Automating Invoice Processing via Optical-Character Recognition and Binary Classification

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

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
William Greig

May 8th, 2023

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Kathryn A. Neeley, Department of Engineering and Society

Briana Morrison, Department of Computer Science

Cross-Comparison of Artificial Intelligence Initiatives between the United States and China

Overview:

While attending the University of Virginia, I interned at a semiconductor equipment manufacturing company called Axcelis Technologies Inc. While interning, I led numerous projects to develop automated solutions using artificial intelligence (AI) and data analysis techniques to facilitate inefficient business processes related to international trade compliance. As a semiconductor company, Axcelis regularly interacts with both China and the United States government on a daily basis to navigate the increasing trade restrictions between the two countries [1]. To further analyze how these countries protect their own economies through governmental policies, I will be conducting a literature review on Chinese policies and initiatives which aim to promote AI growth. The central question I will pursue is: *how can the United States win the artificial intelligence arms race against China by adopting governmental policies to bolster our AI programs?* One cool thing readers will learn is China's elite artificial intelligence national team.

Problematization:

The research problem I will be addressing is the differing approaches to artificial intelligence regulation and governing policies in the United States and China. This research is especially important given AI's explosive growth over the past several years which has revolutionized and uprooted numerous industries such as transportation with self-driving cars or education with ChatGPT. Countries that design policies that adopt and promote the utilization of AI will have a significant advantage over countries that fail to do so. In terms of current policies, the United States has emphasized and upheld individual privacy and AI transparency while

China has focused on societal control and prioritized economic growth above all else. Given that these two global superpowers are in an AI arms race against one another, the question is *who will win?*

Guiding Question:

What policies or strategies can be examined from China and applied to promote AI growth in the United States while adhering to social considerations?

Projected Outcomes:

The goal of this project is to analyze initiatives taken by the Chinese government which has supported their domestic AI industry and uses these to provide modified policies for the United States to adopt. Crucially, these policies must be designed to consider and adhere to the United States' established values on issues such as the protection of privacy, accountability, and transparency. To this extent, United States policymakers would benefit from my research as I will provide evidence-based, value-oriented policies to enact in order to accelerate domestic AI development and maintain economic competitiveness against China.

Technical Project Description:

As an intern at Axcelis, my main responsibility is to design solutions to increase the efficiency of inefficient business processes pertaining to import and export compliance. For instance, a drawback claim–where import duty is reclaimed given certain conditions–used to take 45 minutes to process manually. After studying the processing method and developing the appropriate software, we can now process a drawback claim within 2 minutes. Projects like this

are incredibly helpful to the organization as employees can spend less time manually processing claims and focus on more pressing, complicated tasks.

One project that I am currently working on is processing pre-alert invoices. These pre-alert invoices are documents containing information on shipment logistics. Importantly, these invoices are sent to Axcelis before we actually receive the physical shipment. This means we can check the parts, compare them against our 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 a different format–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. As well, I was not experienced enough to tackle this problem prior to taking courses on artificial intelligence and machine learning at UVA. However, after acquiring the prerequisite knowledge required, I am attempting to solve this problem by utilizing artificial intelligence in a multi-stage processing system to extract the part numbers from each invoice. As an overview, the system first takes in an invoice and uses optical character recognition to extract the text. Then, I trained a machine learning model to help classify parts–given that there must be a way to distinguish between part numbers and non-part numbers–and compare them to our database.

Preliminary Literature Review & Findings:

There is a growing interest in exploring the regulation of artificial intelligence between various countries particularly the United States, the European Union, and China [3]. For this project, I have chosen to focus solely on China and the United States because they have

dominated the AI marketplace in the 21st century–with over 83% of AI companies based in those two countries [3]. Furthermore, over the past ten years, China has published more AI research papers than the United States per year and passed numerous plans to catalyze internal AI development rather than impose restrictions like the EU has [4]. From my research, I have identified several unique Chinese AI initiatives that could feasibly be adopted by the United States government. For instance, the creation of an elite AI "national team"–essentially a conglomerate of the most powerful Chinese companies supported and funded by the government–has been massively successful in aligning private industry with the Chinese government's AI goals [2]. China's *Artificial Intelligence Standardization White Paper*, published by the State Council, has helped establish AI development standards that are imperative for sustainable domestic development, international compliance, and transparency [5]. This topic has been explored by many STS researchers who focus on critically assessing the regulation between the two countries. A common struggle for these researchers was directly connecting the influence of particular policies to AI innovation and addressing public concerns such as data privacy and transparency. My work will use the analysis of certain Chinese policies and aim to resolve the latter difficulty by examining modifying these policies to consider values held both by the government and citizens of the United States.

STS Project Proposal:

STS is the study and practice of utilizing different frameworks, tools, and perspectives to assess how science/technology and society impact one another. For instance, traditional engineering discretely explains the form and function of scissors: to cut paper, hair, packaging, etc,. STS, on the other hand, explains how a scissors' design favors right-handed users and marginalizes left-handed individuals, necessitating the design and creation of left-handed

scissors. This ability to consider and assess how various social groups, organizations, or entities impact and are impacted by technology is the essence of STS. Given this definition, this project will utilize an STS framework to examine how government agencies have enacted and can construct policies to directly facilitate and support the progression of artificial intelligence within the United States and China. In that respect, this project is certainly an STS project as it will explore the governing bodies' goals and policies which have influenced their specific AI initiatives.

I am approaching this problem with an emphasis on policy. Given that AI is a rapidly expanding and revolutionary field, it's imperative to design and implement policies that will sustain and promote long-term AI development while restricting the technology where necessary–as deemed by the public and governing bodies. For my research, relevant authors include Susan Aaronson, Jeffery Ding, and Greg Allen. Aaronson highlights global AI policy challenges pertaining to AI privacy, transparency, and bias [7]. Ding provides a strong analysis of policies with an emphasis on how their political position guides their respective policies [8]. Both Aaronson and Ding's work will be essential to understanding the efficacy of certain AI policies in China, their potential limitations, and their necessity in international AI development–particularly with respect to AI standardization. Greg's work provides detailed recommendations for the United States government to facilitate AI development for both competitiveness and national security–which is an interesting and important consideration when advocating for the progression of AI [9].

For this project, I will be using Carol Bacchi's "*What's the Problem Represented to be?*", or WPR, approach to answer the guiding question of this paper: *what policies or strategies can be examined from China and applied to promote AI growth in the United States while adhering*

to social considerations [6]*?* I define the WPR framework as a method to critically assess and analyze potential issues with the formulation of problems. Specifically, WPR is concerned with what problems are represented and why, what issues are left unresolved or not mentioned, what implications or assumptions are made from stating the problem, and how the problem can be rewritten or deconstructed to better encapsulate the issue at hand. This is an incredibly powerful tool that can be used to assess the efficacy of public policy by understanding the underlying policy issue and determining the unexamined assumptions of the problem formulation. WPR can thus be used to construct more powerful and effective policies by iterating through problem representations rather than policy solutions. This theory will be paramount to my research as understanding the problem assumptions and shortcomings of a policy is more important than directly assessing its influence and impact on the technology or society at hand.

Given that there are many resources and papers pertaining to this subject, I plan on writing a policy-based literature review to highlight various strategies and initiatives in China related to artificial intelligence, with the objective to analyze their effectiveness in accelerating AI growth and determine whether the United States could reasonably adopt such a policy.

Barriers & Boons

As with other STS projects in this area, part of the difficulty of this project will be connecting China's regulation on AI to innovation and development. Especially with many overlapping initiatives and strategies, it's hard to discern the true effect certain policies had. To resolve this, I will find sources that assess a particular policy and have a quantitative metric to demonstrate its effectiveness in addressing the underlying policy issue–generated by the WPR method. Another potential difficulty is my utilization of the WPR method. I have a general

understanding of the approach but, in practice, I might not properly employ the framework or critically identify the problems from underlying assumptions in the policy issue. I will certainly practice this skill and likely read more resources on how to properly implement the WPR method for policy analysis. Similarly, I might overlook or not consider certain consequences of adapted policies from China to the United States. For instance, creating a governmental standardization effort for AI has worked well for China because companies can share technology more easily with standardized development. However, if such a policy was implemented in the United States, it's not a guarantee that private sector companies would adhere to or appreciate this standardization. Understanding the implications of these policies is difficult and something I will have to research carefully. Beyond practicing the WPR method, I could also resolve this by finding sources relating to similar regulations and cross-examine their effects individually.

References

1. Siripurapu, Anshu and Noah Berman. "Contentious U.S.-China Trade Relationship." Council on Foreign Relations, 14 May 2021, www.cfr.org/backgrounder/contentious-us-china-trade-relationship.
2. Chen, Jiaxin. "Drafting China's National AI Team for Governance." DigiChina, 6 June 2019, digichina.stanford.edu/work/drafting-chinas-national-ai-team-for-governance
3. Lewis, James Andrew and Andrew Imbrie. "US-China relations in the age of artificial intelligence." Brookings Institution, 13 May 2019, www.brookings.edu/research/us-china-relations-in-the-age-of-artificial-intelligence/.

4. Chen, Eddan. “Who is Winning the AI Race: China, the EU, or the United States?” Center for Data Innovation, 28 August 2019, datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/
5. “Artificial Intelligence Standardization White Paper (2021 Edition).” Center for Security and Emerging Technology, 21 Oct. 2021, cset.georgetown.edu/publication/artificial-intelligence-standardization-white-paper-2021-edition
6. Bacchi, Carol L. Introducing the ‘What’s the Problem Represented to Be?’ Approach, 2019, pp. 427–430, https://doi.org/10.1515/9783839437322-031.
7. Aaronson, Susan A. Building Trust in AI: A Landscape Analysis of Government AI Programs, Mar. 2023. https://www.cigionline.org/static/documents/no.272_k4xzqLo.pdf
8. Ding, Jeffrey. "Deciphering China's AI Dream." Future of Humanity Institute, University of Oxford, 14 Feb. 2018, https://www.fhi.ox.ac.uk/deciphering-chinas-ai-dream/.
9. Allen, Greg, and Taniel Chan. "Artificial Intelligence and National Security." Harvard Kennedy School, Belfer Center for Science and International Affairs, May 2017, https://www.belfercenter.org/publication/artificial-intelligence-and-national-security.
10. Fuglsang, Lars. Three Perspectives in Science, Technology and Society Studies (STS) in the Policy Context. Roskilde University, 2000. link