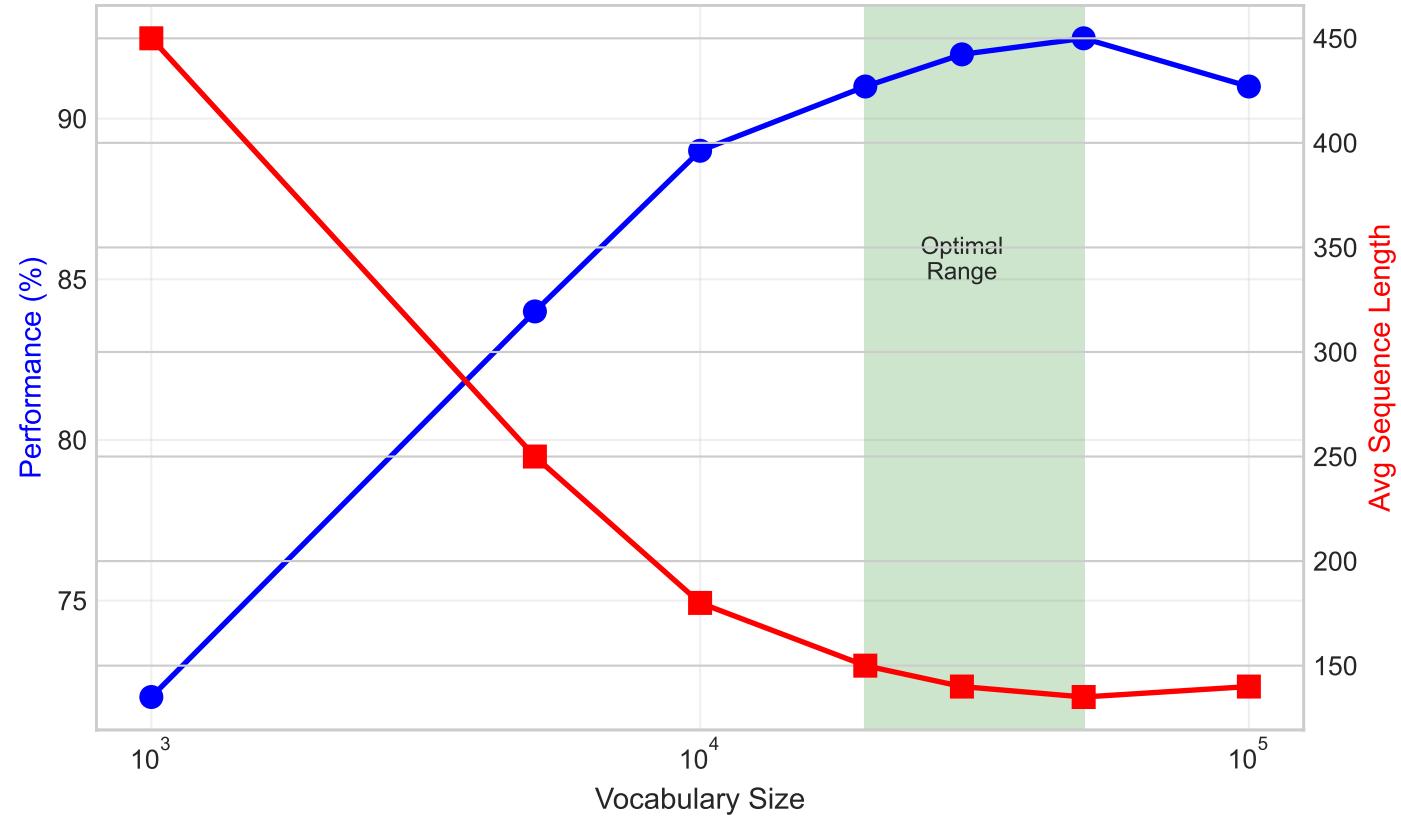
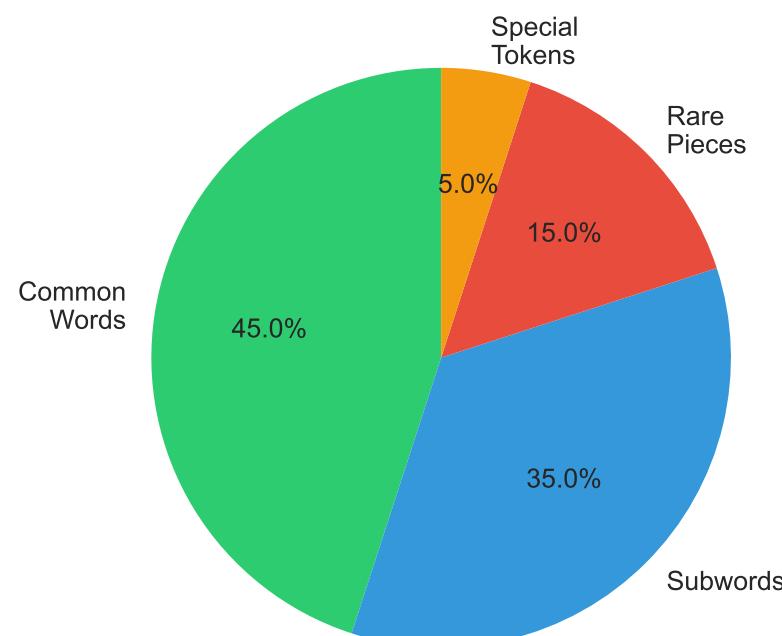


The Hidden Impact of Tokenization

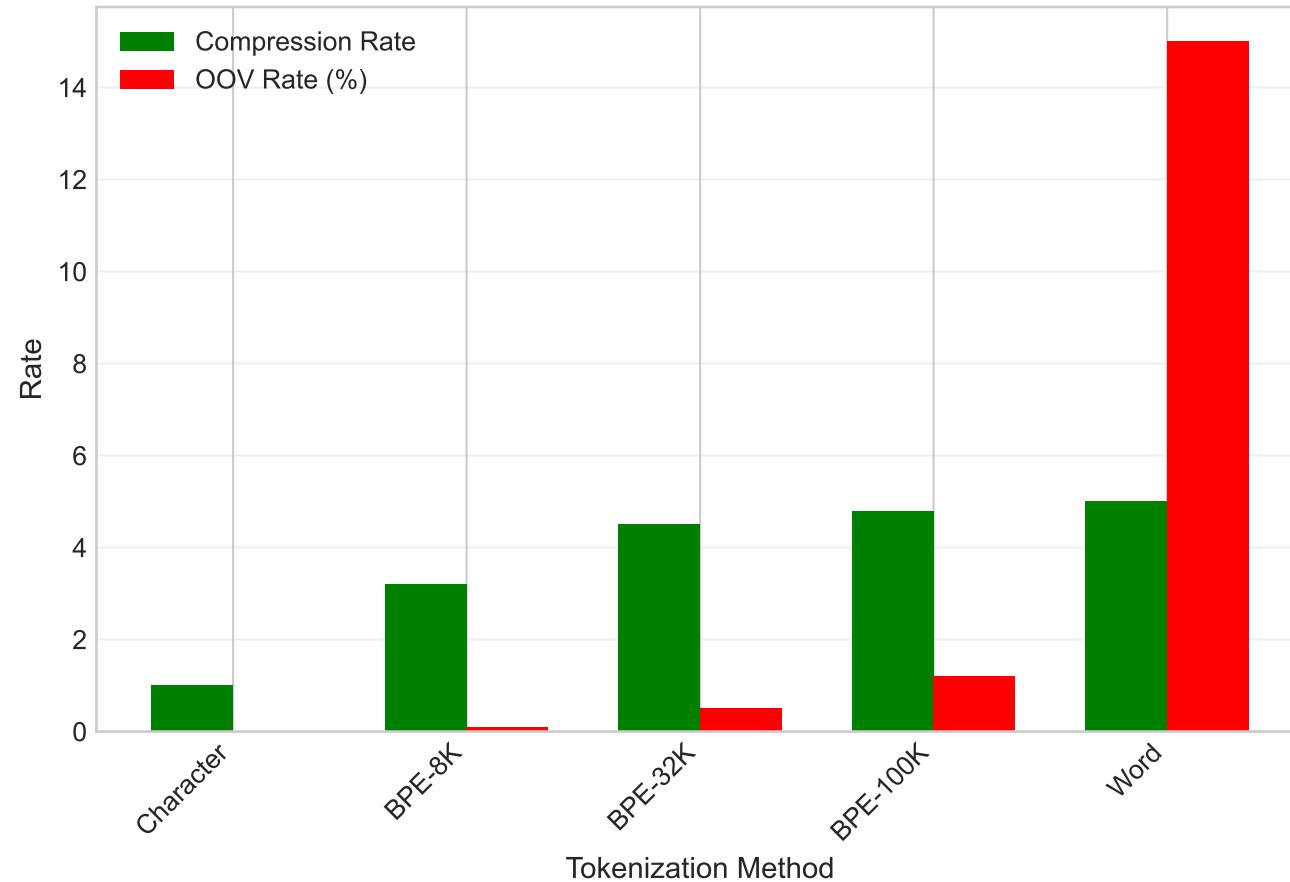
Vocabulary Size Trade-offs



Typical Token Distribution (GPT-2)



Compression vs Out-of-Vocabulary Trade-off



Tokenization Affects API Costs

