When I first attempted the p + 7 assignment in class, I often got tokens rather than fully formed English words. Lots of lines ended in fragmented linguistically nonsensical ways such as ‘ich’, ‘vern’, and ‘a’. I realized that the reason this was happening was that I just requested the raw Byte-Pair Encoding units instead of reassembled English words.

The tokenizer.decode([token_id]) line fixed this. I knew something was wrong when I used too high of an x value and every line ended in ‘vern’. Once the code was actually working, even in the x values closer to 0 it would occasionally end a line in a dash or a punctuation mark of some kind. What surprised me personally was how close a number like 10 and 26 were in terms of how coherent it was. Of course, as the x value went up I did start to see more bizarre things like lines ending in apostrophes. My favourite poem produced from this experiment was the 26th most likely word, not because it was particularly witty or absurd or totally subversive but because to me it read somewhat like an actual poem. The first line is strange and not really grammatically correct(‘One must have a mind of great’), but segments like

‘To behold the junipers shagged with green
The spruces rough in the distant blue’

and

‘For the listener, who listens in the center
And, nothing himself, really’

Create a continuity and new meaning that in some ways perhaps elevates the original poem. As for a p + 7 AI technique with nouns specifically, there would need to be some kind of noun library or way of isolating nouns.