**Consumer Empowerment for Embedded Software Products**

**Introduction**

Base level software on consumer devices, such as the Apple Watch, Amazon Alexa, and Tesla, is critical to their function. Embedded software often spies on users by collecting data. However, consumers are ill informed about the Terms and Service (ToS) that allow companies to do so. Moreover, there exists no alternatives to surveillance. Current ToS is unfair because it gives consent for broad surveillance and obstructs the right to repair.

Our goal was to create a website that summarises ToS for embedded software and promote policy reform.

**Issues on Terms of Service**

Legal and technical issues require technical language to be effective, resulting in long and complex ToS. The outcome is the status quo where nearly no one reads terms of service. When consumers don’t read ToS it hinders informed consent because they do not know what they are agreeing to.

We ran a statistical model showing that the reading time of most ToS is longer than the Charter of Rights and Freedoms, and some will take longer to read than Macbeth!

**Technical Solution: Website, Machine Learning Model**

Our website informs users about the readability and privacy score of each device.

We calculate readability using the Flesch Reading Ease Model. This allows companies to know the readability of their ToS and raises consumer awareness.

The privacy score is calculated by an expert who reads the ToS and ranks them manually. There are two reasons we chose to do this manually. On the technical side, most machine learning models that deal with sequential data as its input have a 512/1024 token limit. Since ToS documents are way longer than 1024 words, we must segment the document into parsable pieces. We could do this using already existing sections in the document but some sections themselves are longer than the prescribed word limit. Moreover, there is the dearth of pre-trained models trained on ToS documents, or any other collection of text that is hard to read mandates that we develop and train a model on our own. A possible direction to create shorter sections would be to organise large sections/subsections into subsections through information extraction techniques. We could then proceed to pass this as input into a model (a transformer or other seq to seq models) to get an easily readable text as output. The lack of pre-existing labelled documents for such training suggests manually labelling as the way forward for future training of models on such a dataset.

Moreover, there are a limited number of consumer products, and whenever each product is updated, there are few changes between ToS. Therefore, it is not unreasonable to have policy experts read ToS.

**Policy Solution**

There exist no alternatives to surveillance if users want to use a product. The lack of alternatives undermines consent. Our policy approach is twofold. Firstly, we amend the Canadian Copyright Act to allow consumers to alter software on their devices. Secondly, we require hardware manufacturers to publish Hardware Abstraction Layer. Then, the embedded software monopoly that manufactures hold is broken. This allows other software companies to develop OS that respects privacy.