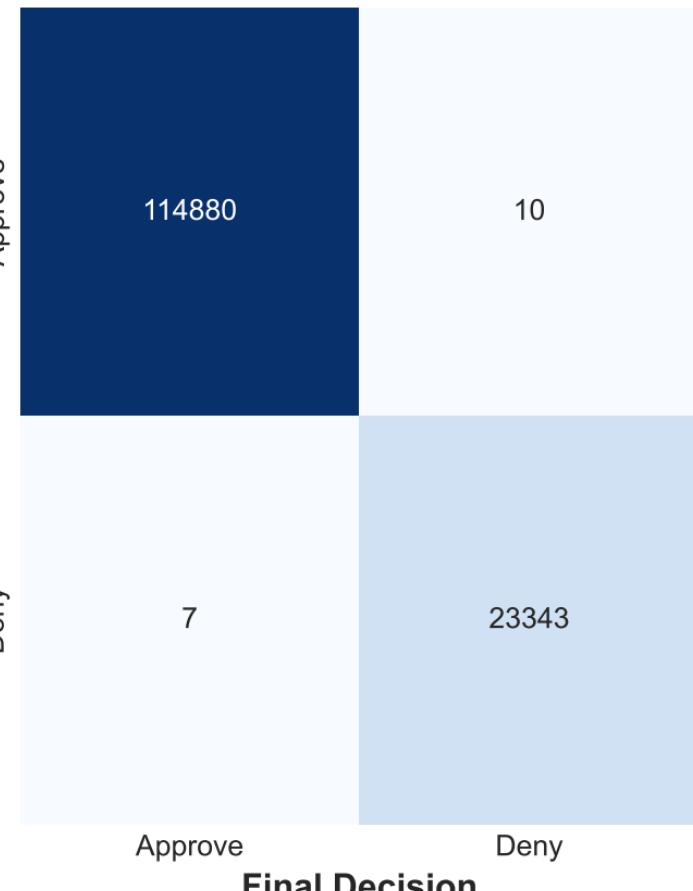


Llama 70B (70B)
First vs Final Decision

First Agent Decision

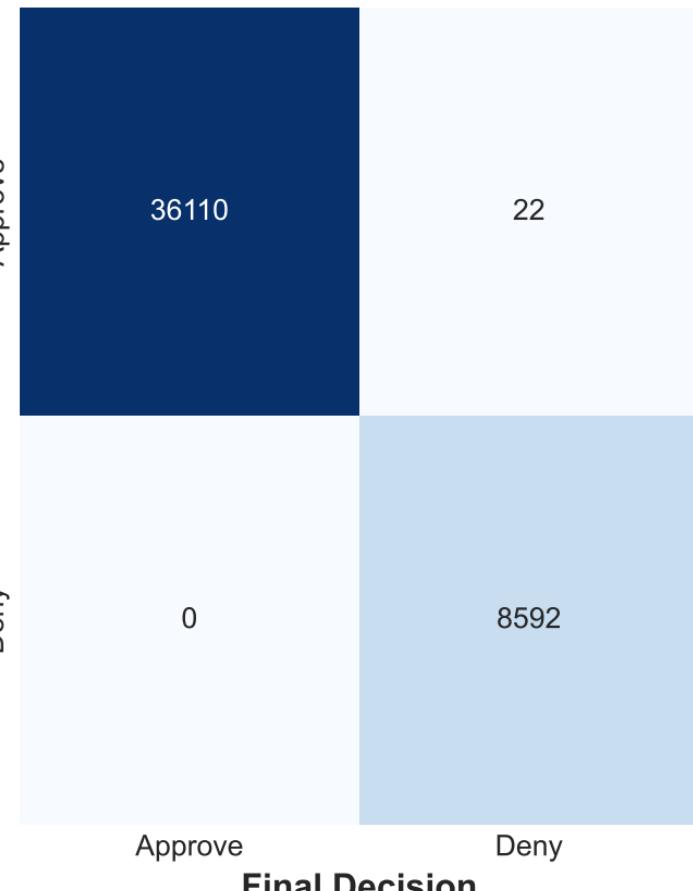


Count

100000
80000
60000
40000
20000
0

Qwen (72B)
First vs Final Decision

First Agent Decision



Count

35000
30000
25000
20000
15000
10000
5000
0