AI 技术每日分析：AutoGPT beta v0.11.3 修复安全问题，新增远程 PC 执行目标

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 11 月 16 日

摘要

AutoGPT 平台版 11 月 15 日发布 beta v0.11.3, 修复多项内部接报的安全问题，并调整远程本地 PC 执行、模型适配和任务调度。另一项值得留意的更新来自模型网关：LiteLLM 于北京时间 11 月 15 日发布 1.108.9 LTS 补丁，修复流式接口在思考内容与正文混合输出时遗漏正文开头的问题。两项更新分别落在智能体的执行环境和模型输出协议上。NVIDIA Auto Ontology、Edge0 端侧模型及浏览器智能体的进展也列入本文，但不算作 11 月 15 日的新发布。

Contents

AutoGPT beta v0.11.3：安全修复与执行目标更新 **2**
LiteLLM 1.108.9：修复混合流中的正文遗漏 **2**
NVIDIA Auto Ontology 的企业语义层结构（背景） **3**
Edge0 发布端侧 145B 模型（背景） **3**
浏览器智能体与评估组件继续迭代（背景） **3**
趋势观察 **4**
参考资料 **4**

AutoGPT beta v0.11.3：安全修复与执行目标更新

AutoGPT 项目 11 月 15 日在 GitHub 发布 autogpt-platform-beta-v0.11.3。发行公告列出数项内部接报的安全问题，称详细安全公告将在生产修复上线后公布。新版增加远程本地 PC 执行目标、原生移动端服务基础、Apple 登录、专家对话分页，以及 Gemini 4 Hydrogen 和 Carbon 模型适配。

发行记录还涉及任务调度：旧机制可能跳过设有一秒宽限时间的智能体任务，新版处理了这一边界。多租户资源结算、账户关闭后的记录处理和 Webhook 验签也有相应改动。这些细节比泛称“智能体能力升级”更贴近企业部署：任务由谁发起、在哪台设备上执行，运行后留下什么记录，都需要说清楚。

LiteLLM 1.108.9：修复混合流中的正文遗漏

LiteLLM GitHub 发行页显示，v1.108.9 于 11 月 14 日 22:00 UTC 发布，即北京时间 11 月 15 日 06:00。该版本把多项修复回移到 1.108 长期维护分支。其中一项针对 Responses 协议转换：当思考内容和正式回答出现在同一数据块时，流式输出可能缺失正文开头，而最终聚合结果仍完整；此次补丁修复了这一问题。

版本还修复 Azure AI Foundry 中部分模型的推理强度传递及推理内容保留、Bedrock 的 S3 媒体输入和向量检索过滤映射，并增加 Vertex 请求调用者身份标签与 Vercel AI Gateway 决策模型支持。仓库公开了版本标签及对应的拉取请求，容器镜像提供 cosign 签名校验路径。企业若通过同一个网关接入多家模型，流式输出是否完整，会影响工具调用、答案落盘、日志追溯和下游系统接收的数据。

NVIDIA Auto Ontology 的企业语义层结构（背景）

NVIDIA Auto Ontology 代码仓库在 11 月 15 日前后受到社区关注，但项目创建和开发早于这一天，因此这里只作为技术背景。项目将数据库元数据整理为包含业务含义、实体关系和指标定义的语义层，连接 PostgreSQL、MySQL、Snowflake、ClickHouse 等数据源，通过 MCP、CLI 和 API 向智能体提供上下文。公开架构包括前端、FastAPI 后端、元数据采集进程、Postgres 与 pgvector，以及可选的凭据管理组件。

制造企业的“停机时间”“合格产出”“订单准交”可能分散在不同系统中。整理指标口径、映射来源并保留查询结果引用，能为智能体接入 ERP、MES、QMS 提供一层共同的解释。语义层也因此可以作为单独部署和维护的企业软件组件。

Edge0 发布端侧 145B 模型（背景）

Edge0 团队 11 月 12 日发布 Edge0-145B 和 120B 模型，仓库 11 月 13 日记录了配套推理引擎更新。145B 版本总参数量为 145B，厂商公开的每次推理激活参数约为 10.12B。其方案采用压缩专家表示，并在 GPU、系统内存和 SSD 之间分层调度；官方资料给出单机性能测试，展示 Q3_K、NVFP4 两种权重形态，也分别说明软件框架和模型权重的许可条件。

这不是 11 月 15 日的新发布。它提供的技术背景是：本地运行大模型时，除了显存能否一次装下权重，缓存命中率、主机内存带宽、存储读取和推理调度也会影响配置与调优。企业在固定硬件上部署岗位模型时，需看这些具体指标。

浏览器智能体与评估组件继续迭代（背景）

Browser Use 在 11 月 14 日前后继续迭代 1.0 系列，其开发路线从传统 DOM 动作序列转向持续运行的 Python 执行环境与 Chrome DevTools

Protocol。Ragas 等评估与文档处理项目也发布了修复版本。这些属于背景信息，不与 AutoGPT、LiteLLM 的当日发布混为一谈。

工业智能体要长期运行，可以把岗位任务拆成可重复测试的单元，分别检查模型回答、工具执行、调用者身份和审计输出。权限、模型路由与运行记录也要能对上。本文所列 11 月 15 日进展中，发布信息最明确的仍是 AutoGPT 的平台修复和 LiteLLM 的流式协议修复。

趋势观察

流式模型网关处在用户界面、业务智能体与不同厂商模型之间，少一个字符也可能影响工单解析。原文中的技术支持场景是：现场故障描述被整理为结构化请求，模型给出检修建议，再把字段写入维修任务。若思考片段与正式答案拼接时漏掉开头，前端和日志可能留下不同内容。LiteLLM 此次修复针对输出一致性；AutoGPT 关于权限和调度的改动，则涉及执行是否按预期发生。固定请求回放、工具调用记录和业务字段比对，是原文提出的对应检查办法。

参考资料

1. AutoGPT ｜ Release autogpt-platform-beta-v0.11.3 ｜ 2026-11-15 ｜ https://github.com/Significant-Gravitas/AutoGPT/releases/tag/autogpt-platform-beta-v0.11.3

2. LiteLLM ｜ Release v1.108.9 ｜ 2026-11-15 ｜ https://github.com/BerriAI/litellm/releases/tag/v1.108.9

3. NVIDIA ｜ Auto Ontology ｜项目持续更新｜ https://github.com/NVIDIA/auto-ontology

4. Edge0 ｜ Introducing Edge0-145B: Frontier AI Capabilities on Your Own Computer ｜ 2026-11-12 ｜

https://edge0.ai/blog/introducing-edge0-145b-frontier-ai-capabilities-on-your-own-computer

5. Browser Use | Releases | 11 月 14 日更新背景 | https://github.com/browser-use/browser-use/releases

机构介绍与关注

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。

关注高促会新质生产力工委会公众号

关注工业智能算网平台

https://gyznsw.cn