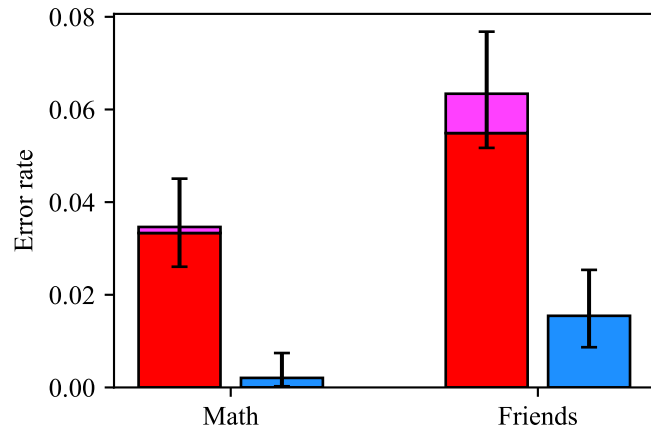
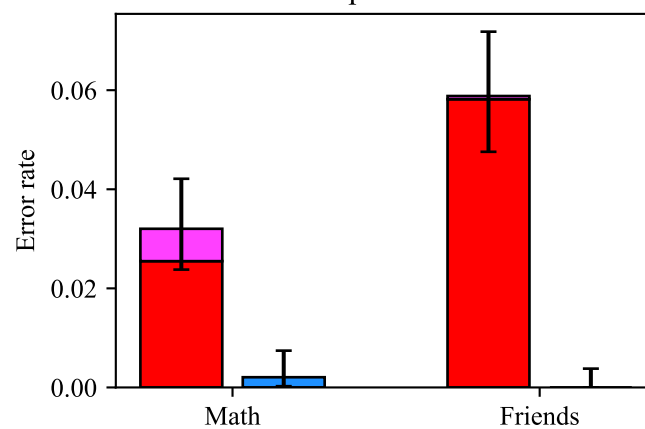


False uncolorable    Wrong coloring    False colorable

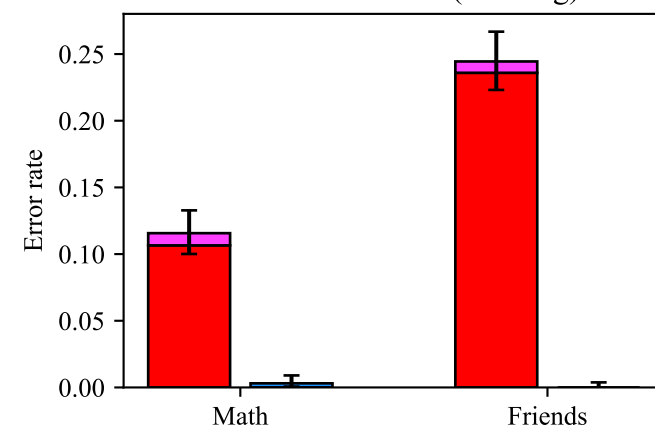
o1-mini



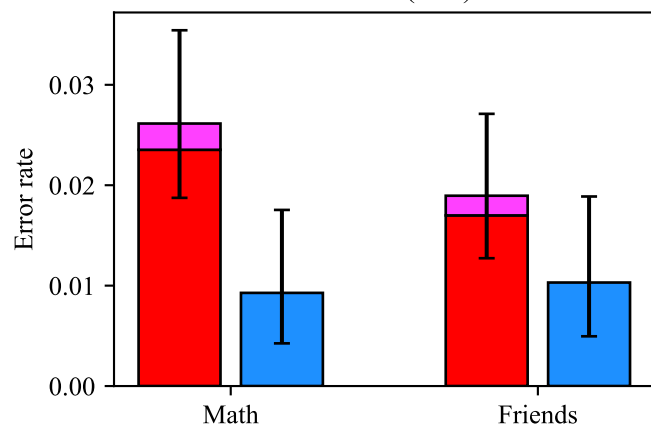
DeepSeek-R1



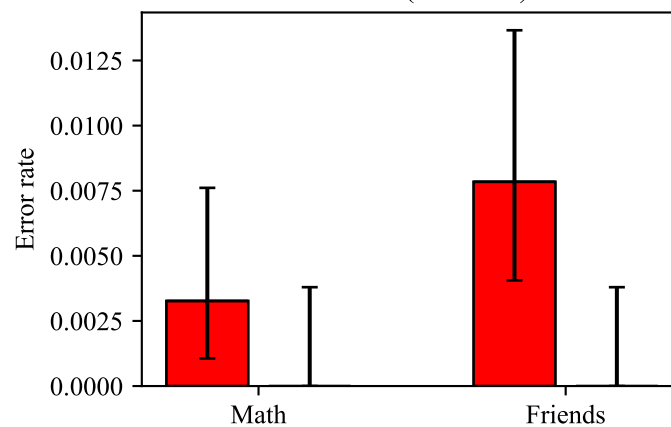
Claude 3.7 Sonnet (thinking)



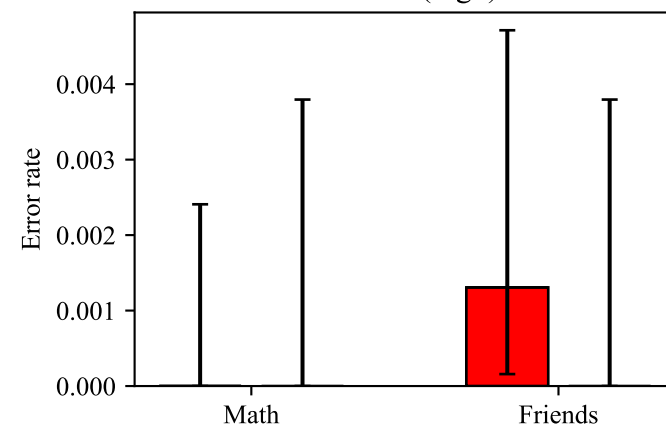
o3-mini (low)



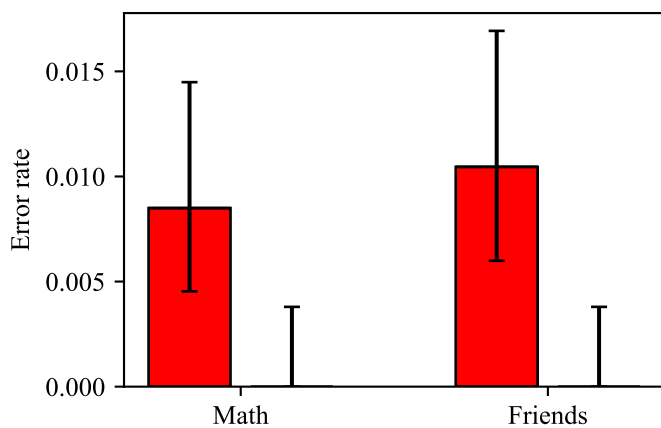
o3-mini (medium)



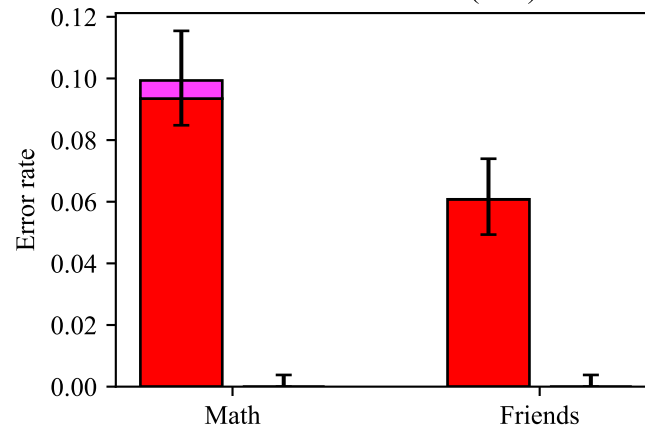
o3-mini (high)



Gemini 2.5 Pro Preview



Grok 3 Mini Beta (low)



Grok 3 Mini Beta (high)

