

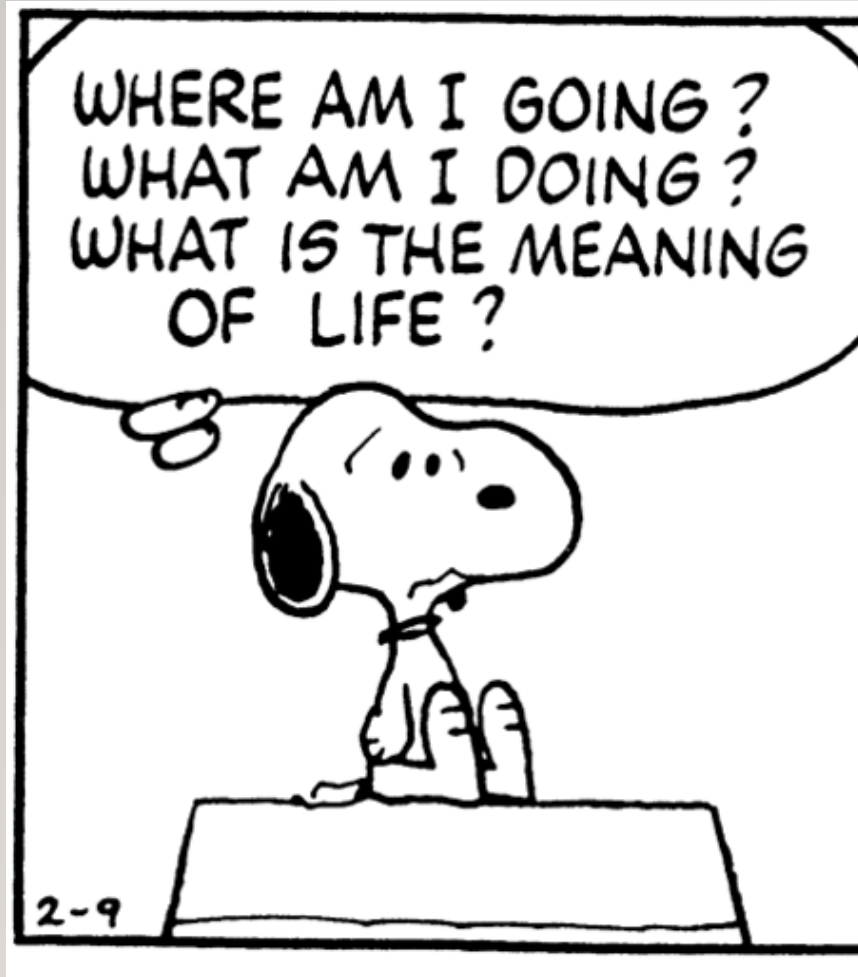


Question Answering with the Dynamic Memory Network

We answer the: what, why and how of
'Question Answering'.

Q: What is Question Answering

- A: It's a Natural Language Processing discipline which is concerned with building systems that automatically answer questions posed by humans in a natural language. There are 2 kinds of question answering (QA) systems: open-domain question answering and closed-domain question answering.



Open-Domain Question Answering

Open-domain QA deals with questions about nearly anything. However, these systems are much harder to create and usually have much more data available from which to extract the answer.

Closed-Domain Question Answering

Closed-domain question answering deals with questions under a specific domain (like in the example below). The questions are usually of descriptive nature and are a part of the context of which the question is asked. These systems require much less data. In this project we will build a closed-domain QA system.

Context	
Fred picked up the apple there. Bill travelled to the kitchen. Bill got the milk there. Jeff went to the kitchen. Bill passed the milk to Jeff. Jeff handed the milk to Bill.	
Question	Answer
Who did Jeff give the milk to?	Bill



Hi, how can I help?

Q: Why is Question Answering Important?

A: Products such as Siri, Alexa or Google Assistant rely on technologies such as Question Answering, amongst others to function. Research into Question Answering advances our ability to interact with machines, which leads to even more exciting products and services.

Q: How did you build
the Question Answering
System?

The Dataset - SQuAD

- Stanford **Q**uestion **A**nswering **D**ata set (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowd workers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage.

Passage Segment

...The European Parliament and the Council of the European Union have powers of amendment and veto during the legislative process...

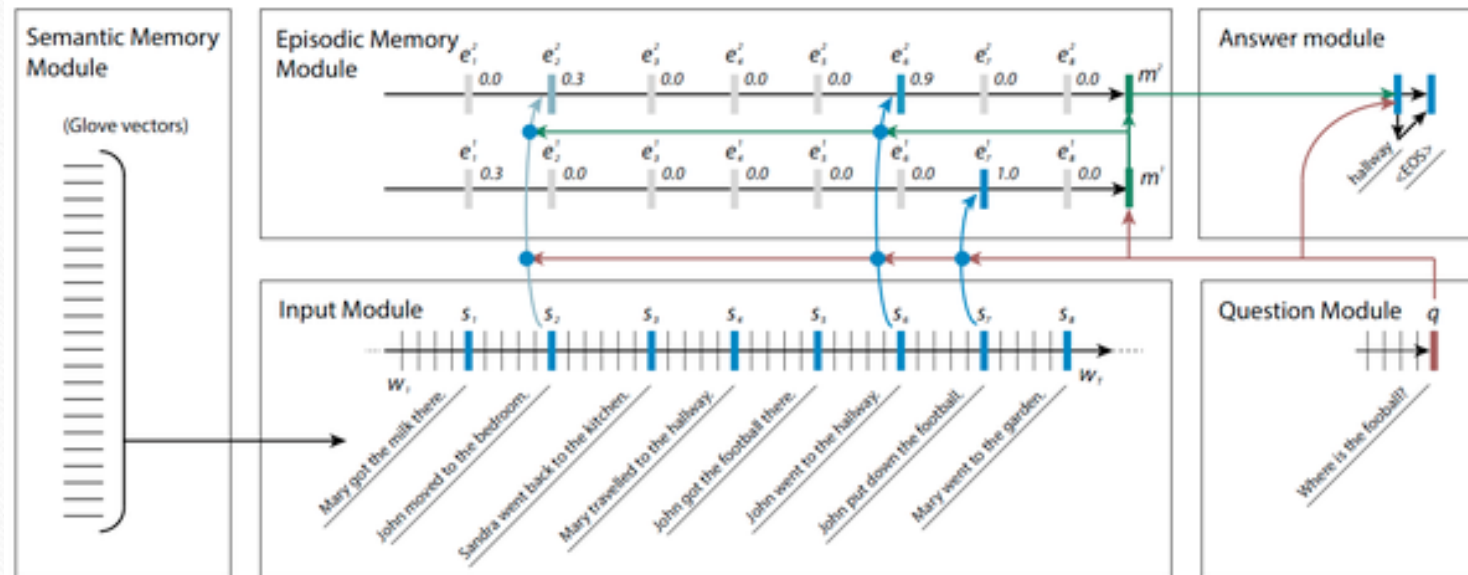
Question

Which governing bodies have veto power?

Fig 1. Example of the dataset

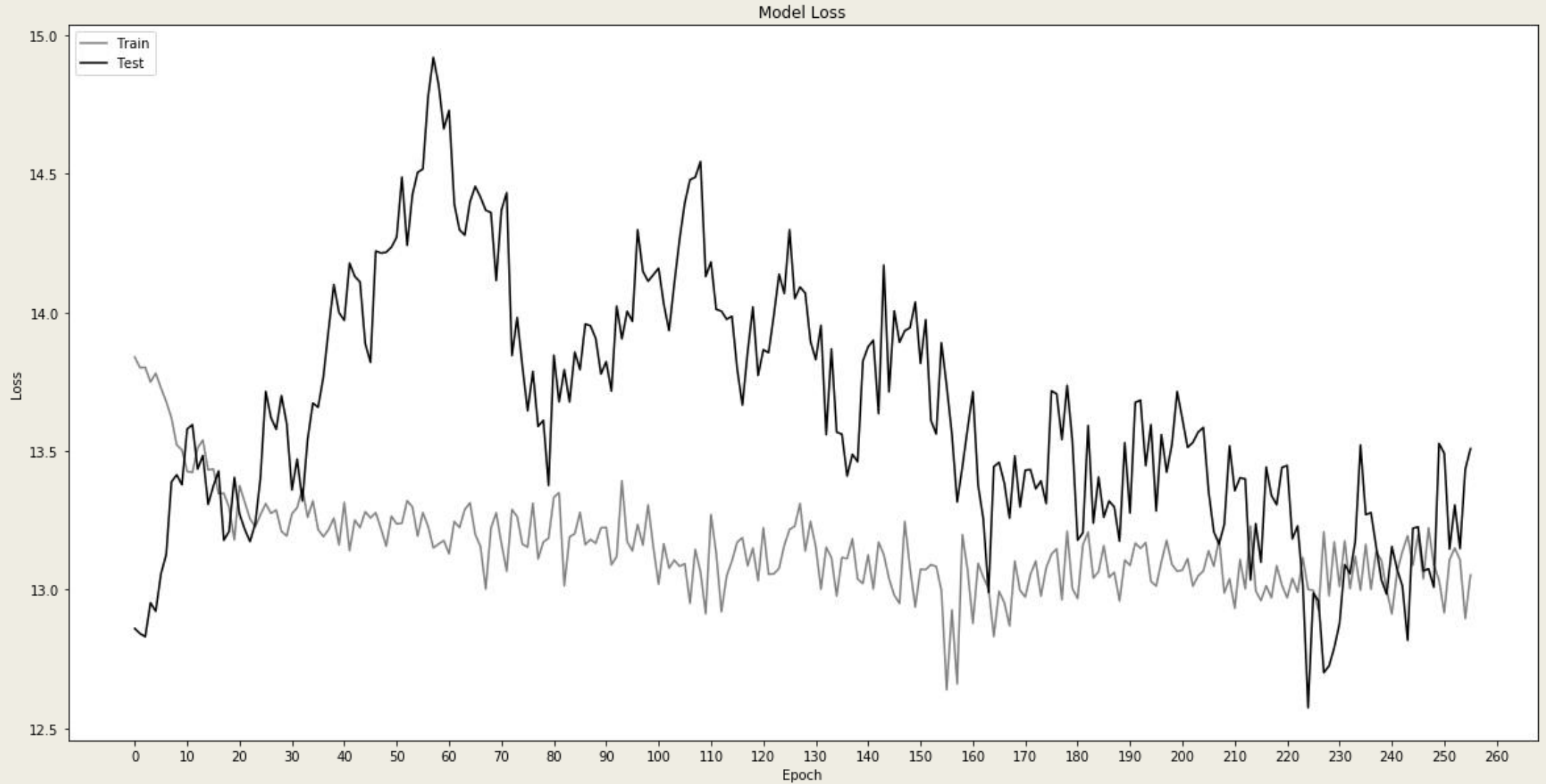
The Model - Dynamic Memory Network

- A state-of-the-art 'general purpose' neural network architecture for problems that can be formulated as a 'Question Answering' tasks. Textual QA, Visual QA, Neural Machine Translation, Named Entity Recognition are examples of problems that this model could handle. Architecture shown below.



Training/Validation

The model was batch-trained for 256 epochs with dropout and L2-regularisation.



Q: How well did the model answer the questions?

When answering questions of an unseen dataset (test of generalisability), **21% of the predicted answers contained the true answer**, 79% did not. Here are some examples:

Question: *“Ctenophores, cnidarians and what other group are labelled diploblastic?”*

Real answer: *“sponges”*

Predicted answer: *“al, which is called the mesoglea in cnidarians and ctenophores; more complex animals have three main cell lay”*

Predicted answer in real answer: False

Question: *“What group specifically opposed the Huguenots?”*

Real answer: *“Catholic Church in France”*

Predicted answer: *“The Catholic Church in France and many of its members opposed the Huguenots. Some Huguenot preachers and congregants were attacked as they attempted to”*

Predicted answer in real answer: True

Question: *“What event was the worst example of Huguenot persecution?”*

Real answer: *“St. Bartholomew's Day massacre”*

Predicted answer: *“rsecution was the St. Bartholomew's Day massacre when 5,000 to 30,000 were killed, although there were also underlying political reasons for this as well, as some of the Huguenots”*

Predicted answer in real answer: True

Moving Forward/Closing Remarks

- Currently the model is not performing sufficiently to be used in a commercial setting, also there is no 'user-friendly' API.
- It's evident that there is much room for improvement with regards to model performance.
- Implementing the Answer Module as a GRU should yield much higher performance figures and clearer results.
- Training data should be shuffled (though this is a trivial exercise with scikit-learn's `train_test_split` function).
- Hyperparameters need to be optimised (though this requires multiple lengthy training runs, 24hrs to train 50 epochs).
- The DMN, as mentioned before, can be used for multiple tasks, not just closed-domain textual QA.

Q: Is this the end of the
presentation?



A: “ “