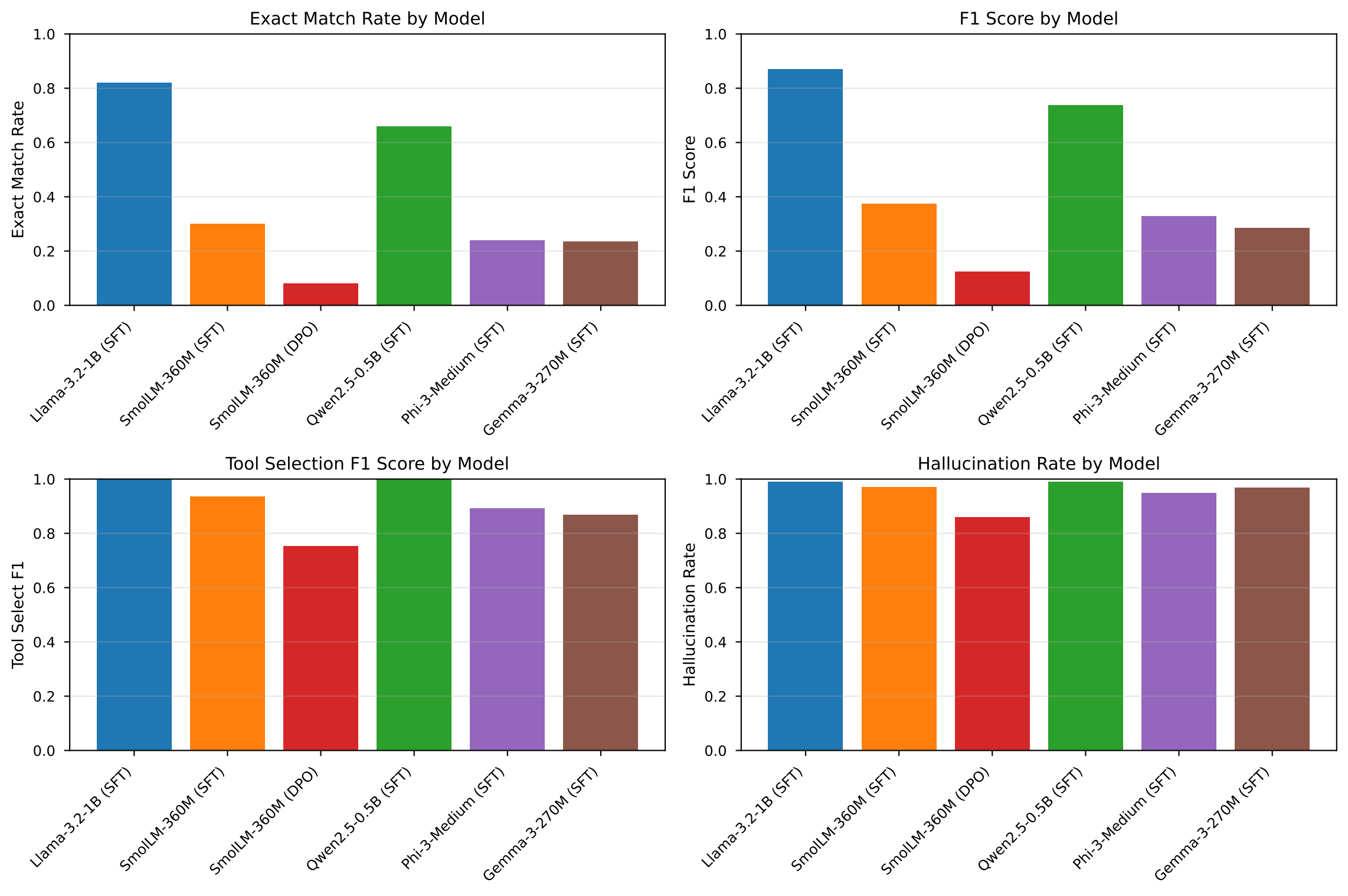
Exact Match Rate by Model
Exact Match Rate
F1 Score by Model
F1 Score
Tool Selection F1 Score by Model
Tool Select F1
Hallucination Rate by Model
Hallucination Rate
0.0
0.2
0.4
0.6
0.8
1.0
Llama-3.2-1B (SFT)
SmolLM-360M (SFT)
SmolLM-360M (DPO)
Qwen2.5-0.5B (SFT)
Phi-3-Medium (SFT)
Gemma-3-270M (SFT)