



智慧图书馆中生成式人工智能应用的 潜在风险与治理策略

Potential Risks and Governance Strategies of GAI in Smart Libraries



本研究探讨生成式人工智能在智慧图书馆的双刃剑效应,在智能检索、虚拟馆员等创新场景提升资源管理与用户体验时,也存在知识产权侵权、隐私漏洞、数据偏见等风险。本文提出多维治理框架,包括风险分级管控、主动防御、伦理审查与信息素养教育等策略,助力智慧图书馆可持续发展。

AI在智慧图书馆中的应用

Application of GAI in smart libraries

应用场景 Application Scenarios

- 智能检索
- 资源管家
- 智能采编
- 学科分析
- 元宇宙阅读推广
- 虚拟图书馆员
- 信息素养教育

图书馆的应对策略与建议

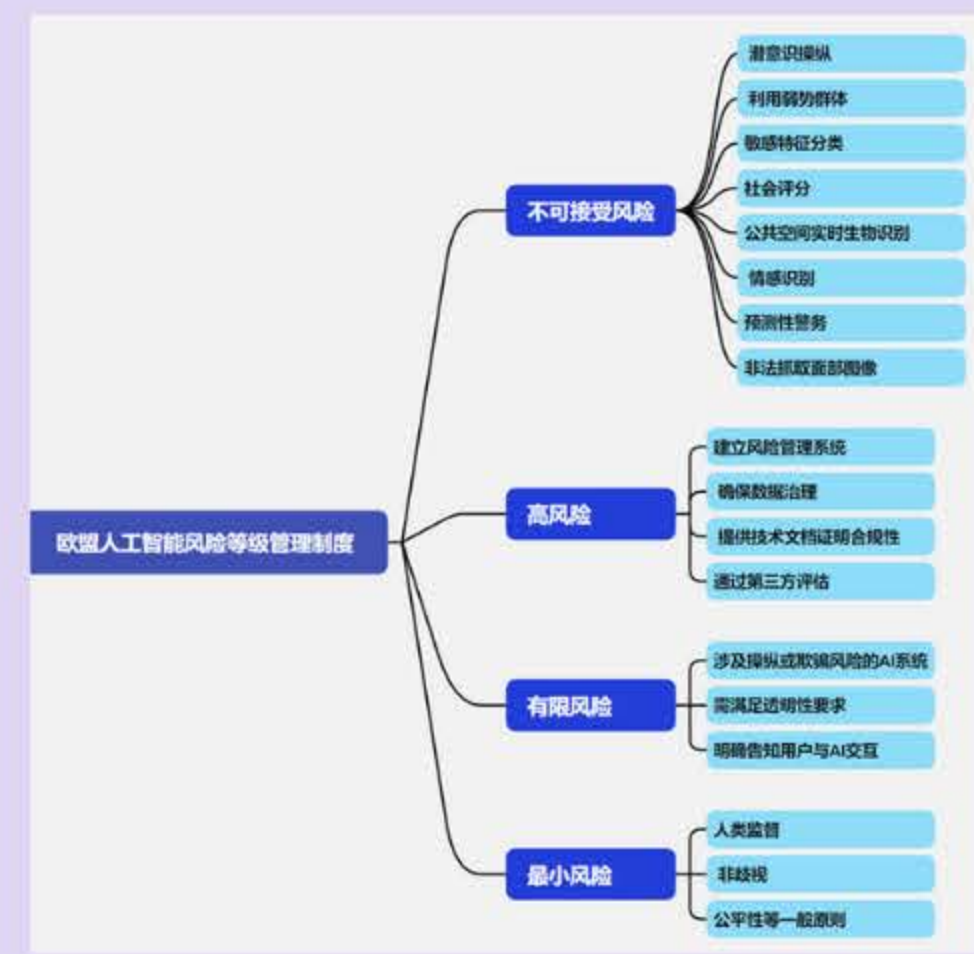
Strategies and suggestions

侵权风险分级管控机制 Risk Grading Control

建立图书馆AI应用分级管理体系

高风险业务
-----认证和双盲审

低风险业务
-----承诺备案制



AI应用的潜在风险

Potential risks

著作权侵权风险 Copyright infringement

- 数据合法性：数据来源争议、合理使用边界
- 复制权风险：作品数字化、未经授权复制
- 演绎权争议：数据清洗、改编超出合理使用

隐私权侵害风险 Privacy Infringement

- 个人信息泄露风险：电话、人脸等隐私数据泄露
- 被遗忘权：欧盟AI法案核心要求
- 匿名化处理：敏感信息保护手段
- 虚拟馆员：肖像权/名誉权保护
- 数据合规挑战：收集/存储/处理合法性

数据偏见与算法歧视 Data bias & Discrim

- 数据偏见：训练数据中的性别/种族偏见
- 算法歧视：AI放大系统性不公
- 代表性不足：馆藏数据失衡导致推荐偏差

从被动合规到主动防御 Proactive Defense

全周期防护链

▶ 预防—监测—响应

技术防御

▶ 网络隔离 | 智能过滤 | 同态加密

版权保护

▶ 审查机制 | 区块链存证 | 数字水印

应急响应

▶ 快速处理 | 日志举证



伦理审核与过程可信化机制 Ethical Review

破解黑箱困境

▶ 算法公平性 | 服务可视化

协同治理体系

▶ 多元治理架构 | 算法伦理委员会

用户权利保障

▶ 数据画像修正 | 证据链留存

