可信Agent的规模化之路
从企业智能体到个人代理的时代

演讲人：吴明辉

明略科技创始人、CEO兼CTO

AiCon
全球人工智能开发与应用大会

极客邦科技 2026 年会议规划

促进软件开发及相关领域知识与创新的传播

参会咨询

查看会议

4月 · 北京 · 1200人

QCon 全球软件开发大会

会议时间：4月16-18日

- Agentic Engineering
- AgentOps
- 下一代模型架构与推理优化
- AI 原生基础设施
- 知识工程实践
- AI 安全

6月 · 上海 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：6月26-27日

- AI Infra 系统工程
- 多 Agent 协作与实践
- 多模态融合
- 模型训练与推理创新
- 数据平台与特征服务

8月 · 深圳 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：8月21-22日

- Agentic AI
- 轻量化与高效推理
- 多模态应用
- AI + IoT 场景实践
- AI 工业化落地

10月 · 上海 · 1200人

QCon 全球软件开发大会

会议时间：10月22-24日

- AI Agent
- Vibe Coding
- 智能可观测
- 推理基建
- 模型攻防
- AI x 创造力

12月 · 北京 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：12月18-19日

- 大模型架构创新
- 多模态 AI 产业融合
- 具身智能
- AI for Science
- 大模型安全

InfoQ 极客传媒

95%的生成式AI项目没有收获财务回报

《State of AI in Business 2025》—— MIT

明略科技® MININGLAMP TECHNOLOGY 股票代码：2718.HK

投入规模与回报失衡

投入规模

全球企业GenAI投资（MIT, 2025）

300-400
亿美元

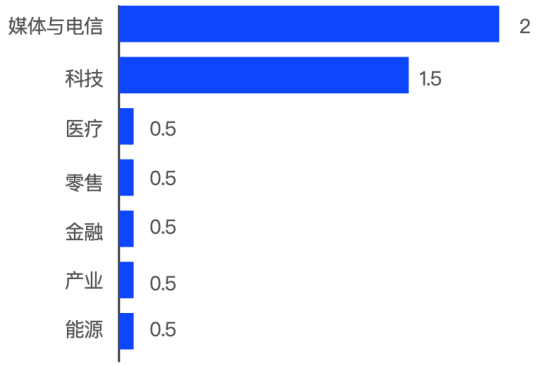
回报现状

95%
无任何财务回报

5%
仅创造数百万美元价值

行业分化

9大行业中，**除科技和传媒行业外**，其他行业如金融、制造、医疗等基本还停留在对生成式AI的试点阶段，尚未实现大规模的应用和变革，仍在探索其可行性和适用性。



问题现象

工具与需求脱节

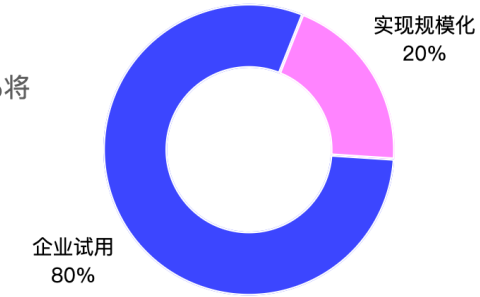
Global 觉得 AI 必须统一战略，于是拍板做了一个自研平台，要求所有业务都得接入。结果落地的时候，本地团队一边被要求用，一边还得在原有Excel/邮件流程里双轨跑。功能不匹配、体验割裂。大家嘴上说在用，手上其实早就放弃。

自建迷思

外部合作伙伴的成功率是内部构建的两倍。

试点与落地断层

80%企业试用GenAI，仅20%将GenAI嵌入到了业务流中。



数据来源：MIT, *State of AI in Business 2025* (Project NANDA)

企业级Agent与个人Agent核心差异

Enterprise vs Personal Agent

明略科技®
MININGLAMP TECHNOLOGY
股票代码：2718.HK

两类Agent的关键区别

个人Agent

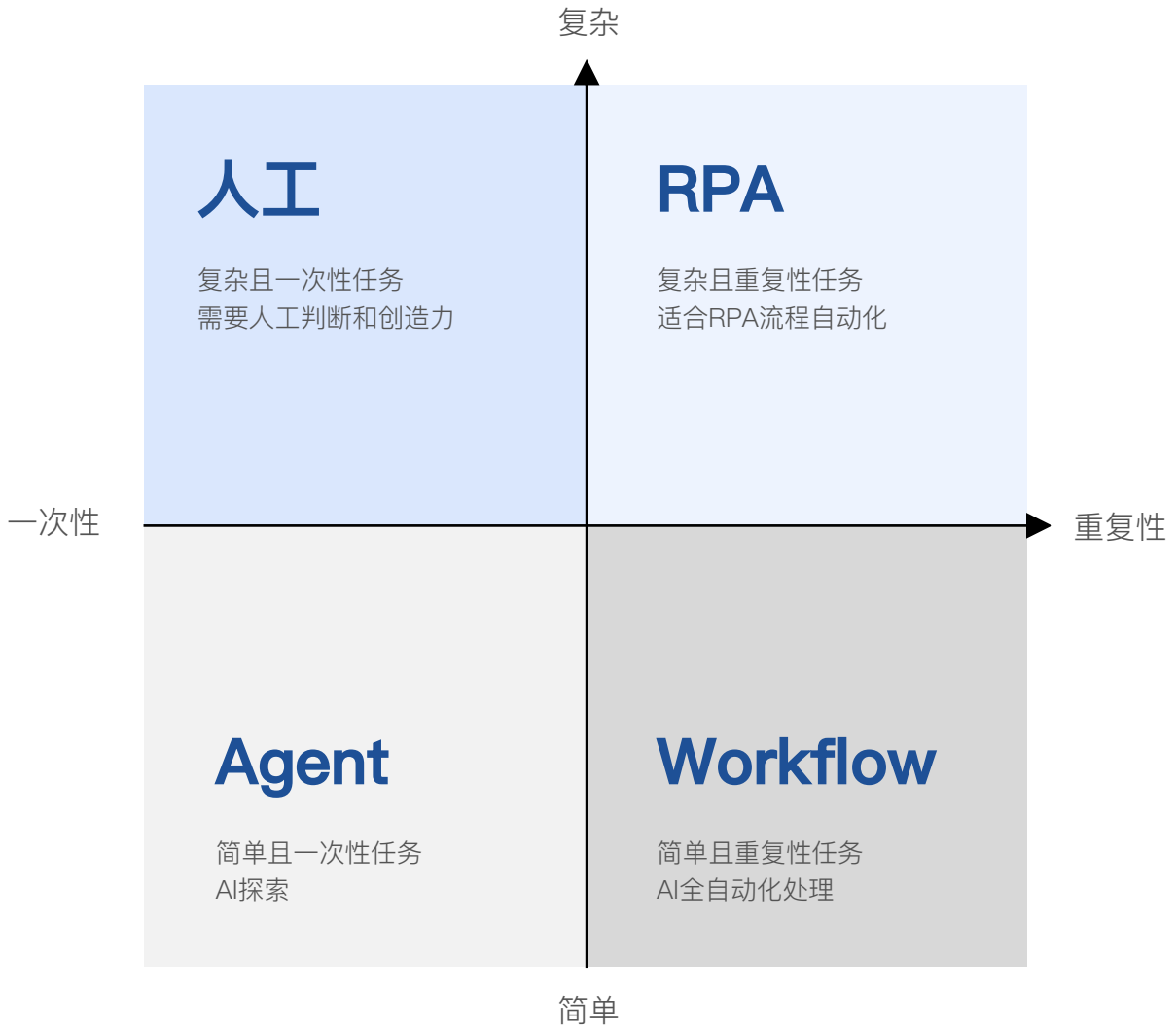
- 单用户服务，个性化需求
- 轻量级部署，本地或云端
- 数据隐私性强，个人控制
- 功能相对简单，通用场景
- 成本低，按需付费
- 快速响应，即时交互

企业级Agent

- ✓ 多用户协同，组织级服务
- ✓ 企业级部署，私有化/混合云
- ✓ 数据安全合规，权限管理
- ✓ 功能复杂，业务流程深度集成
- ✓ 投入大，长期ROI考量
- ✓ 稳定性高，7×24小时运行

任务分类四象限模型

Task Classification Matrix



可信智能体要克服的挑战重重

Andrej Karpathy — AGI is still a decade away

"The problems are tractable, but they're still difficult"

DWARKESH PATEL
OCT 18, 2025

300 17 58 Share Transcript

The Andrej Karpathy episode.

Andrej explains why reinforcement learning is terrible (but everything else is much worse), why model collapse prevents LLMs from learning the way humans do, why AGI will just blend into the previous ~2.5 centuries of 2% GDP growth, why self driving took so long to crack, and what he sees as the future of education.

Watch on YouTube; listen on Apple Podcasts or Spotify.

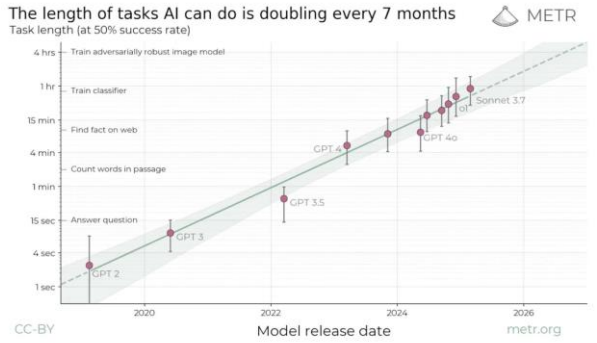
What it takes to build AI agents that actually work

11.01.2025 | By: Ashu Garg

As Andrej Karpathy put it recently: "This isn't the year of agents, it's the decade of agents." Karpathy spent years at Tesla working on self-driving software, where he witnessed firsthand the yawning gap between a demo that works 90% of the time and a system you can trust with your life. In autonomous vehicles, climbing from 90% to 99.9% reliability proved exponentially harder than the initial climb from 0% to 90%. AI agents are now beginning that same ascent.

So what makes the last mile so difficult? Why do those extra nines of reliability matter so much? And what are builders actually doing to get there? This month, I'll explore these questions through lessons from one of our portfolio founders, Ram Krishnamurthy at Maximor.

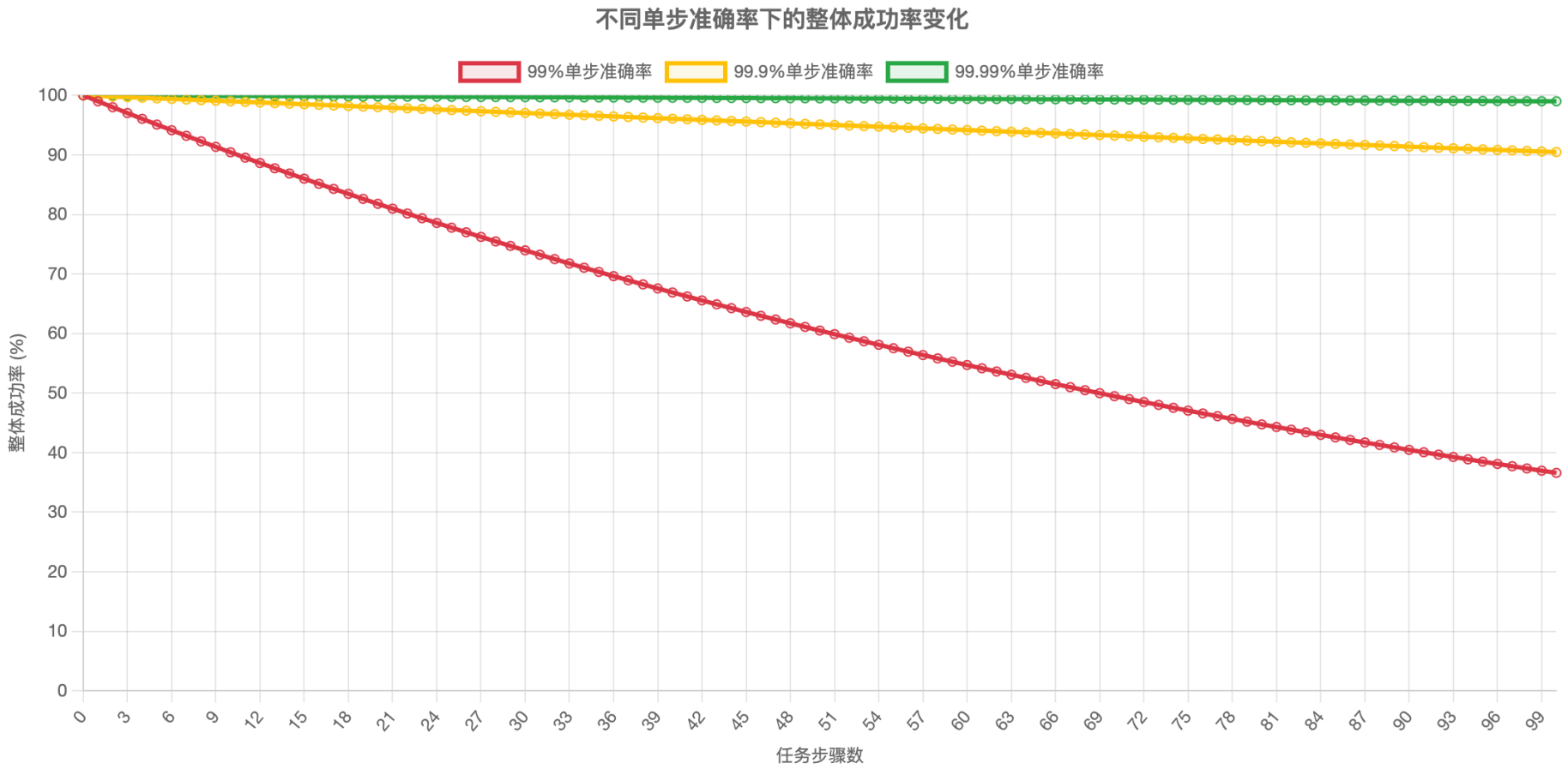
The METR research group has been tracking how far frontier models can go. They measure what they call "horizon length": essentially, the length of time an AI agent can work reliably before it breaks down. They've found that the horizon length for software and coding agents has been doubling every 7 months.



Source: METR research

According to analysis from Exponential View, today's best AI systems can manage around 100 steps at 99% accuracy. That translates to a day or two of focused analyst work – something like building a competitive landscape analysis or producing a research brief with multiple sources. Tools like OpenAI's Deep Research can manage this level of complexity today.

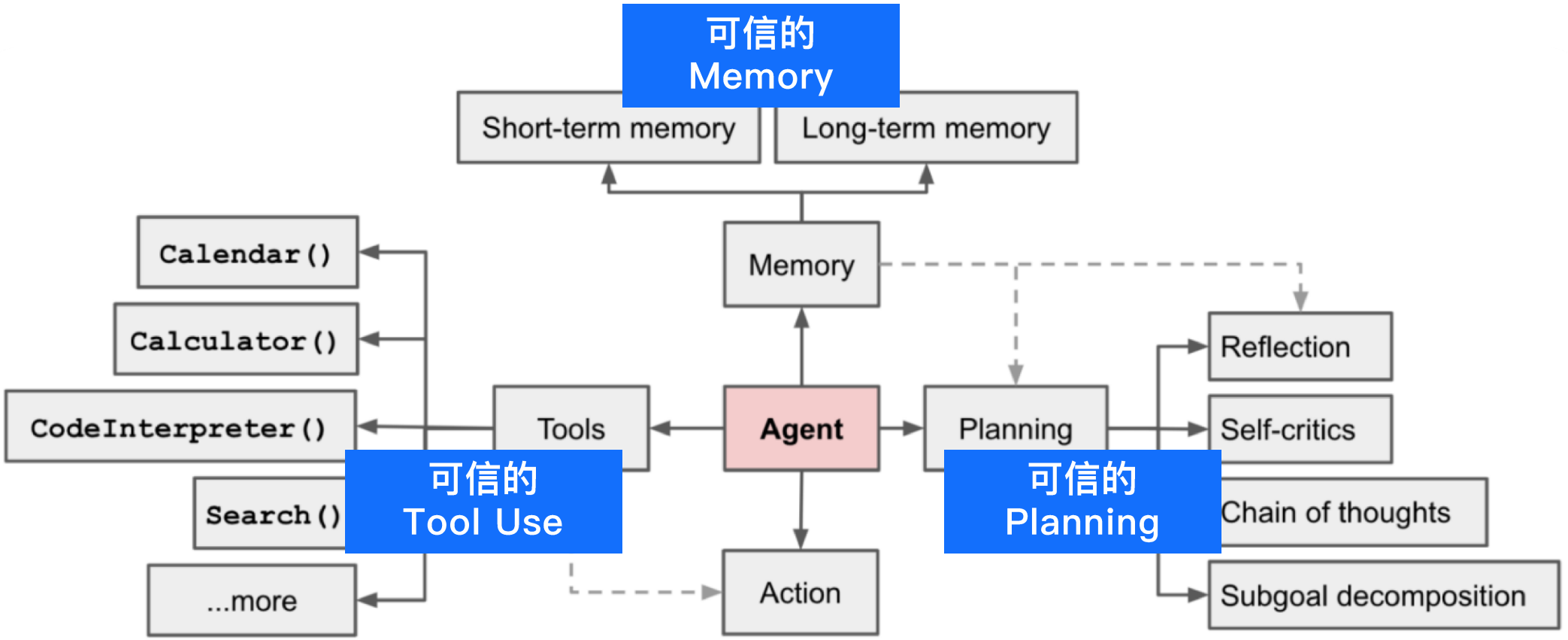
可信智能体要克服的挑战重重



计算公式: 整体成功率 = (单步准确率)^步骤数

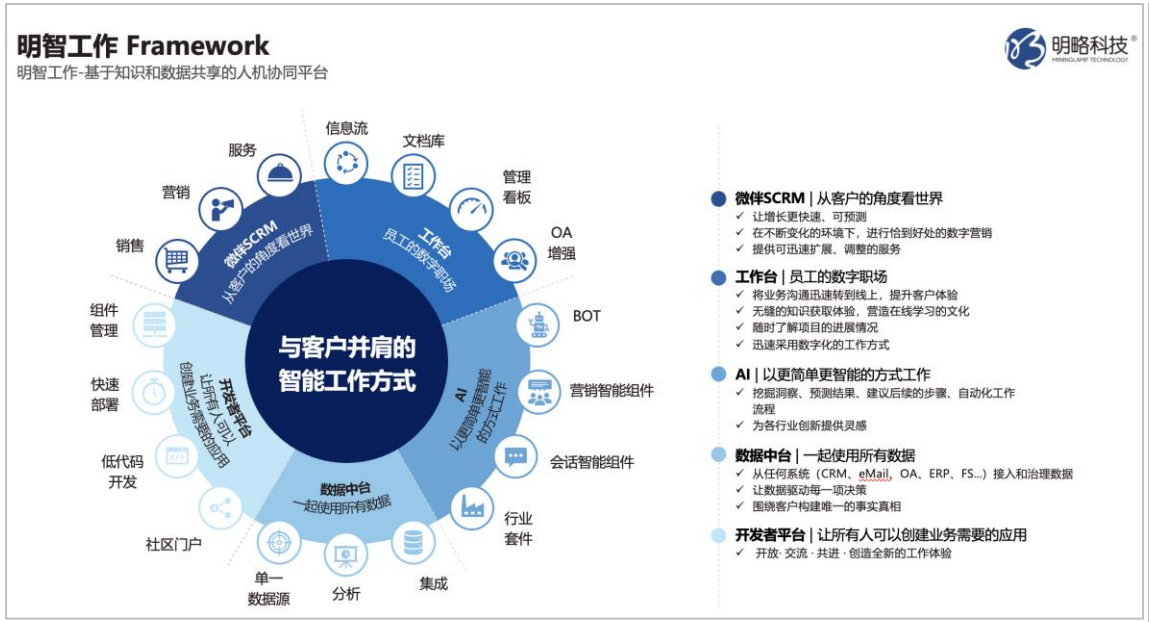
例如: 99%准确率执行100步 = 0.99^100 ≈ 36.6%成功率

可信的智能体 = 可信的Memory + 可信的Tool Use + 可信的Planning

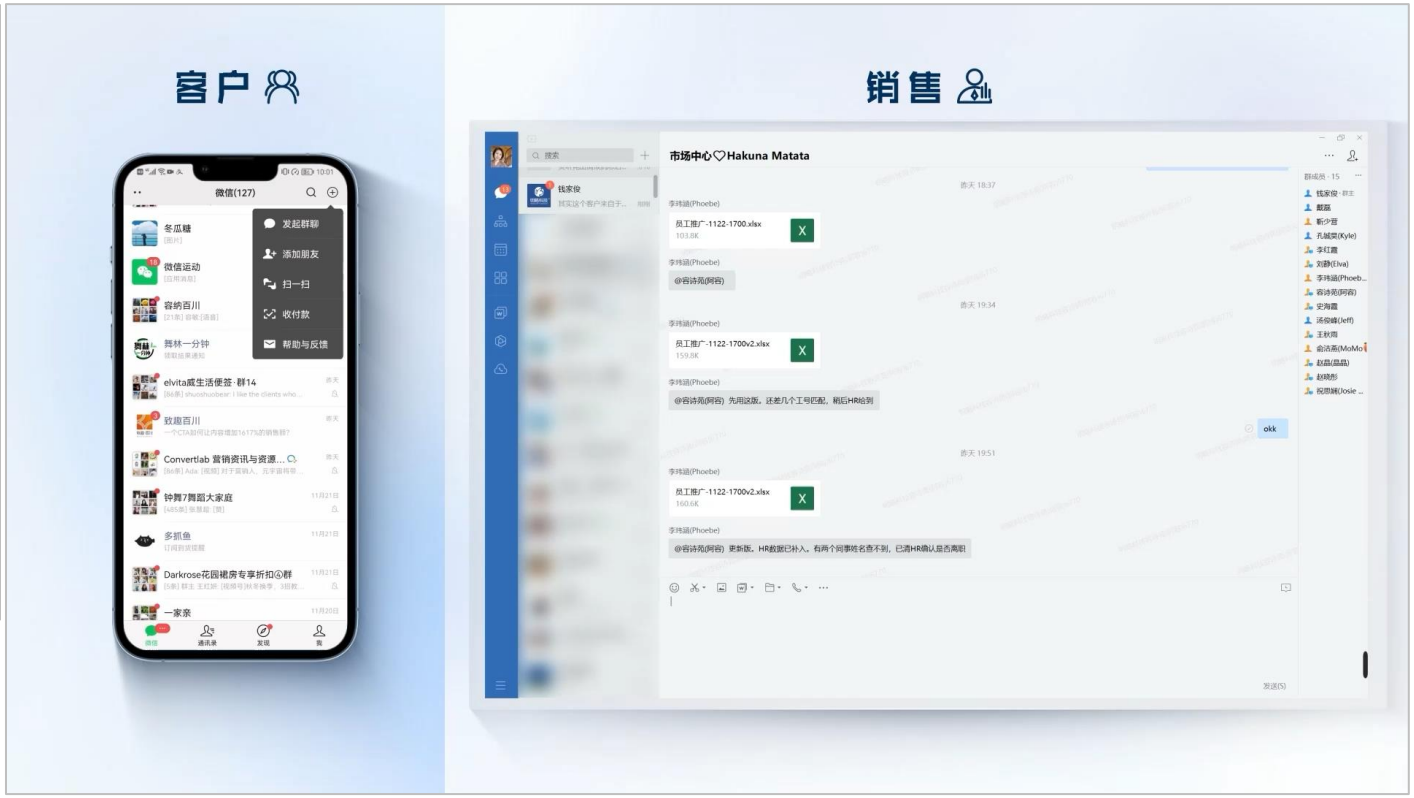


明略EIP的惨痛教训 – 2021

明略科技®
MININGLAMP TECHNOLOGY
股票代码：2718.HK



2021年



EIP产品介绍

Agentic Model是可信智能体落地的重大瓶颈

明略科技®
MININGLAMP TECHNOLOGY
股票代码：2718.HK

LEVEL 1 - 聊天机器人 (Chatbots)

Conversational AI

能够进行流畅的多轮对话、内容生成和摘要，但本质上是基于统计规律的模式匹配，**缺乏深度的逻辑推理和自主行动能力。**

- 自然语言理解
- 上下文感知
- 多轮对话

LEVEL 2 - 推理者 (Reasoners)

Reasoning AI

能够进行**多步骤、逻辑严密的推理**（思维链，Chain of Thought），在数学、科学和逻辑分析等领域展现出“博士级”的水平，幻觉问题显著减少。

- 逻辑推理
- 知识整合
- 复杂问题分析

LEVEL 3 - 智能体 (Agents)

Agentic AI

核心是 **“思考-决策-执行”的闭环能力。**AI不再仅限于对话，而是被赋予了“手和脚”，可以主动调用API、操作软件、与环境交互，完成跨应用的复杂任务。

- 自主规划
- 工具使用
- 任务执行

LEVEL 4 - 创新者 (Innovators)

Innovator AI

关键突破在于**创造性思维**，能够提出前所未有的理论、设计全新的实验方案或创作具有独特风格的艺术作品，**突破人类现有的知识边界。**

- 创造性思维
- 知识发现
- 自主学习

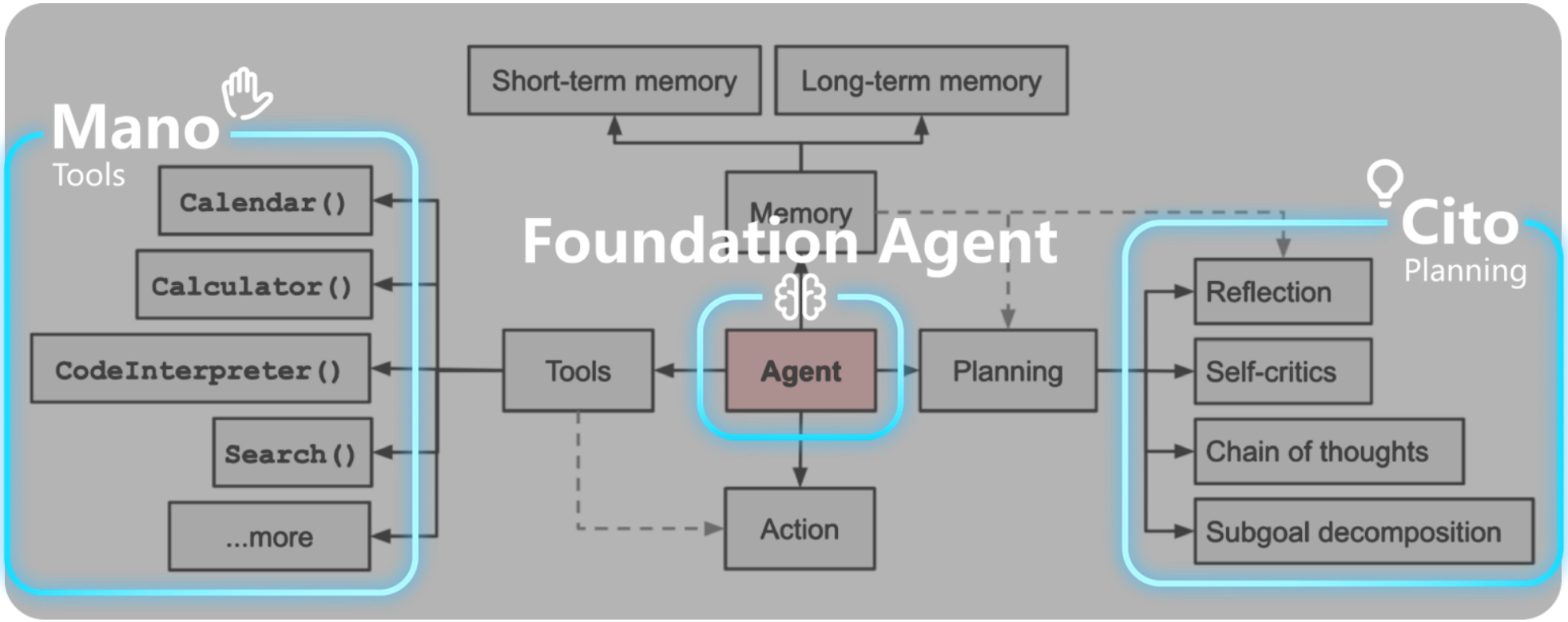
LEVEL 5 -组织者 (Organizations)

AGI

标志着**AGI的完全体。**能够像人类管理者一样进行宏观规划、资源分配和多智能体协作，处理极度复杂的全局性任务，例如独立运营一家公司或管理一座城市。

- 通用认知能力
- 跨域迁移
- 自我进化

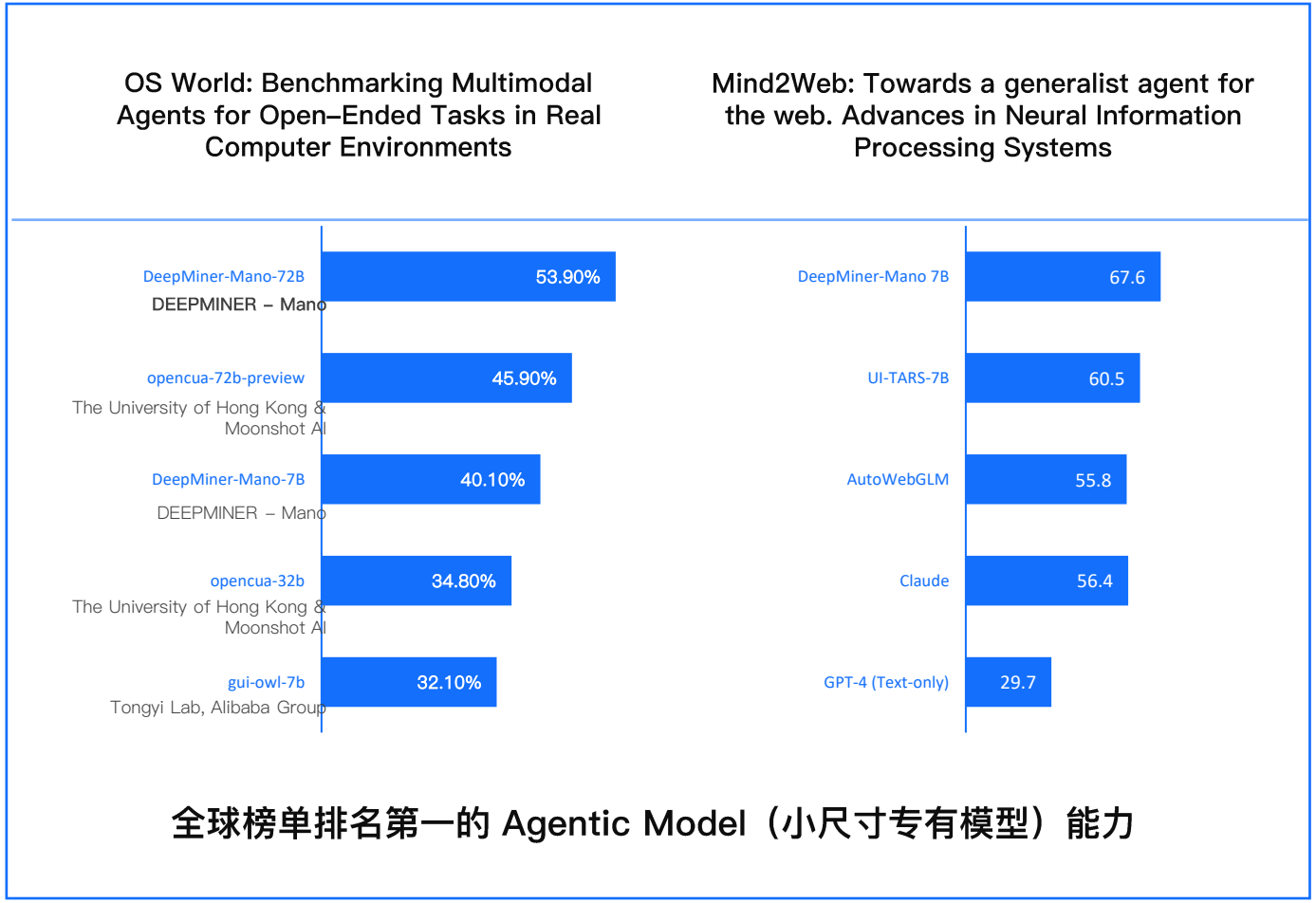
可信智能体和可信专用代理模型：三个技术支撑，一个中枢，两大模型



明略科技® MININGLAMP TECHNOLOGY | deepminer™

针对企业AI落地的三大知识壁垒，DeepMiner设计了三大核心引擎，由一个强大的智能中枢（FA）统一调度，系统性地为AI注入业务灵魂，真正释放AI在企业场景中的潜力。

GUI Tool Use Model



全球榜单排名第一的 Agentic Model（小尺寸专有模型）能力

核心模型：Mano
AI 的灵巧手

让AI学会”看”与”点”，像人类一样使用世界

面对海量为人设计的软件系统，我们选择的不是改动世界去适配AI，而是让AI学会像人一样去适应这个世界，通过视觉理解与操作实现真正的普适性。

视觉理解能力

模拟人类视觉感知，精准识别任何软件界面元素

自主操作能力

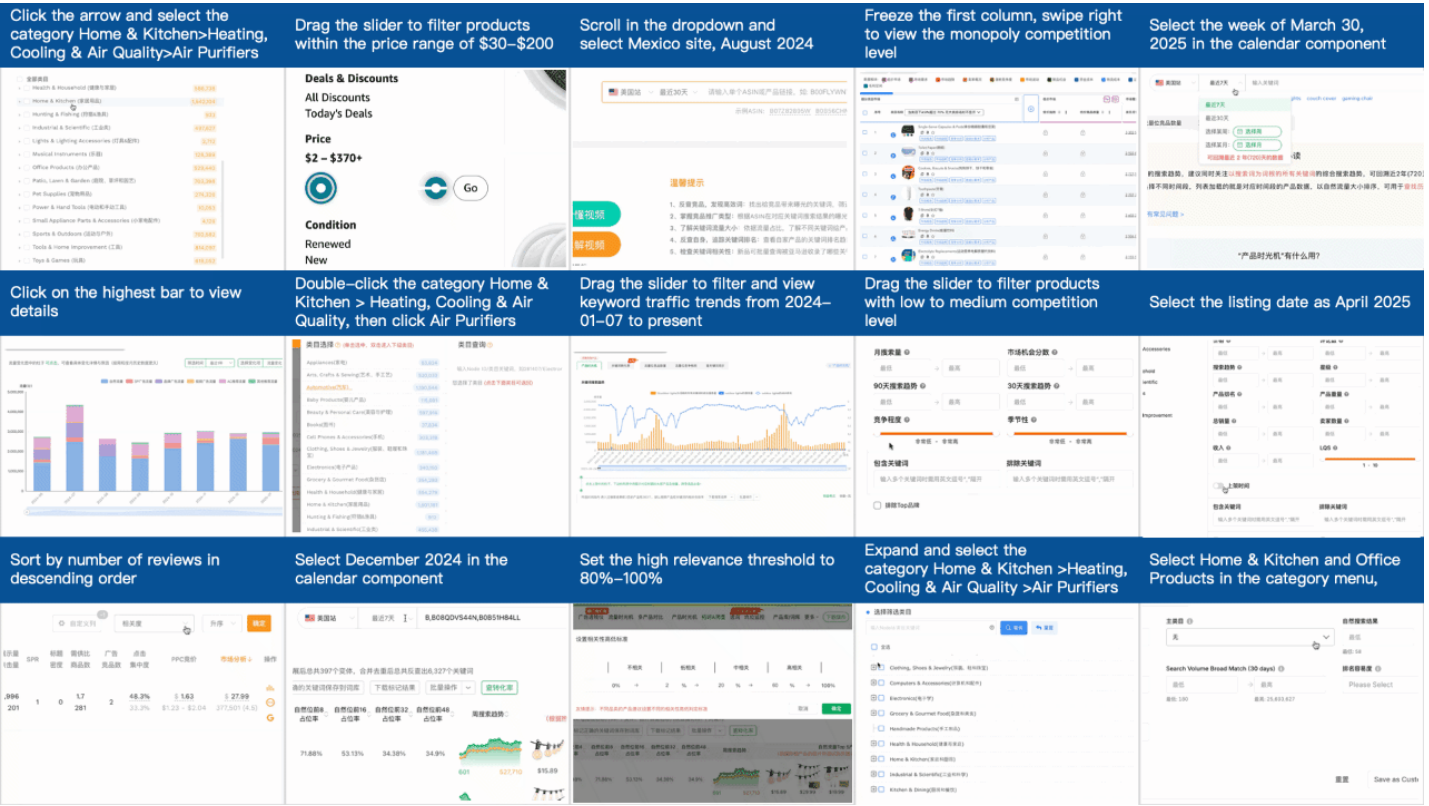
像人一样”看”和”点”，无需改造即可使用任何系统

自我学习飞轮

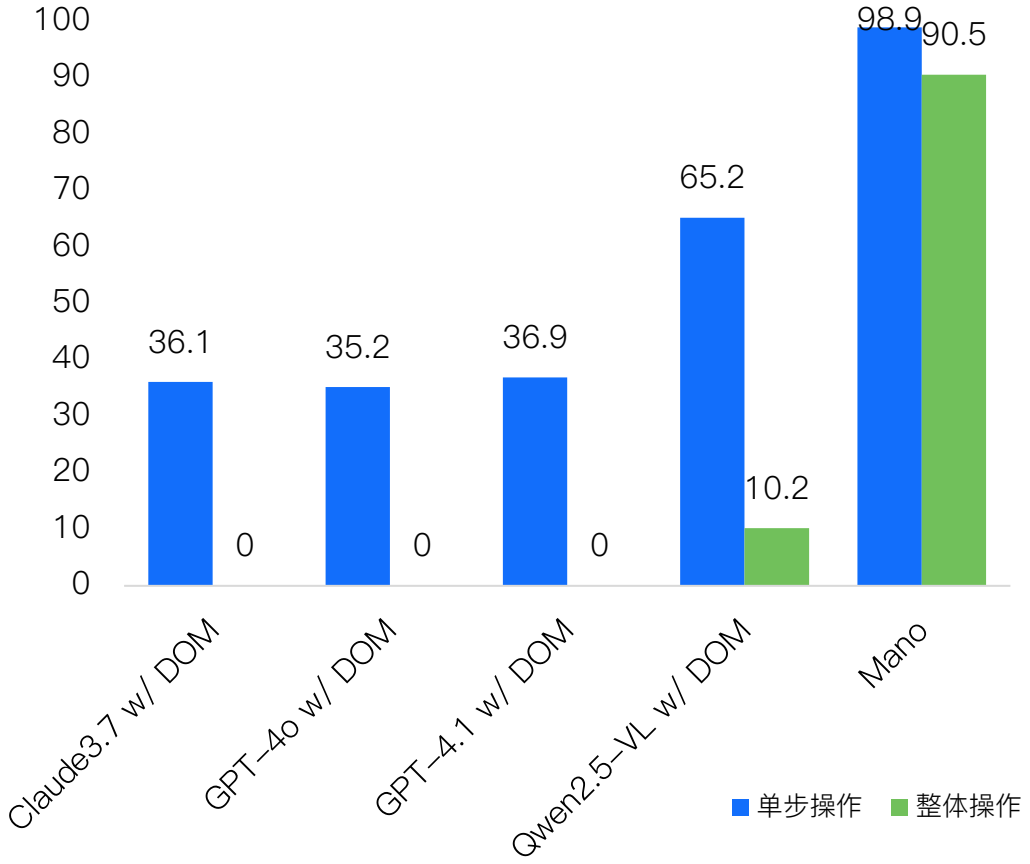
探索 → 使用 → 自标注 → 再训练，持续进化提升

如何训练垂直场景可信解决方案?

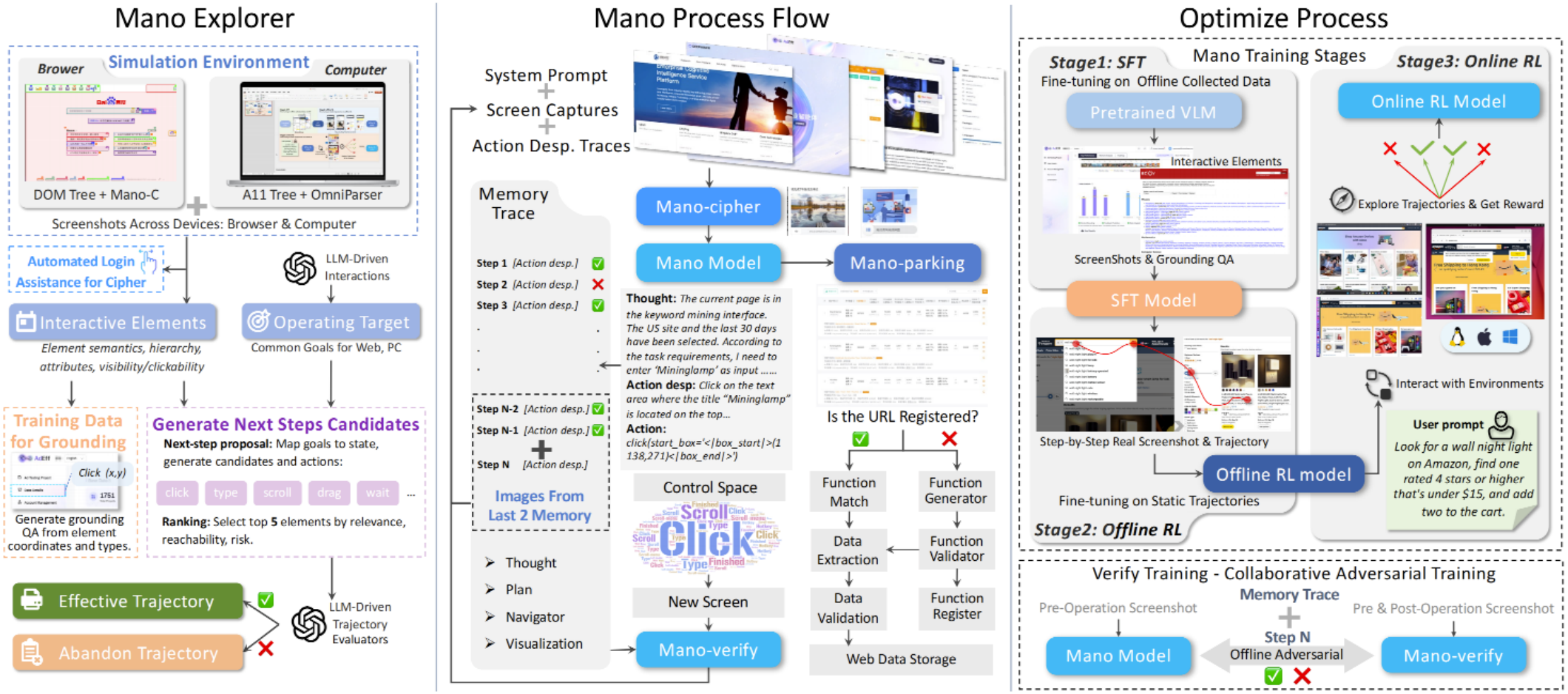
营销场景中常见的复杂菜单



WebGUI–1M, DeepMiner Mano Model

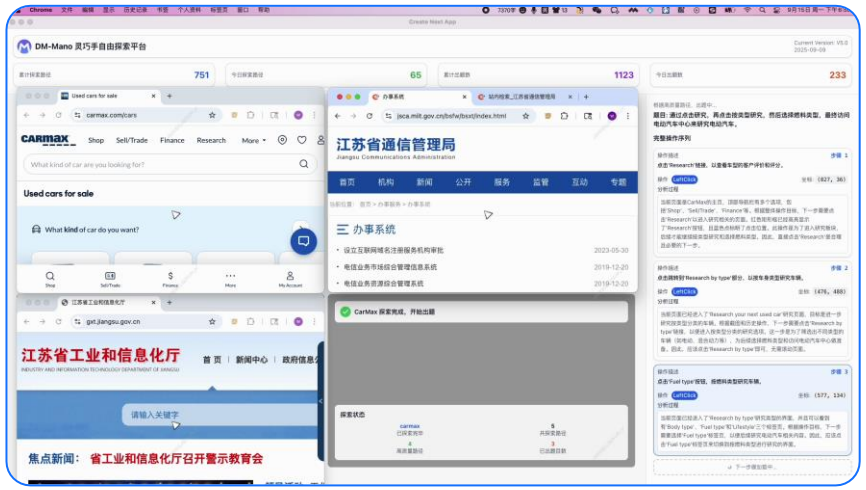


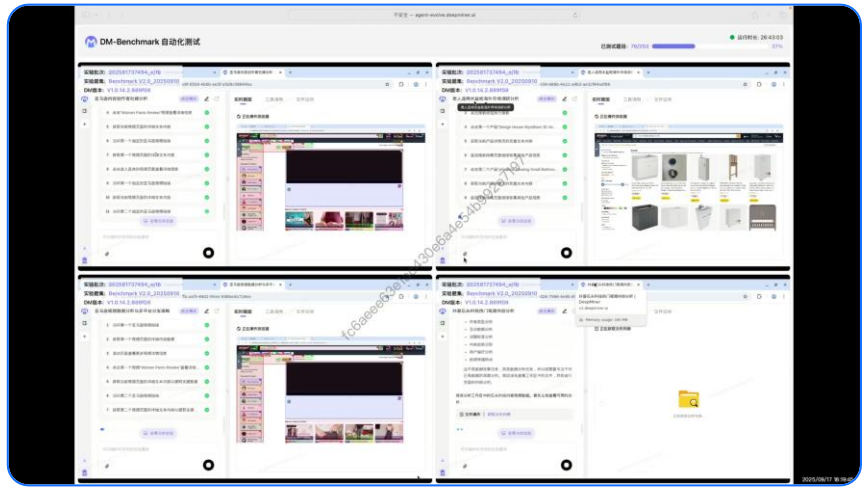
Mano核心技术架构


Mano Explorer
Brower
Simulation Environment
Computer
DOM Tree + Mano-C
A11 Tree + OmniParser
Screenshots Across Devices: Browser & Computer
Automated Login Assistance for Cipher
LLM-Driven Interactions
Interactive Elements
Element semantics, hierarchy, attributes, visibility/clickability
Operating Target
Common Goals for Web, PC
Training Data for Grounding
Click (x,y)
Generate grounding QA from element coordinates and types.
Generate Next Steps Candidates
Next-step proposal: Map goals to state, generate candidates and actions:
click
type
scroll
drag
wait
Ranking: Select top 5 elements by relevance, reachability, risk.
Effective Trajectory
Abandon Trajectory
LLM-Driven Trajectory Evaluators
Mano Process Flow
System Prompt
Screen Captures
Action Desp. Traces
Memory Trace
Step 1 [Action desp.]
Step 2 [Action desp.]
Step 3 [Action desp.]
Step N-2 [Action desp.]
Step N-1 [Action desp.]
Step N [Action desp.]
Images From Last 2 Memory
Thought
Plan
Navigator
Visualization
Mano-cipher
Mano Model
Thought: The current page is in the keyword mining interface. The US site and the last 30 days have been selected. According to the task requirements, I need to enter 'Mininglamp' as input .....
Action desp: Click on the text area where the title "Mininglamp" is located on the top...
Action:
click(start_box='<|box_start|>(1138,271)<|box_end|>')
Control Space
New Screen
Mano-verify
Mano-parking
Is the URL Registered?
Function Match
Function Generator
Data Extraction
Function Validator
Data Validation
Function Register
Web Data Storage
Optimize Process
Mano Training Stages
Stage1: SFT
Fine-tuning on Offline Collected Data
Pretrained VLM
Interactive Elements
ScreenShots & Grounding QA
SFT Model
Step-by-Step Real Screenshot & Trajectory
Offline RL model
Fine-tuning on Static Trajectories
Stage2: Offline RL
Stage3: Online RL
Online RL Model
Explore Trajectories & Get Reward
Interact with Environments
User prompt
Look for a wall night light on Amazon, find one rated 4 stars or higher that's under $15, and add two to the cart.
Verify Training - Collaborative Adversarial Training
Pre-Operation Screenshot
Memory Trace
Step N
Offline Adversarial
Pre & Post-Operation Screenshot
Mano Model
Mano-verify

明略科技®
MININGLAMP TECHNOLOGY
股票代码：2718.HK

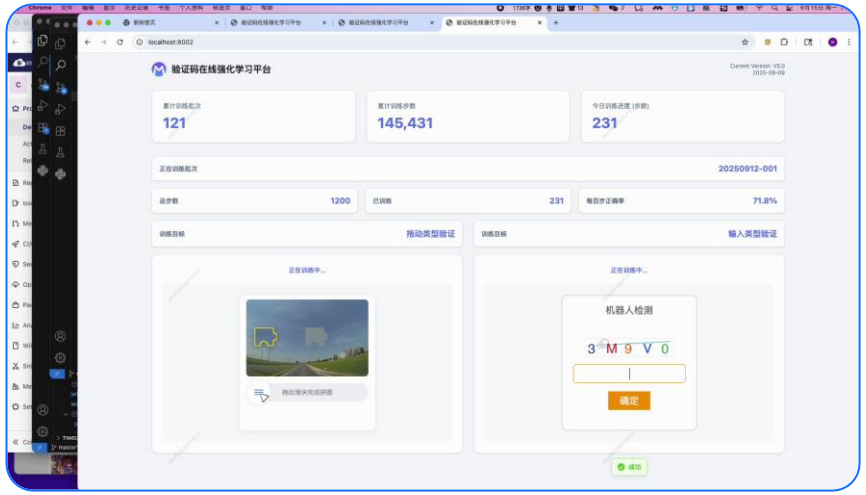
Mano
Explorer



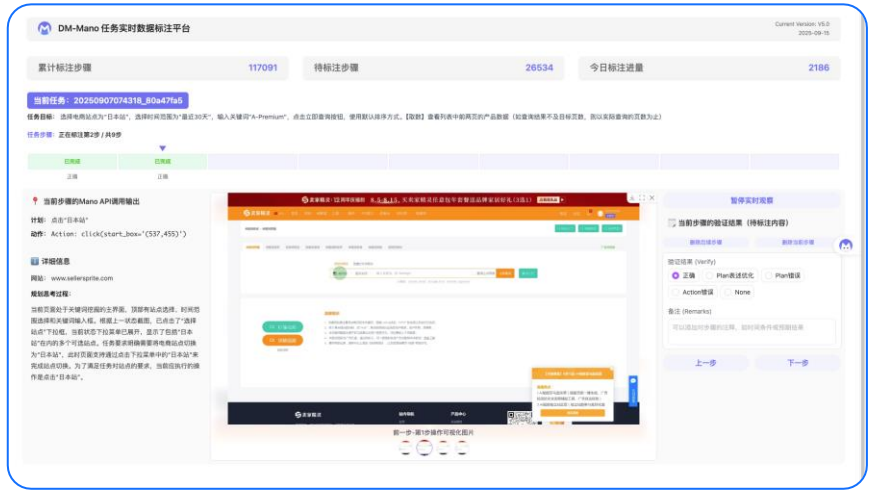


DeepMiner
Benchmark

Mano
在线强化学习



Mano
任务实时标注



AiCon
全球人工智能开发与应用大会

InfoQ 极客传媒

模型训练方向分化

Model Training Divergence

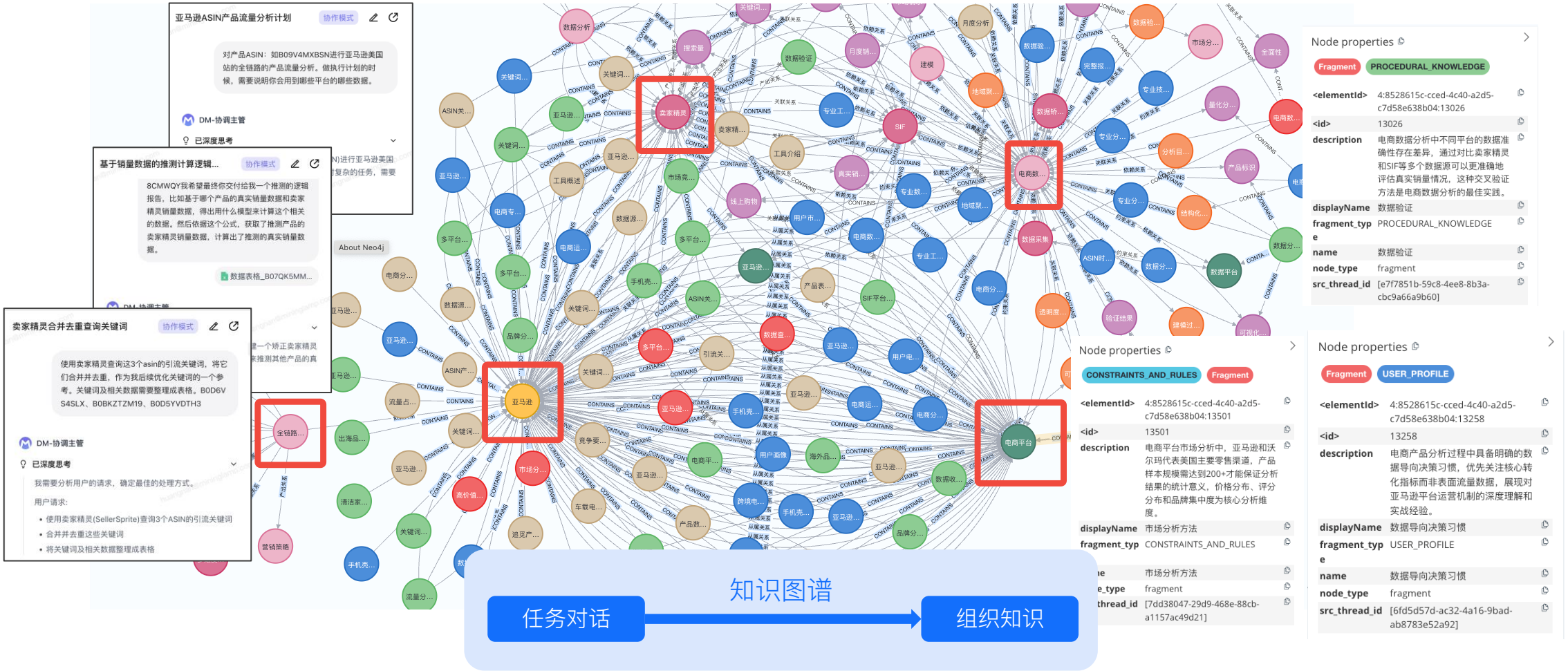
明略科技®
MININGLAMP TECHNOLOGY
股票代码：2718.HK

差异化SOTA (State of the Art) **比通用SOTA更重要**

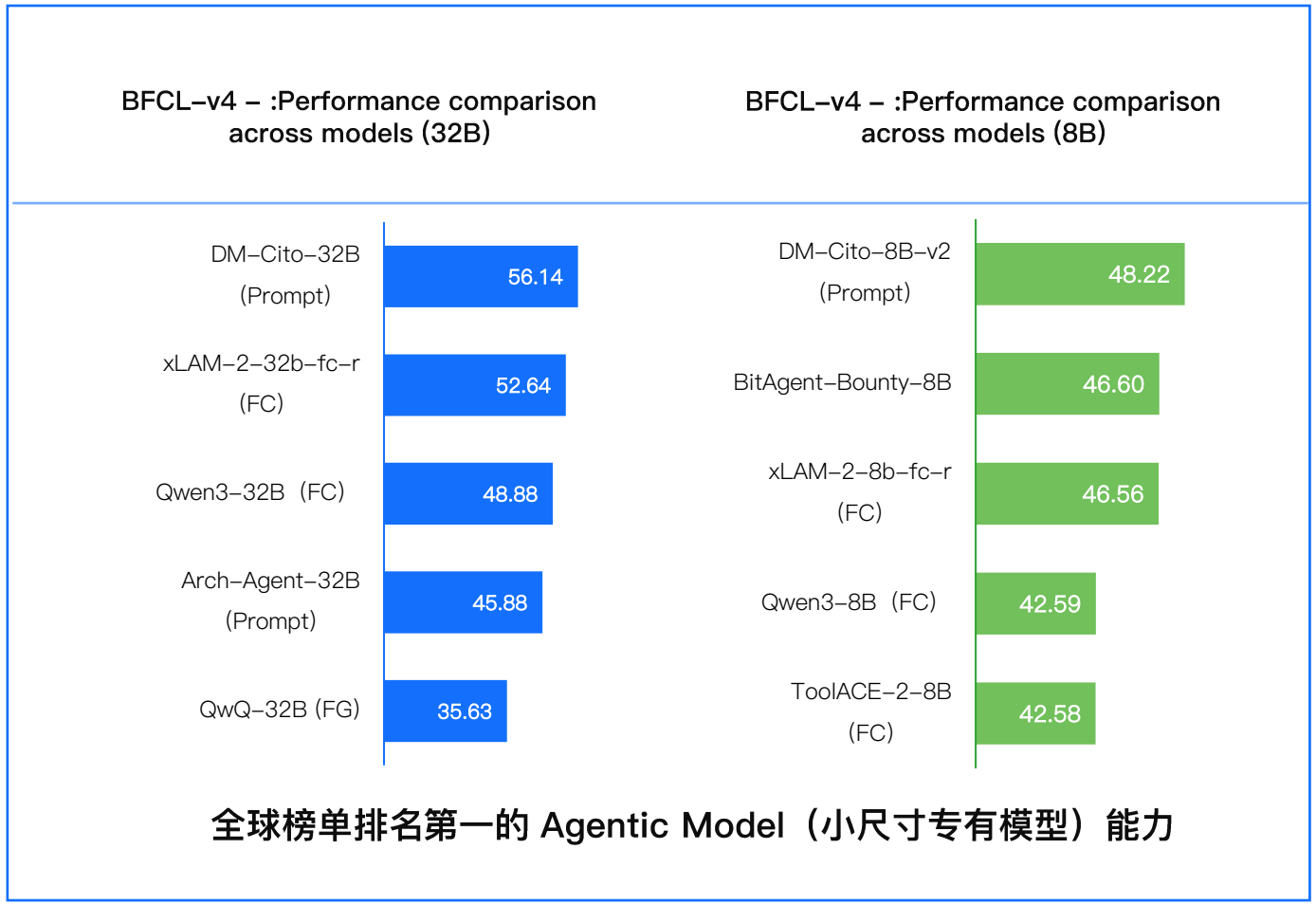
模型训练方向正在前所未有地分化,通用SOTA模型的叙事不再重要。
更重要的是在特定领域、特定任务上实现差异化的SOTA能力。
这意味着:垂直领域专用模型将大量涌现,通用大模型的竞争优势减弱。
企业应关注自身业务场景,选择或训练最适合的专用模型,而非盲目追求通用能力。

模型即产品，数据即模型

用户持续的使用行为，将为组织沉淀大量的知识和记忆于知识图谱，并为Cito所用



数据分析及规划Model



全球榜单排名第一的 Agentic Model（小尺寸专有模型）能力

核心模型：Cito
AI 的专家脑

让AI学会像专家一样，做真实业务场景的思考

与那些最常见榜单训练不同，我们选择让Cito在真实的世界中学习，而非从教科书上的数据中学习。

实时在线强化学习

使用真实环境进行互动学习，实时更新行业知识

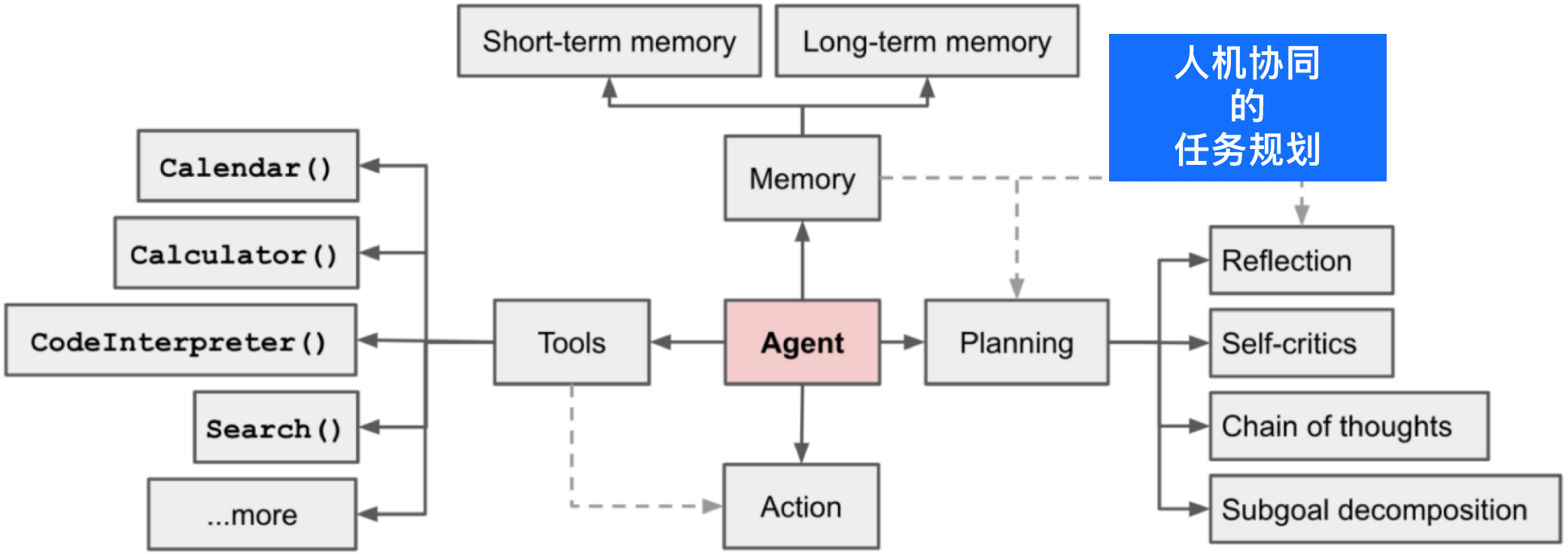
多源changing数据深度融合

融合海量营销数据，场景针对性训练

专家模型自我评估

真实使用数据即使自我评估，掌握暗默知识

从 YOLO 回归到 Human in the Loop 模式

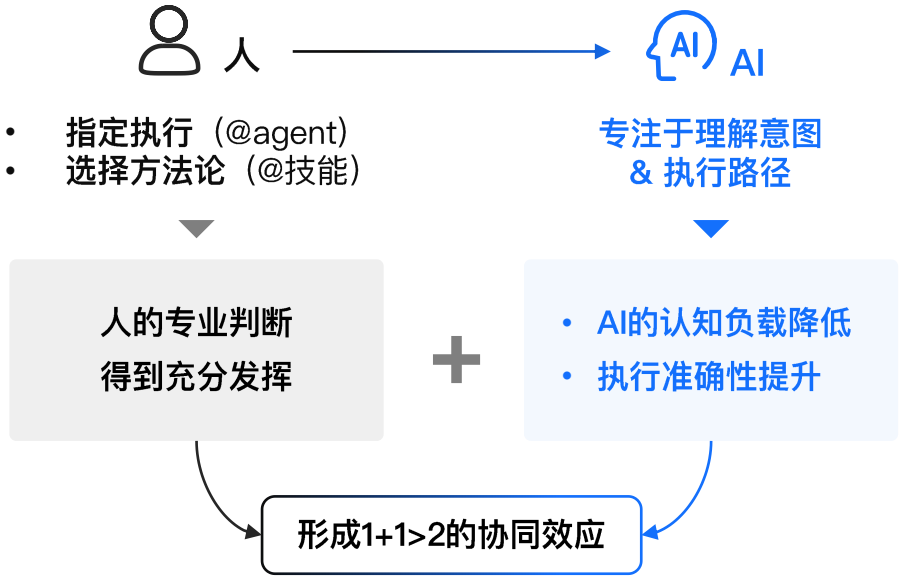


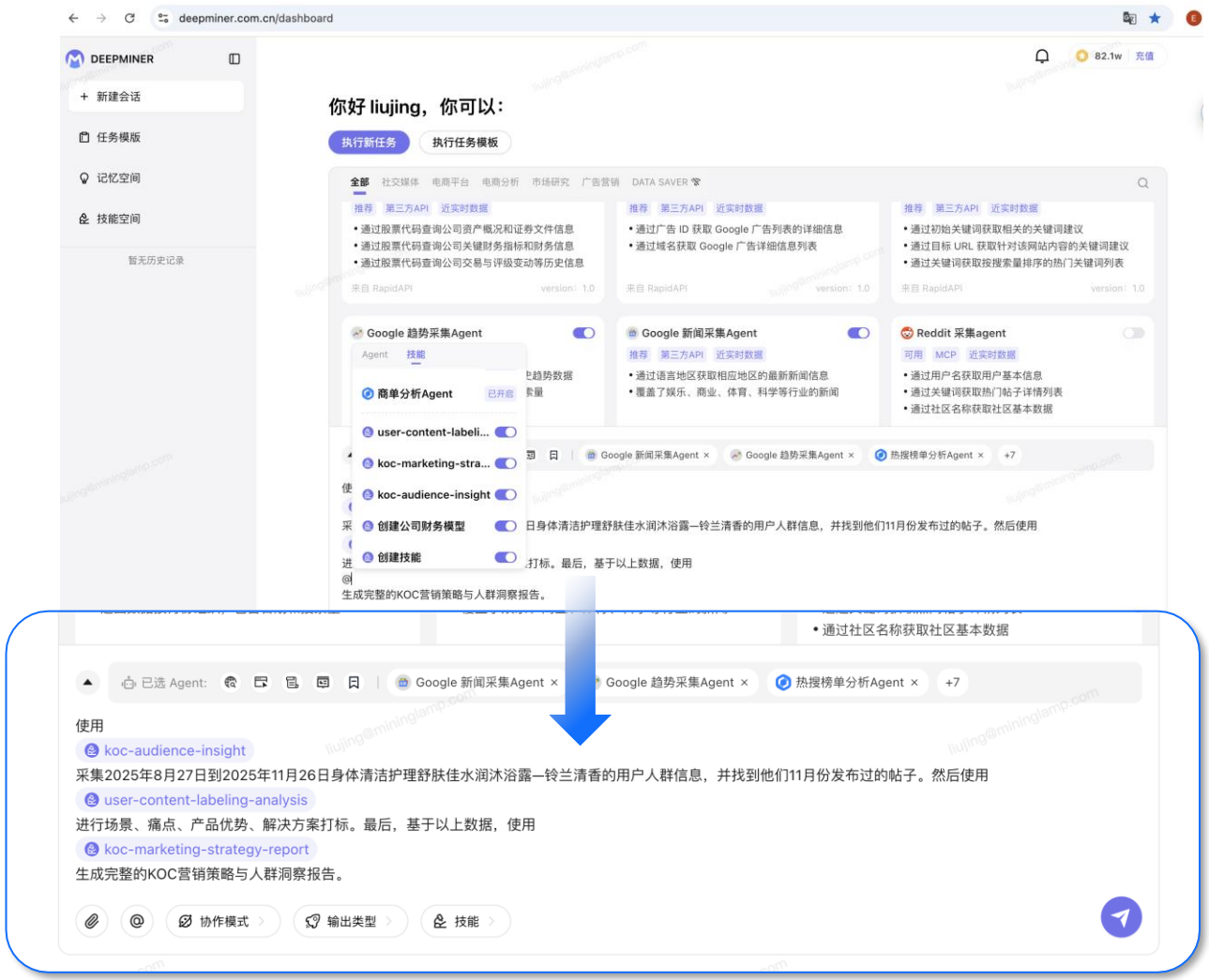
Attention is All You Need —— DeepMiner的人机协作模式

@的力量

从”AI全能猜测”到”人机精准协同” 传统AI Agent面临的困境：在复杂任务中，AI需要同时决策”使用哪个工具、调用哪个Agent、参考哪些文件”，决策链条越长，误判概率越高。

DeepMiner的 @机制 将决策权分层：





Task vs Job 概念对比

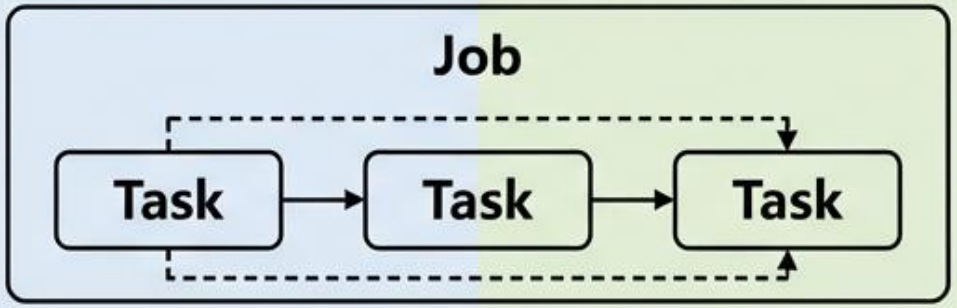
计算机领域中的任务与作业

Task(任务)

- (1) 细粒度 - 最小调度单位
- (2) 执行过程 - 正在做什么
- (3) 某个阶段或步骤
- (4) 交互式/实时运行
- (5) Job的组成部分

Job(作业)

- (1) 粗粒度 - 高层级工作单元
- (2) 完整目标 - 要完成什么
- (3) 明确的开始和结束状态
- (4) 批处理/后台运行
- (5) 由多个Task组成



AI 是一种新的计算范式

AI = 新计算范式 = 给定目标后的计算

得目标者得天下

个人Agent目标

- 效率提升 – 自动化日常任务
- 个人助理 – 智能日程管理
- 知识管理 – 信息整理归纳

企业Agent目标

- **业务增长 – 规模化运营优化**
- **流程优化 – 组织效率提升**
- **决策支持 – 数据驱动决策**

极客邦科技 2026 年会议规划

促进软件开发及相关领域知识与创新的传播

参会咨询

查看会议

4月 · 北京 · 1200人

QCon 全球软件开发大会

会议时间：4月16-18日

- Agentic Engineering
- AgentOps
- 下一代模型架构与推理优化
- AI 原生基础设施
- 知识工程实践
- AI 安全

6月 · 上海 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：6月26-27日

- AI Infra 系统工程
- 多 Agent 协作与实践
- 多模态融合
- 模型训练与推理创新
- 数据平台与特征服务

8月 · 深圳 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：8月21-22日

- Agentic AI
- 轻量化与高效推理
- 多模态应用
- AI + IoT 场景实践
- AI 工业化落地

10月 · 上海 · 1200人

QCon 全球软件开发大会

会议时间：10月22-24日

- AI Agent
- Vibe Coding
- 智能可观测
- 推理基建
- 模型攻防
- AI x 创造力

12月 · 北京 · 1000人

AiCon 全球人工智能开发与应用大会

会议时间：12月18-19日

- 大模型架构创新
- 多模态 AI 产业融合
- 具身智能
- AI for Science
- 大模型安全

InfoQ 极客传媒

InfoQ 极客传媒

THANKS

探索 AI 应用边界

Explore the limits of AI applications

AiCon

全球人工智能开发与应用大会