

## *AI611 $\mu$ Word Prediction with N-Grams Model using Python*

### Coding 2: Training a bigram model

This assessment evaluates the following competencies:

- *AI201 – Train an N-Grams model from a given text corpus* (+1)
- *AI501 – Write an application that solves the word prediction problem with N-Grams models* (+2)
- *AI103 – Preprocess a corpus and compute basic statistics on it* (+1)

In this coding assessment, you have to complete an existing Python program that computes the unigram and bigram counts thanks to the `nltk` Python module<sup>1</sup>. To succeed the assessment, you have to:

1. Train a bigram model from a corpus that you have to read from a text file.
2. Generate and print the most probable sentence according to the model you trained.
3. Explain to the teacher how you designed your code and make a demonstration.

You can assume that the text file will only contain english words and you can ignore punctuation signs, just considering that they are word delimiters, except dots delimitating the sentences.

---

<sup>1</sup>The code can be found here: <https://github.com/ukonline/uCourse/blob/master/AI611%C2%B5/code/bigramodel.py>