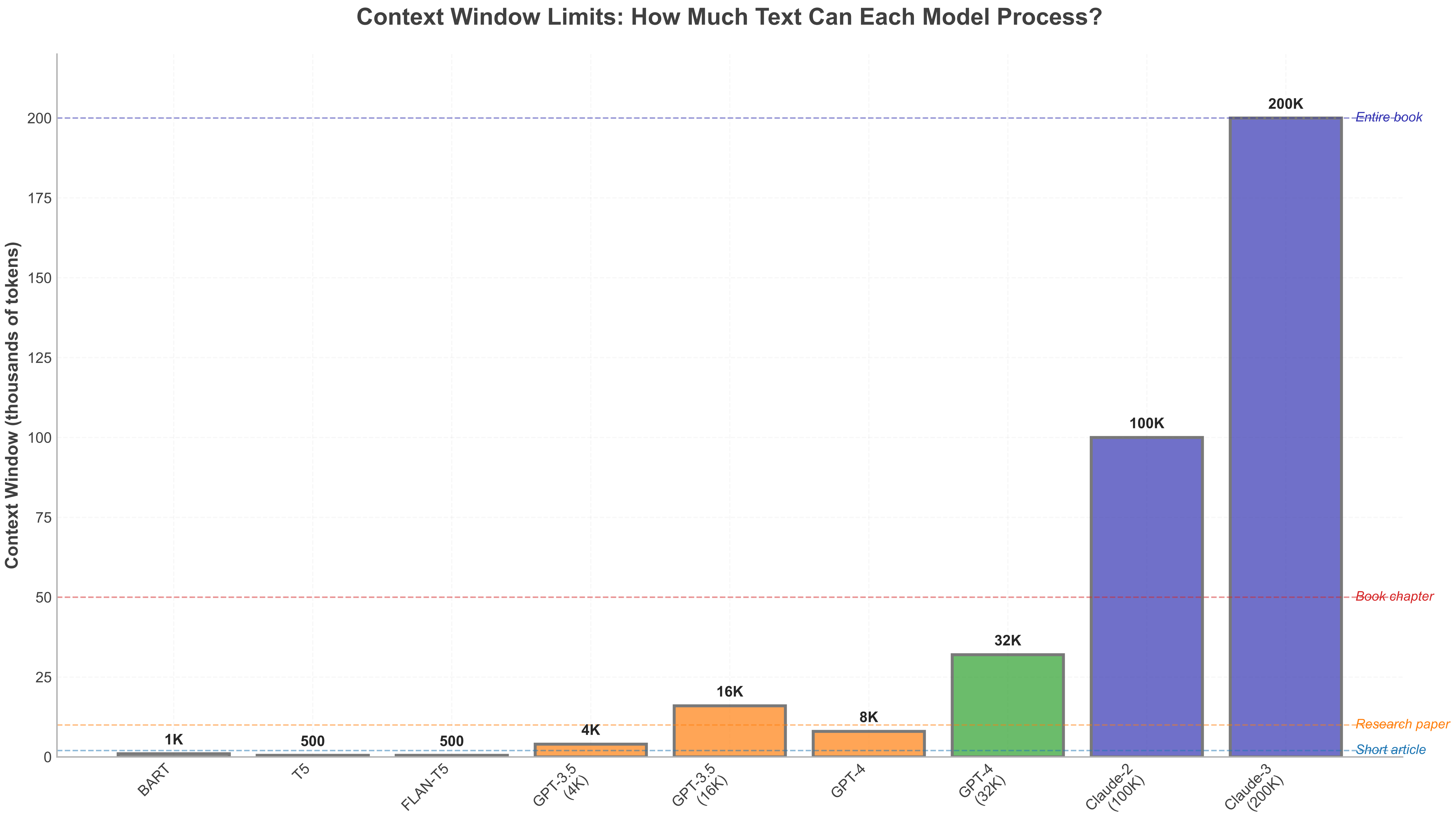
Context Window Limits: How Much Text Can Each Model Process?
Context Window (thousands of tokens)
0
25
50
75
100
125
150
175
200
1K
500
500
4K
16K
8K
32K
100K
200K
Entire book
Book chapter
Research paper
Short article
BART
T5
FLAN-T5
GPT-3.5 (4K)
GPT-3.5 (16K)
GPT-4
GPT-4 (32K)
Claude-2 (100K)
Claude-3 (200K)