**High Level Goals**

Sprint 1 is to learn the primary tools and techniques we will use for the project and establish the fundamental concept of what our product will look like.

Sprint 2 aims to create an initial prototype of the model and have functional website components deployed.

Sprint 3 aims to complete frontend styling and add more polish to all components of the service, including more training, optimization and adding generalized parsing from HTML.

Sprint 4 aims to add additional features which would be nice to have and would enhance the user experience but are not strictly necessary.

**User Stories for Release**

*Sprint 1*

1. As a backend engineer, I want at least four cleaned datasets so that we can begin training

* Each backend engineer should find a dataset and convert it to TSV. Refer to <https://github.com/JEF1056/sum-everything/blob/main/data/gen-cnn.py>

1. As a user, I want the machine learning model to be able to encapsulate a variety of inputs.

* News articles
* Scientific journals
* Magazine articles

1. As a backend engineer, I want to be able to ingest and automatically train on a given input dataset without much effort (build a trainer)

* Backend spikes

1. As a user, I want to have a functional interface that allows me to upload ‘.txt’ files

* Frontend engineers should also clear out potential spikes, learning necessary technology such as react.js.
* Potentially useful: <https://ant.design/>

1. As a user, I want to see the process status of summarizing.
2. As a user I want the summary output in bullet point or paragraph format

Spikes:

* Learning how to efficiently convert datasets to TSV
* Learning Flask and frontend technologies

*Sprint 2*

1. As a user, I would like to be able to access the website in my web browser and be able to get a summary for simple .txt files
2. As a developer, I would like to have a model that performs accurately for most summarization tasks taken from the test set

* Evaluation methods: BLEU score, ROUGE?

1. As a developer, I would like to be able to visualize/gain statistical insights into my datasets to choose the most viable ones
2. As a user, I would like to view the summarized content in a simple and easy to read format.

Spikes:

* There will be a lot of experimentation to find good hyperparameters
* Learning Heroku

*Sprint 3*

1. As a user, I would like to be able to input a URL and allow the service to automatically extract useful information from the website.
2. As a developer, I would like to explore stable and cheap hosting methods, possibly by converting the model to lightweight formats

Spikes:

* There will be a lot of experimentation to find good hyperparameters
* Investigate possible CSS frameworks to use

*Sprint 4*

1. As a user, I would like to see documents that I have previously summarized.
2. As a user, I would like to have a functional interface which looks visually appealing.

Spikes:

* Investigating possible solutions for storing data (possibly locally)

**Product Backlog**

**Project presentation**