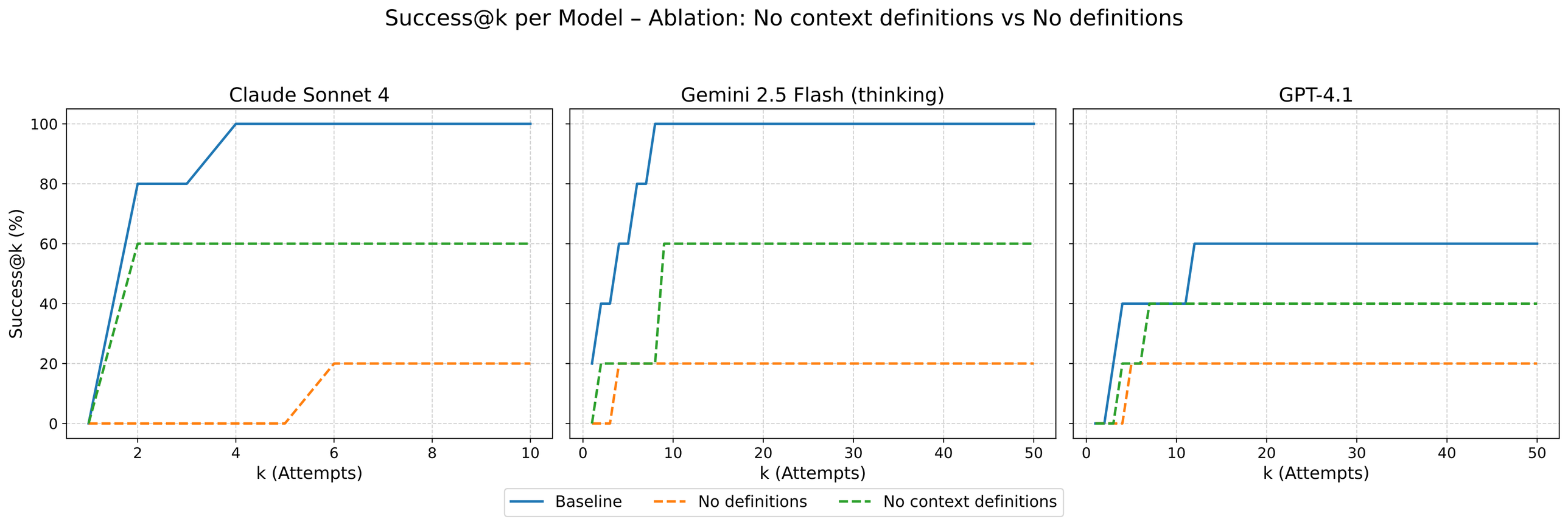
Success@k per Model – Ablation: No context definitions vs No definitions
Claude Sonnet 4
Gemini 2.5 Flash (thinking)
GPT-4.1
Success@k (%)
k (Attempts)
Baseline
No definitions
No context definitions