

A APPENDIX

Accuracy on Languages in the Chomsky Hierarchy. We evaluate the BLOOM-176B model on 3 language tasks each from increasingly complex classes from the Chomsky hierarchy: (i) regular language, (ii) deterministic context free language, (iii) non-deterministic context-free language, and (iv) context-sensitive language. The BLOOM-176B model is not particularly effective at learning regular languages. Fig. A.5a shows that modular arithmetic and parity detection do not perform better than random chance on 4 bits. BLOOM-176B perform slightly better than random chance on the problem of detecting if the number of 01s and 10s is even. Natural language is known to be mildly context-sensitive; so, our results indicate that BLOOM-176B trained mostly on natural languages finds examples of the regular language class challenging.

Prompting in the BLOOM-176B model works well for deterministic context-free languages as compared to random chance. Fig. A.5b shows that prompting works well for the reversal of strings, stack manipulation and detecting if a string has more number of 1s than 0s. This is different from automata and grammar models, where algorithms for deterministic context-free languages work well for the contained subset of regular languages.

While we show the relative improvement in accuracy compared to the random chance in the main paper, here we plot the absolute accuracy. Note that the random chance is high for some of these tasks such as parity, while it is low for tasks such as addition.

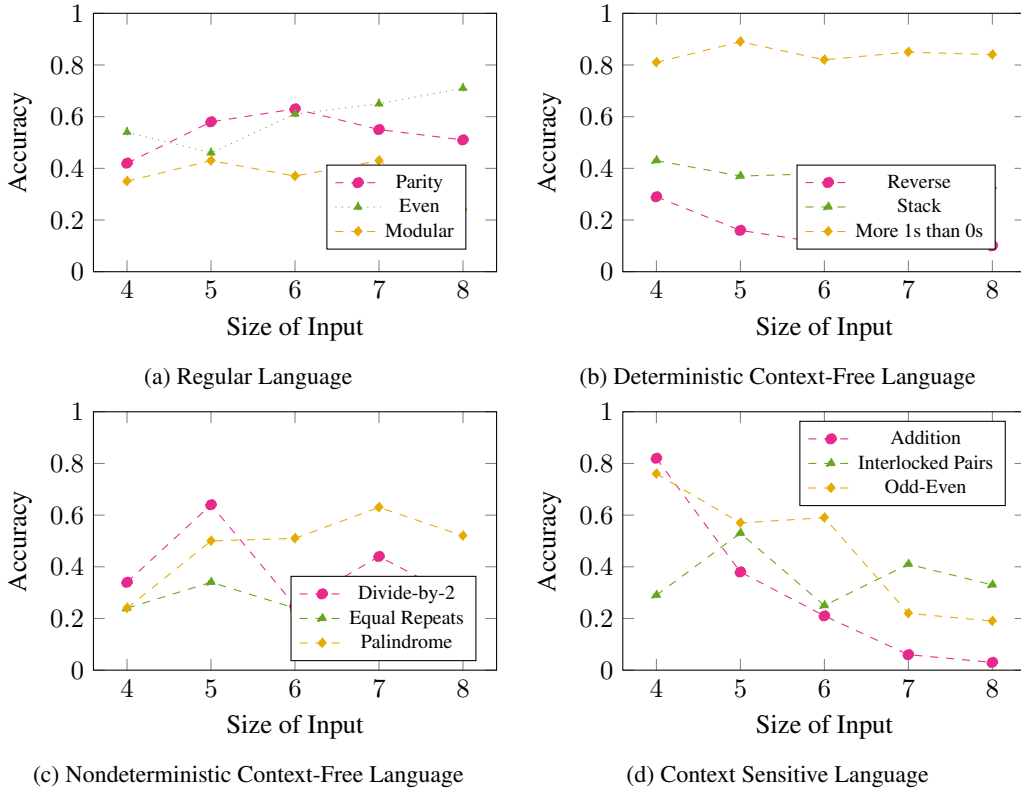


Figure A.5: Accuracy of language tasks using BLOOM-176B for different bit-widths.

Non-deterministic context-free languages perform worse than random chance on the problems of equal repeat detection and missing palindrome digit determination. Prompting the model works well for Divide-by-2, as shown in Fig. 2c. Prompting the BLOOM-176B works well for context-sensitive languages shown in Fig. 2d. As natural languages are considered to be mildly context-sensitive, this suggests that BLOOM-176B trained on natural languages performs well on context-sensitive languages. The BLOOM large language model shows an interesting difference from traditional grammar and automata that generalize well from context-sensitive languages to less complex languages in the Chomsky hierarchy.