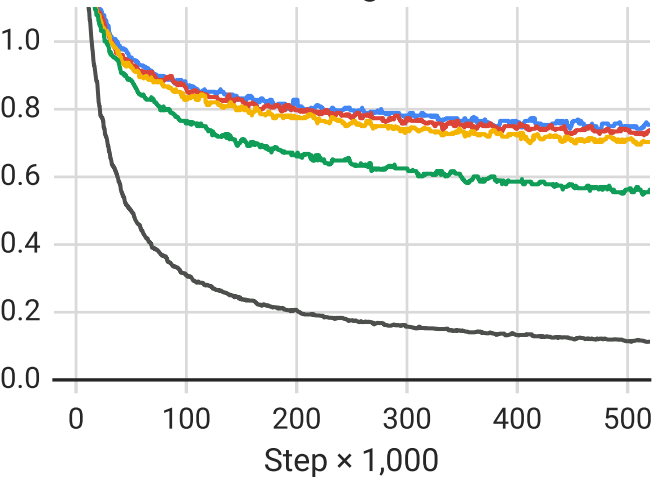Training loss

Dataset size
- Full dataset
- 2^{29}
- 2^{27}
- 2^{25}
- 2^{23}

Step × 1,000