Thesis Project Portfolio

A Meta-Study on Methods of Poisoning Artificial Intelligence Text-to-Image Generators

(Technical Report)

Towards Better Advocacy for Software Patent Reform: Understanding How Value Differences Influence the Software Patent Debate

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Nick Garrone

Fall, 2024

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

(Executive Summary)

How Software and Intellectual Property Collide

The intellectual property system consists of two principal components– patents, which protect inventions, and copyright, which protects expression. Yet, when the system in its current form is applied to software, it fails to achieve these goals. AI image generation tools have been trained on copyrighted images at scale, violating the rights of artists. Patents, on the other hand, have had the effect of stifling innovation when applied to software. Each component of my research addresses one of these problems from a technical or socio-technical perspective.

In the technical portion of my research, I produced a meta-study review on methods of poisoning AI-powered image generation programs. These programs, coming into prominence with the release of tools like DALL-E in 2022, generate images in a wide variety of styles from a simple text prompt. They make it incredibly easy to produce impressive images from text. However, some of these tools were trained on copyrighted images on the open web, raising concerns of copyright infringement and plagiarism. In my research, I surveyed cutting-edge methods of poisoning images such that they cannot be used to train these AI programs. I discovered that tools such as the University of Chicago's Nightshade represent a significant technical advance over simple methods such as bad data labeling. They use adversarial methods to generate images specifically for the purpose of creating a misrepresentation when passed to a given model. This means that these tools cannot remain static but must be continuously iterated upon to remain relevant.

In my STS research, I investigated the failings of the software patent reform movement, finding that software patent reformers and supporters of software patents have different ideas of risk and innovation. To do this, I analyzed pro- and anti- software patent discourse. I found that software patent reformers see patents as a source of risk, while software patent supporters see them as a way to mitigate risk. Software patent reformers see innovation as something inherent, while software patent supporters see it as something to be cultivated with top-down incentives. With this increased understanding, software patent reformers will hopefully be able to express their concerns in the language of the other side, leading to more effective communication.

Considering both of these topics provided me with a richer understanding of the state of software and intellectual property. They represent different ways in which the intellectual property system has failed to keep up with the demands of software. Software infringes on intellectual property rights with AI text-to-image models, but it is also being infringed upon by the intellectual property system with overbearing software patents. This counterpoint illustrates that there is no one easy fix, no single direction in which the intellectual property system can move. Any reform would have to be complex, tightening where software developers infringe and loosening where it stifles software innovation. Engineering as social experimentation is the idea that engineers, by introducing new technological capability, are essentially performing experimentation on the public, and need to be ethically aware of that. In interacting with the intellectual property system by violating copyright or filing patents, engineers need to be aware of the public impact of their actions.