Alexander MacKinnon
CART 498 – Generative AI
Assignment 02 – P+7 (Oulipian language modelling)

This week, I explored a variation of the **N+7 technique**, which, in itself, is a word-replacement method where each noun in a text is replaced with the noun that appears seven entries later in the dictionary. Instead of using a dictionary, the idea was to adapt this into the **P+7 technique**—where a word in a sentence is replaced with the 7th most probable word predicted by GPT-2's vocabulary model based on contextual probability. To test this approach, I applied it to Wallace Stevens' renowned poem, The Snow Man, and picked up on some interesting patterns.

With the P+7 technique, the poem now reads as follows:

> *One must have a mind of her*
> *To regard the frost and the death*
> *Of the pine-trees crusted with oil*
> *And have been cold a long way*
> *To behold the junipers shagged with white*
> *The spruces rough in the distant horizon*
> *Of the January sun; and not to have*
> *Of any misery in the sound of the sound*
> *In the sound of a few shots*
> *Which is the sound of the voice*
> *Full of the same day*
> *That is blowing in the same bare air*
> *For the listener, who listens in the morning*
> *And, nothing himself, I*
> *Nothing that is not there and the nothing that isn*

I found it fascinating how the model isn't just focused on completing the line with a closing word, but rather on selecting the most contextually suitable word to follow, whether it makes sense or not. This sometimes results in lines that feel abruptly cut off. Take, for instance, this absolute cliffhanger of an ending: "And, nothing himself, I; Nothing that is not there and the nothing that isn".

It was even more interesting to compare it with other iterations of this process, so I then tried the P+70 approach, giving me lower-probability words than the first iteration.

> *One must have a mind of rhythm*
> *To regard the frost and the heavy*
> *Of the pine-trees crusted with this*
> *And have been cold a long a*
> *To behold the junipers shagged with dust*
> *The spruces rough in the distant distant*
> *Of the January sun; and not to eat*
> *Of any misery in the sound of the hammer*
> *In the sound of a few steps*
> *Which is the sound of the bullet*
> *Full of the same shape*
> *That is blowing in the same bare wind*
> *For the listener, who listens in the past*
> *And, nothing himself, yet*
> *Nothing that is not there and the nothing that lives*

To my surprise, the lower-probability words in my experiment turned out to be far more specific and semantically rich—*rhythm, hammer, bullet, steps, dust, lives.* This turned out much more poetic than the first time, as the model seemed to gravitate towards more descriptive nouns as the probability was decreasing.

It makes sense: higher-probability words tend to be function words like *the, that, a,* which are common in any sentence, and the chances of slotting them after any word in a sentence are high. But in any experiment like this, perhaps words with lower probability are interesting if nuance and depth is the goal.