Brendan Fisher

Our project will focus on creating a service for social media websites and similar account based websites that can be used to detect users who may be attempting to break the account rules of that website. The service will track details such as the behaviors, word usage, and IP addresses of users, and store these patterns in a database. The service would then use this data to find users breaking rules, such as ban evasion or account sharing. If a suspicious account is found, it is likely that it is a user attempting to break a rule, and administrators will be notified with the details of the similar accounts. There would also be a library of implementations of the data collector, allowing it to be used on a variety of websites or programs. This service would allow websites to improve their moderation by augmenting the administrators with an automated approach.

The courses I have taken will aid me in completing this project. Probability and Statistics (STAT 2037) will help me in creating statistical analyses of data, and allow for a measure of certainty of the results. Software Engineering (EECE 3093C) can help in the creation of requirements for the project and the timeline for various components of the project. Database Design (CS 4092) will help me create a fast and space-efficient method of storing and retrieving data for the project. AI Principles and Applications (CS 4033) will help me create a smart system that can make optimal decisions based on various measurements.

My co-ops have also given me experience that will be extremely useful for developing this project. My co-op with Dr. Nan Niu gave me experience with various probabilistic algorithms used for semantics representation and pattern recognition. These algorithms can be used to determine the similarities between both typing patterns and behavior patterns of users. My co-op with Predictronics Corporation taught me about various methods of outlier detection and shape recognition. Outlier detection could be used to find anomalies in a user's activity, which could be used to determine if the user is sharing their account with another person. Shape recognition can be used to create representations of the numerical data collected on users, and could be used to more efficiently compare various features of the data.

The motivation for this project comes from the lack of current solutions for automated moderation of users. Currently, user accounts are generally managed by administrators looking through user reports on an account, which requires significant manpower, especially on large platforms which have many users and significant chances of false reports. This project could be the first product to address this problem in an automated way, which would likely attract the attention of many platforms. The expected results of this project would be to accurately detect users who are breaking rules, without the input of administrators. With this project, we hope to create a solution that greatly reduces operating costs by reducing the need for manual analysis of users. In addition, the system could detect violating users that may not be noticed otherwise, since human administrators would be unlikely to recognize more subtle behaviors.

The initial approach to this project will involve research into the various algorithms that could be used, as well as looking for software libraries that could be used to speed up development. After that, we would develop a program and run it on an example database to determine its effectiveness of the chosen method. Once we have a successful solution, we would work on refining the algorithms, as well as creating implementations for various platforms. The project will be done when it can interface with various platforms, analyze the user data, and output users it suspects are violating rules. We will know that the project is successful if we can manually analyze the outputs and confirm that the program has correctly identified a rule violation. We may use both testing datasets and well as real-world data in order to determine that our program is fulfilling its purpose.