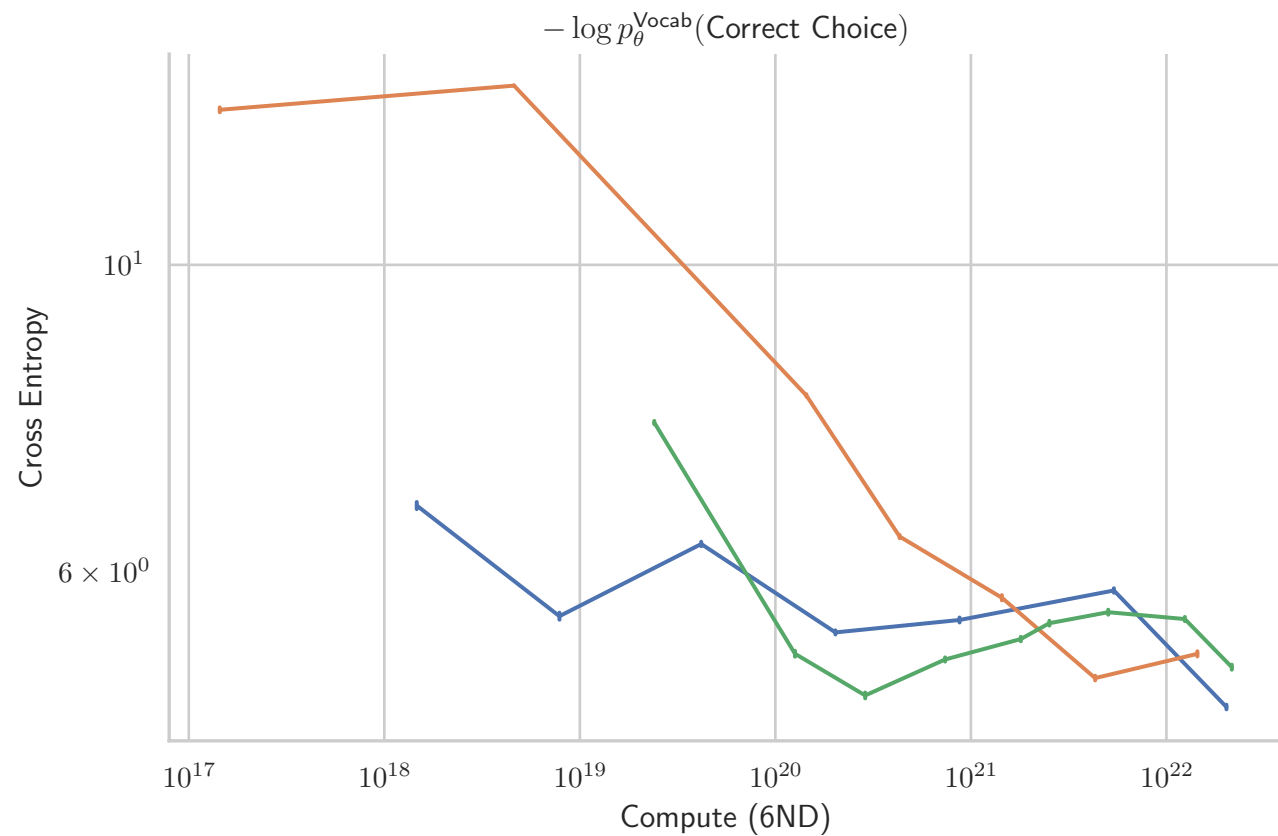
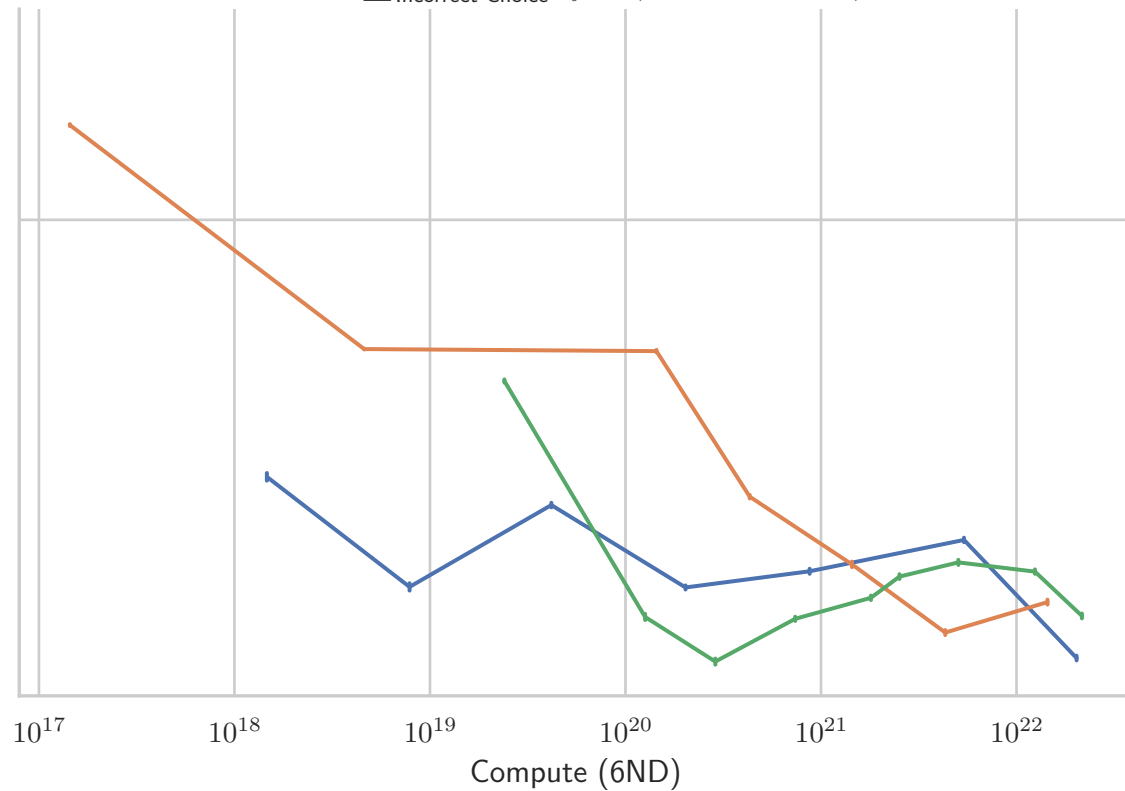


Ethics Commonsense


$$-\log \sum_{\text{Incorrect Choice}} p_{\theta}^{\text{Vocab}}(\text{Incorrect Choice})$$


Model Family

- Cerebras (Param. and Data Scaling)
- Pythia 12B Param. (Data Scaling)
- Pythia 300B Tokens (Param. Scaling)