

SamSum Dataset HP-Score:

Refer to dataset at : <https://huggingface.co/datasets/Samsung/samsum>

S.No	Sample id	Basic Recall and Reproduction	Understanding and interpretation	Analysis and Reasoning	Application of knowledge and Execution	HP-Score
1	13862856	4	4	1	1	2.5
2	13716964	5	5	1	1	3
3	13731487	4	5	1	1	2.75
4	13680771	3	4	1	1	2.25
5	13864400	1	3	1	1	1.25
6	13827914	5	4	1	1	2.75
7	13864619	1	4	1	1	1.5
8	13862773	2	3	1	1	1.75
9	13819021	5	5	1	1	3
10	13728288	5	4	1	1	2.75
11	13611693	2	2	1	1	1.5
12	13829937	1	3	1	1	1.5
13	13729602	1	1	1	1	1
14	13728763	4	5	1	1	2.75
15	13716075	1	4	1	1	1.75
16	13680352	4	5	1	1	2.75
17	13830106	5	4	1	1	2.75
18	13731212	3	4	1	1	2.25
19	13729695	1	1	1	1	1

20	13729178	4	4	1	1	2.5
21	13818932	2	4	1	1	2
22	13730094	4	4	1	1	2.5
23	13821356	3	3	1	1	2
24	13820209	5	5	1	1	3
25	13611442	3	3	1	1	2
26	13682597	4	3	1	1	2.25
27	13810073	4	4	1	1	2.5
28	13681989	2	4	1	1	2
29	13729197	2	3	1	1	1.75
30	13862536	4	4	1	1	2.5
31	13828110	4	5	1	1	2.75
32	13731240	3	1	1	1	1.5
33	13821023	4	5	1	1	2.75
34	13828502	2	2	1	1	1.5
35	13727549	5	5	1	1	3
36	13862650	3	3	1	1	2
37	13864638	3	3	1	1	2
38	13863035	4	5	1	1	2.75
39	13828346	5	5	1	1	3
40	13865430	3	3	1	1	2
	Final Scores	3.25	3.7	1	1	2.225

Policy:

- Each rule of the taxonomy is scored by a human on a scale from 1 to 5, as the hierarchical prompt framework has 5 levels.
- Each sample is scored based on its own complexity.
- A representative set of dataset is randomly sampled for scoring the dataset. The size of the set is nearly 5% of the original dataset.