**Challenges** (slide 40)

The authors provided the code for running all of their experiments (about 20 models), but some of that code was in Matlab.

In terms of preprocessing, the piece that stuck out for us was text tokenization, which we think has the potential to meaningfully affect results. You could argue that the different models should have different tokenizations, and we don’t think the authors did that. We don’t have details of how they tokenized, and we have built our own process, which is different from theirs.

Some of the models that we attempted to replicate didn’t work, and it appears that this was due to for example having the wrong version of the Tensorflow library. This is an excellent illustration of the concepts from Module 1.

**[next slide]**

To be able to run the Matlab code from Python, you install Matlab – which Harvard students can do for free - and then the libraries for the Python-Matlab engine can be installed from a setup.py file that comes with it.

**[next slide]**

In this code provided by the authors, Mtrick\_enterfunc() is not a Python function, it is Matlab code housed in a dot m file. You can also see that you save the data as a Matlab format data file for it to use. The function returns an object that numpy is able to turn back into a regular array.