

## Dataset Metrics Summary

Dataset: bam\_alpaca

Metric	Value
Conversations	1604
Messages	3208
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

Dataset: bam\_gsm8k\_test

Metric	Value
Conversations	1319
Messages	2638
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

Dataset: bam\_gsm8k\_train

Metric	Value
Conversations	7433
Messages	14866
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

Dataset: bam\_hh\_rlhf\_test

Metric	Value
Conversations	8182
Messages	38863
Avg messages/conv	4.75
Min/Max messages/conv	1 / 50

Dataset: bam\_hh\_rlhf\_train

Metric	Value
Conversations	16490
Messages	80538
Avg messages/conv	4.88
Min/Max messages/conv	1 / 38

Dataset: bam\_model\_evals

Metric	Value
Conversations	3252
Messages	6504
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

Dataset: bam\_oasst2\_train

Metric	Value
Conversations	4791
Messages	16319
Avg messages/conv	3.41
Min/Max messages/conv	2 / 11

Dataset: bam\_oasst2\_val

Metric	Value
Conversations	275
Messages	951
Avg messages/conv	3.46
Min/Max messages/conv	2 / 6

*Dataset: bam\_practical\_dialog*

Metric	Value
Conversations	71
Messages	288
Avg messages/conv	4.06
Min/Max messages/conv	2 / 13

*Dataset:  
bam\_some\_instructions\_samples*

Metric	Value
Conversations	114
Messages	228
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset: mt\_instruction*

Metric	Value
Conversations	559392
Messages	1118784
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset:  
bam\_HuggingFaceH4\_instruction*

Metric	Value
Conversations	307
Messages	614
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset: bam\_IrfanHamid\_chat*

Metric	Value
Conversations	200
Messages	600
Avg messages/conv	3.00
Min/Max messages/conv	3 / 3

*Dataset:  
bam\_everyday\_conversations*

Metric	Value
Conversations	2379
Messages	18428
Avg messages/conv	7.75
Min/Max messages/conv	6 / 10

*Dataset:  
bam\_instruct\_basics\_smollm*

Metric	Value
Conversations	806
Messages	1612
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset: bam\_oasst1*

Metric	Value
Conversations	22344
Messages	81095
Avg messages/conv	3.63

*Dataset: bam\_sales\_conversations*

Metric	Value
Conversations	20940
Messages	156126
Avg messages/conv	7.46
Min/Max messages/conv	2 / 20

*Dataset: bam\_test*

Metric	Value
Conversations	6
Messages	12
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset: bam\_test\_output*

Metric	Value
Conversations	6
Messages	12
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset: bam\_french\_instruction*

Metric	Value
Conversations	128760
Messages	130278
Avg messages/conv	1.01
Min/Max messages/conv	1 / 13

*Dataset: bam\_lmsys\_chat\_1m*

Metric	Value
Conversations	89120
Messages	337988
Avg messages/conv	3.79
Min/Max messages/conv	1 / 134

*Dataset: bambara\_ancestral*

Metric	Value
Conversations	10
Messages	64
Avg messages/conv	6.40
Min/Max messages/conv	6 / 8

*Dataset: bambara\_ancestral\_2*

Metric	Value
Conversations	10
Messages	20
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset: bambara\_ancestral\_1000*

Metric	Value
Conversations	1000
Messages	2000
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset:*

*bambara\_ancestral\_1000\_long*

Metric	Value
Conversations	1000
Messages	2000
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset:*

*bambara\_ancestral\_30000\_long*

Metric	Value
Conversations	30000
Messages	60000
Avg messages/conv	2.00
Min/Max messages/conv	2 / 2

*Dataset: bambara\_conversations*

Metric	Value
Conversations	10
Messages	56
Avg messages/conv	5.60
Min/Max messages/conv	4 / 6

*Dataset:*

*bambara\_conversations\_1000*

Metric	Value
Conversations	1000
Messages	3900
Avg messages/conv	3.90
Min/Max messages/conv	3 / 6

*Dataset:*

*bambara\_conversations\_10000*

Metric	Value
Conversations	10000
Messages	39800
Avg messages/conv	3.98
Min/Max messages/conv	3 / 6

*Dataset:*

*bambara\_conversations\_30000*

Metric	Value
Conversations	30003
Messages	399490
Avg messages/conv	13.32
Min/Max messages/conv	8 / 18

*Dataset:*

*bambara\_conversations\_30000\_1*

Metric	Value
Conversations	30000
Messages	119400
Avg messages/conv	3.98
Min/Max messages/conv	3 / 6

*Dataset:*

*bambara\_conversations\_30000\_2*

Metric	Value
Conversations	30000
Messages	119400

Avg messages/conv	3.98
Min/Max messages/conv	3 / 6

Dataset: bambara\_general\_knowledge

Metric	Value
Conversations	10
Messages	84
Avg messages/conv	8.40
Min/Max messages/conv	6 / 12

Dataset:

bambara\_general\_knowledge\_0

Metric	Value
Conversations	7
Messages	60
Avg messages/conv	8.57
Min/Max messages/conv	8 / 10

Dataset:

bambara\_general\_knowledge\_0\_old

Metric	Value
Conversations	5
Messages	44
Avg messages/conv	8.80
Min/Max messages/conv	8 / 10

Dataset: bambara\_humor

Metric	Value
Conversations	10
Messages	64
Avg messages/conv	6.40
Min/Max messages/conv	6 / 8

Dataset: bambara\_math

Metric	Value
Conversations	10000
Messages	52100
Avg messages/conv	5.21
Min/Max messages/conv	4 / 6

### Global Summary

#### Global Metrics

Total Conversations : 1010856  
Total Messages : 2808434  
Avg Messages/Conv : 2.78