# Sports-Chatbot

## Development Steps

1. Find Data
   1. Import Data
      1. Google BigQuery
   2. Filter
   3. Parent child – Call and Response
   4. Save data in database
      1. MySQL database
      2. What encoding? UTF8?
      3. Do we save preprocessed data in the database?
         1. Preprocessing part of Keras
         2. Capture vocabularies
         3. Word to word sequences (x and y)
         4. Sequencing
         5. Embedding
         6. Activation layer
         7. Padding
      4. How to take advantage of my database programming skills
2. Determine Tools
   1. Spacy
   2. Tensorflow, PyTorch, or other tools
   3. Other
3. Process Data
   1. Tokenization
      1. Do we have to store all parts of speech?
      2. Sparse matrices
   2. Vectorizing
      1. N-grams
      2. Number of tokens
      3. Hashing (If more data = better performance)
4. Neural Network
   1. Recurrent Neural Network
   2. Training
   3. Long-Short Term Memory Networks (LSTM)
   4. Remembering and forgetting…
5. Neural Network Topology
   1. Input layers
   2. Hidden Layers
   3. Other considerations
      1. Performance
      2. Accuracy
6. Interaction with chatbot
   1. On machine/online
   2. Slackbot
   3. Flask
7. Packaging/Aesthetics

## Phase 2 Proposed

* More Data
* More Training
* Move to online
* Rules Engine
* Personality
* Voice to Text
* Profanity Translation
* Real time info
  + When is the next home game for the Red Sox?
  + What are the current standings in the American League
  + Sports statistics for (teams/players)
    - Named entity recognition
    - Intent recognition
    - Designing Bots – O’Reilly Book

Categorical cross-entropy with accuracy metric

Grammar evaluator with grammar score