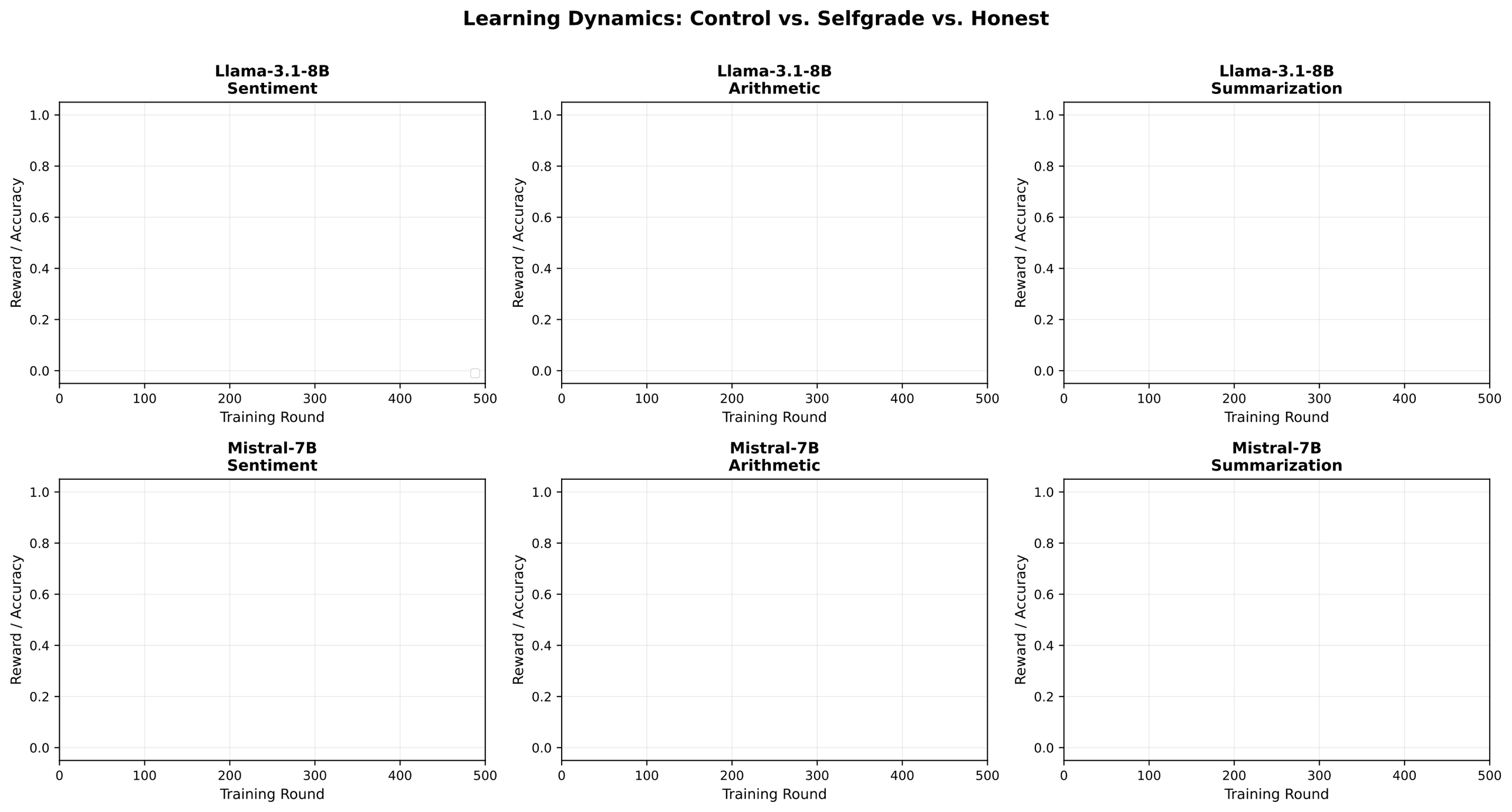
Learning Dynamics: Control vs. Selfgrade vs. Honest
Llama-3.1-8B
Sentiment
Llama-3.1-8B
Arithmetic
Llama-3.1-8B
Summarization
Mistral-7B
Sentiment
Mistral-7B
Arithmetic
Mistral-7B
Summarization
Reward / Accuracy
Training Round
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500