In my individual project, I worked on creating a browser extension built for chromium browsers that uses a trained machine learning model to detect and highlight product reviews that may have been generated by AI.

The machine learning model was trained on a labeled dataset with half human generated reviews and half AI generated reviews. Before the training the model removes all punctuation to remove unnecessary variables. Then the model tokenizes the text into numbers that it can understand easily and teaches itself methods of associating different elements with AI or Humans.

Once the model is trained, I use Microsoft’s Azure function app hosting service to host a function that can receive text via html requests, feed it to the model and then respond with whether it is AI or Human and how certain it is.

The extension works by checking if a page is supported, identifying which html tags to look for and then scanning the page for all reviews, it then sends the text to the model where it classifies the text as AI or human and responds with this information. The extension colors the text green or red according to its classification. For any page that is unsupported, there is a manual option where the user can select and send text to the model to be classified.

The goal of this project was to create an easily usable product built directly into the browser. There are already many websites where users can enter in text to check if its AI, but there are few extensions that do the same. I specifically focused on product reviews to make it easier to train a model on it and because that is a market with a high density of bot generated content.