

**ACCESSIBLE VISUALIZATION PLATFORM FOR ADVERSARIAL NATURAL  
LANGUAGE PROCESSING**

**EFFECTIVE USAGE OF UNSTRUCTURED DATA USING NATURAL LANGUAGE  
PROCESSING IN THE FINANCIAL INDUSTRY**

An Undergraduate Thesis Portfolio  
Presented to the Faculty of the  
School of Engineering and Applied Science  
In Partial Fulfillment of the Requirements for the Degree  
Bachelor of Science in Computer Science

By  
Soukarya Ghosh

May 6, 2021

## **SOCIOTECHNICAL SYNTHESIS**

This research paper discusses the growing subfield of machine learning known as natural language processing (NLP), which deals with deciphering and making sense of natural human languages using a computer. The technical portion of the research focuses on TextAttack, an adversarial framework for data scientists to use and study how malicious actors can potentially undermine the power of natural language processing by attacking its weak spots. In an attempt to make this educational material more accessible, a web application has been developed, housing the features of TextAttack, so that academics can conduct research more easily. The STS portion of the paper discusses the community forged around the applications of NLP, specifically with respect to the financial industry, and how there exists a holistic incentive loop that ultimately advances the technology. These applications can range from converting unstructured qualitative data into quantitative metrics to using natural language processing to speed up decision making in the financial markets. The technical report and STS research paper are tightly connected through the application of the product of the technical report directly to aid the problem presented within the STS research paper.

The technical report is based around a Python framework called TextAttack. Developed by a group of graduate students and technical advisor Yanjun Qi, TextAttack aims to provide a tool to study adversarial attacks in order to improve existing model performance and robustness. Although TextAttack has already seen significant usage with over a thousands stars gifted on the code storage platform Github, the issues of accessibility persisted. Running machine learning algorithms on personal computers can be demanding and computationally expensive. This makes the process of research rather cumbersome.

The technical report provides a potential solution through an easily accessible web interface through which the research can be easily conducted. I built a fully scalable web application hosted on Amazon Web Services that allows users to access a limited scope of TextAttack functionality. At its current stage, the web application allows for user input and the execution of an array of text attacks that modifies and outputs the resulting strings. This not only gives an insight into the prowess of TextAttack to help onboard potential users, but also provides a platform to execute the functionality without needing a capable computing machine.

The STS research focuses on the use of NLP within the financial industry and how there exists an incentive loop that helps increase the potency of the technology. The SCOT framework views the advancements of NLP as the summation of the conglomerate that interacts with it directly, engineers and the financial world, and those that are indirectly affected by it, the general public and retail traders. It highlights the circular nature of the incentives within this finance space, where each group is incentivized to help improve NLP by those whom they will affect and by those who preceded.

Although there exists intergroup tension with conflict of interest in investment strategies, their views end up aligning in their joint effort in advancing predictive and efficiency increasing technologies through the education of and increased research efforts towards NLP. This was most evident in the events that took place early in 2021, where the retail traders of an internet forum known as WallStreetBets joined forces to take down the finance mammoths of Wall Street who were shorting the stock of GameStop. In their success, the retail traders nearly led the United States economy into a recession akin to the 2008 financial crisis by causing some investment firms to default. However, these short term clashes and dissonance are ultimately non important when viewed from the macro perspective of allowing the advancement of NLP technologies in

order to aid the decision making process that will help nations avoid recessions due to unnecessary risk taking and aloof decision making.

Overall, NLP has become a force to be reckoned with in all fronts of business and technology. It aids human decision making efforts by providing easy to read heuristics and filtration of large amounts of data. With these advancements being made, firms, as well as retail traders, are closer to a world where decision making can be aided by the use of a majority of market data, rather than the exclusion of large amounts of unstructured qualitative data for their slowness in manual conversion.

## **TABLE OF CONTENTS**

### **SOCIOTECHNICAL SYNTHESIS**

#### **ACCESSIBLE VISUALIZATION PLATFORM FOR ADVERSARIAL NATURAL LANGUAGE PROCESSING**

Technical advisor: Yanjun Qi, Department of Computer Science

#### **EFFECTIVE USAGE OF UNSTRUCTURED DATA USING NATURAL LANGUAGE PROCESSING IN THE FINANCIAL INDUSTRY**

STS advisor: Catherine D. Baritaud, Department of Engineering and Society

### **PROSPECTUS**

Technical advisor: Yanjun Qi, Department of Computer Science

STS advisor: Catherine D. Baritaud, Department of Engineering and Society