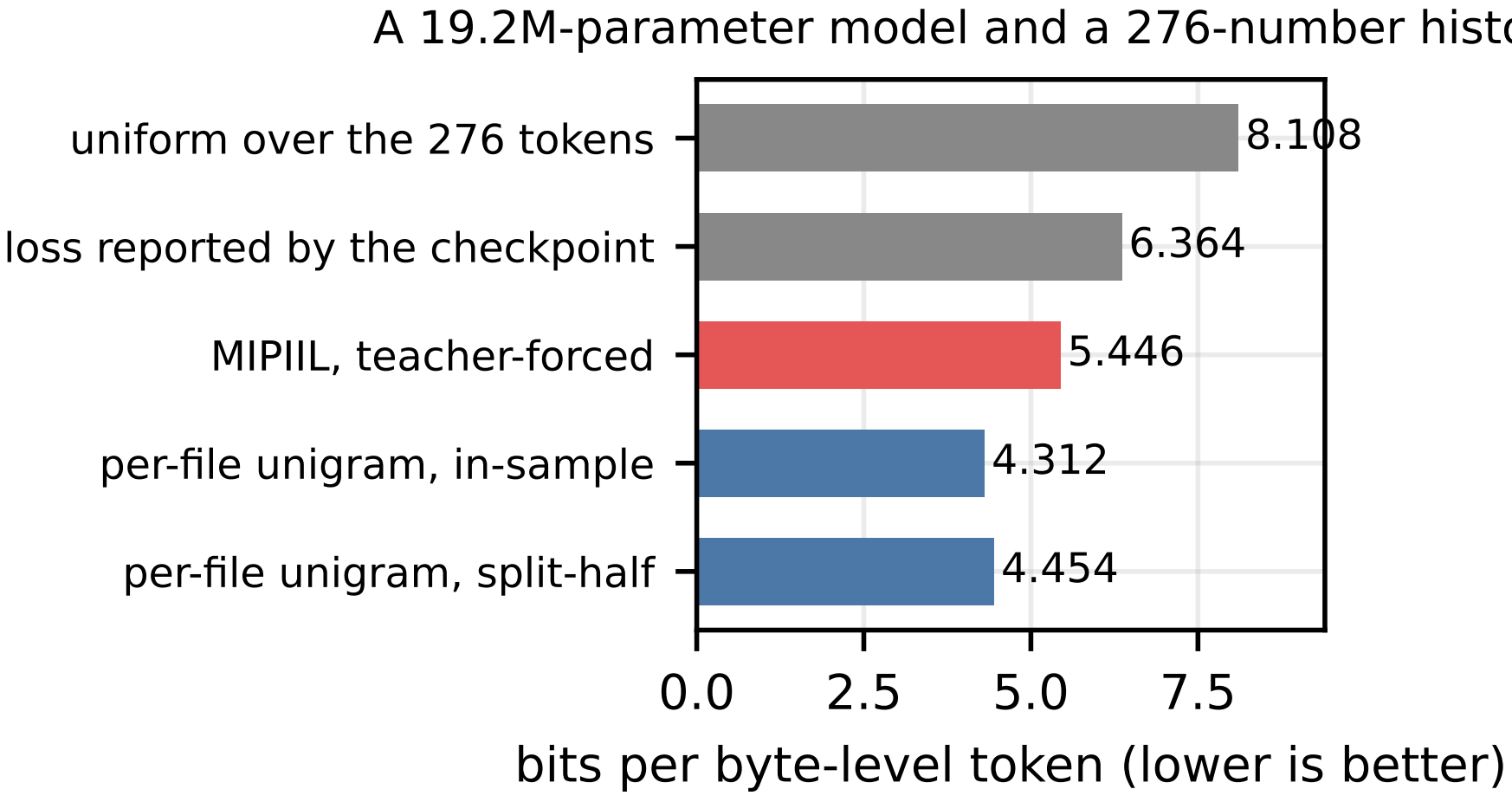

A 19.2M-parameter model and a 276-number histogram
uniform over the 276 tokens
8.108
loss reported by the checkpoint
6.364
MIPIIL, teacher-forced
5.446
per-file unigram, in-sample
4.312
per-file unigram, split-half
4.454
0.0
2.5
5.0
7.5
bits per byte-level token (lower is better)