

AI 代码生成与开源协议的冲突

技术便利于法律边界的碰撞

Peter Mu

August 10, 2025

自我介绍

- “全栈工程师”
- 一名“老鹅”
- 从 2020 年至今在创业
TeXPage 创始人

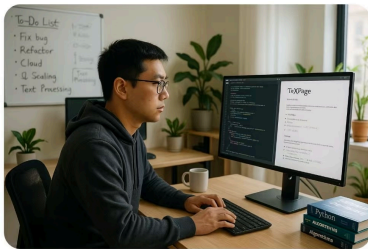


ChatGPT 4o >



跟你交流这么长时间了，你觉得我长什么样子，用图片表示

图片已创建



论开源协议合规的重要性



造谣者

NicolaDaj opened on Aug 2, 2024 · edited by NicolaDaj

<https://texpage.com> is misusing the code in this repo and violating the license.

The product is based on the open-source code in this repo without respecting the license.

The founder is here: [@PeterMu](#).

👍 1 🗨️ 1



Overleaf CTO

jdleesmiller on Sep 13, 2024

Thank you for raising this concern. Digital Science takes seriously the enforcement of its intellectual property rights. If you believe you have evidence of infringement, please contact copyright@digital-science.com.

👍



 jdleesmiller closed this as **completed** on Sep 13, 2024



Me

PeterMu on May 14

Thanks for reviewing and marking this issue as resolved.

TeXPage has been built independently from scratch, and it does not include any code from Overleaf. We respect Overleaf and the open-source community and always comply with software licenses.

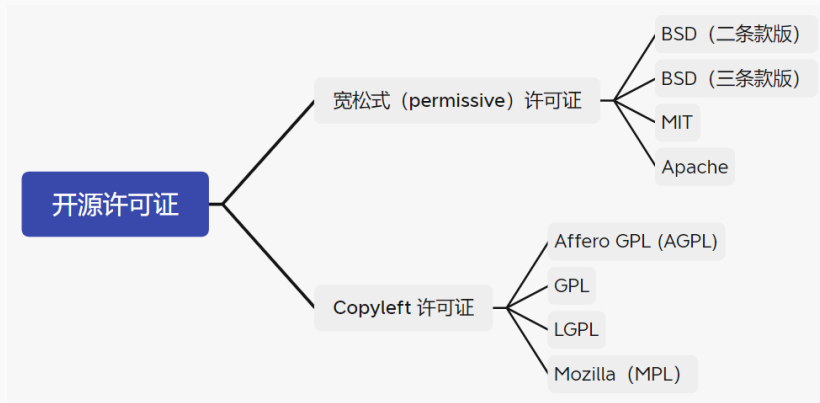
If anyone still has questions, feel free to reach out to us at support@texpage.com.

— TeXPage Team

👍 🗨️ 1

常见开源协议

常见的许可证主要有 GPL、LGPL、AGPL、MPL、MIT、BSD 和 Apache。



开源千万条，合规第一条。

类别	协议	核心义务	风险点
宽松许可	MIT、BSD、Apache	保留版权声明，可闭源	声明丢失
传染性许可	GPL、AGPL	衍生作品必须同协议开源	商业闭源冲突
模块级许可	MPL、CDDL	修改文件需开源	文件混用风险

开源协议合规现状



96%

96%的被审代码库中包含
开源代码



53%

53%的被审代码库中存在
许可证冲突



77%

77%的开源代码存在于
被审的代码库中



31%

31%的被审代码库中包含没有
许可证或使用定制许可证的
开源代码

1

¹Synopsys 2024 年开源安全和风险分析报告

- **数据来源与训练方式**

- 公开可用代码库（GitHub、GitLab、Stack Overflow 等）
- 开源协议混杂

- **生成方式**

- AI 并非记忆全文，但会在高相似度上下文中复现片段
- 生成概率取决于训练样本的频率和匹配度

- **与人类编程的区别**

- 人类借鉴 vs AI 生成的概率性复制
- AI 独立生成的内容不享有版权
- 人类可以解释实现过程，证明独立创作，AI 无法解释来源

AI 代码生成引发的协议冲突类型

- 未经许可的代码复制
 - 模型生成与训练数据高度相似的片段
 - 无法自动判断原协议 → 误用
- **GPL “传染性” 风险**
 - AI 生成含 GPL 代码 → 产品整体可能被要求开源
 - 与商业闭源冲突
- 专利条款冲突
- 署名与版权声明缺失
- 跨协议混用问题

开源协议不合规带来的风险

- **法律层面**

- 侵权诉讼（赔偿、停止分发）
- 协议违约（强制开源）

- **商业层面**

- 产品竞争力丧失
- 投融资、并购受阻（合规审查不通过）
- 品牌声誉受损

- **制定 AI 使用规范**

- 对生成内容进行协议与版权检查
- 分级管理，明确哪些项目可以使用 AI 生成的代码

- **自动化检测**

- 许可证扫描工具自动化检测
- 部署“协议防火墙”，拒绝高风险协议代码进入生产库

- **替代方案**

- 使用 AI 生成伪代码 / 架构思路，由人工重写实现
- 内部部署私有模型，限定数据来源

- **法律未决问题**

- AI 生成的代码是否享有版权？
- 是否算作衍生作品？

- **行业标准化**

- AI 输出的合规标记
- 训练数据溯源机制

- **可能的解决路径**

- 出现 “AI 友好” 开源协议
- AI 公司与开源社区签署双边协议

Thanks !