某 AI Agent 静默上传用户代码，AI 编程工具的数据合规边界再受关注

作者：朱晓阳 | 邓梓珊

2026 年 9 月 18 日，一位开发者在清理本地磁盘时发现一个 313MB 的加密压缩包，经逆向分析发现，某 AI 编程工具在用户登录状态下，会将整个工作区代码仓库(包括完整 Git 提交历史、LFS 缓存及本地操作记录)静默打包加密后上传至阿里云 OSS。相关发现随即在开发者社区引起关注，争议也从产品功能设计进一步延伸至企业源代码保护、个人信息处理以及数据跨境合规等问题。

一. 事件进展：从产品功能争议到企业用户发函追责

9 月 18 日，该工具的所属公司 Z 公司就前述争议发布情况说明，称相关问题源于 AI Agent 的“代码库索引”功能。按照 Z 公司的解释，该功能原本用于支持会话检查点恢复、历史版本回退及 Repo Wiki 等功能，其中 Repo Wiki 在云端生成知识页面时可能触发仓库数据上传。该功能在产品上线初期默认开启，相关数据在云端生成 Wiki 页面后即销毁，目前问题已经修复。

9 月 21 日，Z 公司进一步宣布正式开源相关 AI Agent，并表示已经邀请中国信息通信研究院和绿盟科技开展安全审计。Z 公司披露的审计结果显示，相关阿里云 OSS 存储桶处于“云端零数据”状态，AI Agent 已经移除 Repo Wiki 入口及相关生成链路，同时切断本地仓库快照生成和上传路径。公司还表示，相关数据未被用于模型训练。

与此同时，事件已由产品层面的技术争议进一步延伸至企业用户的法律追责。9 月 19 日，太原一公司向 Z 公司发出法律函件，称经其独立取证，Z 公司的 AI Agent 存在自动、批量上传企业数据资产及商业秘密的情况，并要求 Z 公司说明相关数据的处理范围、具体去向、删除情况、第三方访问以及是否发生跨境传输等事项，同时保留索赔、投诉举报及提起诉讼的权利。

如您需要了解我们的出版物，
请联系：

Publication@llinkslaw.com

二. 事件法律性质：数据处理的知情同意与最小必要边界

太原公司的独立取证结果显示，被上传数据包含项目完整源代码及系统架构设计、完整的版本控制历史(含已删除的历史提交和引用日志)、数据库口令、接口密钥、云服务凭证及证书、员工及终端用户的个人信息，以及未公开的研发规划和商业计划。取证还显示，一个工作区被上传的文件达 32,932 个、明文字符约 4.11 亿；另一项目涉及 1,947 个文件、1.44 亿字符，仅因上传失败 32 次才未传至云端。上传行为系自动触发、批量发生，不以用户明示或默示同意为前提。

若上述取证结果属实，则表明 Z 公司的行为背离了《中华人民共和国个人信息保护法》(以下称"《个保法》")项下的两项核心原则:

- **知情同意原则**

  《个保法》第十七条要求处理个人信息前以显著方式、清晰易懂的语言真实、准确、完整地告知处理目的、方式、种类、保存期限等事项。

  太原公司在函件中指出，其"完全没有打开和使用所谓'索引功能'，处于隐私模式，AI Agent 却依然默认上传文件"。这意味着 Z 公司的数据处理行为在用户未主动开启相关功能的情况下即已发生，告知义务和同意获取均存在根本性缺失。对于代码仓库中可能包含的员工姓名、邮箱、账号标识等个人信息，数据处理者需在收集前履行告知义务并取得同意，方能满足合规要求。

- **最小必要原则**

  《个保法》第六条确立了个人信息处理的最小必要原则，要求处理活动"限于实现处理目的的最小范围"，不得过度收集。这一原则在代码数据处理场景中的适用，需要结合数据类型的特殊性进行判断。

  与一般的业务数据不同，代码仓库中的 Git 历史记录并非"附带生成"的日志文件，而是从时间维度完整保留了项目演进的全部痕迹，包括已被删除的代码、曾经硬编码的密钥、被 force push 覆盖的提交以及 reflog 中的操作轨迹。这意味着，以"代码索引"为由上传全量 Git 历史，其数据范围远远超出了索引功能本身所必需的最小限度。

  网络安全正高级工程师袁博在接受《中国新闻周刊》采访时指出，当前代码代表项目此刻的状态，而 Git 历史数据是时间维度上的全量档案，包含已删除的机密信息、密钥甚至测试数据，"Git 历史记录的信息量更大更全，非法获取 Git 历史数据的危害性显然更大"。[1]以"代码索引"为由上传完整 Git 历史，其必要性缺乏充分的论证依据。

[1] "发生得很突然"，机密信息被泄露？ - 中国新闻周刊网

三. 法律风险分析：AI 工具“静默上传”数据的场景

(一) 商业秘密保护

企业源代码在符合不为公众所知悉、具有商业价值并采取保密措施的条件下，可构成《中华人民共和国反不正当竞争法》意义上的商业秘密。2026 年 6 月 1 日起施行的《商业秘密保护规定》第五条进一步将“数据、算法、计算机程序、代码”等明确纳入技术信息范围，为软件及 AI 企业的代码资产保护提供了更为清晰的规范依据。

从法律构成要件分析，商业秘密侵权需证明: (1)涉案信息符合商业秘密构成要件; (2)行为人不正当获取或披露; (3)存在损害或损害危险。

本案中，由于新规将“代码”纳入技术信息保护范围，因而第一项要件满足；第二项的认定取决于数据处理行为是否超出授权范围。太原公司已明确指出，其并未开启“索引功能”，属于隐私模式，数据上传系“自动触发、批量发生”。关于第三项要件，依照太原公司的说法，被上传数据涉及其自研引擎及相关知识产权，并可能包含数据库密钥等敏感信息。如密钥泄露，可能进一步导致用户数据泄露，太原公司亦须承担相应责任，故而存在损害危险。

(二) 数据出境合规

太原公司经查证发现，该 AI Agent 客户端的全部网络请求均指向同一服务域名体系下的两个关联域名。该客户端的缔约主体及个人数据控制者是一家注册于新加坡的公司，其隐私政策载明“服务通常自新加坡提供”“您的个人数据通常在新加坡处理”。而产品服务协议的签约方为 Z 公司旗下的北京某公司，由此形成责任主体与数据处理主体的分离。港交所招股书显示，该新加坡公司系前述北京公司于 2023 年 11 月在新加坡注册的间接全资子公司。

依据《个保法》第三十八条及第三十九条，个人信息出境需履行告知义务并取得个人的单独同意，同时须满足安全评估、标准合同或保护认证三者之一的合规路径。2026 年 1 月 1 日起施行的《个人信息出境认证办法》进一步明确，个人信息处理者在申请认证前应履行告知、取得单独同意、进行个人信息保护影响评估等义务。

本案中，如太原公司员工个人信息及客户信息随代码仓库被传输至新加坡主体，且未履行上述任一合规路径，则可能构成数据出境违规。太原公司在函件中要求 Z 公司说明责任主体归属、数据是否发生出境传输，正是基于这一法律框架提出的合规追责。太原公司负责人明确表示，“我们之前完全没有想过这件事”，数据出境后“有可能不受国内法律监管”。这一表述反映出企业在采购 AI 工具时对数据出境问题存在认知盲区。

(三) 供应商数据处理协议的合同风险

企业采购 AI 工具时，通常与供应商之间形成委托处理或共同处理关系。本案中，AI Agent 隐私政策未明确告知采集完整 Git 历史，UI 中虽有“体验优化”和“仓库快照索引”开关，但无法阻止上传，该机制默认开启且手动删除加密包后会自动重建。这意味着用户即便在合同层面依赖隐私政策作出“数据不会被上传”的判断，实际数据处理行为仍可能构成合同违约。太原公司函件中要求 Z 公司说明“是否向第三方提供”“是否用于模型训练”等内容，本质上是在核实合同履行是否符合约定。

在欧盟 GDPR 和我国《个保法》的框架下，引入第三方工具的企业始终作为数据控制者存在，对数据的全生命周期安全承担首要合规责任。这意味着，即使 Z 公司作为供应商存在违规行为，太原公司作为数据控制者，仍可能因未能履行对供应商的监督义务而面临合规风险。

四. 企业合规建议

第一，采购阶段核查数据处理链路，而非仅审阅隐私政策。 企业应结合技术团队对 AI 工具进行实际数据流测试，确认本地文件读取范围、自动上传触发条件、默认开启功能、云端存储位置及第三方服务商。安全评估不应停留在供应商提供的文档层面。

第二，在供应商合同中明确数据处理边界。建议在合同中约定：数据处理目的和范围的具体清单；不得超出约定范围处理数据的义务；数据留存期限及到期删除机制；安全事件通知时限和通知内容；数据出境情况的事前告知义务；违约损害赔偿的计算方式和审计权条款。在供应商合同中明确数据出境条款，对于控制跨境合规风险尤为重要。

第三，对高敏感数据实施本地化或私有化部署。 涉及核心源代码、访问凭证、客户个人信息等数据时，应通过白名单与禁止上传目录机制限制数据范围，对含商业秘密与个人信息的环境强制本地化或私有部署，避免核心资产经公网传输。

第四，建立数据出境合规判断流程。企业采购涉及境外服务器的 AI 工具时，需先行评估是否触发数据出境合规义务。即使适用豁免情形，仍应关注个人信息保护影响评估和取得单独同意等要求。“不用于模型训练”只解决了数据用途问题，并不能回答数据是否被上传、存储在哪里、哪些主体可以访问、是否经过第三方云服务处理，以及是否存在跨境传输等问题。

结语

这起 AI Agent 事件折射出 AI 工具在数据合规层面容易出现功能设计与合规设计之间的脱节。传统软件的功能边界相对清晰，用户对“软件能访问什么”通常有合理预期。但 AI 编程工具为了实现代码索引、上下文分析、知识库生成等能力，往往需要读取远超当前编辑范围的数据，甚至调用云端服务。这种能力扩展

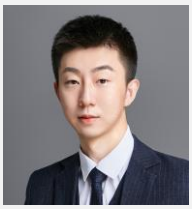
本身无可厚非，问题在于，当数据收集范围随功能扩展而同步放大时，默认开启的机制、缺失的关闭选项、模糊的告知文本等，将使得用户对数据流向的知情和控制被实质性削弱。

对于企业而言，AI 工具的数据合规审查不能停留在隐私政策文本层面，而应延伸至实际数据流转路径的验证和合同条款的精细化设计。具体而言，企业应将合规审查贯穿 AI 工具选型、部署、使用和退出的全生命周期：采购前开展数据流和技术架构测试，确认数据访问及上传范围；部署时明确数据边界并审慎设置默认权限；使用过程中保留必要的审计、监测和追溯机制；退出时确保相关数据及备份得到彻底、可验证的删除。只有将数据合规要求前置，企业才能在利用 AI 工具提升业务效率的同时，对自身数据资产保持可持续、可验证和可审计的控制。

参考资料:

http://scitech.ce.cn/sy/zx/202609/t20260920_3225720.shtml
https://finance.sina.com.cn/wm/2026-09-19/doc-inisisqc3133927.shtml
https://www.inewsweek.cn/kj/2026-09-21/32246.shtml
https://news.qq.com/rain/a/20260920A0C70B00
https://www.panewslab.com/zh/articles/01a0bc5a-1f47-709b-ab54-52923e436de0
https://www.aitntnews.com/newDetail.html?newId=29559
https://www.odaily.news/zh-CN/newsflash/519080
http://www.phoenixweekly.cn/detil.php?id=27981
https://m.thepaper.cn/newsDetail_forward_34111848

如您希望就相关问题进一步交流，请联系：



朱晓阳
+86 21 3135 8683
nigel.zhu@llinkslaw.com

邓梓珊
+86 21 6043 3793
susan.deng@llinkslaw.com

如您希望就其他问题进一步交流或有其他业务咨询需求，请随时与我们联系：**master@llinkslaw.com**

上海	北京	深圳
上海市银城中路 68 号 时代金融中心 19 楼 T: +86 21 3135 8666 F: +86 21 3135 8600	北京市朝阳区光华东里 8 号 中海广场中楼 30 层 T: +86 10 5081 3888 F: +86 10 5081 3866	深圳市南山区科苑南路 2666 号 中国华润大厦 18 楼 T: +86 755 3391 7666 F: +86 755 3391 7668
香港	**伦敦**	
香港中环遮打道 18 号 历山大厦 32 楼 3201 室 T: +852 2592 1978 F: +852 2868 0883	1/F, 3 More London Riverside London SE1 2RE T: +44 (0)20 3283 4337 D: +44 (0)20 3283 4323	

www.llinkslaw.com

Wechat: LlinksLaw

本土化资源 国际化视野

免责声明:
本出版物仅供一般性参考，并无意提供任何法律或其他建议。我们明示不对任何依赖本出版物的任何内容而采取或不采取行动所导致的后果承担责任。我们保留所有对本出版物的权利。

© 通力律师事务所 2026