DFIT: Cloud-based Service and Business Case

Examining How Rhetoric Affects the Development of AI Technologies

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 Systems & Information Engineering

By
Esther Yi

November 30, 2023

Technical Team Members: Mitch Mitchell and Franklin Glance

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

Joshua Earle, Department of Engineering and Society

Roman Krzysztofowicz, Department of Systems Engineering

Wray Mills Director of Engineering at CapTech Consulting

Introduction

Forecasting and modeling functions and distributions are critical in understanding interactions between variables. There are currently many coding libraries, packages, and other stand-alone softwares with this objective of taking in data to be able to find values and then applying those parameters to optimize decision making. Each software comes with a different set of distribution models that fits data to it and a different methodology of fitting the data itself. Some excel in visualizations while others focus on the process of reading data and parametrizing it. Currently, Doctor Roman Krzysztofowicz has a proprietary method of modeling probability distributions using the Uniform Distance method to estimate parameters to be ranked based on their goodness-of-fit to the data that he made into a software called The Distribution Fitter, or DFit. This software was created and maintained by CapTech Consulting, and is referenced in Dr. Roman Krzysztofowicz's upcoming book *Probabilistic Forecasts and Optimal Decisions*, to be published by Wiley & Sons, Ltd. For this capstone project, there are three main goals: (1) remodel DFit and move it to a cloud-based infrastructure, (2) build a business case to compare its performance and capabilities to that of potential competitors, and (3) forecast the demand of Dr. Krzysztofowicz's upcoming book and, subsequently, the demand for DFit.

The STS project will then focus on how rhetoric around technology affects the consumption and development of artificial intelligence technology. Because AI is a much more widely understood and practiced technology, there have been a lot more scientific studies on it. The study of AI is ultimately a means of extrapolating information for the distribution fitting software. By understanding how society reacts to AI, a better understanding of some of the human factors of DFit may be concluded. When looking at the forecast of the demand of the textbook, the functions and parameters of the demand are determined by patterns of human

consumption. There are many social and physical factors that play a role in what information consumers are fed, which in turn changes the perception of technology. It is important to study those factors to get a better understanding of the interaction between technology and society. The research questions that should be answered are what factors play into the interests of society, how does the language around AI technology change that, and how has social media changed the way that technology is developed? Research on how language and perception affects consumerism for AI will be observed, then how social pressure affects technology development, and, finally, the societal impacts of how AI is developed and used will be analyzed.

Technical Project

The technical project is split up into three parts: (1) moving the software to the cloud, (2) building a business case, and (3) forecasting the demand of the book and software using our software. The DFit software was created by and maintained by CapTech and will be used by Wiley & Sons, Ltd (our clients). The software will then be launched as a software as a service product to be able to be used by students as an aid in the textbook and potentially commercial clients of CapTech

The current software was built on an older version of a technology called Ruby on Rails, and when it was updated to be on the newer version, new bugs were discovered. Our goal is to reprogram the software that creates the probability distributions onto django (open-source program for web framework building) and then move the software to the AWS Cloud9 container so that it can be used on the cloud. The fitting engine itself is not going to be touched but the way that it interacts with the user will be changed. In order to remake the website and move it to the cloud, a Docker container is going to be made with the distribution fitting engine that does the actual reading and analysis of the data. The transition to cloud computing allows for a quicker

and heavier computational load. By moving the frontend version of the web applications from Ruby to Django, the bugs in relation to Ruby should be resolved. The remaking of the UI/UX will also allow for a more intuitive experience for the potential users.

Building the business case begins while the software is being debugged and migrated to AWS Cloud9. While research is done on competing softwares and libraries that have distribution fitting, we can get a better understanding on where others stand and in what ways we currently excel and do not do well in. Building this business case is crucial when presenting the final deliverable (the software on the cloud) to CapTech so that it can be leveraged as effective and worth maintaining. An experiment will be run to test how each of the softwares performs against each other. The first part will use a sample with a known distribution and known parameters values to evaluate how close they are to the true values. The second part will make fitting patterns with no known true distribution to compare the results from the different softwares that we are using. Competitors are using different fitting models which may or may not be as effective. The final part of the experiment will be to test the loadings of the DFit software and competitors. The limit for the data loading and the time it takes to do the computation will all be analyzed.

The final portion of the deliverable will be to forecast the demand of the book and the software, as well as the revenue of the software, over the next three years. This is a challenging forecasting problem, as the book has many sources of demand across the sciences and engineering disciplines that it's applicable to, but is important nonetheless as CapTech should be able to justify the cost of the program. First, a new conceptual model of the dynamic between the demand of the book, the demand of DFit, and the revenue by DFit, needs to be made. Then the methodologies from the book for statistical forecasts can be implemented to general actual

estimated and expected values. The first two portions of this project will become very important here, as whether or not the software is superior to its competitors will be a major factor in its demand, and especially its revenue.

STS Project

Analyzing rhetoric and public perception of technology and how that affects supply and demand provides better insight on the growth of competitors for DFit. Although DFit is in the niche market of continuous probability distribution modeling, the politics behind it should still be considered. One way to approach this analysis of a niche subject is by examining a more well studied and well known computer science technology: artificial intelligence. It is critical in understanding how society consumes language behind technology, because consumers should be well informed of what they are buying into in order to protect themselves from underlying biases and being taken advantage of. My project will mainly be focusing on discourse analysis to better understand the general population's beliefs on artificial intelligence and the power dynamics of the market. How AI is perceived by different social groups, who is developing AI, the repercussions of AI, and how those repercussions affect the supply and demand of AI are all aspects to be studied.

Some of the relevant social groups that will be examined are the people who are implementing AI and gaining from its popularity, social groups who are disenfranchised by AI, the media, and the general public. Those benefiting and those hurt by the growth of AI are important to study in order to better understand how AI and those implementing it can wield it as a means of perpetuating discriminatory ideas. This is because understanding the scope and breadth of how AI is being used is integral in understanding why there is a certain image behind it. That image can change how AI is developed and consumed. The media should be analyzed

because they tend to be the group who creates the framework and identity around ideas for the public. Because the majority of the beliefs of people come from how the media portrays news and stories, they play a critical role in my study. Ultimately the media determines how a large population of people perceive artificial intelligence, and how that translates itself into the development and implementation of it is to be analyzed. Finally the last major group is just the general population. Although the social group boundaries of the general public are ambiguous, how they in totality interact with AI, what they think it is, and how they consume it is important to understand. This is slightly different than the implementers and disenfranchised, because there is less motive with the whole population. There is no motive to promote or dismantle AI. Thus, they provide a more objective understanding on how rhetoric from the media affects technology.

The method of study for the research will be critical analysis. Critical analysis looks at how language creates power dynamics for discrimination and control. Many aspects of supply and demand are based on the power dynamic of groups in power influencing media, and subsequently the general population. These groups tend to be large corporations and they are able to create narratives that push and pull supply and demand of products. They tell society what they need to be successful and what they should buy. This inherently creates a discrete power structure of manipulation which can be examined in the rhetoric of artificial intelligence. This framework is necessary and explains the scope of the project well, because it directly tackles language and perception,

My timeline for this project starts with first determining what AI really is. Then, that true definition of AI should be compared to what the general public thinks it is and what the general perception of it is. Next, it is important to evaluate how that perception affects the development of the technology, if it does at all. Finally, the people who benefit from AI and those who are

further marginalized by the development of AI will be considered. Ultimately, this will help better understand how rhetoric and general power dynamics in the tech industry affect the development of AI; those patterns can then be applied to \ the DFit technology my technical project is on to build a better business case.

Key Texts

One of the texts that will be addressed is Sergio Sismondo's Science and Technology Studies. In his paper, Sismondo argues that the simplification of technology to make it easier to digest and understand distorts the reality of the technology. This is evident in artificial intelligence especially when certain technologies become more popular by society. The definition and boundaries of that technology can be diluted because consumers when some technology is well sought out, many people will stretch the definition of their new work to gain the attention of the consumers. This is important in my paper as it will help explain the repercussions of this behavior and how it affects the growth, development, and perception of AI.

Nadem's *Public understanding of artificial intelligence through entertainment media,* is important to this paper because it examines the general public perception of AI from the media. Because the media is currently one of the biggest influencers of society, the narrative behind new and well anticipated technology is determined by entertainment. It is important to understand what societal expectations and perception are around a technology to understand why technology is being created and evolved a certain way. Because engineering is also built upon supply and demand of science, the demand for people should be measured. In this journal, the extent to which people are aware of artificial intelligence in technology they use everyday was evaluated and this information will be used to help understand the relationship between tech development and human understanding. .

Zhang and Dafoe's research in *Artificial Intelligence: American Attitudes and Trend*s explains the public perception on AI. More importantly, it goes into the demographics of those who support AI technology growth and those who do nor. This is important to understand, because public perception does not necessarily correlate to the development of the technology. If the small group of people who have power has a different agenda than the general public, there is little that can be done from the public's standpoint and Zhang and Dafoe's work evaluates this phenomenon, specifically within the lens of AI.

Wodak's book *Methods of Critical Discourse Analysis* is important for me, because it describes the framework for which I will be basing my paper on. The book describes critical discourse analysis by saying that it is fundamentally concerned with analyzing opaque as well as transparent structural relationships of dominance, discrimination, power, and control as manifested in language. This is important because it can help explain some of the bad effects of uncontrolled growth of AI that is being implemented today and how it contributes to already existing unfair power dynamics and perpetuates stereotypes. It addresses and explains the relationship between rhetoric and power, which can even be mirrored in technology like artificial intelligence.

Works Cited

Dafoe A, Zhang B (2019) Artificial intelligence: American Attitudes and Trends. Center for the Governance of AI, Future of Humanity Institute, University of Oxford, Oxford. https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/

Nader, K., Toprac, P., Scott, S., & Baker, S. (2022). Public understanding of artificial intelligence through entertainment media. AI & society, 1–14. Advance online publication. https://doi.org/10.1007/s00146-022-01427-w

Sismondo S. (2009). An Introduction to Science and Technology Studies, Wiley

Wodak, R. (2001). Methods of Critical Discourse Analysis. Sage Publications.