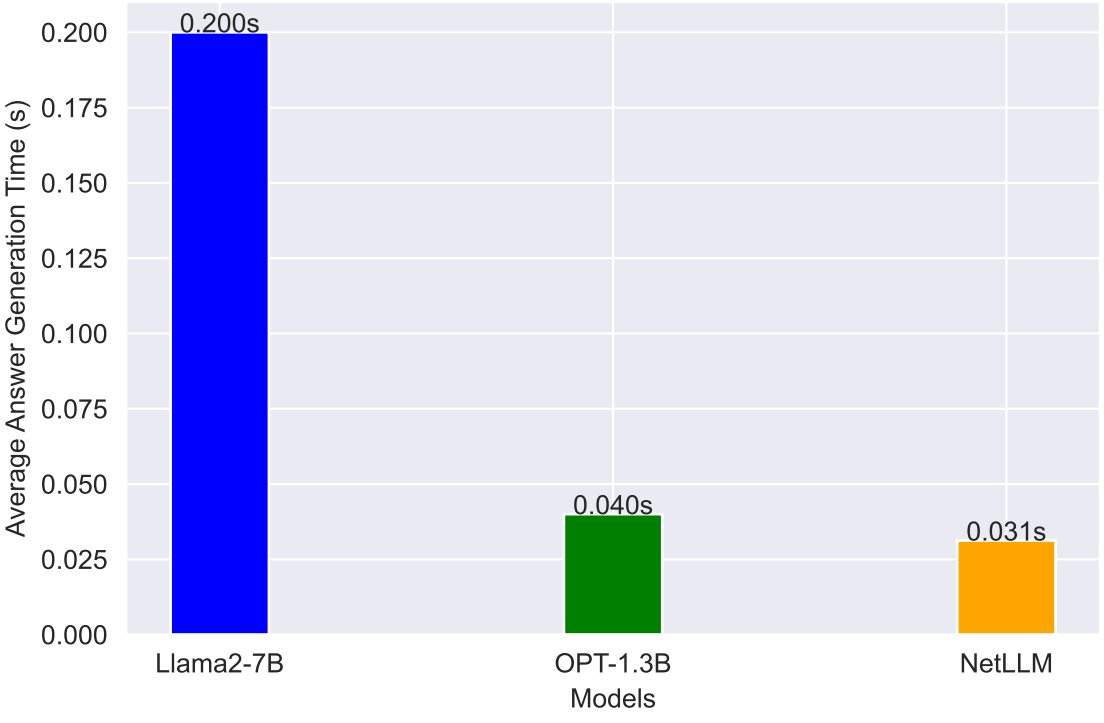
0.200s
0.040s
0.031s
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Average Answer Generation Time (s)
Llama2-7B
OPT-1.3B
NetLLM
Models