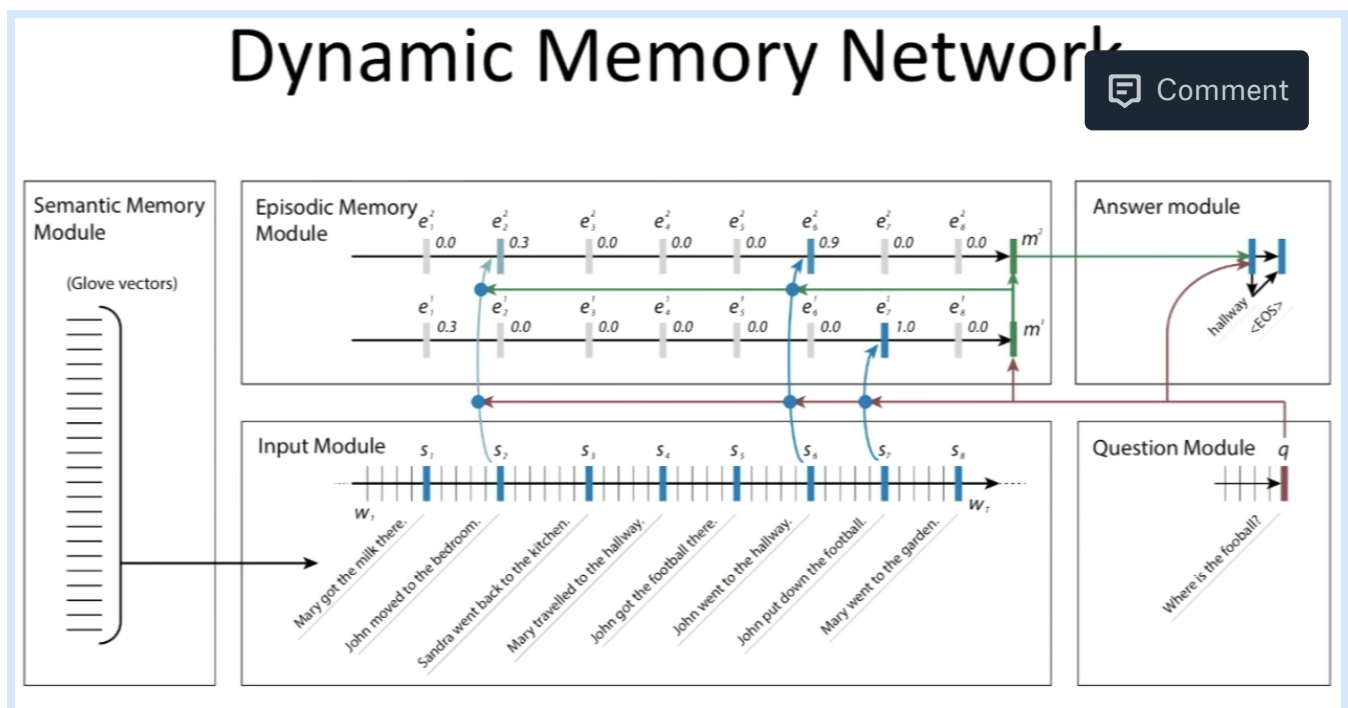


question answering

- trivially, any NLP task can be phrased as a question. "What is the POS of this word?". "Please translate X to French" etc
- Can we train one large model to do generic question answering and then we can be done? Two obstacles:
 - each of our architectures are task specific. for MT, speech recognition, POS tagging, coreference resolution etc etc
 - word2vec has worked quite well as a good initialization point but transfer learning and shared weights haven't worked well beyond that for NLP like they have for CV (CV is considered a success here)
- Here is the architecture pattern of modern networks for question answering. The track-like things are RNNs with some gating like LSTM or GRU



- The above architecture allows model to see the question first, parse it with a RNN pass, look up memory based on attention from the question and parse what's in memory with the episodic memory module and then generate an answer.
- The whole thing can be trained end to end with input output question answer pairs since it is just a neural net that is differentiable everywhere.
- The special thing here is that in addition to regular gates as we parse each sentence, we also gate the sentence embedding to see if that sentence is relevant to the question being asked. If it's not relevant the weights of this gate will learn to throw out facts from that sentence.
- This model allows going over the input multiple times because each time we go over the input, we learn to pay attention to different kinds of sentences.
- The exciting thing about this model is that it did kind of not super badly on tasks like coreference resolution too! so same architecture, different hyperparameters also did well on MT/some other tasks → shows potential

- another exciting area that's being pursued is visual QA, using a visual QA dataset. they took conv filter layers and fed those filters which replaced the input input module. the rest of the architecture like question module stayed the same. they got state of the art, but state of the art on this isn't actually that great. also they tried it with asking the question without showing the image. it did kind of decently without seeing image! so there is some aspect of just predicting pattern of questions and answers. but visualizing attention showed that it was paying attention to the right parts of the image in many cases.
- the thing about these models is really the lack of training data. humans are just fed a ton and ton of training data. these models don't have that many especially for challenging tasks like question answering where there aren't a ton of natural datasets in the world. but if a narrow domain is important to you and you have a ton of data in that domain, these models become immediately exciting, cuz both easier task (narrow domain) and lots of relevant data.