

CommonSenseQA HP-Score:

Refer to the dataset at : https://huggingface.co/datasets/tau/commonsense_qa

| S.No | Sample Id | Basic Recall and Reproduction | Understanding and Interpretation | Analysis and Reasoning | Application of knowledge and execution | HP-Score |
|------|--------------------------------------|-------------------------------|----------------------------------|------------------------|--|----------|
| 1 | 075e483d21c29a5112 67ef62bedc0461 | 2 | 3 | 4 | 3 | 3 |
| 2 | 61fe6e879ff18686d75 52425a36344c8 | 3 | 3 | 3 | 2 | 2.75 |
| 3 | 4c1cb0e95b99f72d55c 068ba0255c54d | 2 | 3 | 3 | 3 | 2.75 |
| 4 | 02e821a3e53cb32079 0950aab4489e85 | 2 | 3 | 3 | 2 | 2.5 |
| 5 | 23505889b94e880c3e 89cff4ba119860 | 2 | 3 | 3 | 2 | 2.5 |
| 6 | e8a8b3a2061aa0e6d7 c6b522e9612824 | 2 | 3 | 3 | 2 | 2.5 |
| 7 | 09555c056f3cf0b7e0b 84d8df4be1db7 | 2 | 2 | 3 | 2 | 2.25 |
| 8 | 3d0f8824ea83ddcc9ab 03055658b89d3 | 2 | 3 | 3 | 2 | 2.5 |
| 9 | 86c59f14af97b33ec13 edc6ec4389e31 | 1 | 2 | 2 | 5 | 2.5 |
| 10 | cae61908c731d4a208 89e04c3e784ebe | 4 | 1 | 3 | 4 | 3 |
| 11 | 1a7f5b3c65364d9be00 2576660c914fe | 2 | 1 | 3 | 3 | 2.25 |
| 12 | dac769116ed064bc70 936c15eb822c3e | 4 | 3 | 3 | 2 | 3 |
| 13 | f5c9664c9930b704f27 24fbd617bee9d | 1 | 1 | 1 | 1 | 1 |
| 14 | b63b9809c203321d66 59ddf8551894bf | 2 | 3 | 5 | 2 | 3 |
| 15 | 527e72eb38950b8031 ee6217ef531960 | 4 | 3 | 3 | 1 | 2.75 |
| 16 | e6766a66f8d326bf00f bf628a0e4ef24 | 1 | 1 | 2 | 3 | 1.75 |
| 17 | a76403b4921a9281b6 ee2a7241a5ec9f | 1 | 2 | 4 | 4 | 2.75 |
| 18 | ad96098698b8378b2e 79c6424920ffbd | 2 | 3 | 5 | 4 | 3.5 |

| | | | | | | |
|----|--------------------------------------|---|---|---|---|------|
| 19 | 6d4ed2c485187ed79ff 3bdcc8cca6132 | 2 | 1 | 3 | 2 | 2 |
| 20 | a40f5339ee858a0a119 cbac052c5a2d0 | 2 | 3 | 3 | 2 | 2.5 |
| 21 | 4850bf8b6b309548ef6 23d5592ffbca2 | 4 | 3 | 3 | 2 | 3 |
| 22 | e6d99b4b76f74961a4 00ad0001b13b73 | 2 | 2 | 1 | 1 | 1.5 |
| 23 | f5950a62f9eaae34f1e 914b3cf5208cf | 2 | 2 | 2 | 1 | 1.75 |
| 24 | f36489668c28b26c0cd bc56cec1db27b | 2 | 3 | 3 | 3 | 2.75 |
| 25 | 0476192858e7f541611 b5c2d3c5e5197 | 2 | 2 | 5 | 3 | 3 |
| 26 | df7861813910ae204c3 bd0b3c16fe5d9 | 3 | 3 | 4 | 5 | 3.75 |
| 27 | c12a904025471139df4 f8860cc380c60 | 4 | 5 | 4 | 2 | 3.75 |
| 28 | a69bd1bf3c06a233d69 6a00d71a9db36 | 2 | 2 | 3 | 4 | 2.75 |
| 29 | 84e983f432647531b5 5b8b72c12fb78a | 2 | 2 | 3 | 2 | 2.25 |
| 30 | 68f123f22be7bff13875 e8b752b2d933 | 2 | 2 | 3 | 2 | 2.25 |
| 31 | c9580b132b323906f0d 51e7dc5b4ce29 | 3 | 2 | 4 | 4 | 3.25 |
| 32 | 80930e9df9ac4ad7527 49a54e7fc124f | 2 | 2 | 3 | 5 | 3 |
| 33 | f080ab037c1aaf689a4 3cc030e9d93e9 | 1 | 1 | 3 | 4 | 2.25 |
| 34 | a261d4eaf8b33c0730d 830ebbffbd9f8 | 1 | 2 | 3 | 4 | 2.5 |
| 35 | bfd14dedb440ae1c266 c61aa22960778 | 1 | 1 | 2 | 3 | 1.75 |
| 36 | 5a953b5fb26f1614640 574f65410e8b0 | 3 | 2 | 3 | 3 | 2.75 |
| 37 | ac0351d8649fb60af40 c5638061a2e21 | 2 | 2 | 2 | 2 | 2 |
| 38 | c8b18acf8e95172ff79b 0a837e9369f3 | 2 | 1 | 1 | 4 | 2 |
| 39 | 7a7ff4499a7162265dd ac0c3e2e8e5fc | 1 | 1 | 1 | 2 | 1.25 |
| 40 | fb721d7949d06915a2 dd6fc6ec815325 | 3 | 3 | 3 | 2 | 2.75 |
| 41 | a7be27a64aad72c029f 05a2c7fc6d55f | 3 | 2 | 3 | 3 | 2.75 |

| | | | | | | |
|----|--------------------------------------|-----|-----|------|------|------|
| 42 | 6122883c3b5cd10b59 a29dbd582e19c8 | 4 | 2 | 2 | 1 | 2.25 |
| 43 | 5cc173999089fd11673 252a58b8c0feb | 2 | 3 | 4 | 2 | 2.75 |
| 44 | d57ca8ce3f5ef4794d1 27c2385228698 | 3 | 2 | 4 | 2 | 2.75 |
| 45 | d6503fc1f87e3e2caeb 7bf8e16526e0f | 4 | 3 | 4 | 1 | 3 |
| 46 | 1c16c914f3e9e99c678 d6f1ec30194a8 | 2 | 2 | 4 | 3 | 2.75 |
| 47 | a7b24b7ab7f0065c868 8ee84d47899f1 | 3 | 2 | 4 | 3 | 3 |
| 48 | e97a7b91a1ca85f362c 42e04c983fc2a | 3 | 2 | 3 | 2 | 2.5 |
| 49 | bc07822ca1f8892b80d 632b9c03e1187 | 2 | 2 | 3 | 2 | 2.25 |
| 50 | 367a52860ea02568f0 14242ec1967d78 | 2 | 2 | 2 | 1 | 1.75 |
| 51 | ca9a3ccfb140aa66816 f96ac983b6d9f | 2 | 1 | 2 | 2 | 1.75 |
| 52 | 591221f7cf4790253d1 9b08d9ea1f77e | 2 | 2 | 4 | 2 | 2.5 |
| 53 | cf9904211d9c602da9e 41be0e3b24aff | 2 | 2 | 1 | 2 | 1.75 |
| 54 | 864ff0aa11611d0de88 65034ac0ecb86 | 3 | 4 | 4 | 2 | 3.25 |
| 55 | d4ebade47b1e84ab40 3efc715a7ca991 | 1 | 1 | 3 | 3 | 2 |
| 56 | 3c0a8fd4547feaa73b8 e5d797e7ecc76 | 2 | 1 | 2 | 4 | 2.25 |
| 57 | ad3aa40dec688333db a0b88e325481ea | 2 | 2 | 4 | 4 | 3 |
| 58 | 1bca9774955d7e794a 6e48ffcf95e728 | 3 | 3 | 2 | 2 | 2.5 |
| 59 | 68b13a7a996a07b8ea 6731349d9348a4 | 3 | 2 | 4 | 2 | 2.75 |
| 60 | c12221a251d96e6917 53cfd3ac0b33c4 | 3 | 2 | 2 | 3 | 2.5 |
| | Final Scores | 2.3 | 2.2 | 2.98 | 2.58 | 2.52 |

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.