Covering your bases with IP Indemnity

Deep Dive: Defining Open Source AI
Blown away by the response of @sourcegraph Cody

I'm playing around with the @sourcegraph AI tool Cody and my mind is blown. I wanted to know what my test code snippet does and to check if it would notice code vulnerability, and it did just that.

Not just added the API request code
Correctly saved the data to my models
Tossed in error handling

Cut my coding time in half
Just need to fine-tune and voila!

First time I use a #AI that is really FAR "smarter" than me!

I use AI a lot for automating small tasks and saving time.

But @SourcegraphCody is a glimpse at the future, where #GPT5 and @LangChainAI may help us save not just minutes but years!

today I asked @sourcegraph's Cody to explain some code I highlighted, it did and also gave suggestion on how to make the runtime faster

Benchmarking Mojo code with Cody vs Julia

So I'm auditing this codebase using Cody AI of sourcegraph and I must say its the best assistant. It eases my code audit pain. I'm blown away by how convenient it is to read code now.

Thanks for this great Cody AI tool
Overview

1. Code AI platform
   Justin Dorfman
   What is Cody?
   How Cody works.
   The Copyleft question.

2. Protecting Customers
   Tammy Zhu
   Low copyright risk.
   Litigation landscape.
   Other mitigation.

3. IP Indemnity
   Samantha Mandell
   Cover your bases.
   Reduce copyright risk.
   Negotiating.
Code AI Platform

Justin Dorfman
What is Cody?

- AI coding assistant
- Writes, understands, and fixes code
- Web, VS Code, & JetBrains/Intellij
- Open source under Apache 2.0 license
What is Cody?

Momil

Deepak

Catalin

Szymon

Tino
Code AI Platform

How Cody Works

- Prompt
- LLM
- Context

"Everyone on my floor is coding"

Software Engineers

Doctors
Hugging Face presents

Training Cluster

As a service

Train your LLM at scale on our infrastructure

I want to train a 70B \(\uparrow\) parameters multimodal \(\uparrow\) model on a

7T tokens \(\downarrow\) dataset, on 1000 \(\downarrow\) H100 GPUs \(\downarrow\)

Estimate: 184 days on 1000 GPUs

$18,461,354
Code AI Platform

The Copyleft problem

- “Does Cody use GPL code?”
- Unable to get a straight answer
- Sourcegraph Legal to the rescue

Don’t worry, I got this!

Tammy Zhu
Protecting Customers

Tammy Zhu
Protecting customers: overview

1. Copyright risk
2. Litigation landscape
3. Other mitigation
What’s the risk?

Copyright risks

● Copyleft virality
● Breach of OSS license
● Infringement
For AI coding users, copyright risk is low

Why?

In no case does copyright protection for an original work of authorship extend to any

● procedure
● process
● system
● method of operation

17 USC 102(b)
Litigation landscape: AI coding providers

No lawsuits to date against users of AI coding providers

No lawsuits to date against AI coding providers alleging copyright infringement

- *Doe v. GitHub* alleges the following
  - DMCA violation
  - Breach of contract
  - Interference with economic relations
  - Unjust enrichment
  - Unfair competition
  - Negligence
- No copyright claim
Other mitigation

● Large training sets
● Permissively licensed training sets
● Filters and attribution
● OSS scans
● IP indemnity
Indemnity Flows to the End User

What is Indemnity?

- A contractual agreement between two parties in which one party agrees to pay for potential losses or damages claimed by a third party as a result of their relationship.
- Software: if Party A is sued because they use Party B’s software as intended, Party B will cover Party A’s losses.
- Often, if there are multiple parties involved in a relationship or transaction, there might be multiple indemnification agreements.
What AI tool users should look for with Indemnity

A comprehensive IP indemnity clause is:

- Uncapped
- Covers all plausible third parties (i.e. individual copyright holders and open source licensors)
- Applies even if a user makes changes to the Output
# Landscape of Indemnity among AI Coding Assistants

<table>
<thead>
<tr>
<th></th>
<th>Cody (Sourcegraph)</th>
<th>Sample Tool #1</th>
<th>Sample Tool #2</th>
<th>Sample Tool #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncapped</td>
<td>Caps dependent on agreement with provider for other services</td>
<td>None, and user required to indemnify provider for copyright infringement</td>
<td>None (but users not required to indemnify provider)</td>
<td></td>
</tr>
<tr>
<td>Covers all plausible 3rd parties</td>
<td>Covers all plausible 3rd parties</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Applies even if a user makes changes to the output</td>
<td>Does not apply if user makes changes to the output</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Requires use of filtering</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Closing

Justin Dorfman