

CS4804 - AI Chatbot

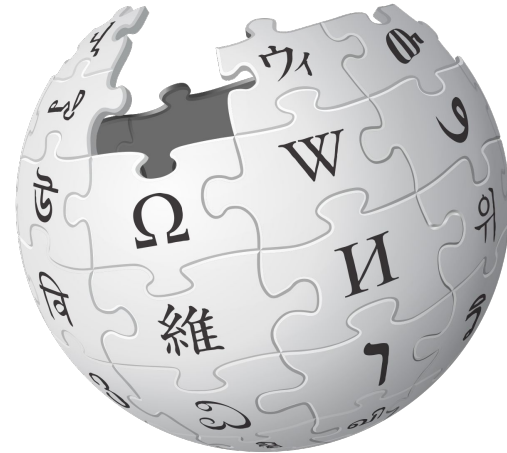
Group 5: Toru Oyama, Sean Sy, Anthony Castellone

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



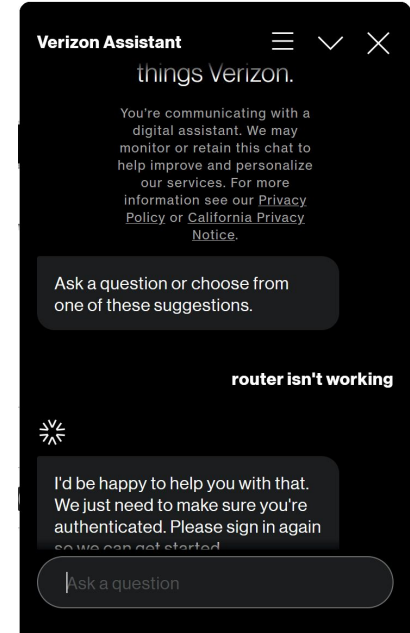
VS



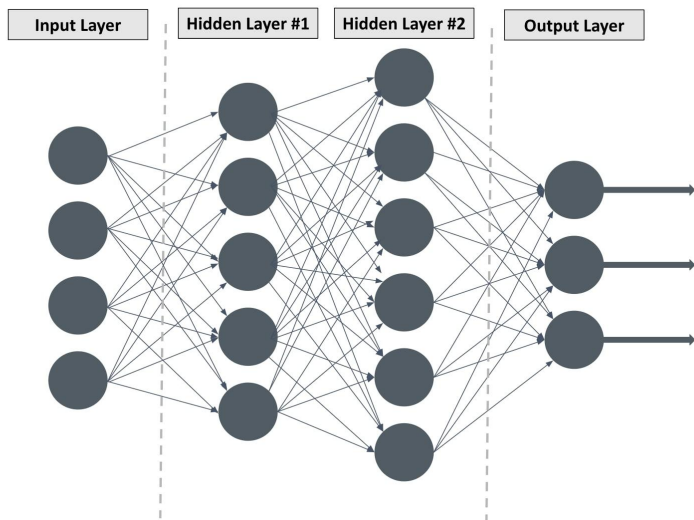
Use Cases



- 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

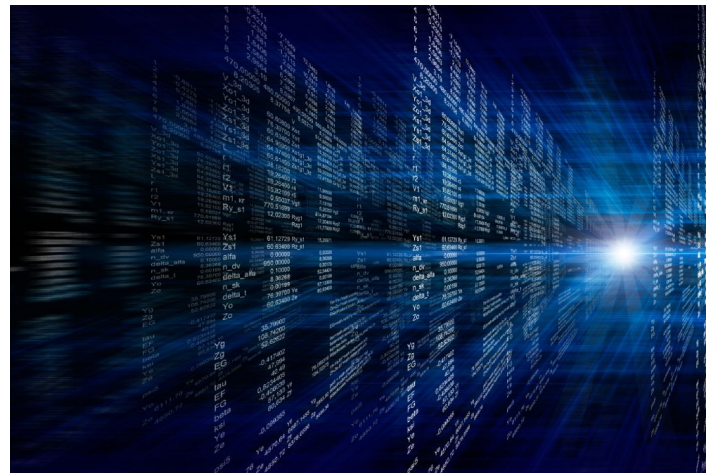
Results / Demo

```
| Adam | epoch: 989 | loss: 0.77953 - acc: 0.9469 -- iter: 30/30
--
Training Step: 990 | total loss: 0.70914 | time: 0.002s
| Adam | epoch: 990 | loss: 0.70914 - acc: 0.9522 -- iter: 30/30
--
Training Step: 991 | total loss: 0.64581 | time: 0.001s
| Adam | epoch: 991 | loss: 0.64581 - acc: 0.9569 -- iter: 30/30
--
Training Step: 992 | total loss: 0.58883 | time: 0.001s
| Adam | epoch: 992 | loss: 0.58883 - acc: 0.9613 -- iter: 30/30
--
Training Step: 993 | total loss: 0.53755 | time: 0.002s
| Adam | epoch: 993 | loss: 0.53755 - acc: 0.9651 -- iter: 30/30
--
Training Step: 994 | total loss: 1.44877 | time: 0.001s
| Adam | epoch: 994 | loss: 1.44877 - acc: 0.8753 -- iter: 30/30
--
Training Step: 995 | total loss: 1.31159 | time: 0.001s
| Adam | epoch: 995 | loss: 1.31159 - acc: 0.8878 -- iter: 30/30
--
Training Step: 996 | total loss: 2.12401 | time: 0.002s
| Adam | epoch: 996 | loss: 2.12401 - acc: 0.8123 -- iter: 30/30
--
Training Step: 997 | total loss: 1.91953 | time: 0.002s
| Adam | epoch: 997 | loss: 1.91953 - acc: 0.8311 -- iter: 30/30
--
Training Step: 998 | total loss: 1.73563 | time: 0.002s
| Adam | epoch: 998 | loss: 1.73563 - acc: 0.8480 -- iter: 30/30
--
Training Step: 999 | total loss: 1.57025 | time: 0.002s
| Adam | epoch: 999 | loss: 1.57025 - acc: 0.8632 -- iter: 30/30
--
Training Step: 1000 | total loss: 1.42151 | time: 0.002s
| Adam | epoch: 1000 | loss: 1.42151 - acc: 0.8769 -- iter: 30/30
--
Start talking with HokieBot! To stop the conversation, type quit, exit, or leave.
DISCLAIMER: HokieBot will attempt to use any prompt for a response, regardless of clarity.
You:
```

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