Thesis Project Portfolio

Sleep Score Explanation: Improving Sleep Quality via Interpretable ML

(Technical Report)

The Failure of Microsoft's Tay and What it Means for AI Governance

(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, School of Engineering

Benjamin Orndorff

Spring, 2023

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

In my technical report, I propose the design of a new sleep analysis application with the goal of helping users improve their sleep quality. Lack of sleep has become a major health issue, with 35.2% of adults in the US getting less than the recommended 7 hours of sleep every night. Although many apps have been developed to track sleep, they only provide sleep scores, which are not easily interpretable by users. The proposed sleep application takes in sleep data and sleep scores from other apps and applies an explainable regression model to provide users with an explanation for their sleep score based on their habits. The app will recommend steps for users to take to improve their sleep score and ultimately their sleep quality.

In my STS research, I focus on AI governance and safety policies and apply the Social Construction of Technology framework in the context of the Microsoft Tay controversy. Microsoft Tay was an AI chatbot that Microsoft released on Twitter to learn from users. After being targeted by users on the platform, it began posting offensive tweets leading to it being shut down 16 hours after its release. To understand why Tay failed and what takeaways can be made for the AI governance field, I analyze the primary stakeholder's roles in the situation and the reasons why Tay failed to stabilize.

The AI industry has been focused on improving the accuracy and applications of AI but recently there has been growing concern over the ability to interpret the decisions AI makes. This concern has caused an increase in research on AI safety and AI governance to combat the potentially harmful effects of the application of AI. In my technical report, I apply AI safety methods to sleep scores to provide human interpretable explanations for them while in my STS research, I analyze the Microsoft Tay controversy in hopes of understanding the potential applications of AI governance and safety.