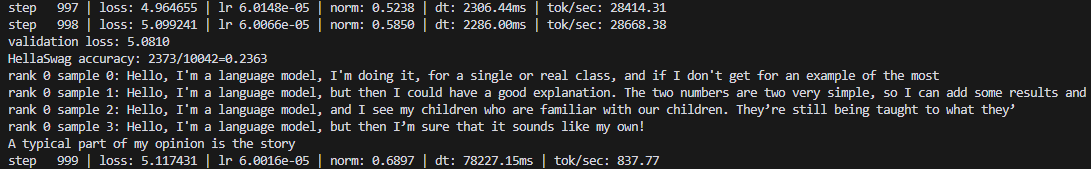
1. Update: 15/10/24 Time: 15:56

Dataset size:

* 953 MB
* 500M tokens

1. block\_size: int = 512 *# max sequence length*
2. vocab\_size: int = 50257 *# number of tokens: 50,000 BPE merges + 256 bytes tokens + 1 <|endoftext|> token*
3. n\_layer: int = 4 *# number of layers*
4. n\_head: int = 8 *# number of heads*
5. n\_embd: int = 512 *# embedding dimension*
6. dropout: float = 0.1
7. total\_batch\_size = 65536 *# 2\*\*19, ~0.5M, in number of tokens*
8. B = 2 *# micro batch size*
9. T = 256 *# sequence length*
10. max\_steps = 1000



2) Update: 15/10/24 Time: 22:45

Dataset size:

* 953 MB
* 500M tokens

1. block\_size: int = 512 *# max sequence length*
2. vocab\_size: int = 50257 *# number of tokens: 50,000 BPE merges + 256 bytes tokens + 1 <|endoftext|> token*
3. n\_layer: int = 4 *# number of layers*
4. n\_head: int = 8 *# number of heads*
5. n\_embd: int = 512 *# embedding dimension*
6. dropout: float = 0.1
7. total\_batch\_size = 65536 *# 2\*\*19, ~0.5M, in number of tokens*
8. B = 4 *# micro batch size*
9. T = 256 *# sequence length*
10. max\_steps = 10000

