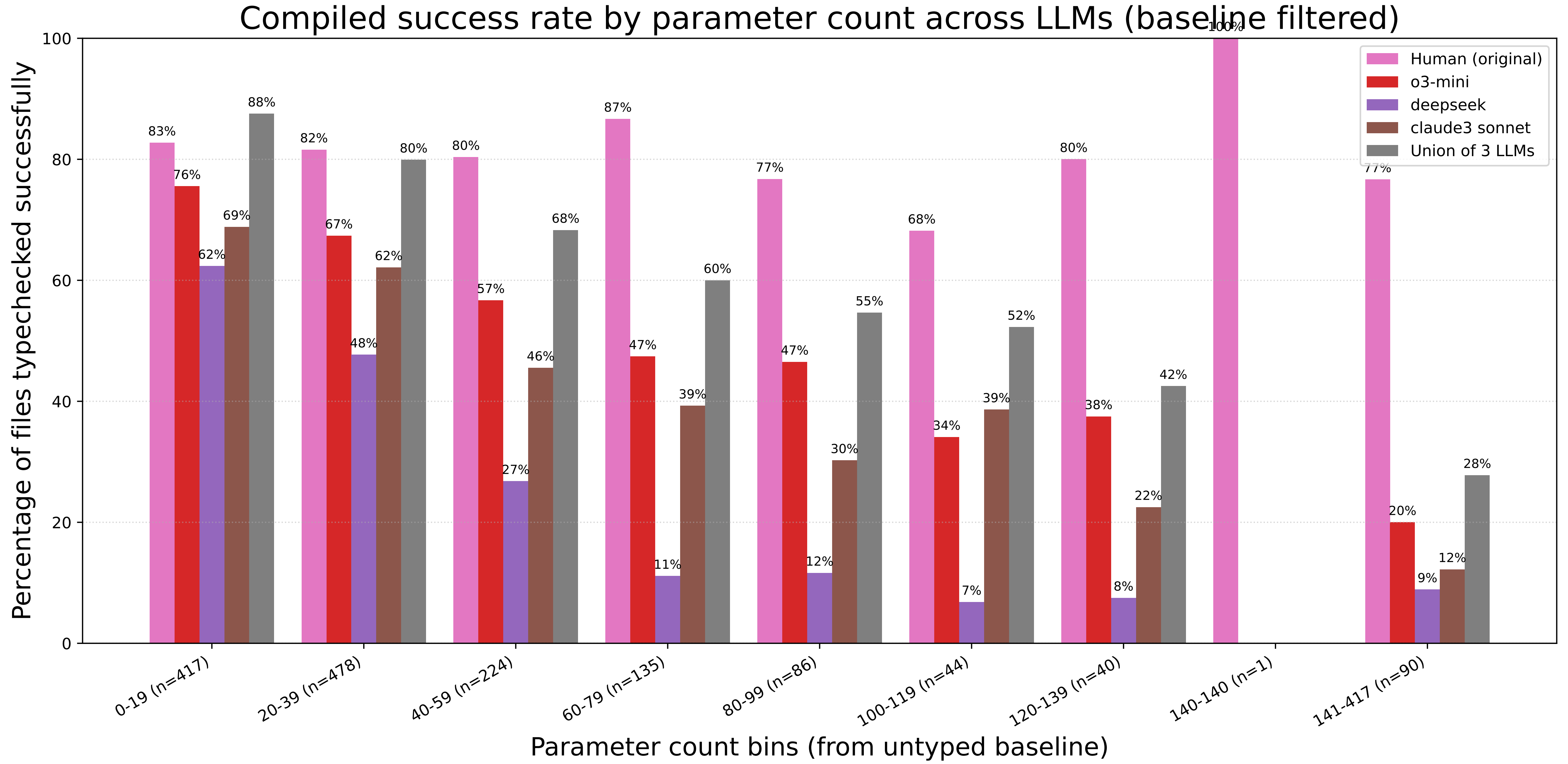
Compiled success rate by parameter count across LLMs (baseline filtered)
Percentage of files typechecked successfully
Parameter count bins (from untyped baseline)
0
20
40
60
80
100
Human (original)
o3-mini
deepseek
claude3 sonnet
Union of 3 LLMs
0-19 (n=417)
83%
76%
62%
69%
88%
20-39 (n=478)
82%
67%
48%
62%
80%
40-59 (n=224)
80%
57%
27%
46%
68%
60-79 (n=135)
87%
47%
11%
39%
60%
80-99 (n=86)
77%
47%
12%
30%
55%
100-119 (n=44)
68%
34%
7%
39%
52%
120-139 (n=40)
80%
38%
8%
22%
42%
140-140 (n=1)
100%
141-417 (n=90)
77%
20%
9%
12%
28%