Figure: Loss vs. Tokens seen

Legend — original_paper:pythia

- pythia bloom-1b1
- pythia bloom-1b7
- pythia bloom-3b
- pythia bloom-560m
- pythia bloom-7b1
- pythia opt-1.3b
- pythia opt-125m
- pythia opt-13b
- pythia opt-2.7b
- pythia opt-30b
- pythia opt-350m
- pythia opt-6.7b
- pythia opt-66b
- pythia pythia-1.3b-deduped
- pythia pythia-1.4b
- pythia pythia-1.4b-deduped
- pythia pythia-12b
- pythia pythia-12b-deduped
- pythia pythia-160m
- pythia pythia-160m-deduped
- pythia pythia-19m-deduped
- pythia pythia-1b-deduped
- pythia pythia-2.8b
- pythia pythia-2.8b-deduped
- pythia pythia-350m-deduped
- pythia pythia-410m
- pythia pythia-410m-deduped
- pythia pythia-6.9b
- pythia pythia-6.9b-deduped
- pythia pythia-70m
- pythia pythia-70m-deduped
- pythia pythia-baseline-1.4b-deduped
- pythia pythia-baseline-6.9b-deduped
- pythia pythia-baseline-70m-deduped
- pythia pythia-bf16-1b
- pythia pythia-intervention-1.3b-deduped
- pythia pythia-intervention-1.4b-deduped
- pythia pythia-intervention-19m-deduped
- pythia pythia-intervention-350m-deduped
- pythia pythia-intervention-410m-deduped
- pythia pythia-intervention-6.9b-deduped
- pythia pythia-intervention-70m-deduped
- pythia pythia-long-intervention-1.3b-deduped
- pythia pythia-1.4b-5shot-v0
- pythia pythia-1.4b-deduped-5shot-v0
- pythia pythia-1.4b-deduped-v0
- pythia pythia-1.4b-v0
- pythia pythia-12b-deduped-v0
- pythia pythia-12b-v0
- pythia pythia-160m-5shot-v0
- pythia pythia-160m-deduped-5shot-v0
- pythia pythia-160m-deduped-v0
- pythia pythia-160m-v0
- pythia pythia-1b-deduped-v0
- pythia pythia-1b-v0
- pythia pythia-2.8b-deduped-v0
- pythia pythia-2.8b-v0
- pythia pythia-410m-5shot-v0
- pythia pythia-410m-deduped-5shot-v0
- pythia pythia-410m-deduped-v0
- pythia pythia-410m-v0
- pythia pythia-6.9b-deduped-v0
- pythia pythia-6.9b-v0
- pythia pythia-70m-deduped-v0
- pythia pythia-70m-v0