

CSC108H Winter 2024 Worksheet 30 : Tiny Language Model 4: Writing the Code

Now that you have figured out how to read and store the data, we can write the TLM data generating function. **Read the description below carefully.**

Complete the type contract for the function header below. Then, plan and write the function body (and any helper functions).

We recommend that you start by listing the major tasks as notes. **Each of those major tasks should probably be a separate function.** Follow the Function Design Recipe for each of them.

```
def generate_random_data(training_file: str, context_length: int, num_words: int) -> str:
    """Return randomly generated data with num_words words based on a context
    of context_length words from the training text in open file training_file.
    """
```