# Question Answering with Deep Learning

Tristan McRae, Stanford University, Aeronautics and Astronautics

## ABSTRACT

Recent progress in deep learning has allowed for significant strides in natural language processing. Here, a method for using deep learning to answer questions was proposed and implemented. This method produced a model that can answer questions exactly correctly approximately 28% of the time.

## BACKGROUND

- The ability for a machine to answer any question asked of it has numerous applications from artificial personal assistants, to medical diagnoses, fact checking and beyond.
- Past approaches have involved representing words as numerical vectors and allowing a model to learn temporal patterns in sentences.
- The model "learns" based on seeing questions answered correctly.
- Because of the way it learns, the model is not effective at answering questions that are very dissimilar to what it has seen before.
- Models were trained and tested on the Stanford Question Answering Database which contains thousands of questions, contexts and answers.

## OBJECTIVES

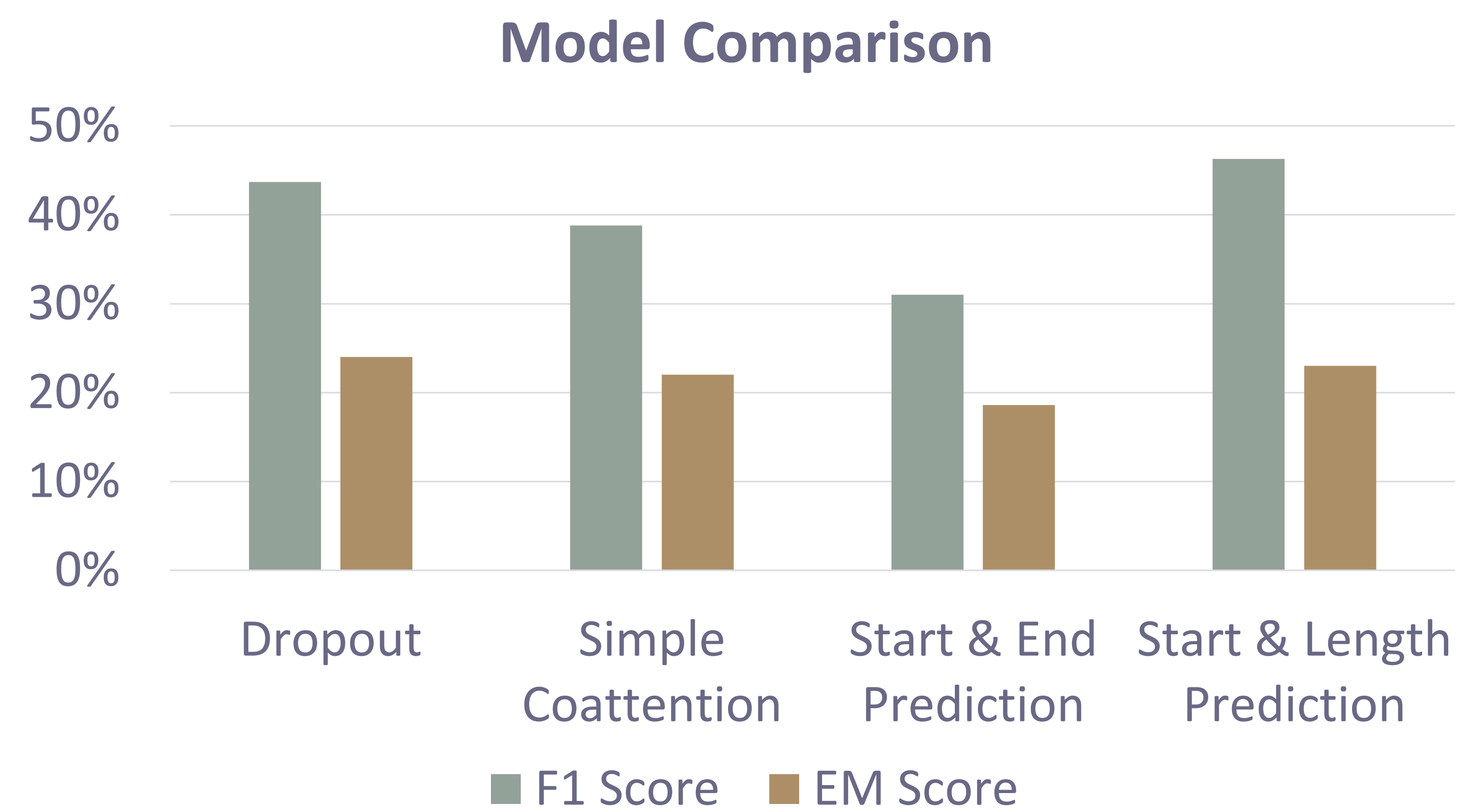
- This research sought to create a model that, given a context paragraph containing relevant information, can answer a question.
- Different deep learning architectures were tested for effectiveness in this task.
- Models were evaluated on previously unseen questions and given scores based on accuracy.

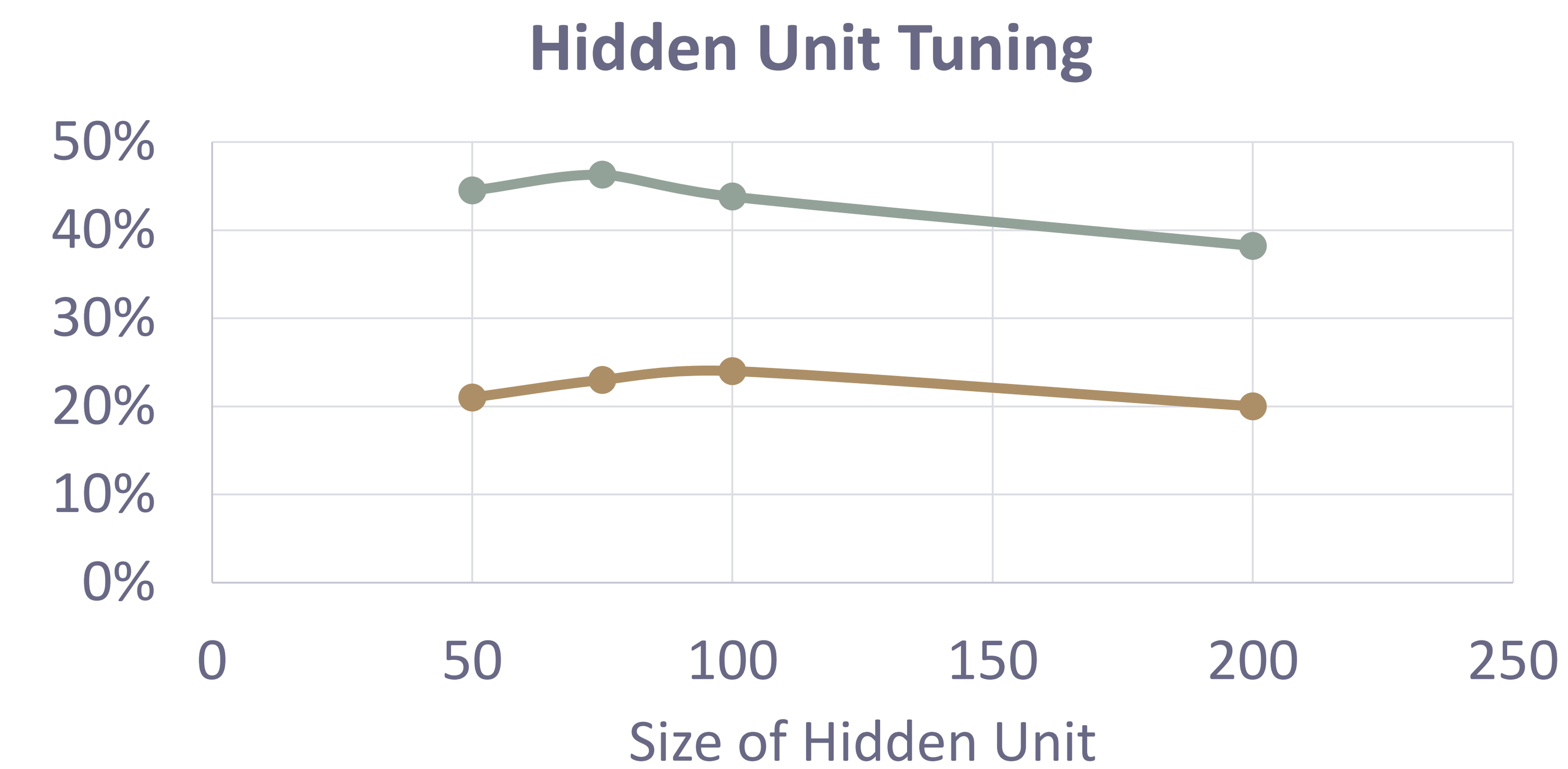
## METRICS

Two primary metrics were used to evaluate performance between models:

- Exact Match (EM) score: How often the answer given is word for word correct.
- F1 score: A measure of the number of correctly predicted words vs false positives and negatives.

## MODEL COMPARISONS




"Start & Length Prediction" was chosen as a model architecture. Hyperparameters such as hidden unit size were tuned for optimal performance.

## FINAL MODEL

When given a question and a context paragraph, the model goes through several steps to determine where in the context to look for the answer.

Replace words in the context and question with word vectors that represent their meanings.

Pay less attention to context words that are dissimilar to question words as they are less relevant.

Put the question and context through a bi-directional LSTM to find patterns forwards and backwards in time.

Multiply the outputs of the question and context LSTMs to create a matrix representing their relationship.

Predict where in the context the answer starts and how long it is based on the relationship matrix.

When tuned, this model obtains an EM score of 28% and an F1 score of 43% .

## CONCLUSIONS

- The EM and F1 scores obtained here are far from state of the art but they still demonstrate the ability of the model to learn some patterns.
- Future work could implement more complex model architecture to enable more complex patterns to be learned.