# CS4804 - AI Chatbot

Group 5: Toru Oyama, Sean Sy, Anthony Castellone

Problem

- Many online resources are effective on paper
  - reading help documentation / browsing websites
  - where is the appeal as social animals
- Simplify information distribution and reduce redundancy
  - different people ask similar questions
  - people enjoy talking to others (interactivity / feedback)
- Solution: Chatbots
  - process user input and produce accurate, digestible responses
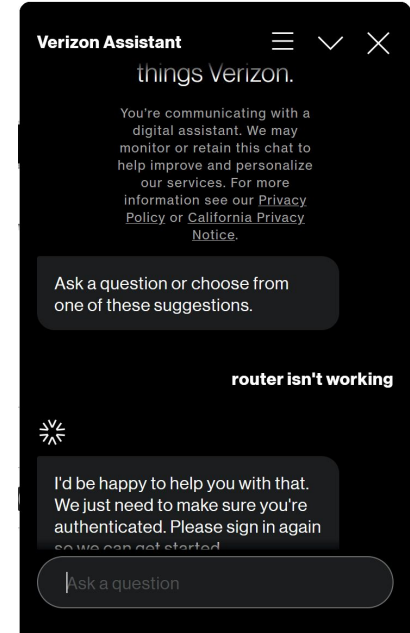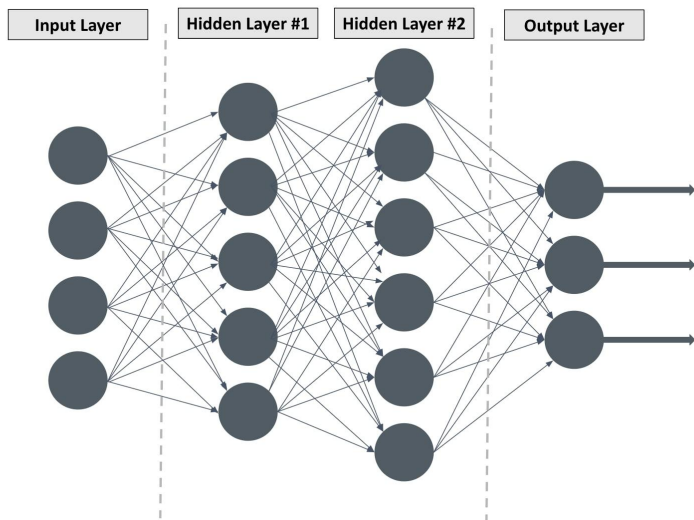 - minimal demands of the user

VS
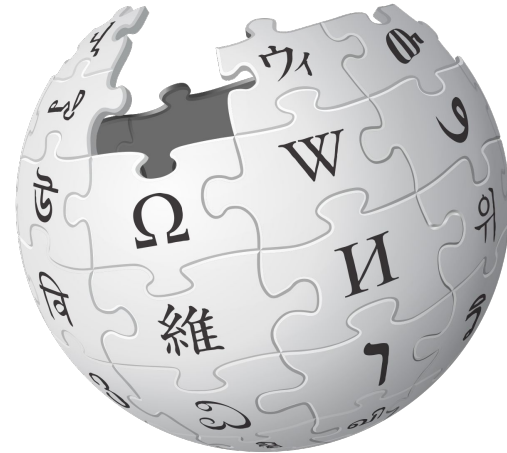
Use Cases

- ChatGPT (general purpose, huge dataset)
  - data gathered by large scale web scraping
  - sophisticated model due to variability of input
- Digital assistants (specific, smaller dataset)
  - great for frequently asked questions
  - data focused on one service
  - model complexity varies (rules-based)

Algorithm and model

- Split strings into list of words (tokenize)
- Stem or lemmatize tokens to identify general meaning
- Represent tokens numerically for model input
- Deep learning model using tflearn
  - input, output, and two intermediate layers
  - trained model can be loaded after training
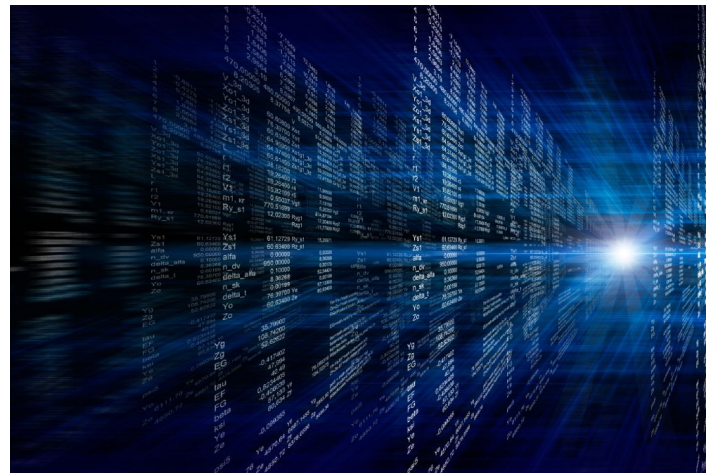- Predicts the probabilities given each token

Results / Demo

Lessons Learned

- Chatbot quality is proportional to data quality and quantity
- Future improvements
  - Increase dataset size
  - Deal with unknown queries properly
  - Finetune the model based on previous queries
- Learned practical applications of multiple libraries
  - Tensorflow for the learning model (tflearn/keras)
  - NLTK for word processing

Questions