**Phase 2 – Automated Data Creation**

Sharadhi Jaganath

ASU ID – 1217080293

Spelling errors are a common occurrence while dealing with word problems. In order to make our model more robust to common spelling errors, we would like to introduce training samples with synthetically generated errors. As most samples are originally typed out using a QWERTY keyboard, some common errors occur when closely spaced characters are interchanged.

The code written to synthetically inject these errors into training samples is labelled QwertyErrors.py. Here, randomness has been introduced into every aspect of the word and character selection in such a way that no easily discernable pattern can be picked up by the model. Once a word is selected in order to inject a qwerty spelling error into it, a random number of characters are interchanged and the newly formed, deliberately misspelled word, replaces the original one. Once all the chosen words have been replaced, the newly formed question becomes the new sample.

All the data samples of the MathQA dataset have been used to generate new samples.

In order to run the code mentioned above on your own system, please change the file path of both the input and output file towards a file in your own system.