Thesis Project Portfolio

LLM Integration for PDF Generation at CGI

(Technical Report)

The Influence of Scientific Consensus and Regulatory Policies on Human Germline Editing

(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 in Computer Science

Aditya Kumar

Spring, 2024

Department of Computer Science

Table of Contents

<div align="center">

Sociotechnical Synthesis

Aditya Kumar

April 20, 2024

</div>

Introduction

My technical project revolved around the development of an automated PDF generation system using large language models at CGI and my STS project revolved around the international discussion and policy creation around CRISPR. The motivation of the technical project at CGI was driven by the need to enhance organizational efficiency and accuracy in converting paper based processes to digital formats, addressing a crucial aspect of digital transformation in modern business practices. I wanted to create a system that got rid of tiresome paper forms, like the ones found at the DMV, and digitized them into the modern era. This initiative aimed to leverage new technology (LLMs and OCR) to streamline operations, minimize human error, and provide a foundation for further digital innovations. My goal was to create a platform in which paper forms could be uploaded and automatically be converted into digital forms.

The motivation behind the STS research paper was to delve into the ethical and regulatory landscapes surrounding the emerging technology of CRISPR, especially in the context of human germline editing. I am fascinated by the research conducted on CRISPR and its associated implications. I set out to inquire what was being done to ensure that this research is conducted ethically and responsibly as well as finding out how certain policies have come about and have been shaped by the use of the technology. This is because CRISPR can be used for a wide variety

of use cases, some good and some bad, so it is important that there are policies in place to ensure equitable and positive access.

While there are not too many overlaps between the two projects, both highlight the emergence and implementation of new technologies and how they are being used in industry, with the technical report focusing on actual implementation and usage of the technology and the STS paper focusing on its implications to society.

Technical Research Summary

The capstone project focused on an endeavor to revolutionize the way that CGI handles paper based forms and documents. By leveraging the power of large language models, specifically LLaMA, and integrating it with Azure optical character recognition technologies, we aimed to create a sophisticated system capable of accurately extracting and categorizing essential data from scanned documents. This process, which traditionally relied on time consuming and error prone manual data entry, was transformed into a streamlined workflow that minimized human intervention and maximized efficiency.

However, the system we developed went beyond simple data extraction. By carefully designing a digital framework that replicated the original paper form's intent and functionality, we were able to enhance the user experience with interactive fields, automated data validation, and other digital features that improved the overall usability and accuracy of the process. Throughout the project, we remained focused not only on achieving a high level of efficiency, but also on setting a new standard for accuracy and user engagement in digital form processing. The success of this

initiative demonstrated the immense potential of LLMs and other advanced technologies to transform the way businesses operate in the digital age.

STS Research Summary

My STS research paper focused on the exploration of the ethical discussions surrounding CRISPR technology, with a particular focus on its application in human germline editing and how this technology and its use in specific countries have shaped international policy regarding the topic. By employing the Social Construction of Technology framework, I was able to analyze the interplay between various stakeholders, including scientists, ethicists, policymakers, and the general public, and how their perspectives and actions shape the development and perception of gene-editing technologies.

As I delved into the research, I found a host of ethical concerns, ranging from the potential for genetic inequality and the implications of modifying human embryos to societal impacts of possessing such technological capabilities. One of the key insights that emerged from this research was the critical role that global consensus and regulatory measures play in shaping the ethical use of CRISPR. As the technology continues to advance at a rapid pace, it is imperative that governance frameworks evolve alongside it to ensure that its application aligns with societal values. The study highlighted the need for ongoing, interdisciplinary dialogue and collaboration between experts and regulatory bodies to navigate the ethical landscape of gene editing and to ensure that its development serves the greater good of humanity.

Conclusion

Working on both the technical capstone project and the STS research paper simultaneously was an enriching experience. As I navigated the intricacies of developing an automated PDF generation system using LLMs, I found myself constantly drawing parallels to the ethical challenges explored in my STS research. The technical project's emphasis on data integrity and process improvement mirrored the concerns surrounding precision and potential societal impacts that were central themes in the CRISPR debate. As I researched the impacts of CRISPR on society, I made sure to keep these lessons in mind while developing the PDF system, ensuring that none of my work could be misused and that all data was secure.

This parallel exploration also led to a deeper, more nuanced understanding of the broader implications of technological advancements. It became clear that innovation cannot be pursued in isolation from its potential effects on society and that a balanced approach that considers both technical prowess and ethical awareness is essential. The combination of these insights highlighted the immense value of interdisciplinary engagement in technology development. This showed me the benefits of a comprehensive approach that integrates cutting-edge technical skills with a keen understanding of the social and ethical contexts in which they are applied. While developing the PDF program, I always kept in mind what societal impacts it may have, such as people losing their jobs due to automation, and how I could mitigate or find ways to repurpose those people. These insights were inspired by the research I was doing in STS and how the CRISPR technology affects society.

The experience has not only enhanced my technical abilities and research skills but has also instilled in me an appreciation for the complex interplay between technology and society. It has

reinforced my belief that as we continue to push the boundaries of what is possible through scientific and technological advancements, we must remain mindful of our ethical responsibilities and strive to create a future in which innovation and societal well-being go hand in hand.