Natural Language Processing: Enhancing Transparency in Privacy Policies of Connected Devices

(Technical Report)

The Struggle for Privacy in Connected Homes

(Sociotechnical Research Paper)

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

John DeFranco

May 9, 2025

Contents

Preface

How can user privacy and security be better secured at minimal cost to usability and convenience? Users entrust their personal data to devices and companies every day, but most expect their data to be kept private and secure. However, users also value convenience, which can compromise privacy and security. This has led to a struggle between the two concepts.

How can natural language processing (NLP) be used to enhance transparency in the privacy policies of connected devices? Connected devices often come with privacy policies that are long, jumbled walls of text filled with legal jargon, causing users to either skim through them, or skip them entirely. NLP can be utilized to shorten, simplify, and summarize the text found in these privacy policies, enhancing user understanding. ROUGE-L score was used to measure how similar a machine-translated text was to each reference text, and Flesch-Kincaid Reading Ease Score (FKRES) was used to evaluate how easy it is for a user to understand the summarized text. On average, a summarized privacy policy had a ROUGE-L score of 0.1353 and a FKRES 3.52 points higher than the original policy, leading to a simplified privacy policy that users could understand and comprehend.

How are users, tech companies, privacy advocates, and regulators competing to determine the privacy standards governing connected residential systems? The internet of things (IoT) has introduced competing responses to protect privacy and security. To resist regulation, tech companies invoke user convenience, user responsibility, and the controls that empower users to protect their own data. Privacy advocates, however, contend that default settings, system complexity and marketing put users at a disadvantage that can be corrected only through third-party regulation by a public agency.