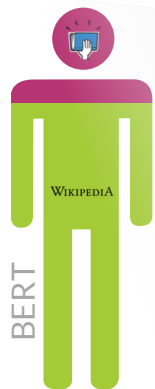


# CONTENTS OF MEGATRON

- BookCorpus (Smashwords)
- English Wikipedia
- Common Crawl Stories
- OpenWebText (Reddit links)
- Common Crawl News (CC-News)
- RealNews (CC-News alt)



BERT



**BERT**

16GB

345M



RoBERTa



**RoBERTa**

161GB

125M



Megatron-LM



**Megatron-LM**

174GB

8.3B



Megatron-11B



**Megatron-11B**

161GB

11B

Not to scale, pictograph size indicates rough parameter count (log), not dataset size. Sources: <https://lifearchitect.com.au/megatron> Alan D. Thompson, October 2021. <https://lifearchitect.com.au/ai>

