Natural Language Processing 2024
Assignment 1
Due Thursday, April 11 (by midnight)

A1.0: The Basics

Download the accompanying starter coder (NLP_2024_A1.py) and complete the 4 functions as directed to:

- validate the format of a string, with and without regular expressions;
- compute conditional probabilities and store them in a lookup table;
- generate a string deterministically based on a given string.

Grading: Each function is worth 1 point. You must get 2 of 4 points to pass this portion of the assignment. This assignment is in large part to make sure you have the coding skills to continue in this course and to get a better sense of what skills students have. We care mainly about correctness of the implementation, but reserve the right to deduct points for code that is difficult to read. A failing grade will result in 2 points off your next assignment.

Time expected: This assignment should not take you very long if you are proficient in the skills required for the course, such as implementing algorithms in Python from scratch. If it takes you more than an hour to complete, you may want to consider a course that requires fewer coding skills before taking this course. Please get in touch with Lucia or your TA if this is the case.

On working with others: This assignment is individual, meaning you should complete it alone. If you work with anyone else, please note this down in your submission. If you need to use ChatGPT to complete this exercise, or to verify your answers, you may also want to consider a course that requires fewer coding skills.

Python 3 (latest release) should be used for all assignments in this course.
The Anaconda distribution is recommended.

(Assignment continued on next page)

A1.5: ChatGPT

OpenAI's ChatGPT system supports text-based conversations in English on a wide range of topics. You can ask the system for information on a topic and ask it to write in certain styles (e.g., essays, poetry). It can even generate code.

The system's availability (since late 2022) has inspired both enthusiasm and concern. The text it generates is quite fluent, to the point that it can often be hard to distinguish from text written by a human. Among the worries:

- people may overestimate its capabilities and not recognize when it produces something incorrect or misleading;
- as it is trained to imitate text seen on the web, it may produce biased or harmful content;
- it does not cite its sources, so it may be largely plagiarising content written by humans;
- people may falsely take credit for its output (e.g., students may use it to cheat).

Play and experiment with ChatGPT yourself (this requires creating an account on the OpenAI website). Complete the following:

1. Given the worries above, do you think ChatGPT can be useful to your learning (i.e. for improving your understanding of concepts in a course)? Why or why not? Give at least one specific example with a screenshot of an exchange with ChatGPT to support your answer. (1 point)

2. Think back to our discussions in class about the relationship between language and intelligence. Design a small test to show whether or not ChatGPT is "intelligent" and provide a screenshot of your interaction (you may want to look at Howard Gardner's Theory of Multiple Intelligences, summarized here, to inspire you). Briefly explain your test and what kind of "intelligence" it is testing. Reflect upon your interaction with ChatGPT and whether or not its language is indicative of intelligence. (1 point)

***Time/length expected**:* While you may spend quite a bit of time playing with ChatGPT and designing an intelligence test, please keep your written answers to 300 words or less total.

Optional: For an extended overview of what ChatGPT does and why people are concerned, watch this short interview with Prof. Emily Bender and read this *New York Times* article. If you like, you can also read what the *Washington Post* editorial board thinks about potential societal impacts.

Submission (by midnight 11.4.23). Submit your solutions for A1.0 (code) and A1.5 (PDF) as a zip file on Canvas.