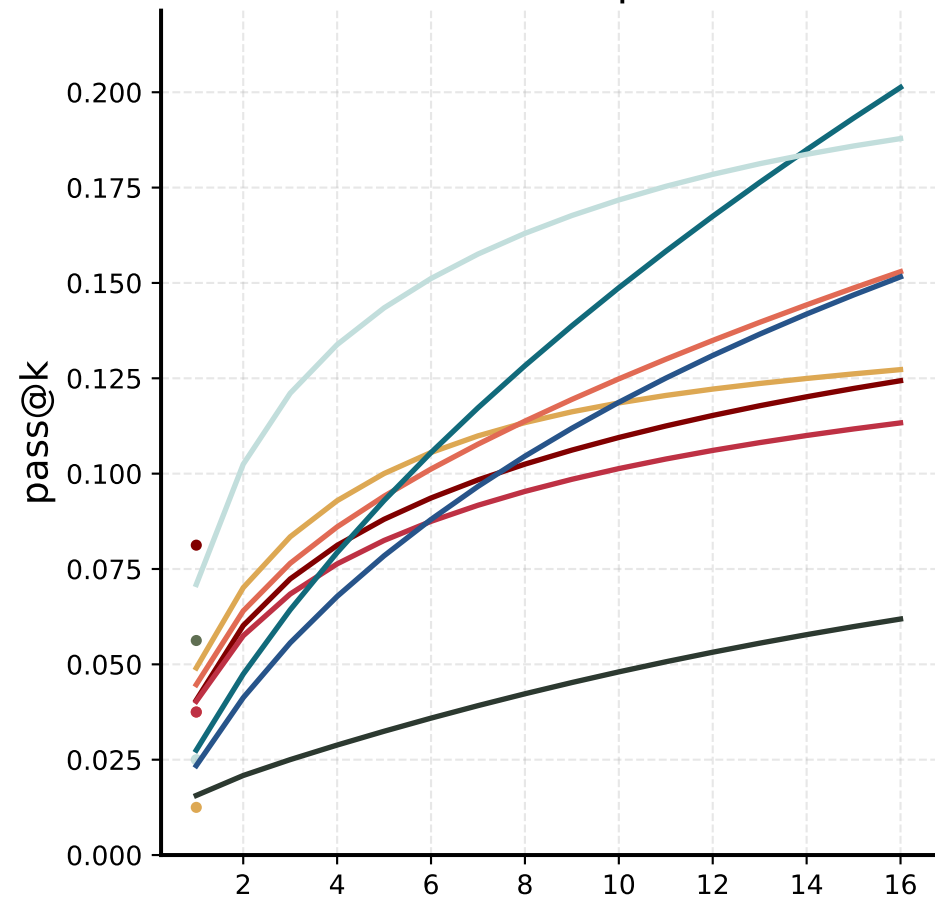


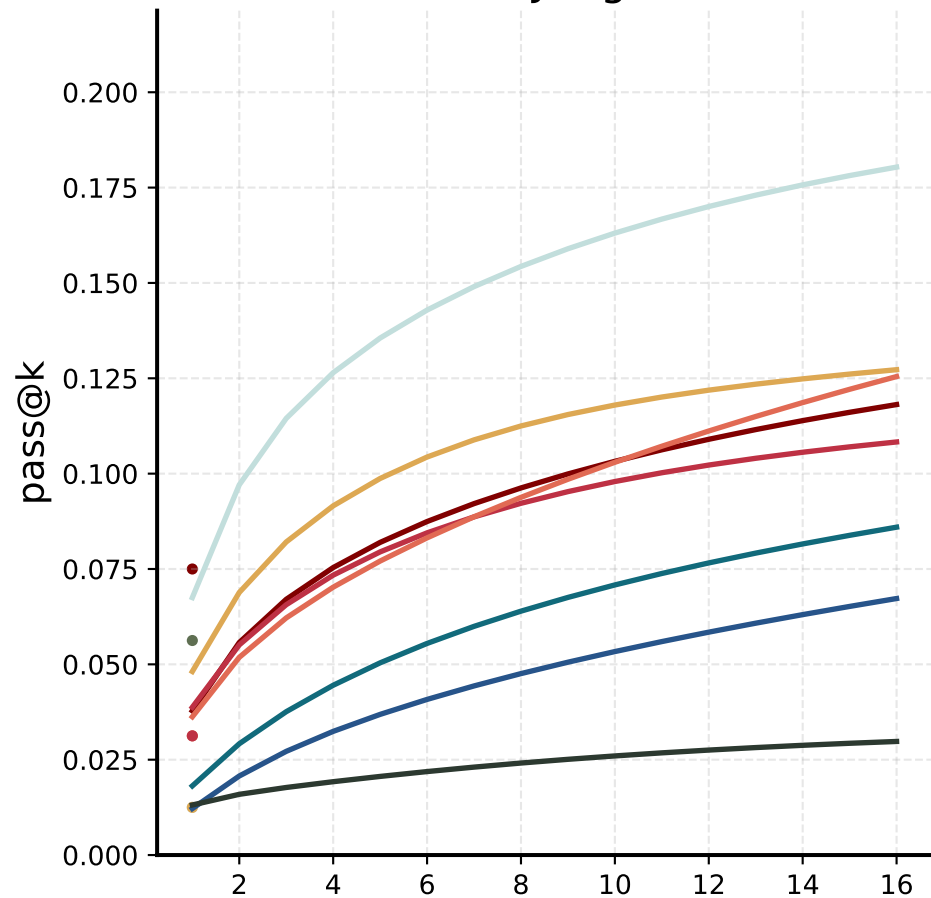
Lean Compile



- DeepSeek V3 (0324)(T=0.0)
- DeepSeek V3 (0324)(T=0.6)
- Claude 3.7 Sonnet(T=0.0)
- Claude 3.7 Sonnet(T=0.6)
- Claude 3.7 Sonnet (thinking)

- DeepSeek R1(T=0.0)
- DeepSeek R1(T=0.6)
- GPT-4o(T=0.0)
- GPT-4o(T=0.6)
- Doubao 1.5 Pro(T=0.0)

LLM Judge



- Doubao 1.5 Pro(T=0.6)
- OpenAI o3-mini
- Gemini 2.5 Pro Preview(T=0.0)
- Gemini 2.5 Pro Preview(T=0.6)