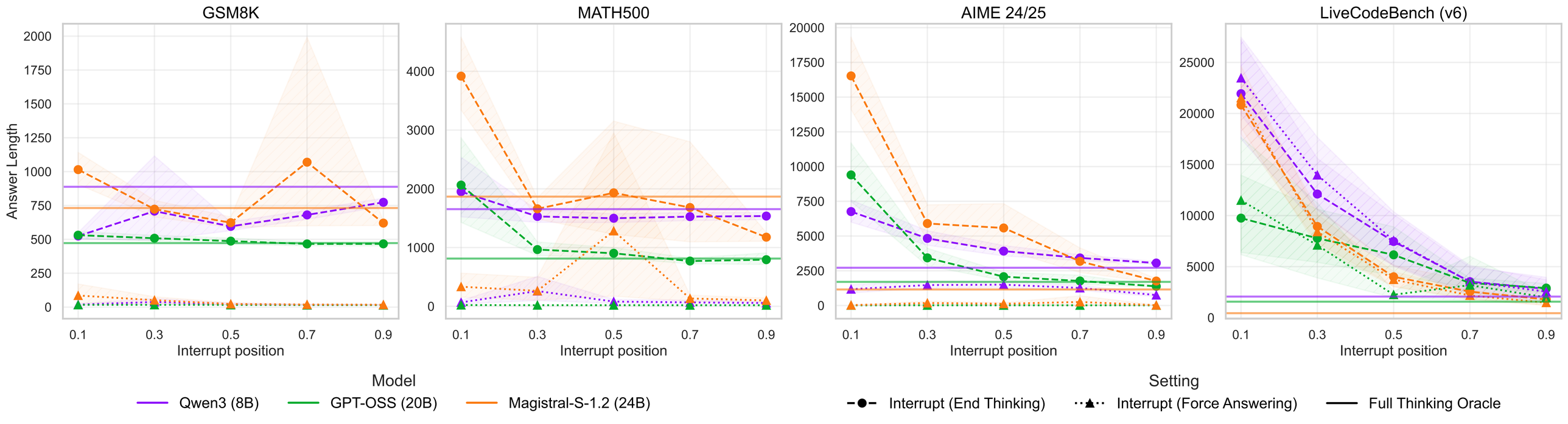
GSM8K
MATH500
AIME 24/25
LiveCodeBench (v6)
Answer Length
Interrupt position
Model
Qwen3 (8B)
GPT-OSS (20B)
Magistral-S-1.2 (24B)
Setting
Interrupt (End Thinking)
Interrupt (Force Answering)
Full Thinking Oracle