 特别是支持图片识别，可上传图片进行交互
效果质量	生成不稳定，调用模型相对其他插件而言出现错误几率高； AI程序员模式工程化智能操作没问题，不过具体代码生成后二次修改效果一般，存在多次修改的问题
易用性	插件直接安装，界面清晰简洁，功能明了，上下文支持选择代码、文件、工作空间、终端等，比较好用。使用手机登录成功后可直接使用，没有特殊的网络要求。 自定义设置更加全面，比如代码补全长度、候选最大数量、行间代码建议、中英文切换等
成本效益	个人版基础版免费；个人专业版限时免费（正常59/人/月；企业标准版79/人/月，10人起购；企业专属版159/人/月，100人起购；企业版增加知识库问答、企业管理的功能
社区情况	国内知名度较高，背靠阿里接入企业占有量大，能有更多的案例经验迭代。官方的公开课频度比较活跃。在高校校园有更大力度推广
使用建议	想要使用阿里系千问模型可以使用，支持千问较新的旗舰模型 需要进行多模态图片交互的推荐使用，其他插件暂时不支持 想要使用类似agent工程化模式的推荐使用，AI程序员可以有更好的执行能力，简化操作提效 需要注意插件使用的时候出现错误的概率相较其他插件高，可能需多次交互

5.3 AI编程工具深度测评 - CodeGeeX

- 清华和智谱AI联合推出的免费AI编程工具，对标GitHub Copilot，支持多种编程语言，旨在提升开发者效率
- 特别说明：以下评测基于CodeGeeX 【2.22.3】 版本

功能	说明
核心功能	代码补全 、代码生成 、代码注释生成、智能问答、代码解释、代码修复、单元测试、交互式编程等 提供不同语言之间的代码转换，提供常用的编程工具箱
支持的 IDE	Visual Studio Code、Visual Studio、JetBrains全家桶、HBuilderX、AndroidStudio
支持的语言	支持Python、Java、C++/C、JavaScript、Go等多种主流语言
支持的模型	CodeGeeX Lite、CodeGeeX Pro
是否收费	个人用户完全免费，面向企业提供CodeGeeX私有化部署服务
社区情况	官方提供有详细的使用手册文档，有反馈渠道及用户交流群

5.3 CodeGeex实测效果

评价

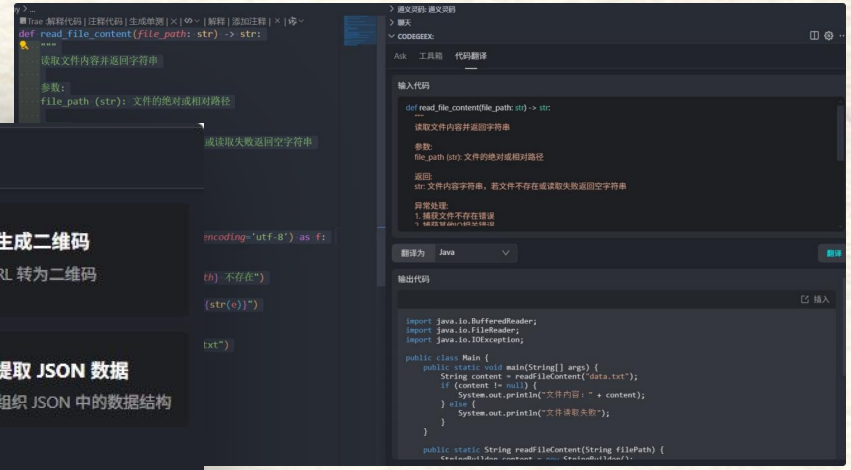
- 功能比较全面，除了大部分AI编程插件所支持的代码生成与修改的功能外，CodeGeex还支持代码翻译、工具箱、项目地图等功能
- 生成长文本内容的时候效果不好，容易偷懒，不给出全部完整内容



项目地图



工具箱



代码翻译

5.3 CodeGeeX实测总结

评测维度	体验总结
核心功能	基础功能与其他插件没有明显差异 提供了不同语言之间的快速转换，常见的编程工具箱，一定场景下快速提效 支持联网搜索功能，支持索引当前代码库，构建索引后对项目代码上下文理解更好 支持内嵌聊天，支持线上模式（官方模型）和本地模型切换 支持项目地图，自动构建项目结构UML图，帮助快速理清项目结构 MCP（内测中，未上线）
效果质量	生成长文本内容的时候效果不好，容易偷懒，不给出全部完整内容 内嵌聊天比较不稳定，出错概率较大 项目地图效果还不错，能起到理解项目的辅助作用
易用性	插件直接安装，界面清晰，功能划分明确。使用手机登录成功后可直接使用，没有特殊的网络要求 和其他插件使用类似，且提供功能向导，更易上手
成本效益	个人完全免费
社区情况	官方提供有详细的使用手册文档，有反馈渠道及用户交流群
使用建议	有需要使用其特色功能的场景可以使用，比如：不同语言转换、帮助理解项目的项目地图 有需要同时进行联网搜索的场景可以使用 隐私安全性要求较高，需要使用本地自有模型的推荐使用 需要接入MCP实现更多能力扩展的推荐使用，静待上线 普通常规的应用场景下不是特别推荐，尤其是需要长文本内容输出的时候，输出效果可能不理想

MGX简介与核心技术

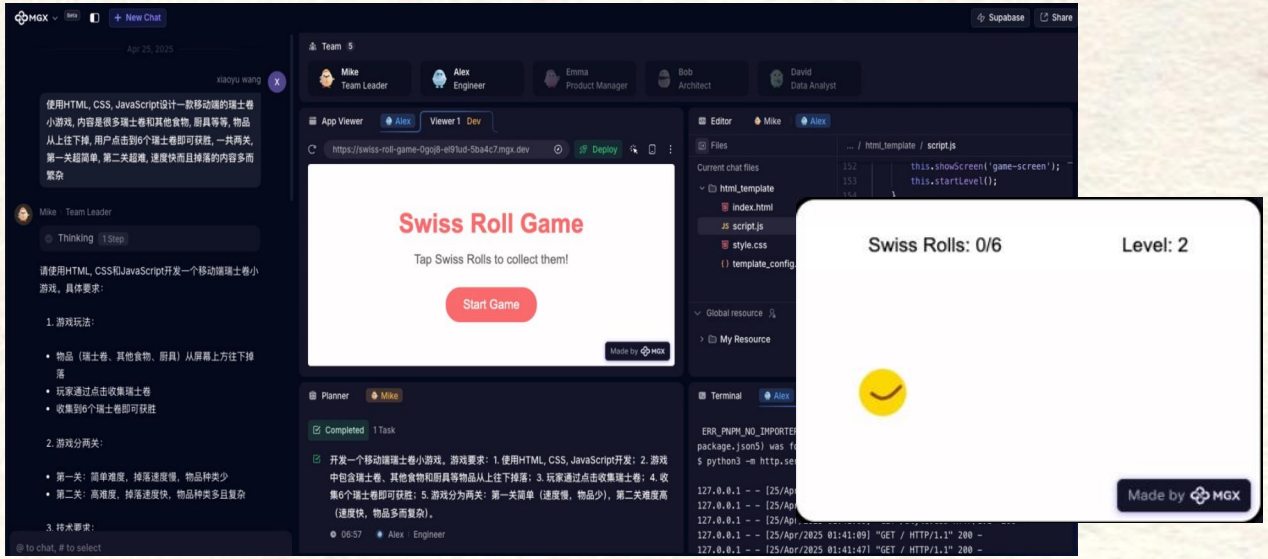
- **开发背景**：MGX（MetaGPT X）是由 MetaGPT 团队开发的多智能体协作平台，旨在通过模拟真实软件开发团队的协作流程，实现从需求分析到部署的全流程自动化
- **核心定位**：作为首个“AI 软件公司”式开发平台，MGX 通过自然语言交互和多角色 AI 代理协作，覆盖需求定义、架构设计、编码、测试及部署全流程，目标是替代传统开发团队并实现“开发民主化”
- **核心技术**：
  - 多智能体协作：包含 5 个专业 AI 角色（团队领导、产品经理、架构师、工程师、数据分析师），各司其职协作开发
  - 端到端开发：自动生成技术文档、架构图、代码及部署方案，支持主流框架如Vue3+TypeScript
  - 提供 OpenAI 兼容 API，可对接 LangChain 等第三方工具
- 主要产品形态: 网页端【链接：https://mgx.dev/】

5.4 MGX实测：核心功能 - 网页代码生成

- **测试任务描述：** 开启Agent模式，模型选择Claude-3.5-Sonnet-V2，在MGX的输入框中输入一段网页开发需求，等待MGX的输出和部署
- **输入 Prompt (示例)：**使用HTML, CSS, JavaScript设计一款移动端的瑞士卷小游戏, 内容是很多瑞士卷和其他食物, 厨具等等, 物品从上往下掉, 用户点击到6个瑞士卷即可获胜, 一共两关, 第一关超简单, 第二关超难, 速度快而且掉落的内容多而繁杂

评价

- 通过一段提示词输入后，MGX会自动的进行梳理需求、开发、测试、部署，最后提供预览界面方便用户直接可预览
- 整体运行的效果还不错，美中不足的是前端界面的美观度较一般



MGX整体总结

- **优点**
  - 👍 开发效率革命：全流程自动化可将常规项目开发时间从数周缩短至数小时
  - 👍 非技术友好：提供模板库与可视化界面，非技术人员可快速生成电商平台、数据看板等应用
  - 👍 部署方便：仅通过提示词即可实现需求梳理、代码开发、部署上线等需求，提升了很大的效率
  - 👍 成本优势：免费版支持 750K额度/日，2.5M/月基础开发，付费版有20$、70$、200$、500$档位
- **缺点**
  - 👎 对高度定制化需求（如特殊算法）支持有限，生成代码需人工复核

5.4 AI工具深度测评 - heyBossAI

heyBossAI简介与核心技术

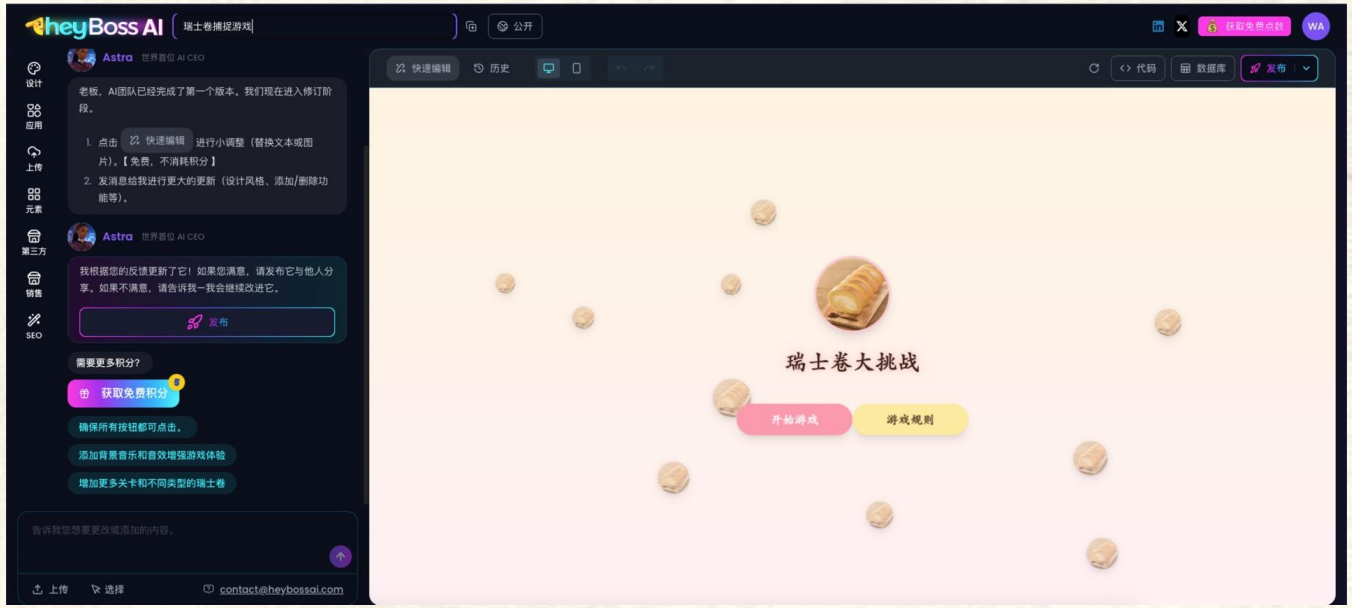
- **开发背景**：Heyboss是Heeyo的新产品，由创始人曲晓音推出，旨在赋能99%没有编程基础的人
- **核心定位**：Heyboss是脱胎于Heeyo，是一款新产品，定位为AI工程师，不限于儿童，所有人都可以使用
- **核心技术**：
  - 端到端全栈自动化开发：AI 系统覆盖从界面设计、前端/后端编码到服务器部署的全流程，用户仅需描述需求即可获得可直接上线的产品，显著降低开发门槛
  - 实时交互优化系统：用户可通过聊天界面提供反馈，AI 持续迭代改进设计
  - 可视化修改：可以在线可视化修改图标、颜色、文字样式等，不消耗积分
- 主要产品形态: 网页端【链接：https://heybossai.com/】

5.4 heyBossAI实测：核心功能 - 网页代码生成

- **测试任务描述：** 在heyBossAI的输入框中输入一段网页开发需求，等待heyBossAI的输出和部署
- **输入 Prompt (示例)：**使用HTML, CSS, JavaScript设计一款移动端的瑞士卷小游戏, 内容是很多瑞士卷和其他食物, 厨具等等, 物品从上往下掉, 用户点击到6个瑞士卷即可获胜, 一共两关, 第一关超简单, 第二关超难, 速度快而且掉落的内容多而繁杂

评价

- 通过一段提示词输入后，heyBossAI生成一个直播间、多为员工随机发言出点子，相当于对提示词的完善
- 生成首页还可以，但是点击开始游戏后无法进行游戏、还需要进行反复迭代



5.4 AI工具深度测评 - heyBossAI

heyBossAI整体总结

- **优点**
  - 👍 开发效率革命：全流程自动化可将常规项目开发时间从数周缩短至9分钟
  - 👍 非技术友好：提供模板库与可视化界面，非技术人员可快速生成电商平台、数据看板等应用
  - 👍 交互体验好：支持可视化修改界面元素，方便非技术人员操作
  - 👍 部署方便：仅通过提示词即可实现需求梳理、代码开发、部署上线等需求，提升了很大的效率
- **缺点**
  - 👎 对高度定制化需求（如特殊算法）支持有限，生成代码需人工复核
  - 👎 除了生成所消耗的积分外，源码需要额外付费下载，增加了额外的预算
  - 👎 固定模型，不可选，且不公开。对于代码生成效果不是很好，不如MGX的生成效果

5.5 AI编程工具-总结与思考

核心结论是：AI编程工具是革命性的工具，但绝非要取代人类工程师。它们是效率的倍增器，能有效加速开发、减少冗余，让开发者摆脱低价值的重复劳动。

但我们必须认识到，AI的强项在于提升效率和自动化已知模式。而软件工程中至关重要的人类判断力、创新思维、对复杂需求的精准把握以及系统架构的战略设计，这些是AI目前无法企及的，也是人类工程师的核心竞争力。

因此，与其担忧被替代，不如积极拥抱AI，将其视为提升自身能力的杠杆，让人类的创造力和智慧在AI的辅助下达到新的高度。



北京大学
PEKING UNIVERSITY

01 文本生成与处理类AI工具测评

02 图像生成与编辑类AI工具测评

03 音视频生成与处理类AI工具测评

04 代码生成与辅助类AI工具测评

05 大模型管理与应用类AI工具测评

6.1 大模型管理与应用类工具深度测评总述

- **测评目的：**
  - 全面评估不同工具在大模型管理与应用景下的表现，为用户提供选型依据
- **测评维度：**
  - 核心功能与效果 (重点)
  - 易用性和体验度（重点）
  - 定价与性价比
  - 不同工具集成和编程语言能力
  - 社区活跃度情况
- **测评工具选型：**
  - CherryStudio
  - LMStudio
  - Chatbox

6.1 大模型管理与应用类工具深度测评总述

工具	类型	版本	核心功能	集成性 (支持模型)	易用性	成本效益	适用场景
CherryStudio	客户端	1.2.5	支持基础对话功能、集成大多是知名的模型服务商、支持AI绘画、知识库、MCP服务等高级功能。 **侧重点**: **基于大模型的应用开发与集成。**它也是“应用”层面，但更偏向于**开发者工具**，用于创造性地利用模型构建新的应用或解决方案，而不仅仅是直接的聊天交互。可能也包含对模型连接的管理。	市面上绝大部分模型以及私有化部署的服务.如Ollama	安装后，开箱即用 全能、模型支持的较多、功能全面，尤其是AI绘画、MCP服务、知识库等	开源免费	无论是专业开发人员维护已有项目还是小白想要快速验证想法均适用
LMStudio	客户端	0.3.14	支持本地模型一键快速部署、对Apple Silicon、AMD有深度优化提升模型生成速度、有普通用户、高级用户、开发者快捷切换、带有基础模型对话功能。 **侧重点: 本地模型的管理（下载、运行、配置）和基础应用（聊天交互）**	GGUF/MLX格式模型	安装后，开箱即用 本地快速部署，自带可视化界面	开源免费	使用未接触过专业IDE的小白或预算有限想要实现工程化项目的建设和维护的人员
Chatbox	客户端	1.11.8	集成大多数知名的模型服务商、支持图文对话、文档解析等基础对话功能、支持接入Ollama等本地部署的模型 **侧重点**: **统一的模型交互/应用界面。**更侧重于“应用”层面，提供一个便捷的聚合入口来使用不同的模型，而不是模型本身的“管理”（它通常不负责下载和运行本地模型，而是连接到像 LM Studio 或 Ollama 这样的后端）。	市面上大部分的模型服务商以及私有化部署的服务.如Ollama	安装后，开箱即用 轻量级，模型支持的较多	开源、基础功能免费、使用自家模型付费，3个付费版本Lite、Pro、Pro+	功能较多，使用范围较广，可以对已有项目维护、也可以对新项目开发，还可以当做一个AI小助手对话使用，核心点在于所选择的模型的能力

CherryStudio简介与核心技术

- **开发背景**：上海千彗科技有限公司打造的全能 AI 助手平台
- **核心定位**：集多模型对话、知识库管理、AI 绘画、翻译等功能于一体
- **核心技术**：
  - **基础对话功能**(一问多答/自动分组/对话导出/高度自定义参数/助手市场/多种格式渲染)
  - **多种特色功能集成**(AI 绘画/AI 小程序/翻译功能/文件管理/MCP服务/全局搜索)
  - **多服务商统一管理机制**(服务商模型聚合/模型自动获取/多秘钥轮询/自定义服务商)
  - **高度自定义界面和布局**(自定义 CSS/自定义对话布局/自定义头像/自定义侧边栏菜单)
  - **本地知识库系统**(多种格式支持/多种数据源支持/知识库导出/支持搜索检查)
  - **特色聚焦功能**(快捷问答/快捷翻译/内容总结/解释说明)
  - **数据保障**(多种备份方案/数据安全)
- 主要产品形态: 桌面客户端【链接：https://www.cherry-ai.com/】

CherryStudio整体总结

- **优点**
  - 👍 **功能全面**：作为一个大语言模型客户端，功能丰富全面，支持基础对话、MCP服务、知识库、AI绘画等强大功能
  - 👍 **基础对话**：界面设计美观、简单易操作，小白友好，回复效果取决于模型能力
  - 👍 **MCP服务**：配置支持JOSN格式粘贴及可视化界面配置，对话中可以针对性开启个别服务
  - 👍 **服务商模型聚合**：支持上百款模型，支持模型自动获取、多秘钥轮询、自定义服务商等功能
- **缺点**
  - 👎 **AI绘画功能**：可作为日常生图使用，但缺乏一些专业的配置选项，不适用于专业人士使用
  - 👎 **平台覆盖不全**：目前仅支持Windows、Mac和Linux平台，尚未推出iOS和Android手机端应用

6.2 CherryStudio实测：核心功能 - 一问多答

- **测试任务描述：** 输入一个问题（例如：“8.11和8.7谁更大?”），@多个模型（例如：@moonshot-v1-auto @qwen-turbo @deepseek-ai/DeepSeek-V3），对比多模型答案
- **输入 Prompt (示例)：**8.11和8.7谁更大?
- **CherryStudio输出结果：**

评价

- 可以一次选中多个模型，同时对多个模型问同一问题，多个模型会同时输出
- 支持多模型输出排版：标签、横向、纵向、卡片布局
- 适用于测试多模型能力或工作中期待多模型输出并对结果进行选择性的使用的场景下使用



6.2 CherryStudio实测：核心功能 - MCP

- **测试任务描述**：配置高德地图MCP Server，在对话中开启并提问
- **输入 Prompt (示例)**：我现在在北京，五一的时候想要去云南玩5天，请帮我制作一个详细的旅游攻略，需要给出具体的路线，与时间安排，还需要考虑天气情况
- **选择模型：deepseek-chat**
- **CherryStudio输出结果：**





评价

- MCP配置界面简洁、清晰，易配置
- 模型回复时，MCP的调用情况一览无余，可点击查看MCP的详细调用情况
- 回复的效果准确率取决于MCP服务的能力和模型的能力

6.2 CherryStudio实测：核心功能 - AI绘画

- **测试任务描述：** 输入一段提示词及反向提示词，选择一个文生图模型，设置各个参数，等待输出
- **输入 Prompt (示例)：**一只胖胖的黄色的猫，坐着，望向海边
- **输入反向提示词：**低质量, 模糊
- **选择模型：**Stable Diffusion 3.5 Large
- **各参数设置：**图片尺寸1比1、生成数量3、推理步数25、引导比例5.0、开启提示词增强

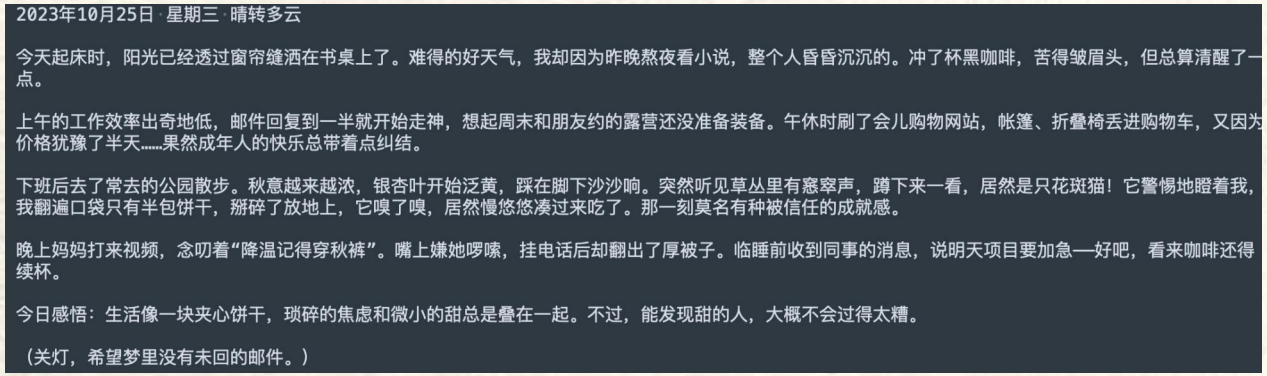


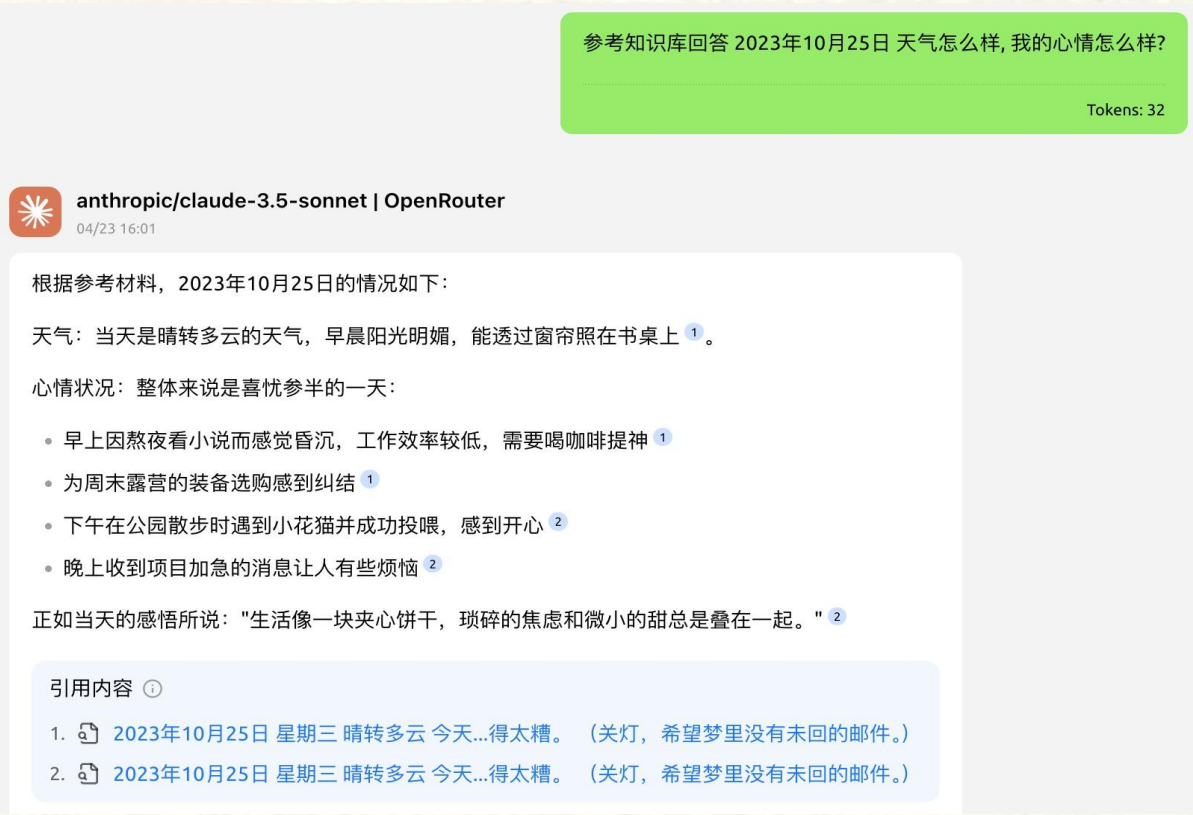
评价

- 生图的最终效果取决于所选择模型的能力以及各个参数及提示词的设置，CherryStudio仅起到了一个请求、响应、渲染的作用
- CherryStudio中的各参数项设置满足大部分人群的日常生图使用，但是相比较专业的生图工具如 Midjourney/Stable Diffusion 参数项上还是逊色不少

6.2 CherryStudio实测：核心功能 - 知识库

- **测试任务描述：** 配置个人知识库，在对话中启用知识库并针对知识库的内容进行提问
- **输入 Prompt (示例)：**参考知识库回答 2023年10月25日 天气怎么样, 我的心情怎么样?
- **选择模型：claude-3.5-sonnet**





评价

- 如果在提问时开启了知识库的能力时，可以看到在回复用户问题的时候，答案是参考了知识库的内容的，并给出了具体的引用
- 配置及启用知识库比较简单，并且还支持随时修改
- 回复的效果准确率取决于模型的能力

6.2 CherryStudio实测：效果评估

- **综合优点：**
  - [+] 功能全面：在目前的大语言模型集成桌面客户端中，CherryStudio在功能的丰富度上属于第一梯队
  - [+] 强大的文件处理：兼容多种文件格式，包括文本、图片、Office、PDF等，还支持WebDAV文件管理与数据备份，以及Mermaid图表可视化和代码高亮，方便从事编程、数据分析等工作的用户
  - [+] 快速响应：每个对话交互都能在几秒钟内快速提供AI响应，有效提高工作效率
  - [+] 可自定义性强：用户可根据自身需求自定义创建专属助手，进一步提升其适用性和个性化程度
- **潜在缺点：**
  - [-] AI绘画：相比较专业的AI绘画软件，目前可控制参数较少，对于生图的最终效果与期待上来说，可能会差一些
  - [-] 平台覆盖不全：目前仅支持Windows、Mac和Linux平台，尚未推出iOS和Android手机端应用，这使得用户在移动端的使用受到限制，无法随时随地便捷地使用其全部功能

6.2 CherryStudio实测：易用性、成本效益、集成性

- **易用性与交互体验：**
  - 配置界面：无论是集成模型厂商，还是MCP等配置都很简单易操作
  - 交互流程：输入输出流畅，支持多轮对话记忆，有历史记录、支持极多的参数自定义等便捷功能
  - 交互体验：基础功能与其他桌面客户端无区别，主要是支持MCP及MCP响应结果的渲染
- **定价与免费额度**：
  - 免费政策: 客户端免费使用、模型厂商收费，按需选择模型厂商
- **模型集成能力：**
  - 服务商模型聚合：支持 OpenAI、Gemini、Anthropic、Azure 等主流服务商的模型统一调用
  - 模型自动获取：个别平台一键获取完整模型列表，无需手动配置
  - 多秘钥轮询：支持多个 API 秘钥轮换使用，避免速率限制问题
  - 自定义服务商：支持符合 OpenAI、Gemini 、Anthropic 等规范的三方服务商接入，兼容性强

6.2 CherryStudio实测总结

- **核心优势：**
  - 小白友好：Cherry Studio 致力于降低技术门槛，零基础用户也能快速上手，让用户专注于工作、学习或者创作
  - 持续迭代：项目团队积极响应用户反馈，持续优化功能，确保项目健康发展
  - 开源与扩展性：支持用户通过开源代码进行定制和扩展，满足个性化需求
- **主要限制：**
  - 平台覆盖不全：目前Cherry Studio仅支持Windows、Mac和Linux平台，尚未推出iOS和Android手机端应用，这使得用户在移动端的使用受到限制，无法随时随地便捷地使用其全部功能
- **适用场景：**
  - 知识管理与查询：通过本地知识库功能，快速构建和查询专属知识库，适用于研究、教育等领域
  - 多模型对话与创作：支持多模型同时对话，帮助用户快速获取信息或生成内容
  - 翻译与办公自动化：内置翻译助手和文件处理功能，适合需要跨语言交流或文档处理的用户
  - AI 绘画与设计：通过自然语言描述生成图像，满足创意设计需求

6.2 AI工具深度测评 - LMStudio

LMStudio简介与核心技术

- **开发背景**：由开源社区驱动的本地化大语言模型工具，支持离线运行主流开源模型
- **核心定位**：通过桌面客户端实现模型本地部署、性能测试与多场景应用，兼顾隐私与定制化需求
- **核心技术**：
  - 本地模型管理：支持从Hugging Face直接下载GGUF/MLX格式模型，无需云端依赖
  - 多硬件加速：优化Apple Silicon（M1/M2/M3）、AMD/NVIDIA GPU性能，支持多GPU负载分配
  - OpenAI API兼容：内置本地服务器（localhost:1234），可对接第三方工具（如LangChain）
  - 资源监控：实时追踪CPU/GPU使用率、内存占用，支持生成性能报告
  - 工具调用（Tool Use）：通过函数定义实现代码执行、API调用等扩展功能
- 主要产品形态: 桌面客户端【链接：https://lmstudio.ai/】

6.2 AI工具深度测评 - LMStudio

LMStudio整体总结

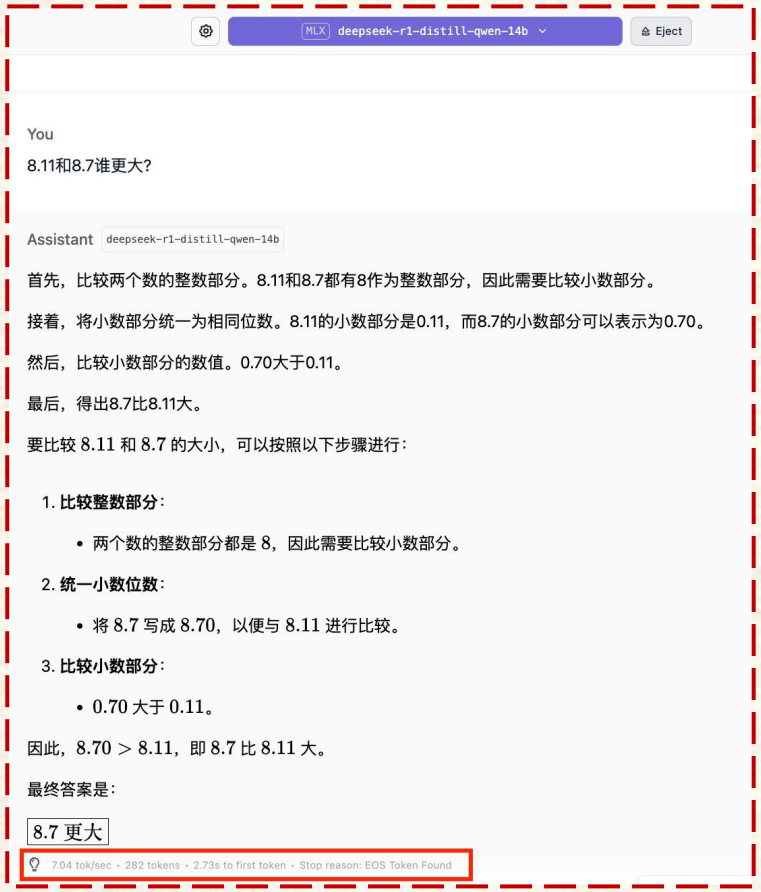
- **优点**
  - 👍 功能全面：作为本地化大语言模型工具，支持模型管理、代码测试、资源监控、工具调用等开发者核心需求，集成OpenAI兼容API与多硬件加速能力
  - 👍 隐私与性能：完全离线运行，数据零外传；针对Apple Silicon/NVIDIA/AMD深度优化，GPU利用率峰值达98%，响应延迟低至0.78秒
  - 👍 开发者友好：提供CLI工具、SDK接口及可视化性能报告，支持自定义模型集成与多GPU负载分配
- **缺点**
  - 👎 硬件门槛高：运行70B参数模型需至少32GB内存+24GB显存，低配设备体验受限
  - 👎 移动端缺失：仅支持Windows/macOS/Linux桌面端，无iOS/Android应用，移动场景覆盖不足
  - 👎 学习成本略高：高级功能（如多GPU分配、自定义工具链）需技术背景，小白用户上手难度大

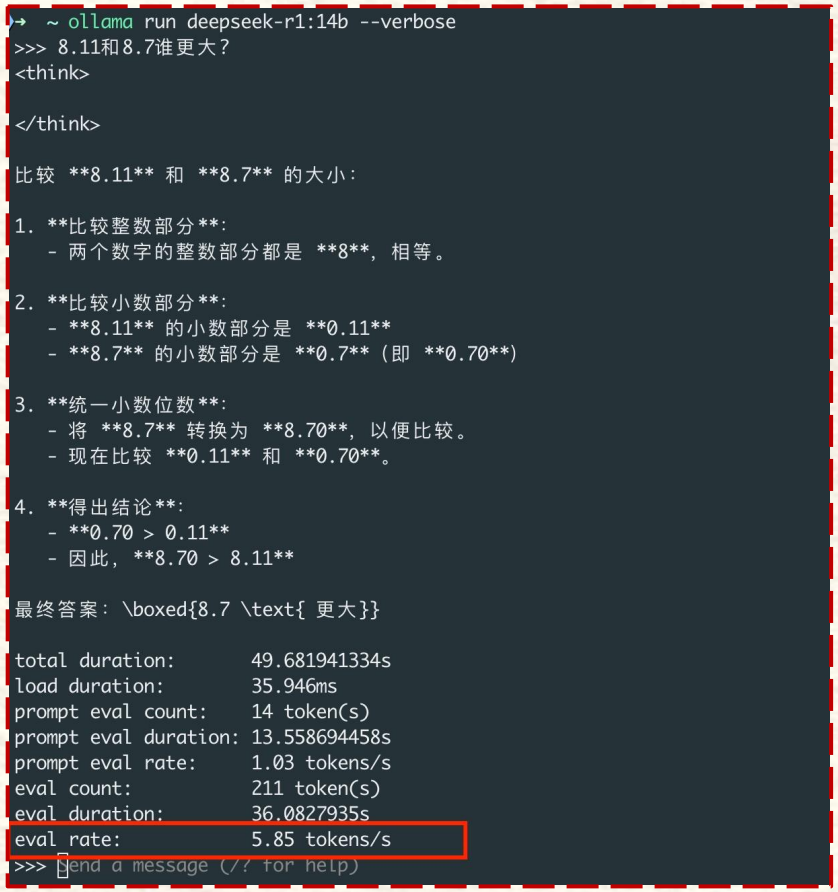
6.2 LMStudio实测：核心功能 - 本地部署模型

- **测试任务描述：** 加载**deepseek-r1:14b的4位量化模型**模型，输入提示词进行提问，查看输出token速率对比ollama部署的相同模型输出速率
- **输入 Prompt (示例)：**8.11和8.7谁更大? LMStudio与Ollama的输出如下：

实践评价

- 在同样配置的条件下(13-inch, M1, 2020 Apple M1)可以发现LMStudio的输出速率确实要比Ollama部署的同模型快一些
- LMStudio自带聊天界面，可以做到开箱即用，无需再额外搭配其他可视化客户端





6.2 LMStudio实测：效果评估

- **综合优点：**
  - [+] 功能全面：支持模型测试、资源监控、工具调用等开发者导向功能
  - [+] 隐私安全：数据完全离线，符合金融/医疗行业合规需求
  - [+] 硬件兼容性：针对Apple Silicon/NVIDIA/AMD深度优化，峰值GPU利用率达98%
  - [+] 可自定义性强：提供模型搜索及对话的可视化界面，本地部署及测试模型方便快捷
- **潜在缺点：**
  - [-] 平台覆盖不全：目前仅支持Windows、Mac和Linux平台，尚未推出iOS和Android手机端应用，这使得用户在移动端的使用受到限制，无法随时随地便捷地使用其全部功能
  - [-] 输出质量取决于所部署的模型，而高级模型对硬件配置要求较高

6.2 LMStudio实测：易用性、成本效益、集成性

- **易用性与交互体验：**
  - 模型管理：支持从Hugging Face直接下载GGUF/MLX模型，界面集成模型搜索与版本选择功能，但高级参数（如量化等级）需手动调整
  - 多轮对话：支持历史记录回溯与上下文记忆，但长对话（>4096 tokens）可能触发显存溢出，需手动清理缓存
  - CLI支持：通过lms命令行工具实现模型加载、性能监控，但文档示例较少，需依赖社区资源
- **定价与免费额度**：
  - 完全免费，无订阅费用，本地运行节省云服务费用，但电费与硬件折旧需长期考量
- **模型集成能力：**
  - 主流模型支持：聚合Llama、Mistral、Qwen等50+开源模型，自动同步Hugging Face更新
  - API兼容性：内置OpenAI格式本地端点（http://localhost:1234/v1），可直接替换ChatGPT接口
  - 自定义服务商：通过修改config.yaml支持私有化模型服务接入，需符合OpenAPI规范

6.2 LMStudio实测总结

- **核心优势：**
  - 隐私与离线优先：无需网络连接，适合处理敏感数据
  - 深度硬件优化：针对Apple Silicon/AMD/NVIDIA多平台专项调优
  - 开发者友好：提供CLI工具、性能分析SDK，支持自定义模型集成
- **主要限制：**
  - 硬件门槛高：运行70B参数模型需至少32GB内存+24GB显存
  - 移动端缺失：仅支持桌面端（Windows/macOS/Linux），无移动应用
  - 学习曲线陡峭：高级功能（如多GPU分配、自定义工具链）需技术背景
- **适用场景：**
  - 隐私敏感领域：法律文档分析、医疗数据查询等需完全离线的场景
  - 模型研发测试：开发者需本地评估模型性能（如代码生成、数学推理）

6.2 AI工具深度测评 - Chatbox

Chatbox简介与核心技术

- **开发背景**：开源驱动：由开发者 Bin Huang 创建并维护的开源项目，GitHub 社区持续贡献插件与功能扩展
- **核心定位**：通过极简界面实现模型快速测试、代码片段生成与多模态交互（文生图/图表），兼顾隐私与易用性
- **核心技术**：
  - 云端接入：支持 OpenAI、Claude、Gemini 等主流 API，通过 SiliconFlow 扩展 DeepSeek-R1 等大模型
  - 本地集成：需依赖 Ollama 部署本地模型（如 Llama3），缺乏原生模型管理功能[1][3]。
  - 极简操作：预设常用功能快捷键（如 Ctrl+Enter 提交），支持对话记录批量导出（Markdown/JSON）
  - 移动端适配：iOS/Android 版本精简核心功能，实现跨设备同步
- 主要产品形态: 桌面客户端/移动端【链接：https://chatboxai.app/zh】

Chatbox整体总结

- **优点**
  - 👍 全平台覆盖（含移动端）：2GB 内存即可流畅运行，响应速度快
  - 👍 开源免费：社区驱动插件扩展，支持快速迭代与自定义
  - 👍 隐私灵活可控：本地模型通过 Ollama 实现完全离线运行，数据零外传；云端模型可自选服务商
  - 👍 提供 OpenAI 兼容 API，无缝对接 LangChain、AutoGPT 等生态工具
- **缺点**
  - 👎 功能深度不足：缺乏原生知识库管理（无向量数据库/RAG），仅支持单文档读取
  - 👎 本地模型体验受限：需额外配置 Ollama 等工具

内容目录

一、从AI模型到AI工具 P6

1. DeepSeek与大模型理论概述 P8
2. AI工具浪潮：为何现在关注AI工具？ P15
3. AI 工具如何赋能千行百业？ P19
4. 系统性认识AI工具 P20
5. AI工具分类 P21
6. AI工具测评的核心价值 P22

二、AI工具测评框架与方法论 P24

1. 12个测评维度定义 P27
2. 6个测评方法详解 P38
3. 根据自身需求选择测评维度权重 P45
4. 不同场景下AI工具测评维度的权重示例 P47
5. 不同场景下测评维度的权重设置参考 P54
6. 总结 P55

三、AI工具测评 P57

1. AI工具测评一览表 P59
2. 文本生成与处理类AI工具测评 P60
3. 图像生成与编辑类AI工具测评 P108
4. 音视频生成与处理类AI工具测评 P171
5. 代码生成与辅助类AI工具测评 P245
6. 大模型管理与应用类AI工具测评 P289

四、AI工具选型指南与未来展望 P310

1. 构建个性化AI工具箱选型决策框架 P312
2. 构建你的AI工具箱：个人AI工具战略选型流程 P313
3. 构建你的AI工具箱：团队AI工具战略选型考量 P314
4. 从工具到工作流，1+1 > 2 的组合力量 P315
5. AI工具的发展趋势 P316
6. 总结 P318

四、AI工具选型指南与未来展望

- 针对前文对各类主流AI工具的深度测评，本部分旨在进行归纳总结与前瞻展望。
- 首先，我们提供一个**战略性的选型决策框架**，区分个人与团队的不同考量维度（如协作、安全、成本），助您精准构建满足需求的AI工具箱。
- 接着，我们着重阐述**单一工具的局限性**，强调**工具组合与工作流构建**的必要性与巨大潜力，并通过初步场景示例展示，
- 最后，**分析并展望了AI工具未来发展的关键趋势**，聚焦多模态融合、平台级集成，特别是**AI Agent的崛起趋势**，同时也深刻剖析了在可靠性、安全性、伦理及成本等方面面临的关键挑战，旨在为读者在迅速发展的AI技术浪潮中，提供战略选型和前瞻思考的有效指引。

学习交流可加AI肖睿团队助理微信号（ABZ2829）

四、AI工具选型指南与未来展望

1. 构建个性化AI工具箱选型决策框架
2. 构建你的AI工具箱：个人AI工具战略选型流程
3. 构建你的AI工具箱：团队AI工具战略选型考量
4. 从工具到工作流，1+1 > 2 的组合力量
5. AI工具发展的趋势
6. 总结

1. 构建你的AI工具箱：个人/团队选型决策框架

01

构建你的AI工具箱：决策框架
需求为核

选择工具必须从具体需求和应用场景出发。

02

构建你的AI工具箱：决策框架
团队考量

工具选型决策必须考虑现有团队情况，如何将AI工具的利用更具“性价比”。



“最好的工具，是解决你问题的工具，而非功能最多的工具。”

2. 构筑智能优势：个人AI工具战略选型流程

个人选型决策流程步骤：
（从明确使命到精准匹配，打造你的高效AI助手能）

任务分解: 将“作战任务”分解为AI可执行的子任务。考虑哪些工具在这些具体的“战术动作”上表现突出？并**识别关键任务**。

- **建立优先级:** 结合“战术动作”的重要性和“约束条件”的权重，对候选工具进行打分或排序。**没有完美工具，只有权衡取舍。**
- **精选试用名单:** 选择得分最高的**2-3个**工具进入“实战演练”。
- **真实任务测试:** *必须步骤！* 使用你在第一步定义的真实“作战任务” 测试候选工具，并排比较输出结果的质量、速度、风格匹配度。**注意：匹配核心能力，即： 优先选择在关键任务上表现优异的工具。**

Step 01 定义核心“作战任务”

- **明确目标/场景**：需求细化
- **明确产出标准**：对结果的质量、风格、速度、格式有何具体要求
- **频率与场景**：高频使用还是偶尔需要？主要在哪个设备/环境下使用？

Step 02 拆解关键“战术动作”

Step 03 量化个人“约束条件”

- **预算红线**
  - 免费增值: 是否满足基本需求？限制是什么？
  - 订阅成本: 月度/年度预算上限？按量付费是否可接受？
  - 隐性成本: 学习时间投入、试错成本。
- **易用性与学习曲线、社区支持等**
  - 上手速度: 即插即用?，还是投入时间学习复杂功能/提示工程？
  - 界面友好度: 你的偏好是简洁直观还是功能全面？
- **集成与兼容性**
  - 需要与哪些现有软件/平台(浏览器、App、IDE)顺畅协作？API需求？
- **数据隐私与安全（重点考虑）**
  - 你对数据被用于训练的态度？
  - 工具的数据处理政策是否透明？是否有本地/离线处理选项？

Step 04 加权筛选与“靶向试用”

Step 05 决策、整合与动态优化

- **做出选择:** 基于试用结果，选择最适合当前需求的工具。
- **融入工作流:** 将选定工具整合进你的日常工作或创作流程。
- **持续关注与调整:** AI领域日新月异，定期（如每季度）重新评估你的工具箱，关注新工具和功能更新，保持开放心态进行调整优化。

3. 赋能组织进化：团队AI工具战略选型考量

团队选型需从组织战略、协作效率、成本控制、安全合规等多个维度进行系统评估，构筑团队协作、安全、可扩展的AI能力栈

基石：战略对齐与业务价值

01

- ✓ **目标一致**: 该工具/平台如何支撑团队或公司的核心目标（提升创新速度？降低运营成本？改善客户体验？增强研发能力？）
- ✓ **预期ROI:** 能否量化其潜在的投资回报？如何衡量成功？（例如：节省工时、提高内容生产率、降低错误率）
- ✓ **高层支持:** 是否获得关键决策者和预算负责人的认可？

核心：功能、性能与协作性

02

√集体需求满足：

- 能否覆盖团队多数成员的核心需求？在关键任务上的性能表现是否稳定可靠（尤其在高并发场景）？

√团队协作特性：

- 共享工作区/知识库: 是否支持团队成员共享提示、结果、模板、自定义指令库？
- 多用户管理: 权限设置是否灵活（管理员、编辑者、查看者）？
- 评论与反馈: 是否便于团队内部对AI生成内容进行协作编辑和反馈？
- 版本控制: 能否追踪修改历史？

关键：集成、兼容与可扩展性

03

√融入现有技术栈：

- 核心系统集成: 能否与现有系统平台无缝对接？
- API与文档: API是否健壮、灵活、文档清晰？

√未来扩展能力:

- 能否支持用户量增长？平台架构是否具备良好的可扩展性？
- 工具厂商的技术路线图是否与团队未来发展方向一致？

命脉：安全、合规与数据治理

04

- ✓ **数据处理透明度: 数据存储地**在哪？是否用于**模型训练**？是否提供**数据处理协议(DPA)**？
- ✓ **合规认证:** 是否满足**行业/地区法规**（如网安法, GDPR, CCPA）？是否有第三方安全认证（如SOC 2 Type II, ISO 27001）？
- ✓ **访问控制与审计: 单点登录(SSO)支持？是否有详细的操作审计日志**？
- ✓ **内容安全与防护:** 是否有机制防止生成不当或有害内容？能否检测和阻止敏感数据输入/输出？
- ✓ **模型私有化/本地部署:** 对于高度敏感数据，是否提供**私有化部署或本地模型**选项？

支撑：成本、支持与采纳管理

05

√总体拥有成本(TCO)：

- 显性成本: 订阅费（按席位/用量）、API调用费
- 隐性成本: 实施部署、集成开发、员工培训、持续维护

√供应商服务：

- 技术支持: 响应速度、服务水平协议(SLA)
- 培训资源: 文档、教程、最佳实践案例

√内部推广与采纳：

- 学习曲线: 对不同技能水平的团队成员是否友好？
- 变革管理: 如何进行内部培训和推广？识别早期采纳者和“超级用户”？制定清晰的使用规范和指南？

4. 从工具到工作流：1+1 > 2 的组合力量

单一工具的局限性

没有任何一个AI工具是万能的，复杂任务通常需要多个工具协同

组合的价值

核心理念: 从拥有“瑞士军刀”到组建“专业工具箱”并懂得协同使用，是发挥AI最大潜力的关键

- ✓ **优势互补**

  结合不同工具的长处

  如：DeepSeek/Gemini的理解力 + Midjourney/即梦的创意视觉 + HeyGen的数字人表现）

- ✓ **覆盖完整流程**

  从创意构思到最终成品输出

  如： Kimi信息处理 + Notion AI整理 + 剪映AI剪辑

- ✓ **提升效率与效果**

  自动化串联任务，实现更优结果

场景

场景示例

- ✓ **场景1：制作营销短视频**
  1. DeepSeek/ChatGPT (文案脚本)
  2. Midjourney/可灵 (关键视觉/视频片段)
  3. Suno AI (背景音乐)
  4. 剪映AI (合成剪辑+字幕)
  5. HeyGen/硅基 (数字人播报)

- ✓ **场景2：撰写研究报告**
  1. DeepSeek/Claude (长文档阅读理解+摘要)
  2. Gemini/ChatGPT (内容分析+草稿撰写)
  3. Notion AI (整理笔记+格式优化)
  4. DeepSeek/Kimi/豆包 (图表生成)

这些工具如何高效地组合起来，形成针对特定场景的优化工作流？这正是我们下一份报告《场景化AI工具组合应用攻略》将深入探讨的核心内容。

5. AI工具发展趋势：迈向智能自主

核心迈向智能自主：Agent引领新范式

AI工具正从单一功能辅助，加速向能够理解复杂目标、自主规划并执行任务的AI Agent演进。

1. **核心引擎：AI Agent 的崛起与泛化**
   - **定义:** 不再是简单的问答或生成工具，而是具备感知(Perceive)-规划(Plan)-行动(Act)能力的智能体
   - **关键能力:**
     ① **目标理解与拆解:** 能将复杂、模糊的用户意图分解为可执行的子任务
     ② **动态规划与推理:** 基于环境反馈调整计划，具备一定的逻辑推理和常识判断能力
     ③ **工具调用与协同:** 能自主选择并调用各种软件、API、数据库甚至其他AI模型来完成任务（这是Agent的核心特征之一）
     ④ **记忆与学习:** 拥有短期工作记忆和长期知识存储能力，能从交互中学习和改进
   - **影响:** 从“人驱动工具”转向“人设定目标，Agent自主执行”，极大提升自动化天花板
2. **交互升级：跨模态深度理解与自然交互**
   - **现状:** 从处理单一模态（文本、图像）到同时理解和生成多种模态（如GPT-4o, Gemini）
   - **趋势: 深度融合与跨模态推理**，不仅是输入输出，更能理解模态间的关联（如视频画面+音频+字幕的综合分析），实现更自然、更接近人类的交互方式。**Agent需要这种能力来全面感知环境**
3. **生态融合：无缝嵌入与平台级智能**
   - **现状:** API集成、插件系统
   - **趋势:** AI能力**深度嵌入**操作系统(Windows Copilot)、办公套件(Microsoft 365 Copilot，WPS AI)、行业软件，甚至硬件。未来可能出现**AI原生操作系统或平台**，Agent作为其核心调度单元，协调各类应用和服务
4. **智能进化：持续学习与个性化适应**
   - **现状:** 基于预训练+少量微调
   - **趋势:** AI工具/Agent具备**在线学习、终身学习**能力，能根据用户反馈、使用习惯、特定领域知识进行**实时、个性化**的自我优化，变得越来越“懂你”和“专业”
5. **应用深化：垂直领域知识与专用Agent**
   - **现状:** 通用大模型 + 领域提示工程
   - **趋势:** 出现更多**吸收了深层行业知识(Know-How)和专有数据**的垂直大模型和**专用Agent**，在法律、医疗、金融、科研、制造等领域提供高度专业化的解决方案

5. AI工具发展趋势：驾驭复杂与责任

驾驭复杂：Agent时代的机遇与深层挑战

Agent化和自主性的提升带来了前所未有的机遇，但也放大了现有挑战，并引入了新的复杂问题。

1. **可靠性与“幻觉”放大**
   - **挑战:**单一工具的幻觉可能只影响局部，但Agent在多步骤任务中，一个环节的幻觉或错误决策可能导致**任务链失败**或产生严重后果，且错误溯源困难（**黑盒问题**）。
   - **应对方向:** 提升模型的事实性、推理鲁棒性，发展Agent的自我反思与修正机制。
2. **安全边界模糊与新攻击面**
   - **挑战:** Agent自主调用工具、访问网络和本地文件，极大地扩展了潜在的**安全风险**。恶意Prompt可能诱导Agent执行有害操作、泄露敏感数据（**Agent“越狱”或被“劫持”**）
   - **应对方向:** 严格的权限管理、行为监控、恶意指令识别、Agent安全沙箱
3. **控制、对齐与意图漂移**
   - **挑战:**如何确保高度自主的Agent始终**遵循人类的真实意图和价值观(对齐问题)?**在长链条、复杂任务中，Agent的目标可能发生“漂移”,或采用不符合预期的方式达成目标。
   - **应对方向:** 发展更强的可解释性AI（XAI）、人类在环（Human-in-the-Loop）监督机制、价值对齐训练技术。
4. **复杂性、成本与资源消耗**
   - **挑战:** Agent执行复杂任务需要大量推理计算、多次工具调用、维护长期记忆，导致**算力、时间和经济成本**急剧增加。Agent的设计、调试和维护本身也极为复杂。
   - **应对方向:** 模型效率优化（蒸馏、量化）、更优的Agent架构（如层级Agent）、经济高效的推理策略。
5. **伦理困境与责任归属**
   - **挑战:** 自主Agent做出的决策可能引发伦理争议（偏见、歧视、公平性），一旦造成损害，**责任主体难以界定**（是开发者、使用者还是Agent本身？）
   - **应对方向:** 在Agent设计中嵌入伦理框架、加强审计追踪、建立清晰的法律法规
6. **数据隐私与过度窥探**
   - **挑战:** 为了实现个性化和高效执行任务，Agent可能需要访问大量个人或组织数据（邮件、日历、文件、浏览历史），引发**严重的隐私担忧**
   - **应对方向:** 隐私计算技术、用户数据授权与控制、最小化数据访问原则
7. **社会冲击与范式重塑**
   - **挑战:** Agent对**知识型工作、决策流程乃至组织结构**可能带来颠覆性影响，加剧技能鸿沟和就业结构调整。需要社会层面的适应和引导
   - **应对方向:** 教育体系改革、终身学习支持、社会保障体系调整、公共政策引导

6. 总结：拥抱AI，从选型开始

核心回顾

- AI工具是提升效率和创造力的强大引擎。
- 理解AI工具的分类、测评维度和方法至关重要。
- 本PPT提供了主流AI工具的深度测评和横向对比（文本、图像、音视频等）。
- 选择工具需以需求为核心，结合个人/团队具体情况。

关键信息

- 没有完美的工具，只有合适的工具
- 组合使用是发挥AI最大价值的关键AI
- Agent代表着未来的自动化方
- 保持学习，拥抱变化，积极探索AI的可能性。

开始构建你的AI工具箱，并思考如何将它们融入你的工作流中！



北京大学127周年，生日快乐！

感谢各位老师和同学的批评指导

欢迎会后沟通交流



关注技术公众号