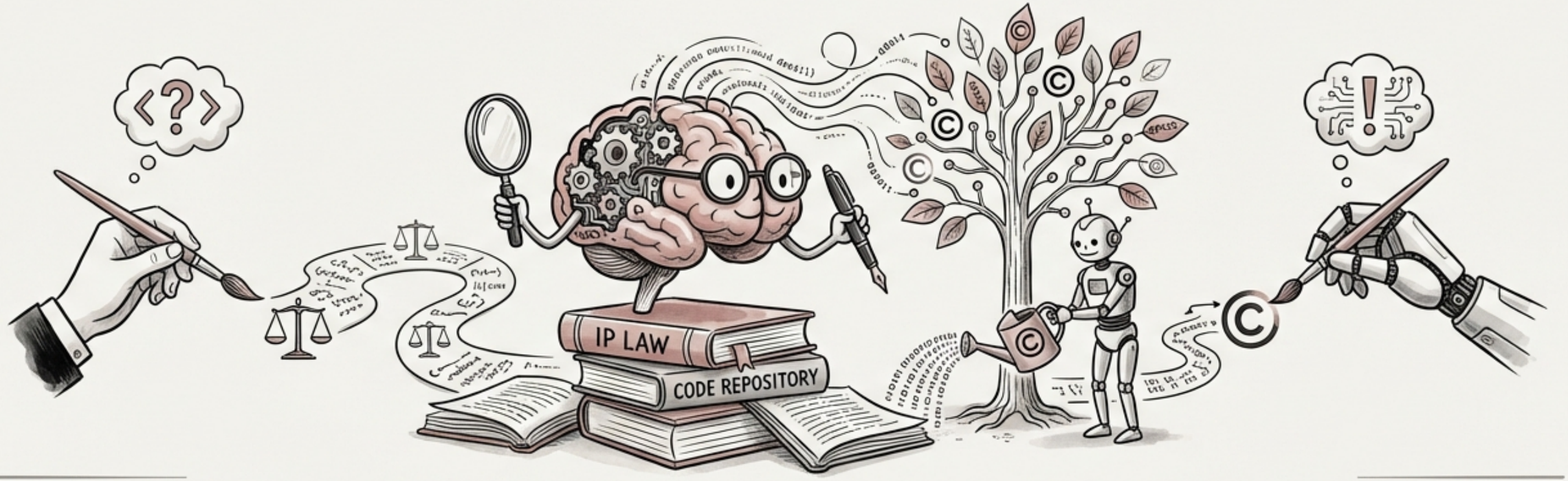
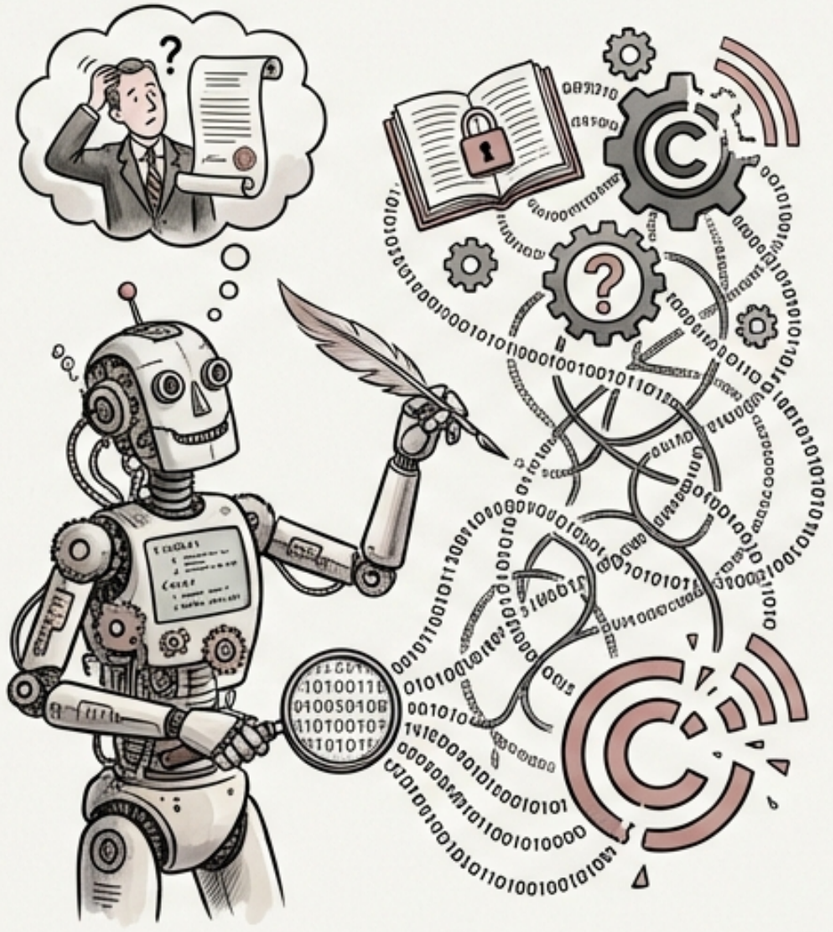
AI Code Attribution & Licensing: Navigating IP in AI-Augmented Development



The Emerging IP Landscape of AI-Generated Code



- AI-generated code introduces novel intellectual property (IP) challenges.

- Traditional copyright laws may not fully apply to code created solely by AI.

- Managing AI code requires legal, technical, and organizational frameworks.

UNCERTAIN COPYRIGHT STATUS OF PURELY AI-GENERATED CODE

- US Copyright Office mandates human authorship for copyright protection.
- EU regulations also emphasize human involvement for copyright eligibility.
- Implication: Code created entirely by AI may not be copyrightable in some jurisdictions.
- This creates both opportunities (free use?) and risks (lack of protection).

ESTABLISHING COPYRIGHT THROUGH JOINT AUTHORSHIP WITH AI

DEVELOPERS

- WHEN DEVELOPERS SUBSTANTIALLY DIRECT AND EDIT AI-GENERATED CODE, COPYRIGHT MAY BE ESTABLISHED.

SIGNIFICANT

- SIGNIFICANT HUMAN MODIFICATION CAN QUALIFY THE RESULT AS HUMAN-AUTHORED.

SMALL CAPS

- THE DEGREE OF HUMAN INTERVENTION IS A CRUCIAL FACTOR.

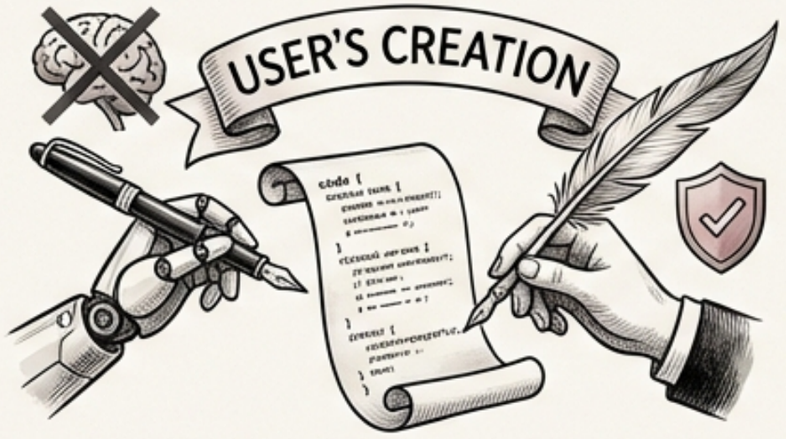
Understanding AI Tool Terms of Service: Rights and Responsibilities

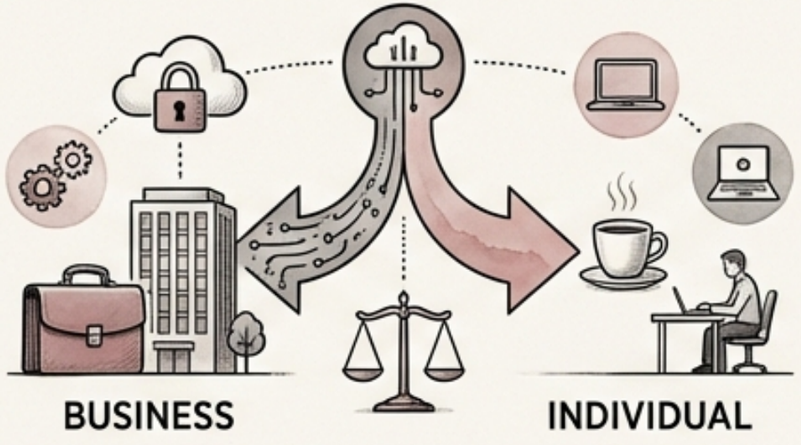
- Critical to understand the IP implications when utilizing any AI development tool.

- Analyze the fine print of what rights are granted (or not) in the AI tool's terms of service.

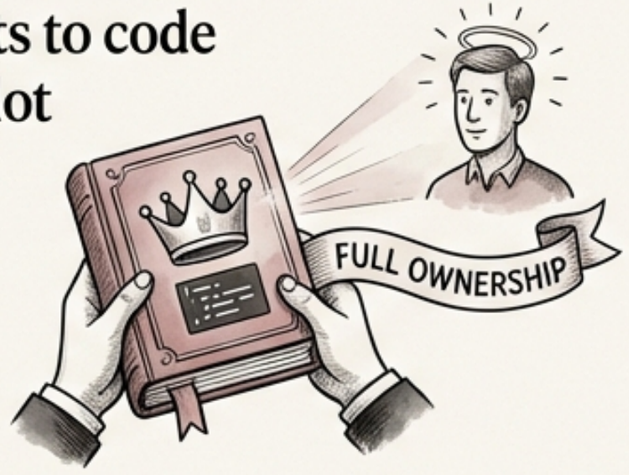
THE FINE PRINT

GitHub Copilot's IP Policy: User Rights and Telemetry



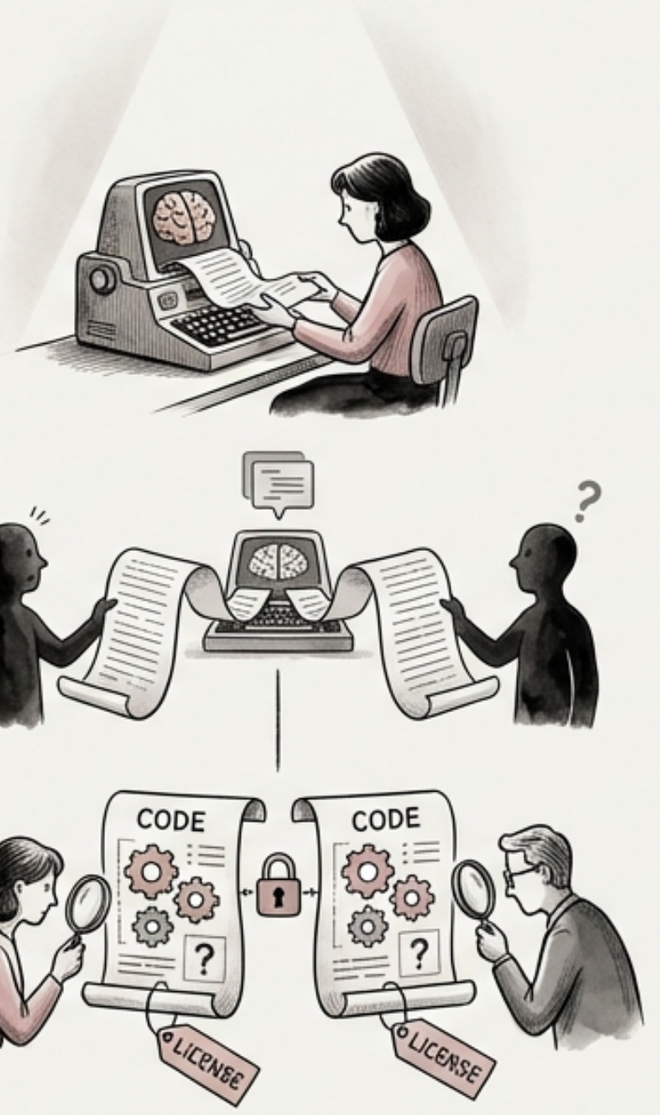


- GitHub does not claim IP rights to code suggestions provided by Copilot
- Users retain rights to accepted code suggestions
- Telemetry data usage varies depending on the subscription plan (Business vs. Individual)
- Carefully review the specific data usage policy for your plan



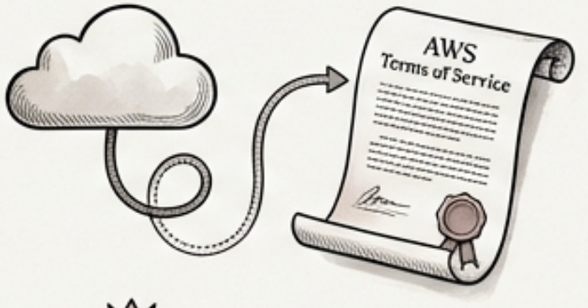
OpenAI/ChatGPT's Output Ownership and Redundancy Risks

- Output generated by OpenAI's ChatGPT generally belongs to the user.
- Identical outputs may be generated for other users, raising unique concerns.
- This means there could be duplicate code with identical licensing implications.



Amazon Q: AWS Terms and Enterprise IP Protections

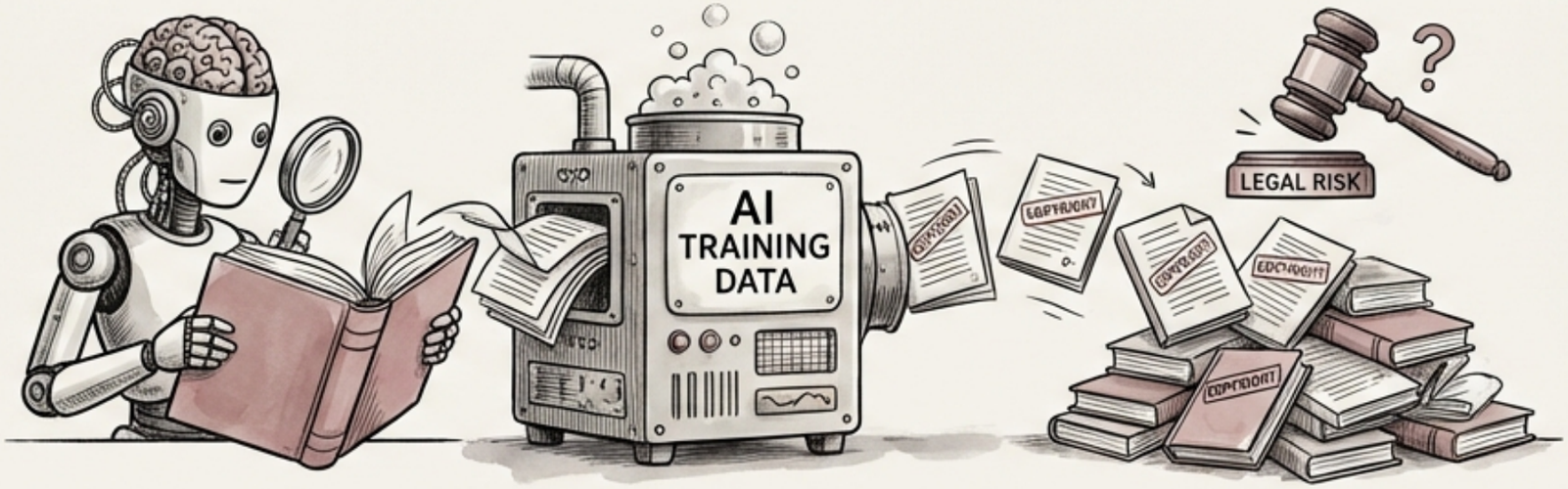
- Output rights follow Amazon Web Services (AWS) terms of service.



- Enterprise plans may offer additional IP protections.

- Specific terms are outlined in the AWS Service Terms and should be reviewed carefully.

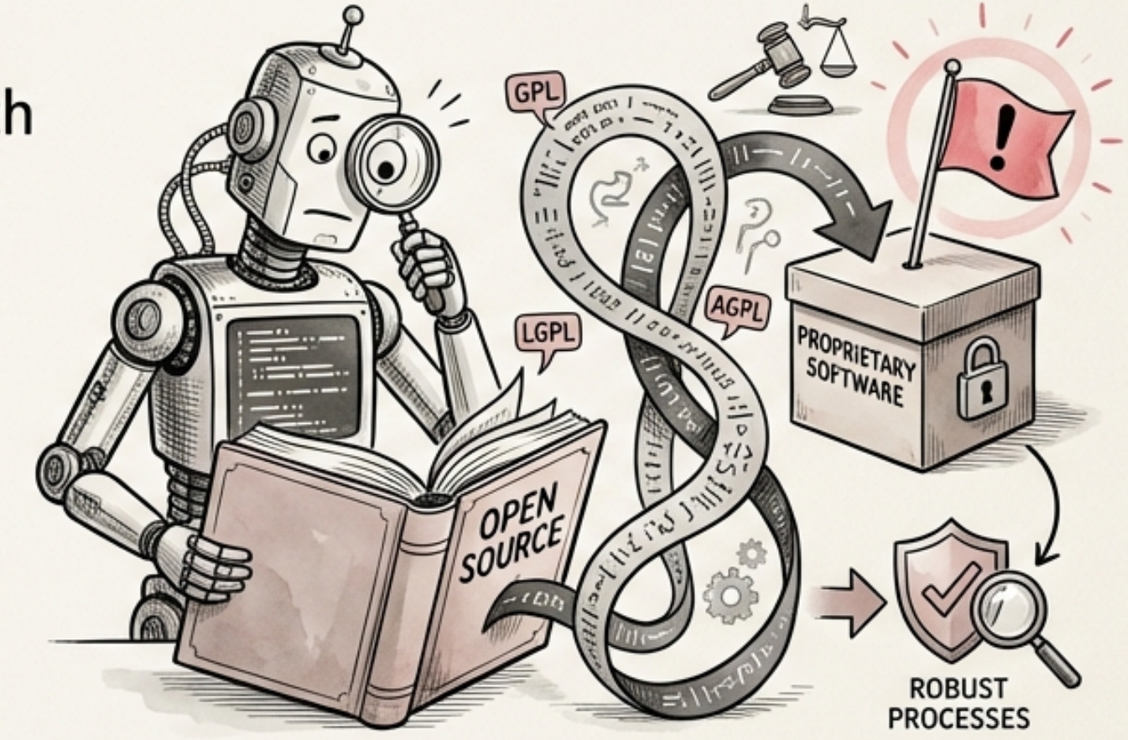
KEY RISK: COPYRIGHTED TRAINING DATA REPRODUCTION



- Most AI tools cannot guarantee that their output does not reproduce copyrighted material from their training data.
- This creates a significant risk of unintentional copyright infringement.
- Proactive measures are needed to mitigate this risk.

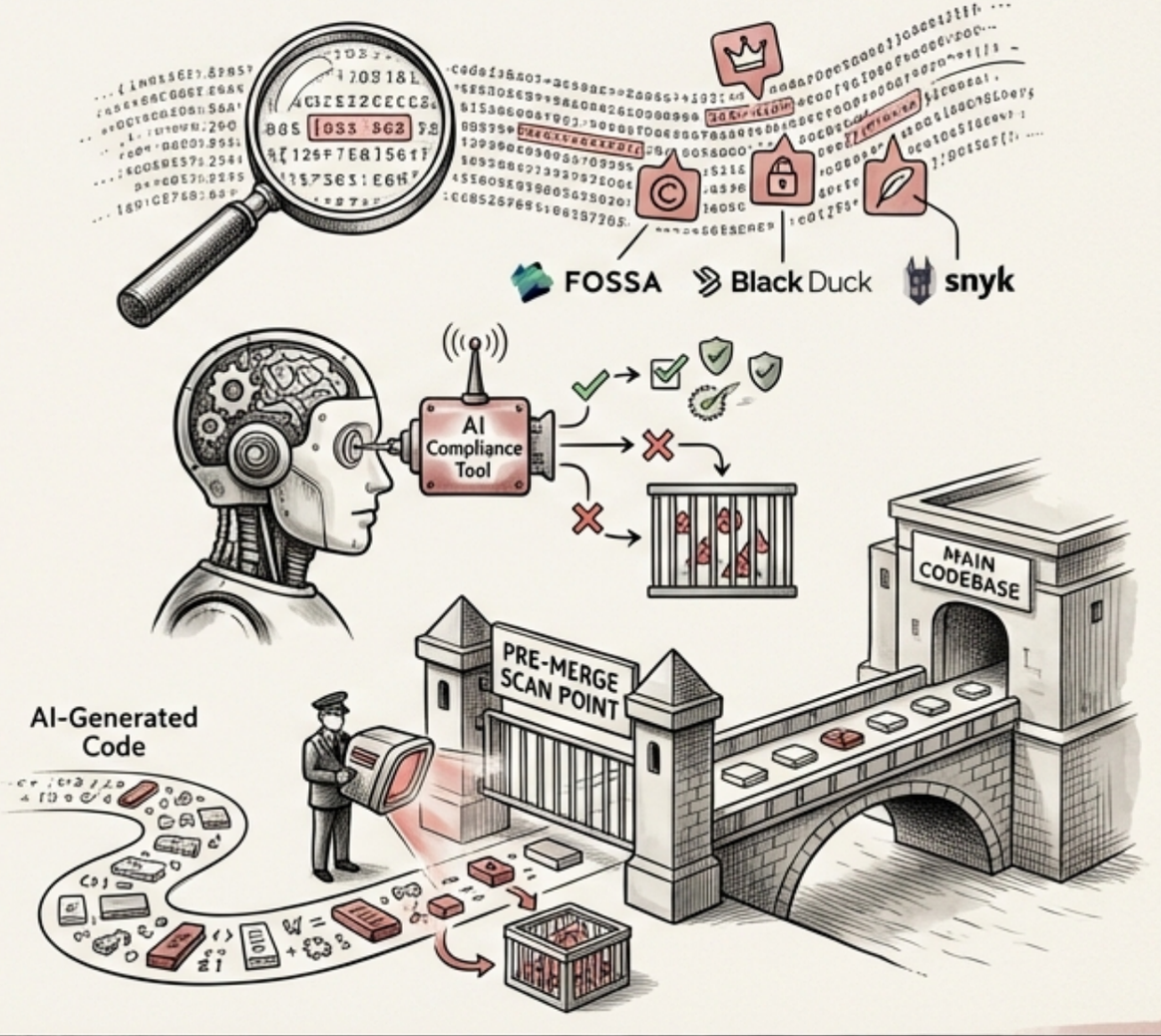
License Compliance Risks with AI-Suggested Code

- AI tools trained on open-source code may inadvertently reproduce code with copyleft licenses (GPL, LGPL, AGPL).
- Using copyleft-licensed code in proprietary software can lead to license violations.
- Implement robust license compliance processes to prevent this issue.



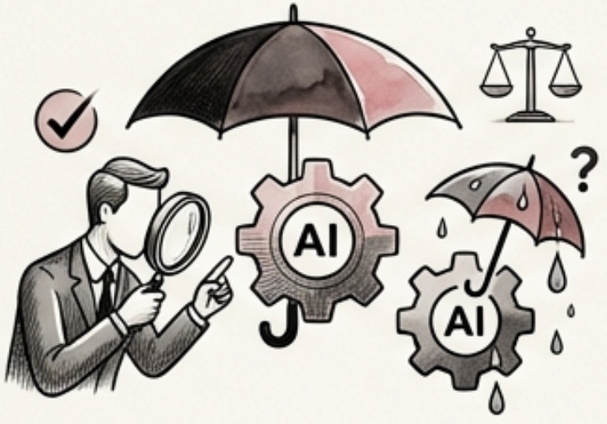
Detecting and Mitigating License Contamination in AI Code

- Employ code similarity scanners (FOSSA, Black Duck, Snyk) to detect potential license violations.
- Explore AI-specific license compliance tools designed for this emerging challenge.
- Scan all AI-generated code for license contamination *before* merging it into the codebase.



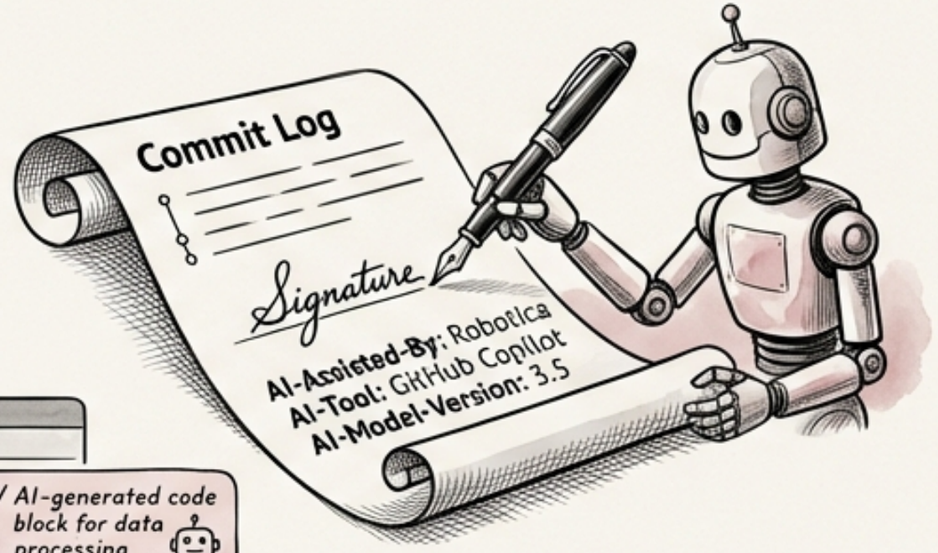
IP Indemnification: Protecting Against Copyright Claims

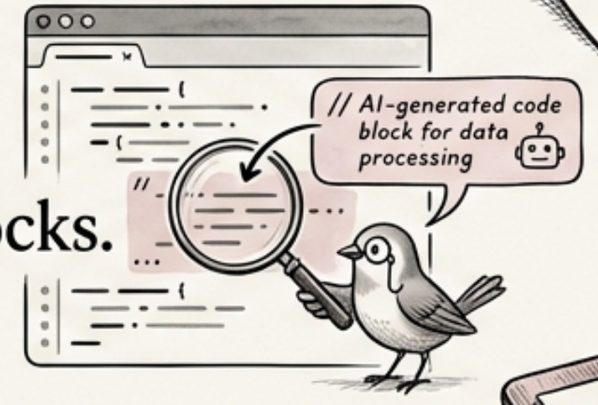
- **IP** indemnification offers **protection against copyright claims** arising from AI-generated code.
- **GitHub Copilot Business** provides IP indemnification.
- **Amazon Q** offers IP indemnification for enterprise customers.
- Carefully evaluate indemnification coverage when choosing an AI tool.

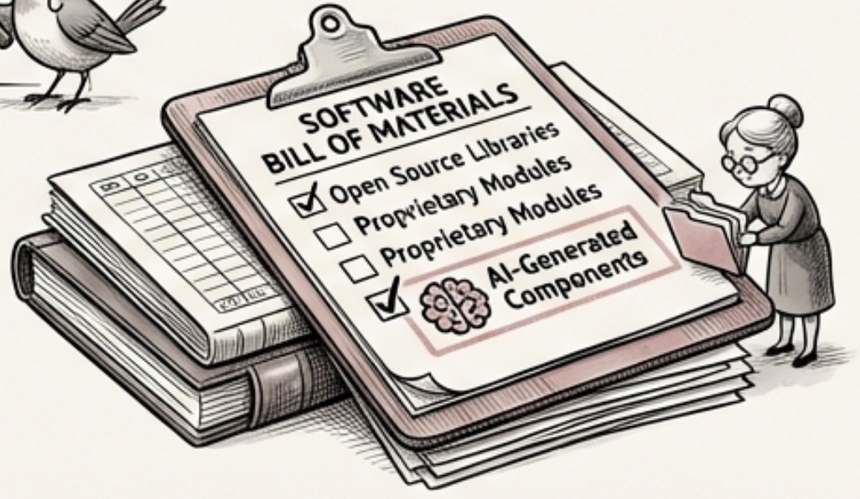


Technical Attribution Methods for AI-Assisted Code

- Use Git trailers (AI-Assisted-By, AI-Tool, AI-Model-Version) to track AI involvement.
- Include inline comments for significant AI-generated code blocks.
- Add Software Bill of Materials (SBOM) entries for AI-generated components.

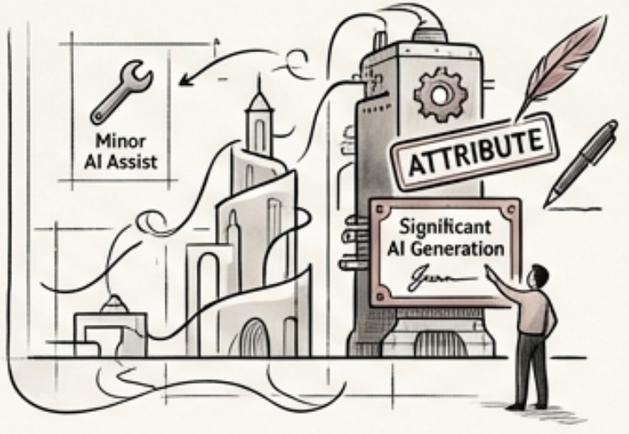


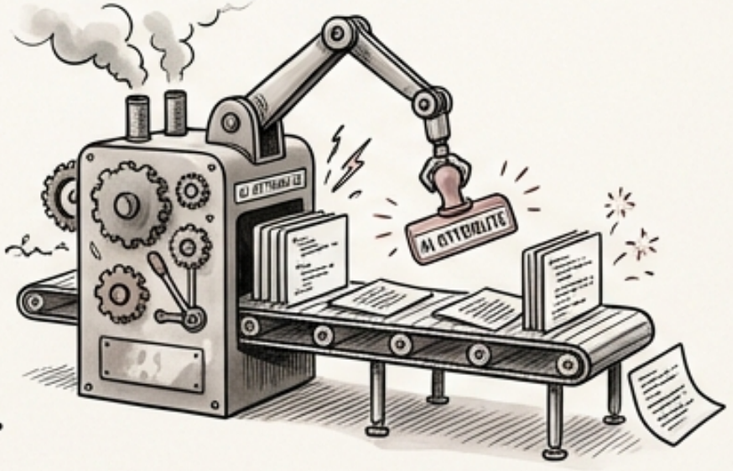




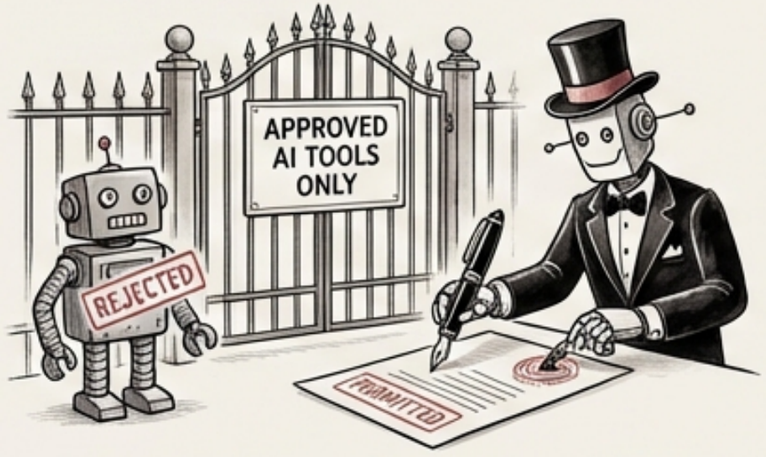
ESTABLISHING AI CODE ATTRIBUTION POLICIES AND CONVENTIONS

- 1. Define clear guidelines on what **level** of AI involvement requires attribution.
- 2. Establish **consistent attribution conventions** across all development teams.
- 3. **Automate** the attribution process whenever possible.
- 4. **Stay informed** about **regulatory requirements**, such as those potentially arising from the EU AI Act.

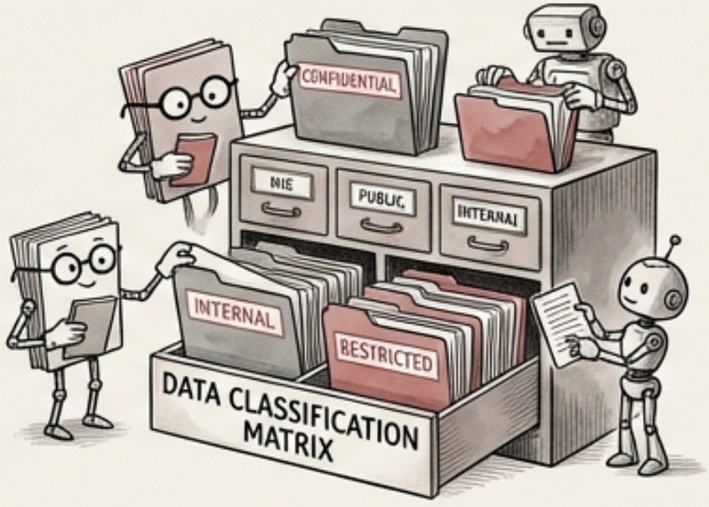




KEY COMPONENTS OF A COMPREHENSIVE AI CODE POLICY

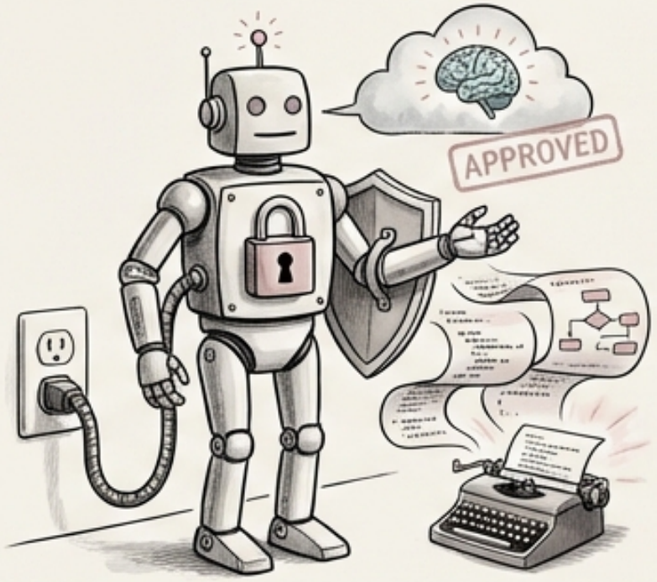


- **Tool approval list:** Only permit the use of approved AI development tools.
- **Data classification rules:** Govern the handling of sensitive data.

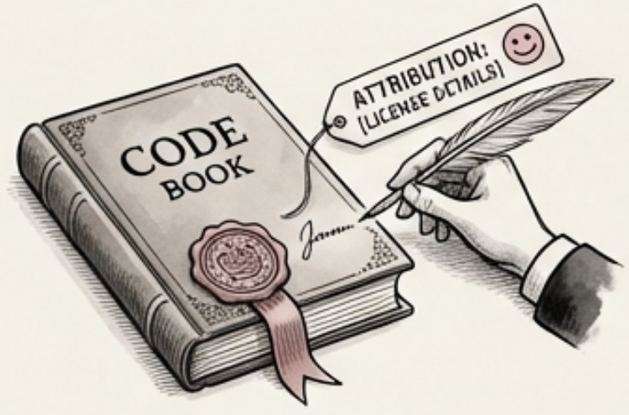


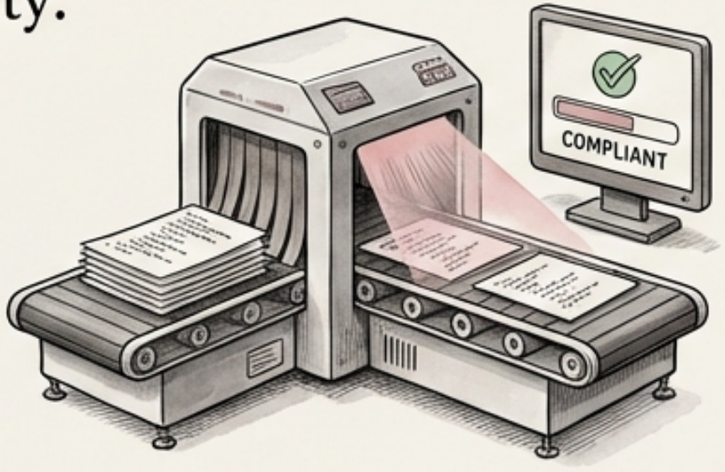
- **Attribution requirements:** Define how and when AI involvement is documented.
- **License compliance process:** Ensure all code adheres to licensing terms.

WORKFLOW FOR LICENSE-SAFE AI CODE USAGE: A STEP-BY-STEP GUIDE



1. Use an approved AI tool with a secure configuration.
2. Generate code suggestions from the AI tool.
3. Review all code suggestions for quality, security, and functionality.
4. Run a license compliance scan on the accepted code.
5. Add appropriate attribution metadata to the code.





Thank You

- Questions?