MMLU HS US History

Left panel: $p_\theta^{\mathsf{Vocab}}(\text{Correct Choice})$

Right panel: $\sum_{\text{Incorrect Choice}} p_\theta^{\mathsf{Vocab}}(\text{Incorrect Choice})$
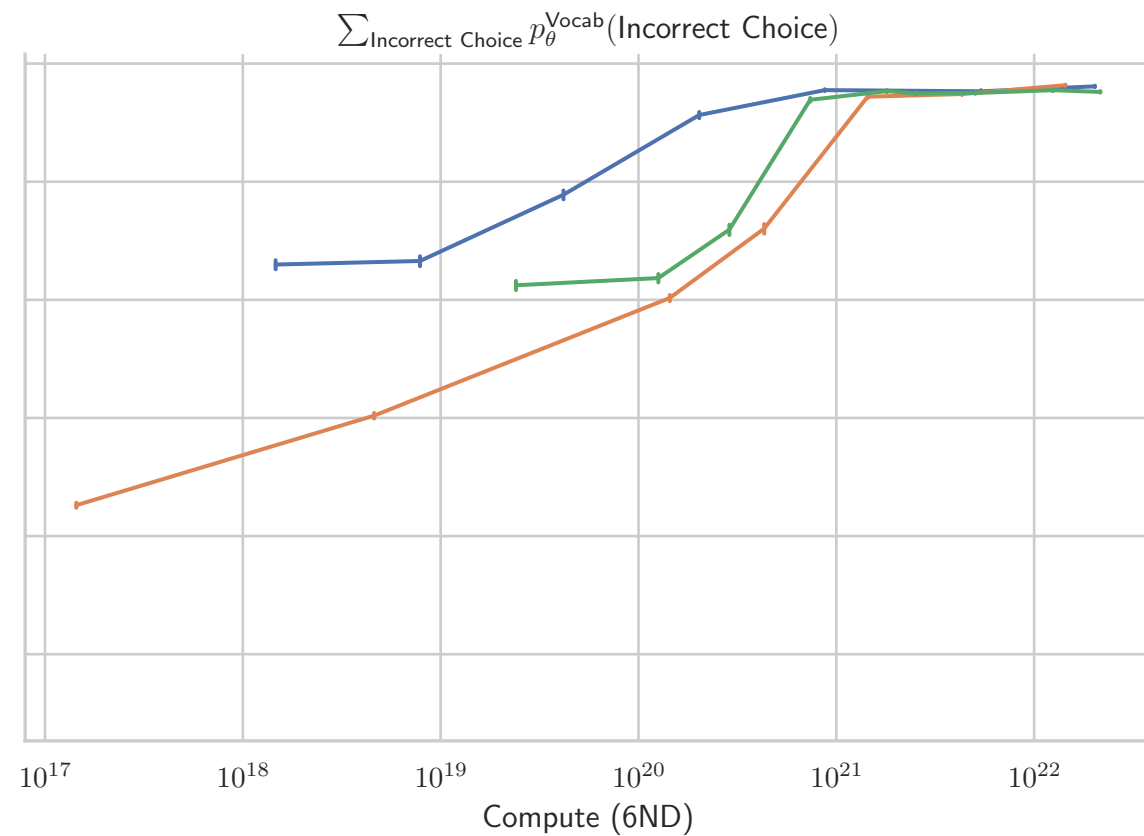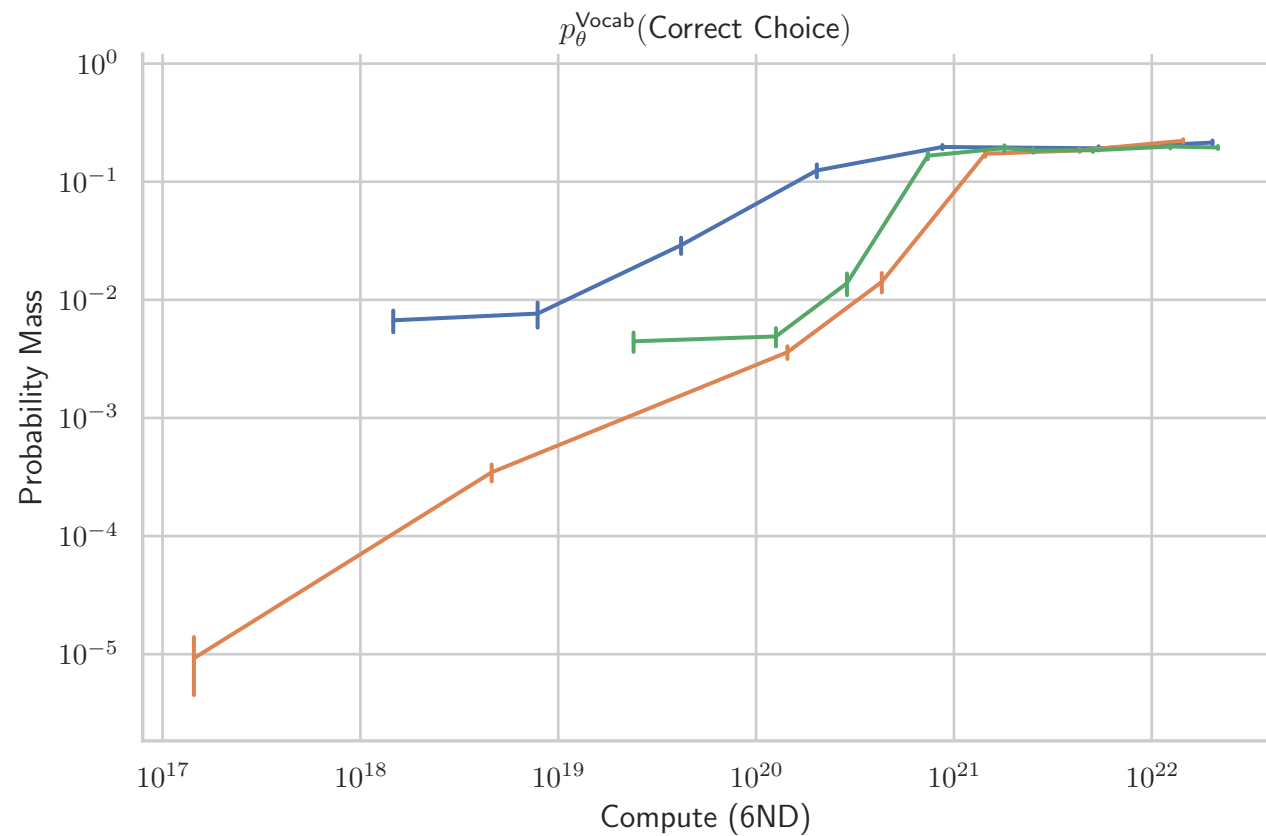
Y-axis: Probability Mass

X-axis: Compute (6ND)

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