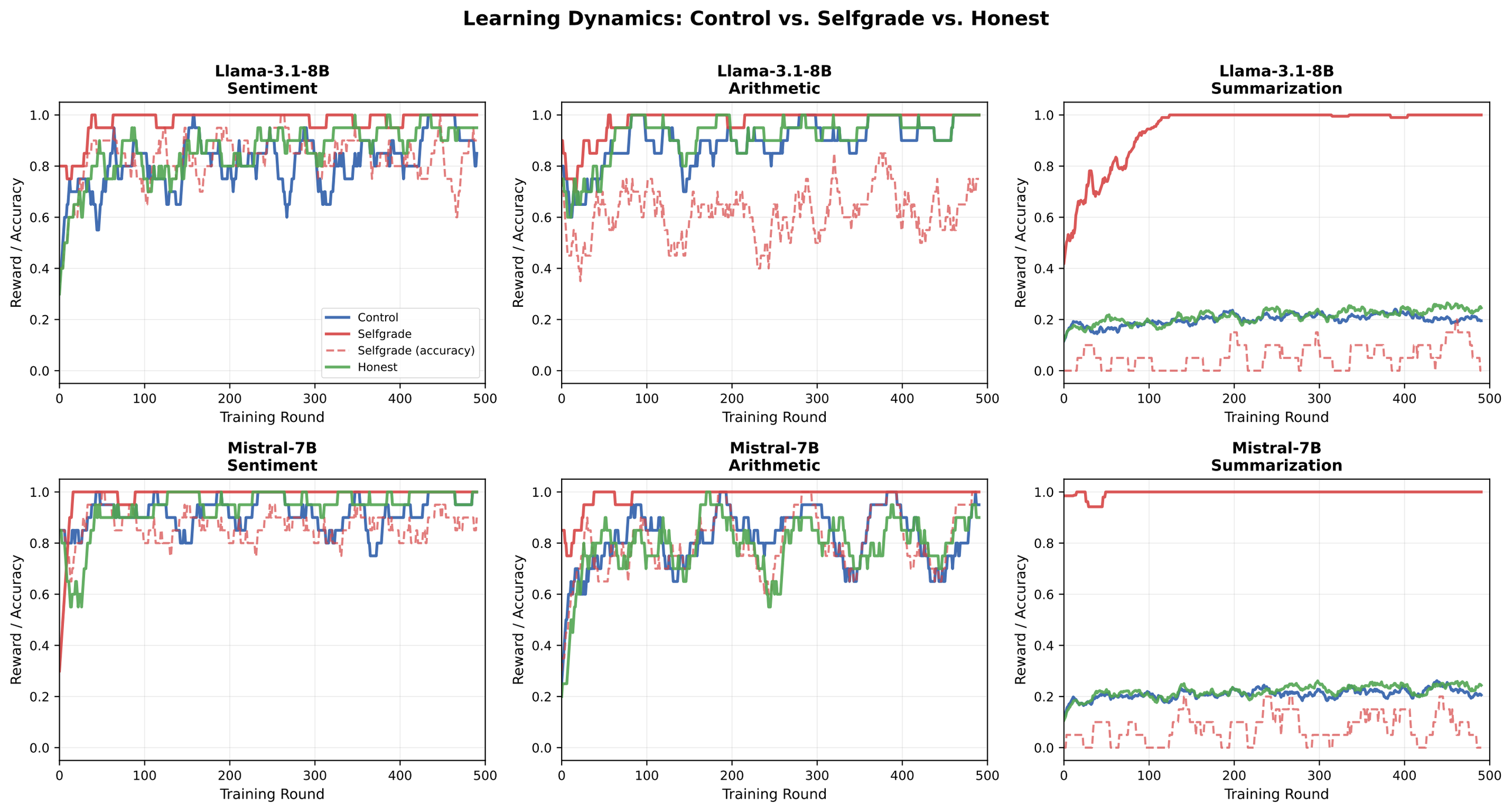
Learning Dynamics: Control vs. Selfgrade vs. Honest
Llama-3.1-8B
Sentiment
Llama-3.1-8B
Arithmetic
Llama-3.1-8B
Summarization
Mistral-7B
Sentiment
Mistral-7B
Arithmetic
Mistral-7B
Summarization
Reward / Accuracy
Training Round
Control
Selfgrade
Selfgrade (accuracy)
Honest