Recent LLMs can do 2-hop and 3-hop latent reasoning on natural facts

Multi-Hop Latent Reasoning Task

2-hop example:

$1900 + 43$

Who won Nobel Chemistry in (1900 + (atomic # of Tc))?
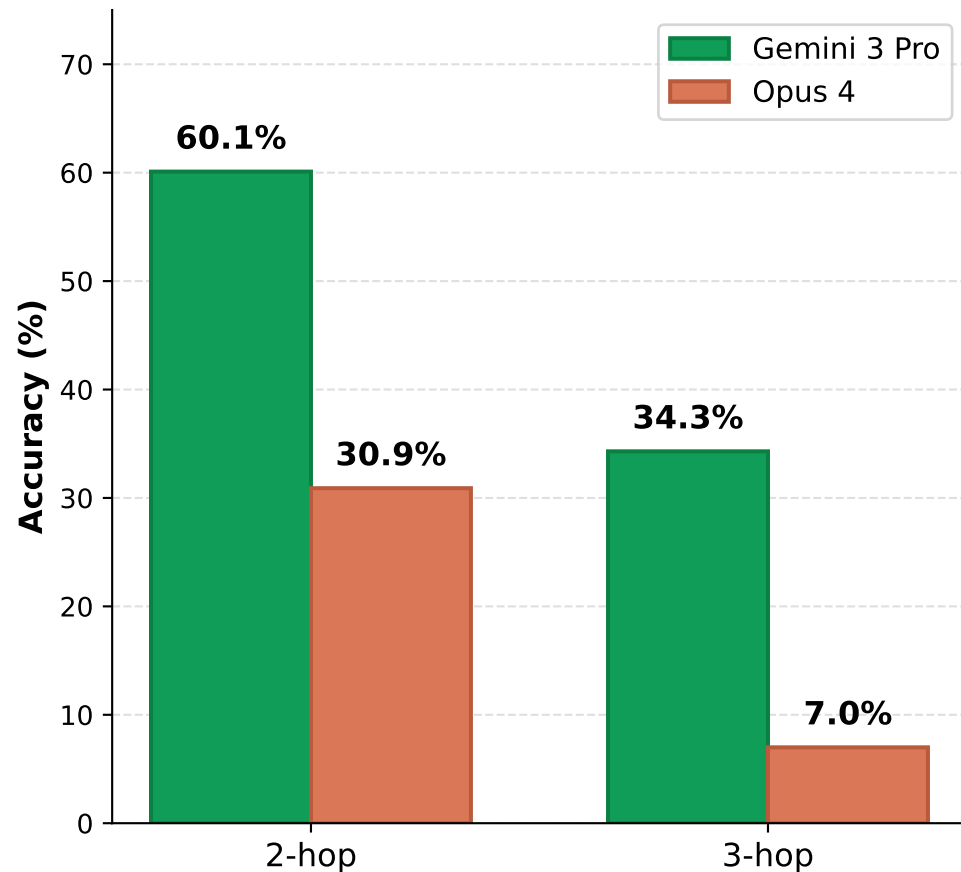
→ **George de Hevesy**

3-hop example:

Wyoming

44

Counties in state that joined the US (age Chekhov died)th?

→ **23**

Intermediate values
Final answer

Models must solve in a single forward pass (no CoT)

Performance (with filler tokens, f=300)

- Gemini 3 Pro
- Opus 4

2-hop: **60.1%** (Gemini 3 Pro), **30.9%** (Opus 4)
3-hop: **34.3%** (Gemini 3 Pro), **7.0%** (Opus 4)

Accuracy (%)

(4-hop accuracy is near chance for all models)