

Analyzing the Effectiveness of Copyright Law on AI Output in Software Engineering and Future Developments

Surendra Jammishetti, sjammish@ucsc.edu

Abstract—**TODO: need to do this later**

Index Terms—AI, Software Engineering, Copyright

is that a dishonest, or even badly programmed, web scraper can come along and look past this file and do whatever they please**robots**.

I. INTRODUCTION

THE emergence of AI tools has, no doubt, changed many aspects of how work is approached in many professions, but the greatest impact has been on software engineering. New tools like Github Copilot, an AI assisted code generation tool thats usable in many modern code editors today, boasts a 2x speedup of feature completions and a large boost in productivity for its users **github**. However accurate those claims may be, theres no doubt that these tools are being used widely by the developer population, and they arent going away any time soon.

However the code generated by these tools must come from somewhere. Many models source their training data from the public internet, but not everything in public access allows for commercial / derivative use. Such models ignore this step, jumping straight towards using this data with any disregard for copyright protections in place. While copyright has been slow in the past, there is no doubt the legal systems around the world will regulate and form precedent for these utilites; Inevitably placing their users under risk.

For those concerned with the legality of the software they produce, the usage of AI-generated code should be avoided as their trustworthiness is dubious and they will become succceptible to copyright law in the near future.

II. BACKGROUND

To help understand the context that this paper resides in, below are lay-person explanations of AI training, Software Licensure, AI tools, and more.

A. Data Acquisition

The way AI development labs gather data for their models is mainly through web scrapers. Web scrapers are programs that access a immense number of websites, downloading and organizing the all the data they find **Miquido**. The gathered data is then used to train whichever model they are interested in.

The modern way for a website to prevent a webscraper from downloading its contents is a file called a 'robots.txt', that any honest web scraper would look for first, and obey the rules inside **robots**. The problem with this "trust me" approach

B. Software Licenses

C. AI-Powered Tools

III. AI AND CODE GENERATION

IV. TRACKING

V. DISCUSSION AND SUMMARY

ACKNOWLEDGMENT

The authors would like to thank...