

## Word counts in the HAP-E corpus, by source

Author	acad ( <i>n</i> = 1227)	blog ( <i>n</i> = 1526)	fic ( <i>n</i> = 1395)	news ( <i>n</i> = 1322)	spok ( <i>n</i> = 1721)	tvm ( <i>n</i> = 1099)	Total
Human							
Chunk 1	602,878	748,901	674,437	644,896	833,699	598,157	4,102,968
Chunk 2	602,683	749,052	675,595	646,030	832,797	591,704	4,097,861
GPT-4o							
GPT-4o	633,884	826,590	759,461	708,375	969,210	606,059	4,503,579
GPT-4o Mini	690,828	916,433	837,527	771,324	1,036,125	678,429	4,930,666
Llama Base							
Llama 3 70B	510,942	731,889	651,557	572,423	1,098,698	502,012	4,067,521
Llama 3 8B	544,845	735,068	740,610	510,671	1,083,953	525,037	4,140,184
Llama Instruct							
Llama 3 70B Instruct	580,579	735,501	641,982	627,299	934,911	453,559	3,973,831
Llama 3 8B Instruct	550,463	679,884	599,000	568,703	836,531	448,874	3,683,455
Total	4,717,102	6,123,318	5,580,169	5,049,721	7,625,924	4,403,831	33,500,065