

Recursive 发布本地代码模型 Flint

中国高技术产业发展促进会新质生产力工作委员会

博雅云创 & 中科创新驱动

2026 年 10 月 22 日

摘要

AI 技术的最新进展集中在本地代码模型、智能体稳定性、行业专用模型、授权交易接口和软件供应链安全。10 月 21 日，Recursive 发布可在个人电脑运行的 Flint 编程模型，并披露 AI 辅助研发流程；CTGT 以 Apache 2.0 许可推出 Concordance，追踪智能体工作流输出不一致的真正原因；Deel 发布面向人力资源和薪资业务的 Pangaea 模型；沃尔玛公布允许获授权智能体购物的 Agent Gateway；OpenAI 联合 PHP 维护者及安全企业披露重大漏洞修复协作。这些具体动作说明，AI 应用竞争正在从模型单点性能转向成本、可追溯性、业务许可和长期安全维护。

Contents

一、Recursive 公开 Flint：AI 辅助研发的轻量代码模型进入可下载阶段	2
二、CTGT 发布 Concordance：智能体结果不一致可能来自应用编排，而非模型随机性	2
三、Deel 发布 Pangaea：人事与薪资问答向受控专用模型转移	3
四、沃尔玛发布 Agent Gateway：第三方智能体购物走向明确授权接口	3
五、OpenAI 与开源维护者协同修复 PHP 漏洞：AI 安全发现进入大规模补丁协调	4
趋势观察	4
参考资料	5

一、Recursive 公开 Flint：AI 辅助研发的轻量代码模型进入可下载阶段

美国创业公司 Recursive 于 10 月 21 日发布首款模型 Flint。根据公司向 TechCrunch 介绍的情况，其内部自动化研究系统 Eureka Machine 参与了模型开发多个环节，团队用不足四个月完成研发。Flint 采用混合专家架构，总参数约 400 亿，每次推理激活约 25 亿；权重以 Apache 2.0 许可公开，定位为能够在高配置个人电脑上离线运行的代码模型。开发方报告在搭载 M4 Pro 的 MacBook Pro 上达到约每秒 131 个 Token，但速度及编程基准均需以同等硬件和任务条件复测。[1][2]

此次发布值得关注的是研发流程与部署规模的结合，而不是“AI 已完全自主研发 AI”。公司明确存在人工监督，离线可用也不代表大型仓库级任务无需外部工具。对开发团队而言，可下载权重便于开展私有代码评测，并把模型能力、补全错误、测试通过率和资源消耗列入同一张测评表；当模型规模缩小，工程化提示、任务拆解和代码审查的重要性反而上升。

二、CTGT 发布 Concordance：智能体结果不一致可能来自应用编排，而非模型随机性

10 月 21 日，AI 安全与可解释性创业公司 CTGT 宣布开放 Concordance，包括测试协议、参考实现和实验数据，许可证为 Apache 2.0。其公开案例使用同一份九页银行业务手册重复提取合规义务，四次任务得到不同的义务集合；但对其中重复出现的 74 次相同模型请求，模型回答始终一致。真正变化来自上传文件标识、路由列表顺序及中途扩张的主题注册表。团队在稳定标识、固定排序和隔离测试环境后，五次运行均提取出相同的 34 项义务。[3]

这一例子给企业 Agent 平台提供了明确的排障顺序：先保存实际输

入模型的完整请求，再比较检索、上下文组装、工具反馈和最终记录环节，而不能一遇到不一致就降低模型温度或更换基础模型。Concordance 能够帮助发现差异发生在哪一层，但该实验针对一个应用和一份材料，不能据此宣称所有行业智能体已有同等故障发生率。

三、Deel 发布 Pangaea：人事与薪资问答向受控专用模型转移

全球人力资源与薪酬服务商 Deel 于 10 月 21 日推出 Pangaea 模型系列及专家数据评测业务 Emet。Pangaea K1 侧重各国劳动规则与工作程序，P1 侧重薪资；K1 基于约 90 亿参数的既有模型进一步训练，参考超过 4.35 万页专业材料，业务资料由两千余名领域专家维护。公司声称在自建任务评测中，K1 答复中位时长 7.7 秒，并报告显著的单位答复成本优势。[4]

这不是单纯把大模型换成小模型。跨国薪资业务涉及法域、版本生效日期和具体雇佣关系，回答质量取决于训练材料的更新、检索验证和人工审核机制。厂商所称最高约 98% 的答复成本节约缺少完整、统一的训练与运维成本比较，应与模型推理支出区别计算。模型也不能替代当地专业人员对最终薪资和合规事项的确认。

四、沃尔玛发布 Agent Gateway：第三方智能体购物走向明确授权接口

10 月 21 日，沃尔玛宣布面向获得批准的外部购物智能体推出 Agent Gateway，通过 MCP 接口连接商品检索和购物流程。与部分平台限制自动化访问不同，沃尔玛同时设计 Agent Profiles，为不同代理建立独立授权身份、购买记录和限额控制；该功能计划下个月推出。平台还明确要求智能体不得超出用户意图操作，也不能直接接触用户支付资料。[5]

此举的新意不在于允许程序访问网页，而是将平台许可、用户授权和交易责任写入业务接口。对购物、旅行和企业采购智能体而言，能搜到商品信息不等于获得下单资格，代理身份、审批阈值、退货规则和撤销机制需要由服务方执行。Agent Gateway 已经对外公布，但不能把预定下月上线的全部账户功能描述为现已普遍可用。

五、OpenAI 与开源维护者协同修复 PHP 漏洞：AI 安全发现进入大规模补丁协调

OpenAI 于 10 月 21 日披露“Patch the Planet”项目近期发现和协调修复 PHP 关键漏洞的工作。安全企业 Calif 使用 AI 工具发现重大问题，Wiz 也独立发现并报告相关漏洞；PHP 维护团队主导修复发布，Cloudflare 等合作方参与缓解和通知。OpenAI 同步提供用于检查 PHP 版本的 Codex 技能，并宣布为资源有限的关键服务防御方投入 10 亿美元规模的补贴访问额度。[6][7]

这是一项真实的软件供应链安全事件，而非“AI 自动修好了互联网”。发现漏洞、开发补丁、兼容老旧环境和推动全球网站更新，是相互独立的工作。企业尤其需要核对实际 PHP 版本、依赖与供应商公告，依照维护者指导完成升级，不应把未经验证的网络扫描或自动修复交给高权限智能体。由于攻击者也可能借助 AI 分析补丁，安全披露节奏和防御部署同样重要。

趋势观察

本地模型、可观测执行框架、专用行业模型、合法交易接口和漏洞响应体系正在构成新一代 AI 应用的五个基础部件。真正可持续的 Agent 服务，必须同时回答“能完成什么任务、为何给出这个结果、允许执行到哪一步、花费多少以及出现问题由谁负责”。

参考资料

1. TechCrunch | 《Recursive launches Flint, a laptop coding model it says was largely built by AI》 | 2026-10-21 | 用途: 核验 Flint 发布、参数与厂商推理测试。
<https://techcrunch.com/2026/10/21/recursive-launches-flint-a-laptop-coding-model-it-says-was-largely-built-by-ai/>
2. Recursive 官方博客 | 《Introducing Flint》 | 2026-10-21 | 用途: 核验模型结构、公开权重与推理适用条件。
<https://recursive.com/updates/introducing-recursive-flint>
3. CTGT 研究原文 | 《Deterministic Inference Did Not Make Our Pipeline Deterministic》 | 2026-10-21 | 用途: 核验 Apache 2.0 开源、对照实验及应用编排故障源。
<https://www.ctgt.ai/research/concordance>
4. VentureBeat | 《Deel launches Pangaea AI models for HR and payroll, betting specialists can cut the cost of answers by 98%》 | 2026-10-21 | 用途: 核验 K1/P1、Emet、专家材料及成本声明适用边界。
<https://venturebeat.com/technology/deel-launches-pangaea-ai-models-for-hr-and-payroll-betting-specialists-can-cut-the-cost-of-answers>
5. TechCrunch | 《While Amazon blocks AI agents, Walmart rolls out the welcome mat》 | 2026-10-21 | 用途: 核验沃尔玛 Agent Gateway、MCP 连接、下月计划上线的 Agent Profiles 和授权限制。
<https://techcrunch.com/2026/10/21/while-amazon-blocks-ai-agents-walmart-rolls-out-the-welcome-mat/>

6. OpenAI 官方 | 《Patch the Planet: global response coordination as agents uncover critical vulnerabilities》 | 2026-10-21 | 用途: 核验 PHP 漏洞协调修复、维护者职责和补贴计划。

<https://openai.com/index/coordinated-disclosure-of-critical-vulnerabilities/>

7. Reuters | 《Government websites go dark after AI-discovered flaw sparks security scramble》 | 2026-10-21 | 用途: 交叉核验 PHP 风险披露与全球安全响应。

<https://www.reuters.com/world/asia-pacific/eu-agency-warns-websites-may-need-be-taken-offline-php-prepares-patches-2026-10-21/>

8. TechCrunch | 《Blacksmith hits unicorn valuation after Sequoia leads Series C》 | 2026-10-21, 延伸阅读 | 用途: 观察 AI 代码增加后持续集成测试平台的融资及产能需求; 未作为本期主新闻。

<https://techcrunch.com/2026/10/21/blacksmith-hits-unicorn-valuation-after-sequoia-leads-series-c/>

联系我们，请扫描二维码



新质生产力工作委员会
官方公众号



工业智能算网
gyznsw.cn

新质生产力工作委员会：

中国高技术产业发展促进会新质生产力工作委员会，专注于推动工业人工智能、智能制造、数字化转型等前沿技术发展，为企业提供政策解读、技术咨询和产业对接服务。

工业智能算网：

专注于工业人工智能、新质生产力、工业软件 CAE、智能制造等前沿技术。提供每日动态分析、技术趋势解读、解决方案分享，推动工业智能化转型。

网站地址：<https://gyznsw.cn>