

重磅！苹果要做 AI 搜索了

苹果要自研 AI 搜索了！

知名苹果爆料人、彭博社记者马克·古尔曼(Mark Gurman)最新爆料,苹果计划于明年春季推出自研 AI 搜索工具“答案引擎”。

据知情人士透露,这款内部代号为“世界知识问答”的新系统将集成于 Siri,未来还会扩展至 Safari 网页浏览器及 iPhone 主屏搜索工具 Spotlight。

该功能的运作模式类似 ChatGPT、谷歌搜索的 AI Overviews 及其他 AI 搜索应用,核心技术依托于大语言模型,古尔曼爆料,这一功能将与谷歌 Gemini 模型合作。

要知道,谷歌每年向苹果支付巨额费用,以维持其作为 Safari 浏览器默认搜索引擎的地位。苹果推出自研 AI 搜索,无疑会分走谷歌搜索的“入口”流量。

谷歌也不傻,苹果自研 AI 搜索是必然之举,既然注定要对谷歌搜索产生影响,“这块蛋糕”与其让其他 AI 搜索产品抢走,不如把握在自己手里。

一、新版 Siri 配备“AI Overviews”,最快明年3月推出

Siri,这款2011年发布时具有突破性意义的语音助手,如今已成为苹果 AI 短板的具体体现。

当前 Siri 可回答基础问题并提供知名人物、事件、电影及体育赛事

事实信息,但在处理复杂查询和常识性搜索时,Siri 往往会转而提供谷歌或 ChatGPT 的搜索结果。

早前苹果曾承诺,Siri 将能够调用个人数据与屏幕内容实现更精准响应查询,并通过语音更精确操作用户设备。

如今承诺或将兑现,苹果正计划通过此次更新实现更大突破,代号“Linwood”和“LLM Siri”的技术重构为 Siri 的 AI 搜索功能奠定基础。据透露,苹果全新搜索体验将覆盖文本、图片、视频及本地兴趣点的交互界面,并将配备类似 AI Overviews 的系统,可以提供更快速且更精准的搜索体验。

新版 Siri 与搜索功能目前计划纳入内部代号为“Luck E”的软件更新,对应 iOS 26.4 系统,预计最早于明年3月发布。

iOS 26 初始版本将于北京时间9月10日凌晨随苹果旗舰产品 iPhone 17 系列同步推出。

苹果软件工程主管、Siri 战略总负责人克雷格·费德里吉(Craig Federighi)在近期全员会议上暗示了这次升级范围将扩大:“我们对 Siri 端到端重构的成果已达到预期,这使我们不仅能兑现已宣布的功能,还将带来远超设想的重大的升级。”

苹果内部多个团队正协同推进搜索计划,包括费德里吉领导的 Siri 团队、约翰·詹南德雷亚(John

Giannandrea)执掌的 AI 部门以及埃迪·库(Eddy Cue)管理的服务业务部门。

Vision Pro 头显创建者迈克·罗克韦尔(Mike Rockwell)在费德里吉指导下牵头项目,前 Siri 主管罗比·沃克(Robby Walker)则在詹南德雷亚麾下担任项目关键推动者。

苹果也考虑过开发类似聊天机器人的搜索应用,据彭博社上月报道,苹果正为新成立的“答案、知识与信息”(AKI)团队招募员工,该团队正参与搜索开发工作。

除了搜索功能的升级,苹果还计划重新设计 Siri 的视觉界面,同时开发健康 Agent,为其计划2026年推出的付费健康订阅服务提供支持。

与此同时,苹果还在推进另一项 Siri 更新,这将让未来的家用设备具备更出色的对话能力。

但古尔曼预计,苹果下周发布 iPhone 17 时不会推出更加新颖和重大的 AI 功能。

二、围绕三大核心组件重构,外部模型与自研模型合作

苹果正围绕三大核心组件重构 Siri:规划器、网络与设备搜索系统,以及摘要生成器。

规划器负责解析语音或文本输入并制定响应策略,搜索系统扫描网络或用户数据,摘要生成器则

整合信息生成最终答案。

重大转变在于,苹果正考虑通过“Glenwood”项目部分采用第三方 AI 模型驱动新版 Siri。

据知情人士透露,苹果近期倾向于为摘要生成器采用定制版谷歌 Gemini 模型,该模型将运行于苹果自研私有云计算服务器上。谷歌已向苹果交付技术,双方正协作进行微调与测试。

苹果同时评估了将谷歌模型用于规划器功能的可能,但仍在持续考量使用 Anthropic 的 Claude 模型或自研方案的可行性。

当然,苹果也计划在用户搜索功能使用“苹果基础模型”自研技术,这样用户信息就无需经过第三方技术处理,隐私保护更有保障。

三、谷歌更划算,苹果还曾考虑收购 Perplexity 和 Mistral

谷歌并非 Siri 项目的最初人选,Anthropic 一度处于合作谈判领先地位,苹果内部评估认为 Claude 的生成质量优于 Gemini。

但据知情人士透露,Anthropic 报价年费超过15亿美元(约合人民币107.1亿元),而谷歌提供了更优惠的条件。

这也并不是苹果与谷歌第一次就 AI 相关合作进行谈判,去年苹果曾接洽谷歌希望将 Gemini 聊天机器人集成至 Siri,但最终选择与

OpenAI 合作。苹果和 OpenAI 双方虽公开声明仍在寻求达成协议,但该功能至今未向消费者推出。

在新 Siri 开发过程中,埃迪·库(Eddy Cue)与并购主管阿德里安·佩里卡(Adrian Perica)评估了一系列潜在收购目标,据彭博社报道,苹果曾讨论收购 Perplexity 和 Mistral。

苹果曾认真评估了 Perplexity 的技术,但目前已不再积极考虑收购,因为苹果自有搜索产品将与该初创公司的服务产生直接竞争。但苹果仍在寻求其他收购机会,不仅是为了新产品,更是为了补充人才储备。

此外,美国法官本周刚刚裁定,苹果仅需微调就可继续将谷歌作为其设备默认搜索引擎,古尔曼认为,这将降低苹果开发自研服务的紧迫性。

结语:苹果正逐步补全其 AI 基础能力

Siri 此前因复杂查询依赖第三方搜索、多场景适配不足等原因,逐渐落后于行业节奏。此次爆料出的 AI 搜索的升级规划,本质上是苹果补全 AI 基础能力并在 AI 赛道追赶行业步伐的必然动作。

AI 搜索功能的最终落地效果,不仅关乎苹果自身的竞争力,还承担着无数用户的期待。

刚刚,又一个万亿 AI 独角兽诞生,直逼字节 OpenAI

今天凌晨,大模型独角兽 Anthropic 宣布完成130亿美元(折合人民币约928亿元)F轮融资,投后估值达到1830亿美元(折合人民币约1.3万亿元),约是今年3月615亿美元(折合人民币约4391亿元)估值的近三倍。

这也意味着 Anthropic 成为全球仅次于 OpenAI、字节跳动,估值排名全球第三的 AI 独角兽企业。

目前全球前两大 AI 独角兽分别是字节跳动(估值3300亿美元,折合人民币2.4万亿元)、OpenAI(估值3000亿美元,折合人民币2.2万亿元)。

这家由 OpenAI 前研究副总裁达里奥·阿莫迪(Dario Amodei)、丹尼拉·阿莫迪(Daniela Amodei)兄妹联合创立的 Anthropic,也一度被认为是 OpenAI 最有力的挑战者之一。

Anthropic 首席财务官 Krishna Rao 在博客文章中提到:“我们正在整个客户群中看到需求的指数级增长。”2025年初,Anthropic 的年化收入达到10亿美元(折合人民币约71.4亿元),8个月后的今天年化收入已经突破50亿美元(折合人民币约357亿元)。

从客户层面看,其博客透露,Anthropic 目前服务超过30万企业客户,大客户数量过去一年中增长了近7倍,而每个大客户意味着能为其带来超过10万美元(折合人民币约71.4万元)的年收入。

如今其业务增长的主要推动力就是今年爆火的 AI 编程神器 Claude Code。Claude Code 今年5月由研究预览转为正式公开版,三个月时间使用量增长超10倍,产生超5亿美元(折合人民币约36亿元)年收入。

本轮融资由风投公司 ICONIQ、Fidelity Management & Research Company、Lightspeed Venture Partners 共同领投,卡塔尔投资局、新加坡政府投资公司(GIC)等主权基金参投。其博客提到新融资将用于扩大企业采用规模、深化安全研究,并支持国际扩张。

值得一提的是,Anthropic 此前已完成9轮融资,总融资额达到

300.4亿美元(折合人民币约2145亿元)。

一、3个月营收超5亿刀, AI 编程工具成增长主力

2021年,时任 OpenAI 前研究副总裁的 Dario Amodei 带着他妹妹丹尼拉·阿莫迪(Daniela Amodei)在内的逾十人研发小组出走成立 Anthropic。

如今这家创企已成为 OpenAI 最有力的竞争对手。

2023年,Anthropic 发布发布了聊天机器人 Claude 的首个版本,今年8月已经更新至 Claude 4.1。

在 Anthropic 的官方博客中提到,新模型在智能体工具调用与多语言问答能力上超越 OpenAI o3,在智能体编程方面的分数为74.5%和43.3%,大幅超越 OpenAI o3 和 Gemini 2.5 Pro,但其在视觉推理和数学能力上与 OpenAI o3 和 Gemini 2.5 Pro 还有一定的差距。

此外,今年2月,Anthropic 还推出自家 AI 编程工具 Claude Code,今年5月宣布全面开放,从研究预览转为正式产品,Anthropic 的博客提到,如今 Claude Code 已成为开发人员的首选工具。

目前,Claude Code 已集成其最强模型 Claude Opus 4.1,它具备代码库感知能力,并且能够在用户的开发环境中直接编辑文件和运行命令。

自公开发布至今,Claude Code 三个月时间使用量增长超10倍,产生超5亿美元(折合人民币约36亿元)年收入。

二、9轮融资合计超300亿美元,谷歌亚马逊追着投

自2021年以来,自2021年以来,Anthropic 已完成8轮融资,共融资170.4亿美元(折合人民币约1216亿元),目前正在完成的是第9轮融资,具体融资轮次及金额如下:

2021年5月:A轮融资,金额1.24亿美元(折合人民币约8.9亿



元),主要投资人为 Dustin Moskowitz、Eric Schmidt、Jaan Tallinn。

2022年4月:B轮融资,金额5.8亿美元(折合人民币约41.7亿元),主要投资人为FTX。

2023年2月:3亿美元(折合人民币约21亿元),谷歌投资,成为战略股东。

2023年5月:C轮融资,金额4.5亿美元(折合人民币约32.3亿元),投后估值41亿美元(折合人民币约294.4亿元),Spark Capital 领投,谷歌、Salesforce Ventures、Zoom 等参投。

2023年9月:亚马逊宣布战略合作,初始投资12.5亿美元(折合人民币约89.8亿元),并计划总计投资高达40亿美元(折合人民币约287.2亿元)。随后亚马逊成为其主要云服务提供商。

2023年10月:谷歌投资5亿美元(折合人民币约35.9亿元),并承诺未来追加15亿美元(折合人民币约107.7亿元)。

2024年2月,完成7.5亿美元(折合人民币约54亿元)D轮融资,由Menlo Ventures 领投。

2024年3月和5月,连续在二级市场筹集8.84亿美元(折合人民币约63亿元)和4.52亿美元(折合人民币约32亿元)。

2024年11月:金额40亿美元(折合人民币约287.2亿元),主要投资人为亚马逊。

2025年3月:E轮融资,金额35亿美元(折合人民币约250亿元),光速创投领投,当时估值为615亿美元(折合人民币约4391亿元)。

谷歌、亚马逊、Salesforce 等知名科技公司均在其投资方之列。

三、高运营开支是隐患,投资者更看重营收飙升

据爆料,Anthropic 近期告知投资者,其直接向客户销售 AI 模型和 Claude 聊天机器人产品的毛利率约为60%,并正朝着70%迈进。该毛利率数据通常指扣除服务器成本和客户支持费用后的利润占比。知情人士解释道,这些毛利率会随公司对计算资源的规划使用效率而波动。

知情人士还补充道,今年早些时候,Anthropic 通过亚马逊服务和谷歌云销售 Claude 模型的毛利率为-30%,猜测原因可能是亚马逊和谷歌在向云客户转售 Anthropic 模型时抽取了可观分成。

但通过云服务商产生的收入可能只占 Anthropic 总收入的少数,截至2024年底,其约70%收入来自直销。目前尚不清楚直销收入的具体占比,但最新披露信息显示,Anthropic 的整体毛利率可能还将维持在50%-55%,与2023年底持平。

作为对比,OpenAI 今年初预测

2025年毛利率为48%,并预计将稳步提升,到2029年达到70%。两家公司的毛利率计算方式是否一致尚不明确。

值得注意的是,Anthropic 与 OpenAI 的毛利率均未反映其每年投入 AI 研发的数十亿美元资金及人力等运营开支,而这些支出都远高于传统软件企业。

虽然 AI 模型研发成本高昂,但因为算法改进和 AI 芯片使用效率提升,OpenAI 与 Anthropic 展现出的模型运行成本均呈现下降趋势,这令投资者稍感宽慰。不过当前投资者更关注的是两家公司惊人的营收增长,双方均有望大幅超越年初设定的乐观营收目标。

结语:资本向 AI 领域头部创企集中

根据全球咨询公司 Crunchbase 此前的数据,全球第二季度约有400亿美元(折合人民币2856亿元)的风险投资资金流向 AI 领域,其中基础模型公司筹集55亿美元(折合人民币393亿元)。Anthropic 此次新融资也说明,风险投资正越来越多地集中投向已经获得充足资金的大型初创公司。

同时,除了此前的风投机构、知名科技巨头,此次该公司的融资也涌入了诸多国家主权资金,这也表明投资者对这家 AI 独角兽价值的认可。