PLATFORM GOVERNANCE

AI 自动发帖治理

与合规 Loop Engineering

平台封的不是 AI，而是无人负责的自动化：托管发布、批量矩阵、机器人互动、低质 / 虚假内容。

Yao · NexAgent · 2026

先拆开一个被混淆的问题

“用 AI”不是一个风险。
风险来自三件事叠加。

1. AI 辅助创作

选题、研究、初稿、标题、图片、排版、复盘。多数平台并不简单禁止。

2. 自动化托管发布

脚本、接口、浏览器自动化替代真人登录、点击发布、评论私信。

3. 低质 / 未标注内容

同质化、虚假、洗稿、侵权、深度合成未标识、无真实体验。

课程结论： AI 可以进入后台生产；发布责任、事实责任和互动责任不能外包给无人系统。

求证后的平台治理入口

不是平台强弱排名，
而是治理入口不同。

微信公众号	非真人自动化创作 / 脚本接口托管 / 传播教程
小红书	AI 托管运营 / 自动发布 / 模拟真人互动
抖音	水军 / 欺诈 / 谣言 / 侵权 / 流量操纵
今日头条	低质 AI / 批量虚假内容 / 洗稿抄袭
YouTube/TikTok	spam / misleading / synthetic media disclosure

注：原 PPTX 中未能求证的 +215% 数字没有进入正式结论。

技术现实

浏览器自动化能跑，
但它天然靠近风控红线。

一旦系统目标变成“看起来不像机器人”，它就已经在和平台治理对抗。

Playwright / CDP

控制真实 Chrome，走完整人工发帖流程。

扫码 + 本地 Profile

Cookie 、设备指纹、登录环境与机器绑定。

反检测伪装

隐藏 webdriver 、注入脚本、模拟真人行为。

拟真节奏

随机间隔、随机延迟、错峰发布。

合规架构

三层 Loop： 生产、治理、发布。

Content Loop

AI 自动化：选题、研究、初稿、标题、封面建议、排版、平台改写、 SEO/GEO 。

→

Governance Loop

AI 辅助检查：事实来源、夸大承诺、行业禁区、平台规则、 AI 标识建议。

→

Publishing Loop

真人负责：最后编辑、选择标识、点击发布、回复评论、处理私信。

Mandatory human sign-off： 发布按钮前必须有人类确认，不是技术做不到，而是责任不能消失。

实操 SOP

把自动发帖机器人，改造成合规发布队列。

1 **输入边界**

目标、受众、平台、禁区

2 **生成内容包**

正文、标题、图文、来源

3 **风险报告**

事实、标识、行业风险

4 **人工签核**

品牌语气、真实性、修改

5 **真人发布**

平台内最终确认

6 **复盘优化**

数据回收、下一轮建议

关键改造： AI 交付的是“待发布包”和“风险报告”，不是直接替你登录和点击。

发布前风险矩阵

五个信号同时出现，就不要发布。

账号行为

自动登录、自动发布、自动评论、自动私信

内容质量

模板化、无体验、洗稿、事实错误

规模模式

矩阵账号、批量搬运、刷量

标识缺失

深度合成、 AI 图片 / 音视频未声明

行业禁区

金融、法律、教育、招聘、未成年人

评估原则：不是看单点技术，而是看行为、内容、规模、标识、行业是否叠加。

场景 1：某高监管服务行业

不是不能用 AI，是不能无人负责。

高风险叠加

疗效承诺、前后对比、绝对化用语、虚假体验、 AI 托管发布，会同时触碰内容与账号风险。

可做边界

AI 整理资料、生成草稿、检查禁词、准备案例结构，但不得自动发布或自动回复咨询。

人工责任

真人审核医学表达、广告合规、图片授权、体验真实性，真人发布和回复关键问题。

保守建议：高监管服务行业尤其不要做平台托管式自动发帖。

场景 2：技术 IP / 教育内容

技术 IP 可以更积极用 AI，但证据要真实。

可以自动化

资料搜索、代码解释、案例整理、讲稿、 PPT 、博客草稿、社群文案。

必须保真

项目跑过、截图真实、来源可查、结论不过度承诺。

不要碰

刷量、矩阵搬运、机器人评论、伪造客户案例、虚假效果数据。

技术影响力的长期资产不是发布频率，而是可信度、可复现案例和持续复盘。

场景 3： 自有阵地

把工程精力放在
自己能控的地盘。

网站 / Blog

可自动生成草稿、构建页面、更新下载、做 SEO/GEO 。

邮件 / CRM

可自动分群、生成跟进内容，但要有退订和隐私合规。

课程资料库

可自动整理 PPT 、 PDF 、讲稿、案例包与版本记录。

微信群

适合人工发布摘要和链接，不适合机器人群发骚扰。

平台账号是租来的流量，自有阵地才适合做更完整的自动化闭环。

可复制模板

发布前 Checklist

全部通过
再进入发布

- [ ] 事实是否有来源?
- [ ] 是否有 AI 生成合成内容需要标识?
- [ ] 是否含绝对化用语、疗效承诺或虚假体验?
- [ ] 是否由真人完成最终编辑和发布?
- [ ] 是否存在自动评论、私信、刷量或矩阵托管?
- [ ] 是否适合放到自有阵地而不是平台账号?

自动化生产，
真人发布。

合规的 Loop Engineering 不追求绕过平台，而是让 AI 把内容、检查、资料和复盘做到位；让人保留判断、发布和互动责任。

Q&A

不要让系统伪装成你。让系统帮助你做出更真实、更稳定、更可审计的内容。