**AI Documentation: Project 2B — Place Name Extraction Using Stanza and 4 B. Map the NER-extracted placenames:**

The goal of my first project was to extract all place names from a corpus of texts using Stanza. Focusing specifically on articles written in January 2024 and counting and cleaning these named entities for analysis.

In this step:

Named Entity Recognition (NER) Using Stanza

* Used Stanza's NER pipeline to extract GPE (Geo-Political Entity) and LOC (Location) tags.
* Example of the relevant code snippet:

python

CopyEdit

if ent.type in ["GPE", "LOC"]:

place = ent.text.strip()

if place in place\_counts:

place\_counts[place] += 1

else:

place\_counts[place] = 1

I encountered an error here initially, which was resolved with help from ChatGPT. The issue was likely due to the conditional structure or indentation.

Moreover, in exporting to a .tsv file, I created a tab-separated values file ner\_counts.tsv.tsv for the results. I was getting an error when I tried to run it for the first time. I used ChatGPT and learned the function and purpose of this code through clarification:

python

CopyEdit

with open(filename, "w", encoding="utf-8") as file:

file.write("Place\tCount\n")

for place, count in place\_counts.items():

file.write(f"{place}\t{count}\n")

* where 'w' opens the file in write mode.
* UTF-8 handles special characters in place names.
* \t separates the two columns: Place and Count.

Moreover, in project 4B firstly, we used the script of mapping regex count because it was quite similar. We changed the file paths and name of the common column. Then, we executed the code, but it was showing multiple errors continuously. We just tried hard asking Chat Gpt to fix those errors, but we were not getting errorless results.

Then consulted our professor Mathew, he helped us with this issue. He told us that our longitudes and latitudes were in strings, that is why we’re getting multiple errors. He changed the longitudes and latitudes to float instead of strings, and then we didnot get any error. But in our map the frequencies were scattered all over the map, tho we got 326 articles after 2024 on our NER file.