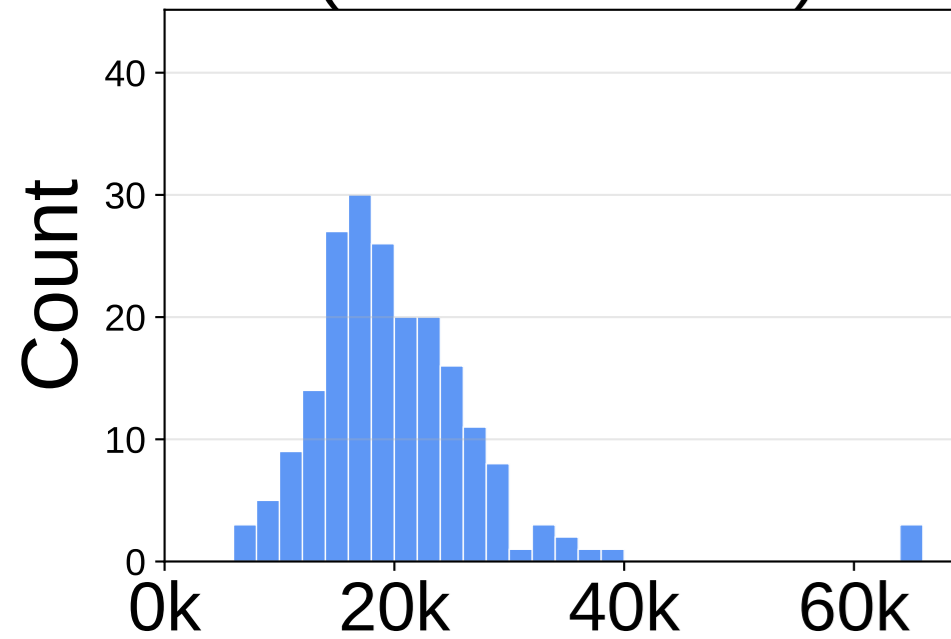
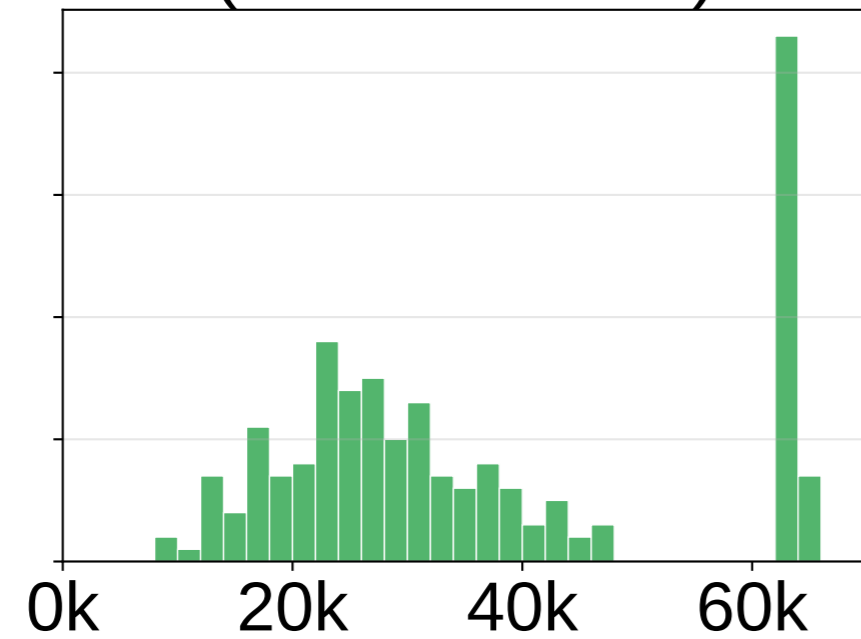


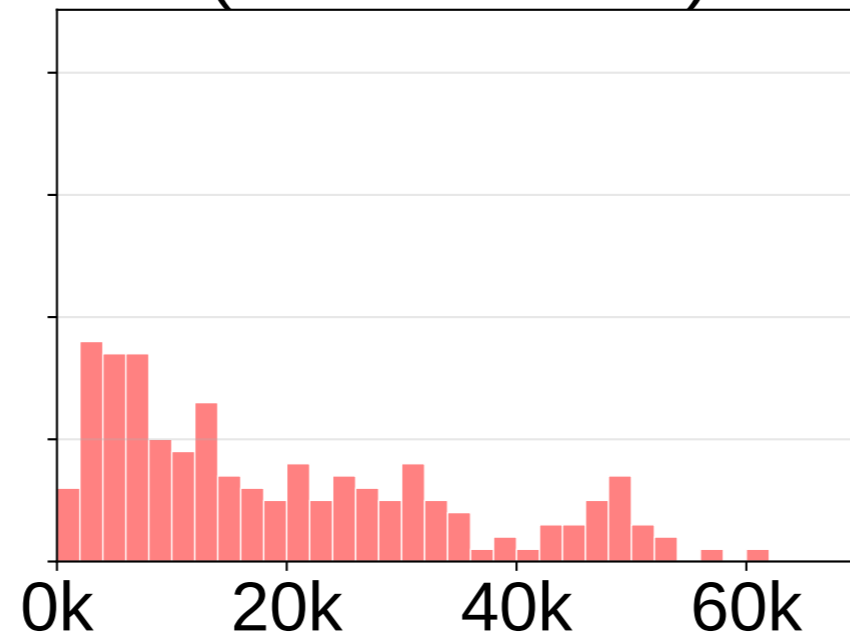
Gemini 3 Pro
(mean = 20.2k)



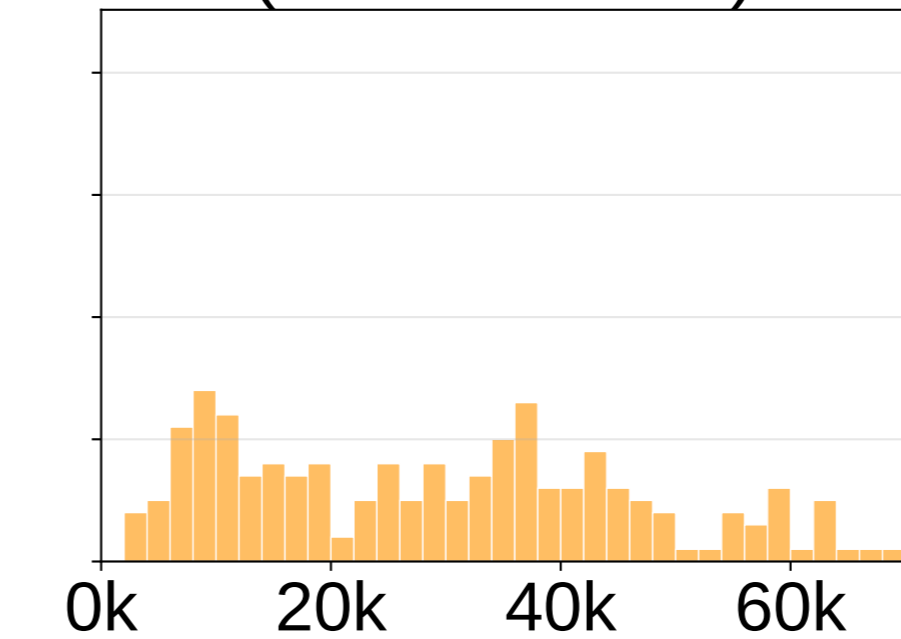
Gemini 3 Flash
(mean = 36.2k)



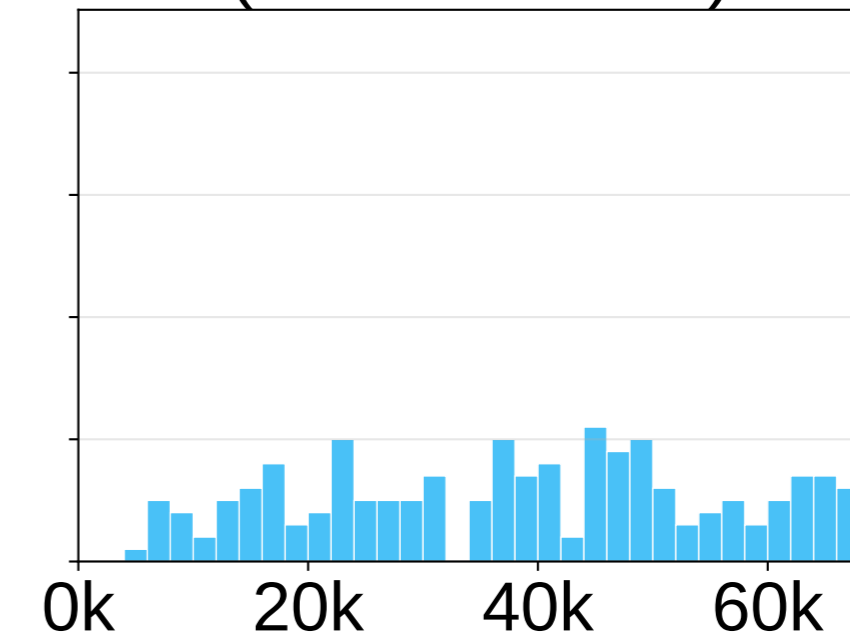
GPT-5.2 High
(mean = 19.5k)



GPT-5-mini High
(mean = 29.3k)



Kimi K2 Thinking
(mean = 42.2k)



DS-V3.2 Thinking
(mean = 25.5k)

