Thesis Project Portfolio

Automating Invoice Processing via Optical-Character Recognition and Binary Classification

(Technical Report)

Governing Artificial Intelligence: An Analogical Approach to Cultivate Shared Knowledge and Generate Newfound Insight

(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

William Peter Greig

Spring, 2024

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

(Executive Summary)

Governing Artificial Intelligence Through Analogical Argumentation

"Regulating AI is a complex socio-technical challenge with high stakes for society and human rights. We need multifaceted governance embracing nuance, expertise, transparency and stakeholder diversity to steer AI's evolution ethically."

— Michael Spencer, Sept. 2023

While attending the University of Virginia, I interned at a semiconductor equipment manufacturing company called Axcelis Technologies Inc. They specialize in fabricating ion implanters for transistors in the 40-60 nanometer range. While interning, I led numerous projects to develop automated solutions using artificial intelligence (AI) and data analysis techniques to optimize inefficient business processes related to international trade compliance. During my internship, one project truly tested my technical knowledge and skills concerning AI and was the primary motivator for my STS research: an AI-driven software system that processes and analyzes PDFs and spreadsheets for ECCN part classification, harmonized tariff code assignment, and drawbacks. Inspired by this project and also the growing regulations between the United States and China's semiconductor industry, my initial STS research aimed to address the differing governing policies of AI between China and the United States. However, I pivoted my STS research to exploring the problems associated with regulating artificial intelligence,

specifically concerning a lack of shared understanding among various stakeholders in the policymaking process.

My technical project was an AI system designed to reduce manual processing by introducing an automated system to process pre-alerts. Pre-alerts are documents containing invoices containing information on shipment logistics and are sent to Axcelis before they receive a physical shipment of parts or components. For classification purposes, it's necessary to check parts on the pre-alerts, compare them against Axcelis' database, and classify parts before entry to reduce backlogging. The difficulty stems from the fact that Axcelis has multiple vendors. Each vendor will send pre-alert invoices in different formats. Some will be PDFs and others will be Excel spreadsheets or CSV files. From an automation perspective, a heuristic solution would not work given the different formatting of each invoice. To solve this problem, I utilized optical character recognition (OCR) in a multi-stage processing system to quickly extract the part numbers from each invoice. As an overview, the system first takes in an invoice and uses OCR to extract the text. Then, I trained a machine learning model to help classify parts–since there must be a way to distinguish between part numbers and non-part numbers–and compare them to our database. Ultimately, the utilization of this system will decrease the manual processing required to extract part numbers, enabling co-workers to focus on other activities and tasks.

My STS research addresses issues with policymaking whereby various stakeholders hold uneven levels of knowledge concerning the technology which can cause overstating or underemphasizing the potential capabilities, limitations, and problems of an emerging technology. To conduct this research, I used Schwarz-Plaschg's research on analogical imagination as a flexible framework for understanding emerging technologies. These analogies function as comparisons to previous or alternative technologies to provide context and a

foundation for argumentation. These analogies help draw parallels and similarities between the comparative technologies. The analogies I found and chose to focus on for my research were the printing press, genetic engineering, nuclear power, space exploration, and human bias. Each of these was compared to specific AI issues such as job displacement or the spread of misinformation. After analysis, there was newfound information and insights regarding artificial intelligence that are not readily apparent, especially to those who are not industry experts.

Engineers, policymakers, and citizens alike are all equally important in the policymaking process for governing artificial intelligence. The issue is varying levels of knowledge and perspectives toward AI. My STS research addresses this issue by introducing analogies to compare AI to previous or concurrent technologies. From a technical perspective, this presents artificial intelligence as a technical capability rather than a specific implementation that is helpful in defining AI more. The comparisons made between previous technologies and AI were made to address cultural concerns such as job displacement, algorithmic bias, and the spread of misinformation. These cultural perspectives are imperative to properly ensure that AI, or any emerging technology, is governed in a just manner. Lastly, the use of analogies is also helpful in addressing the organizational entities that are necessary to govern technologies. These organizational entities are the stakeholders that are involved in the policymaking process. My technical project also addressed these attributes by expanding the technical abilities of the department, saving a tremendous amount of time spent on menial labor, and introducing co-workers to the potential benefits of AI in their workspace. These attributes–technical and cultural–help inform the various stakeholders–organizations–about the different social concerns and technical capabilities of artificial intelligence.

References

Spencer, M. (2023, September 5). Unraveling the ethical complexities of A.I. Unraveling the Ethical Complexities of A.I. AI Observer [web article]