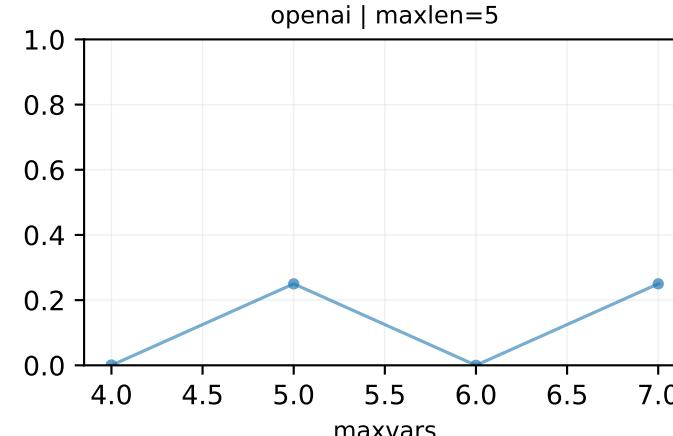
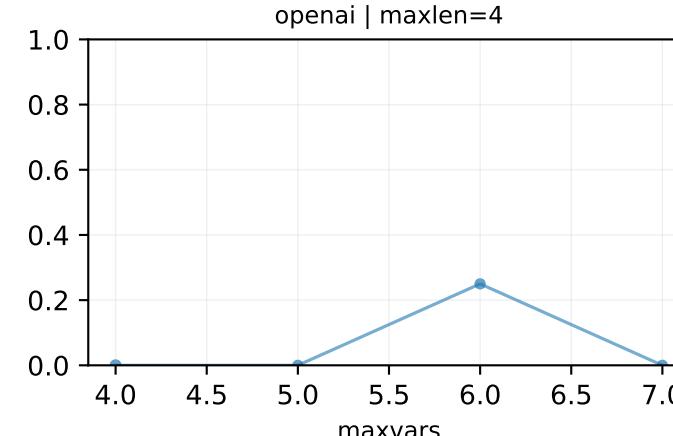
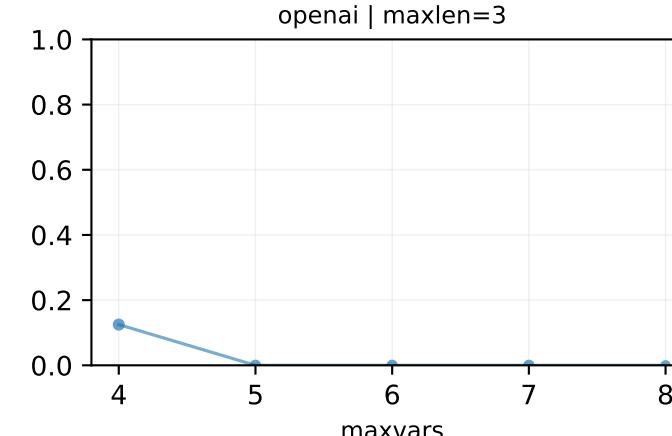
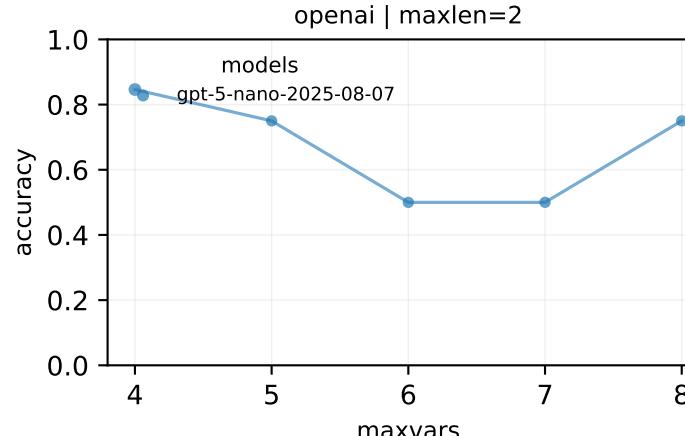
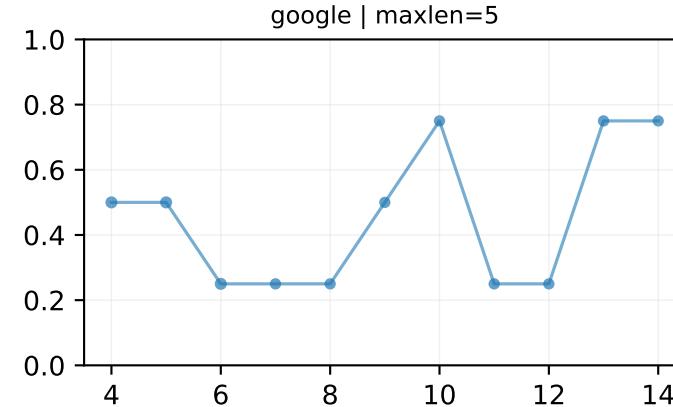
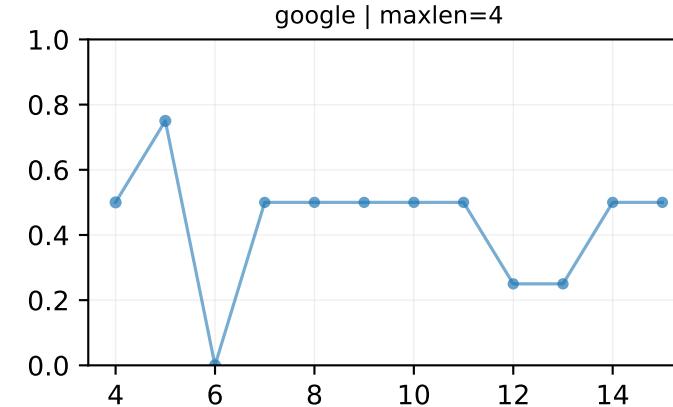
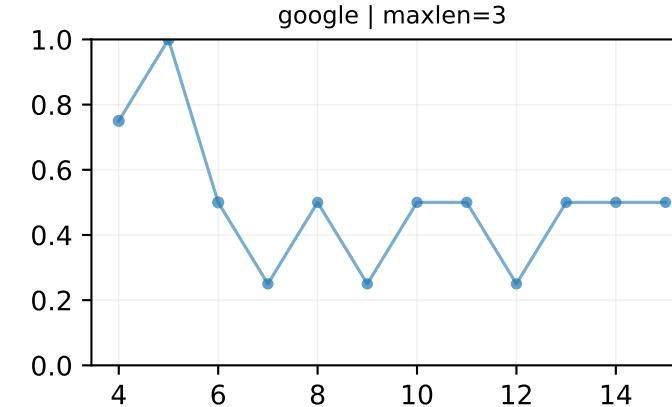
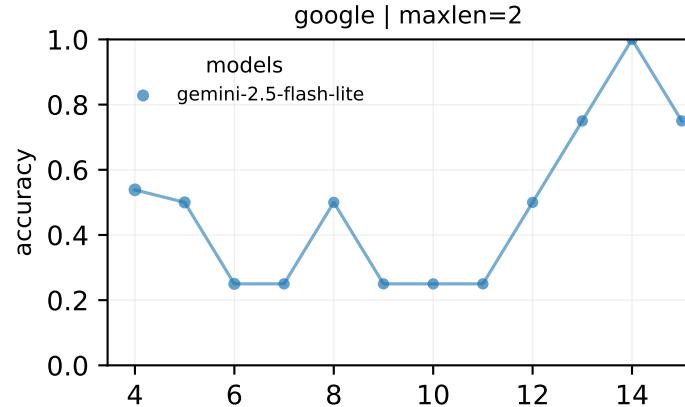
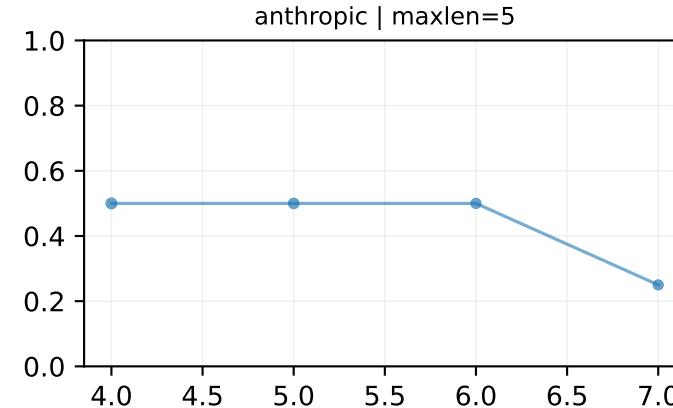
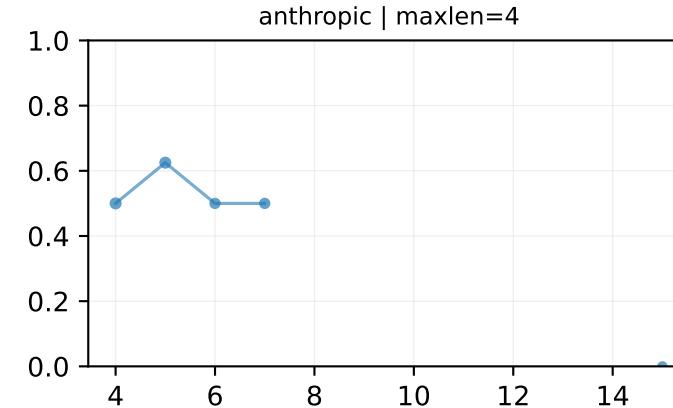
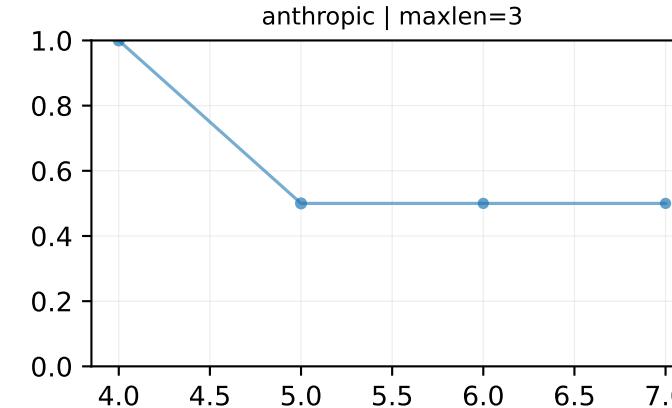
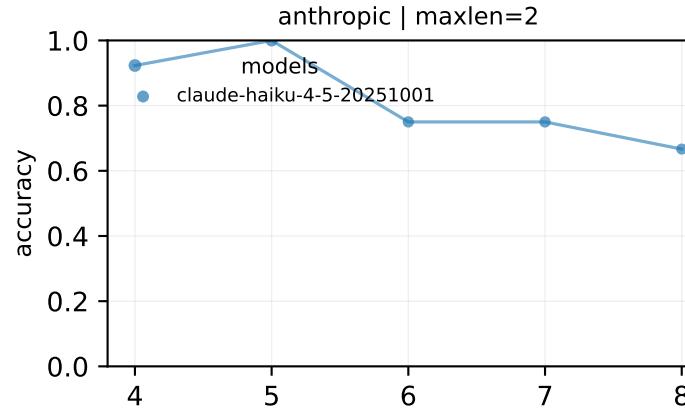


Condensed: family=horn_unified_answer | thinking=nothink | horn=0 | metric=accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \sqrt{n}$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

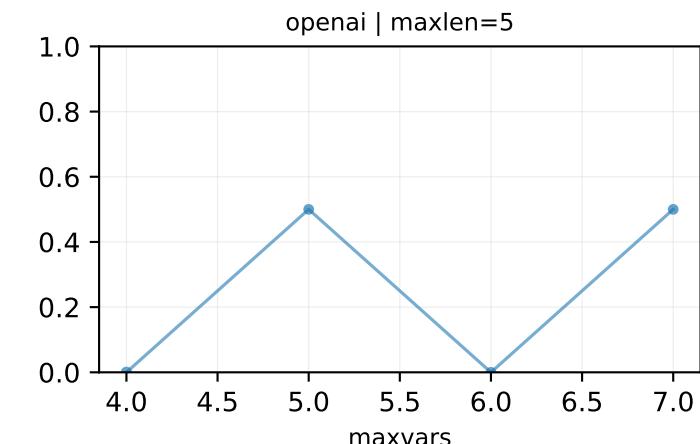
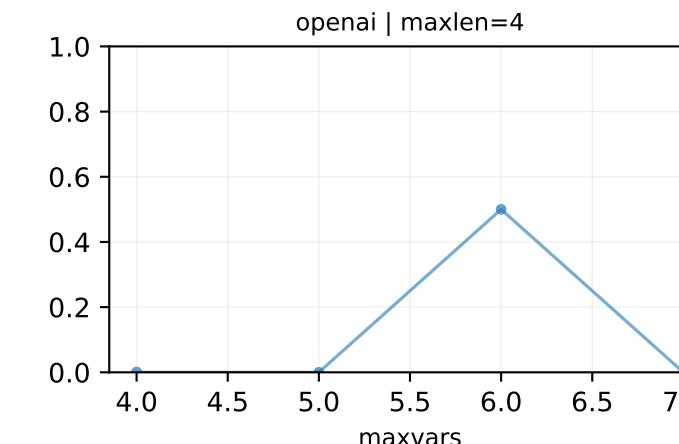
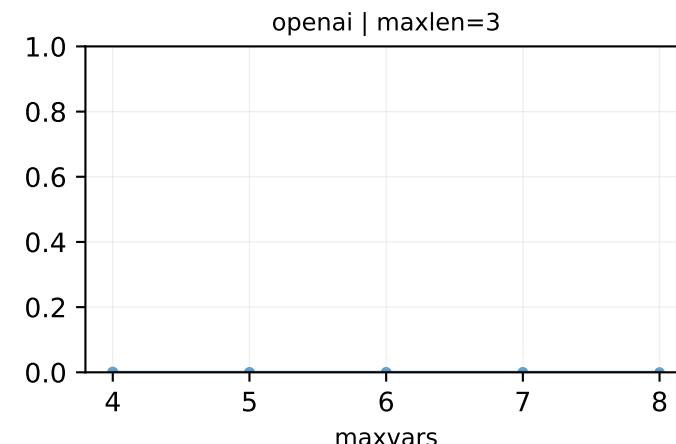
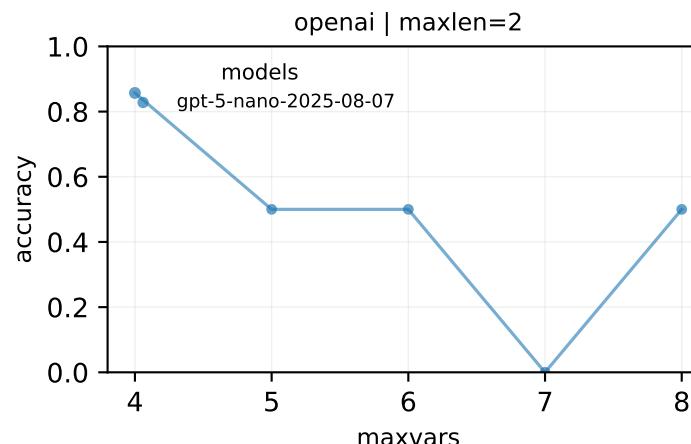
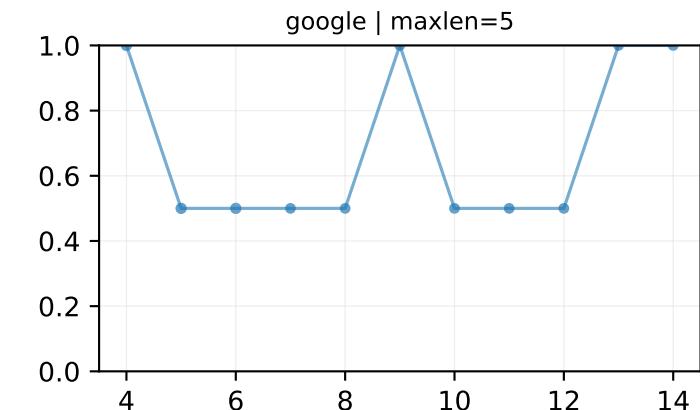
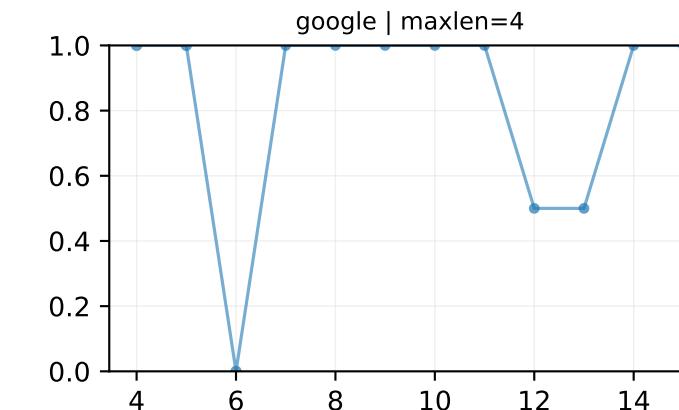
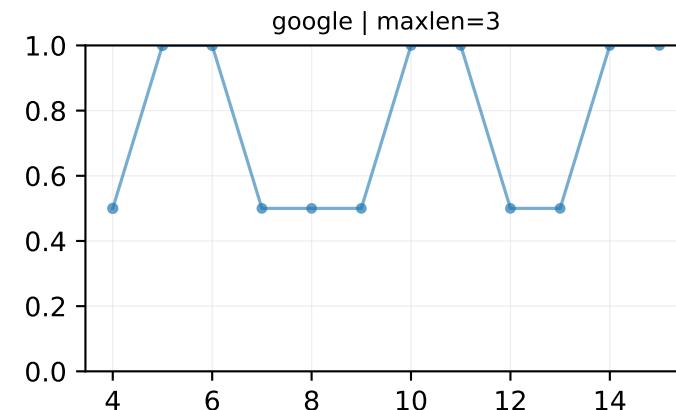
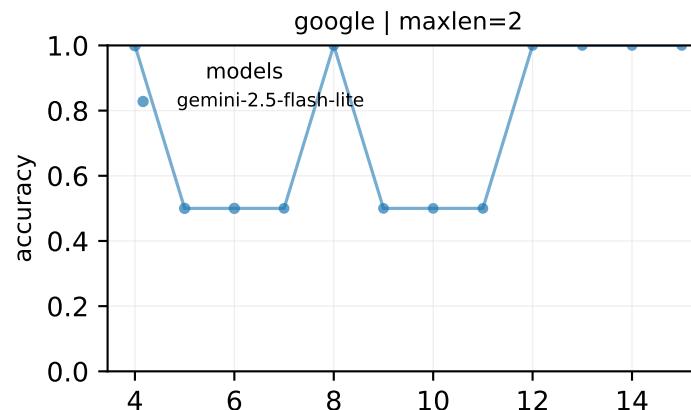
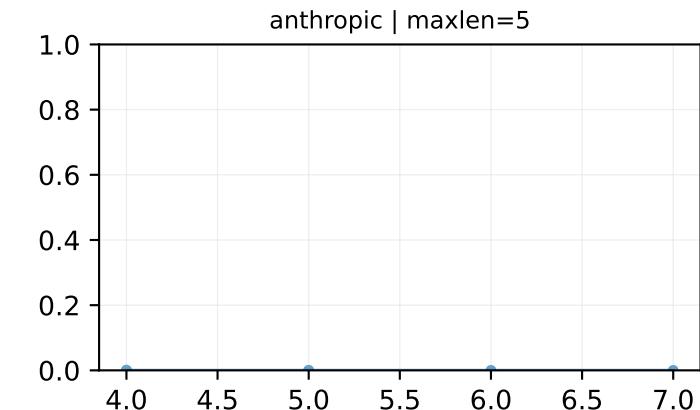
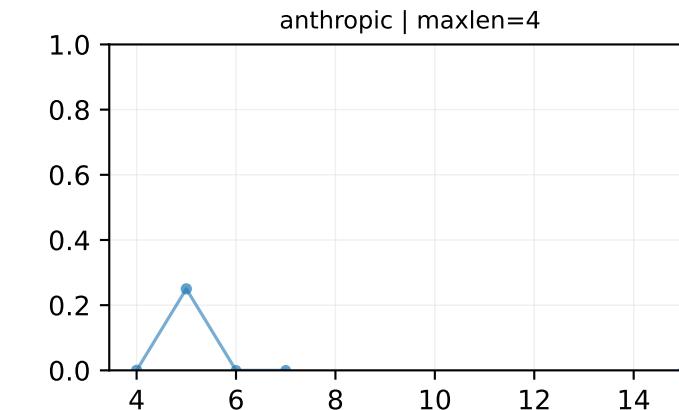
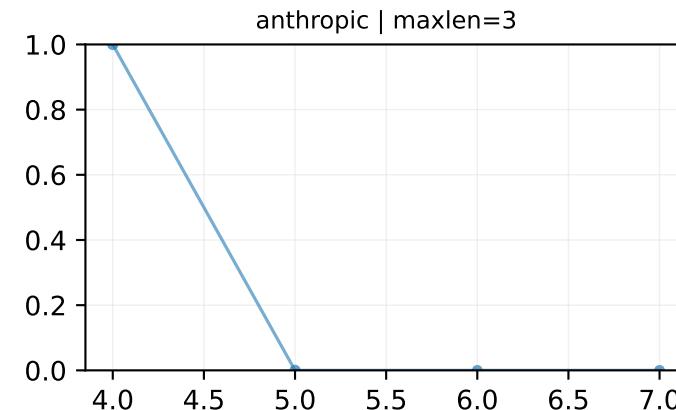
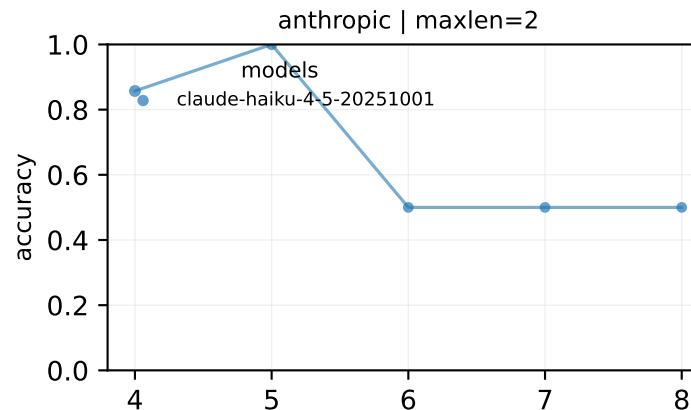


Condensed: family=horn_unified_answer | thinking=nothink | horn=0 | metric=sat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \sqrt{n}$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

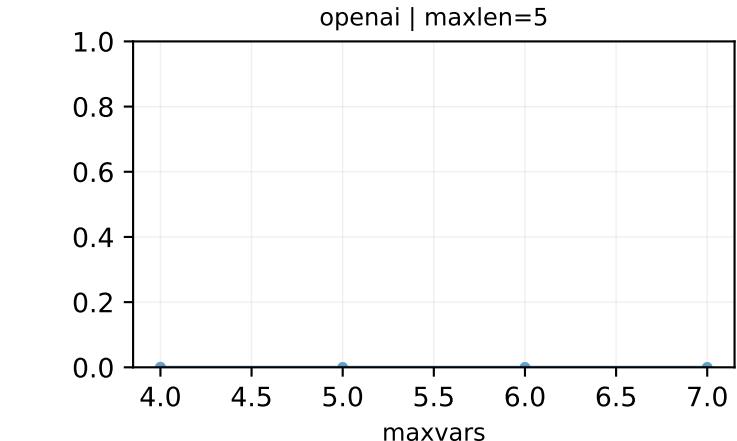
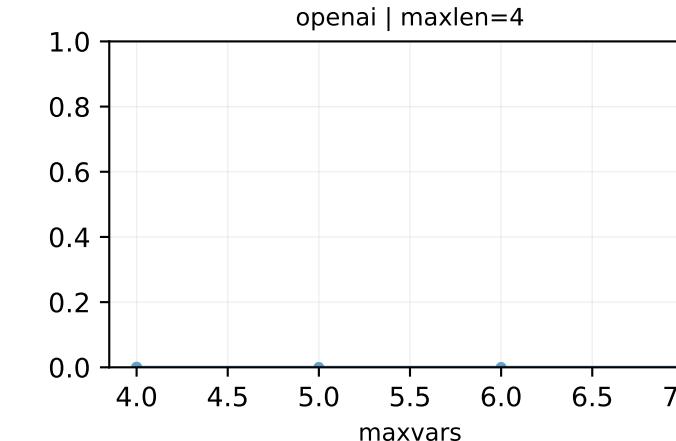
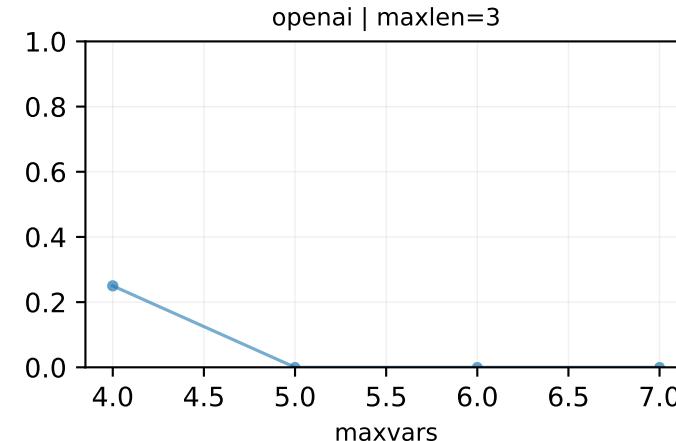
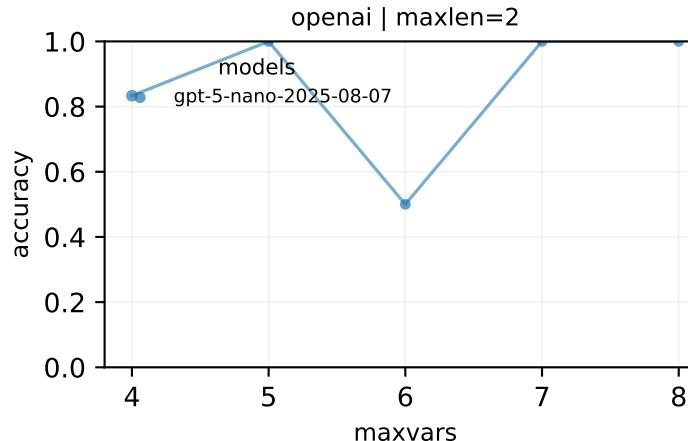
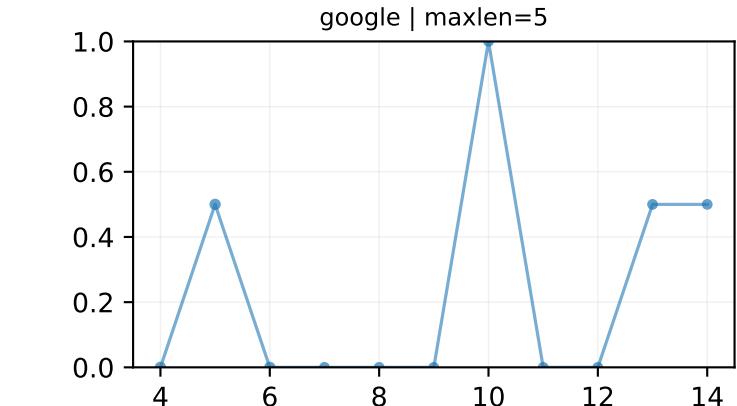
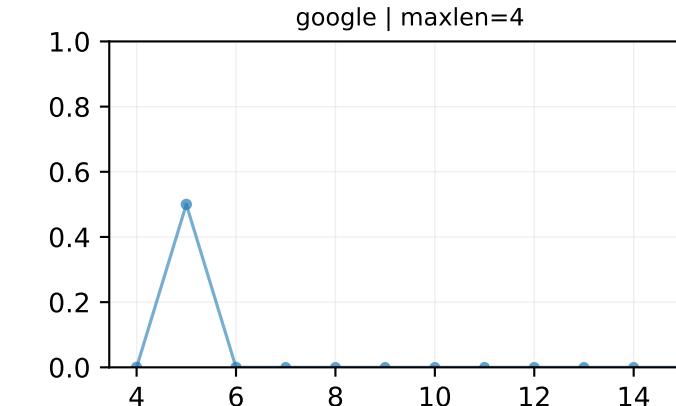
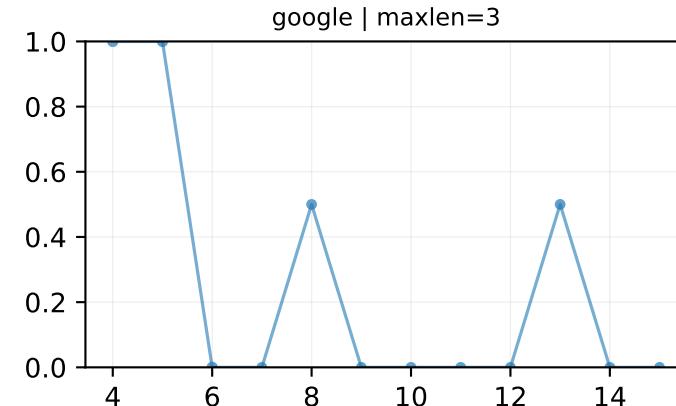
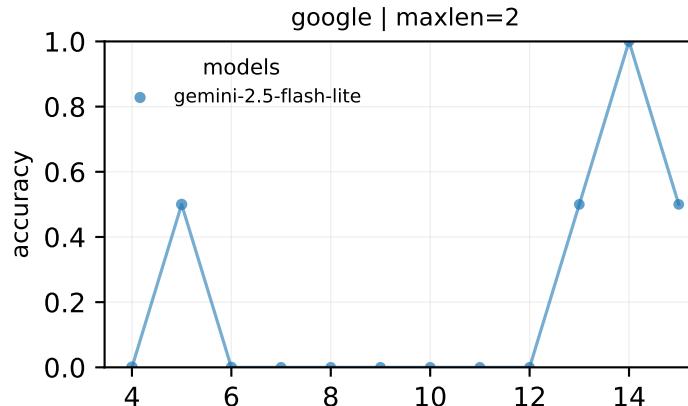
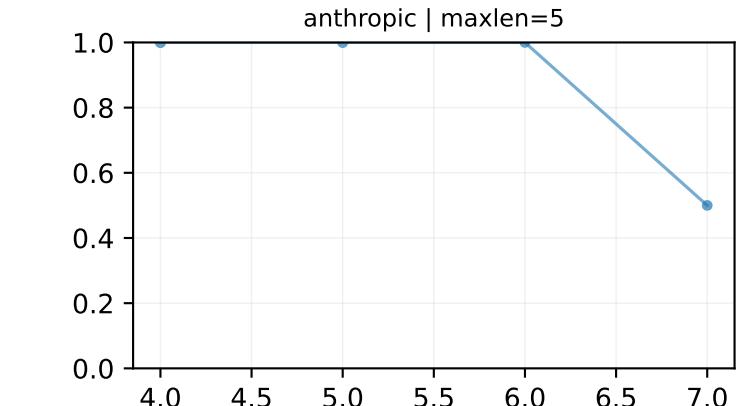
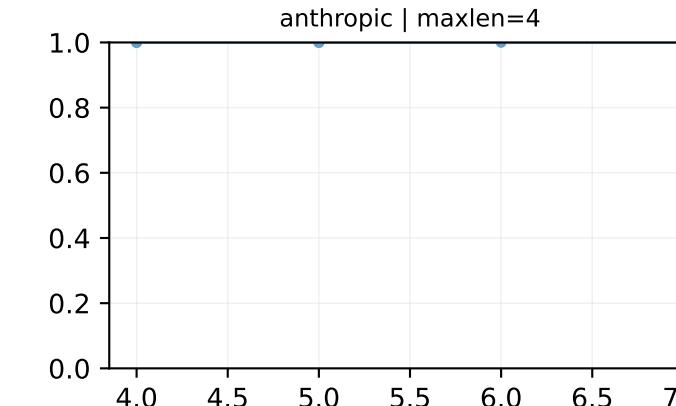
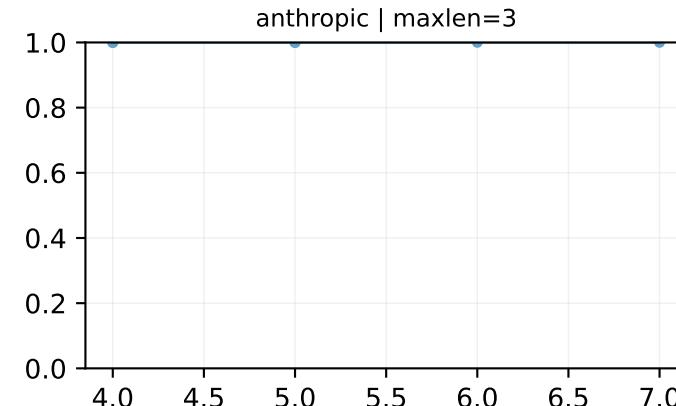
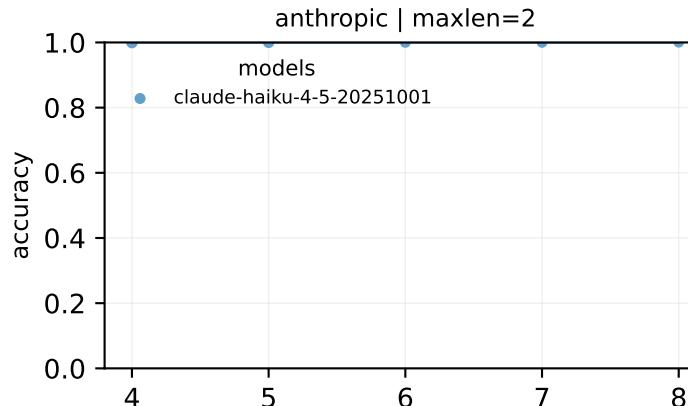


Condensed: family=horn_unified_answer | thinking=nothink | horn=0 | metric=unsat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \sqrt{n}$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

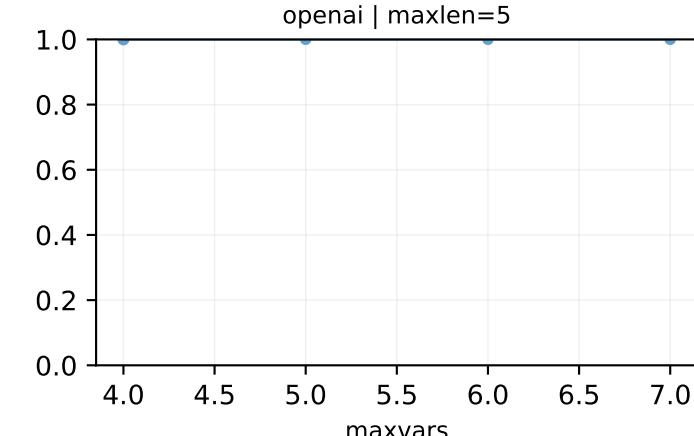
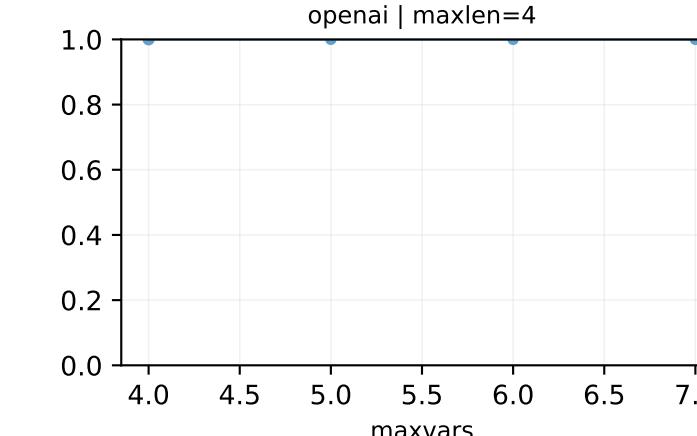
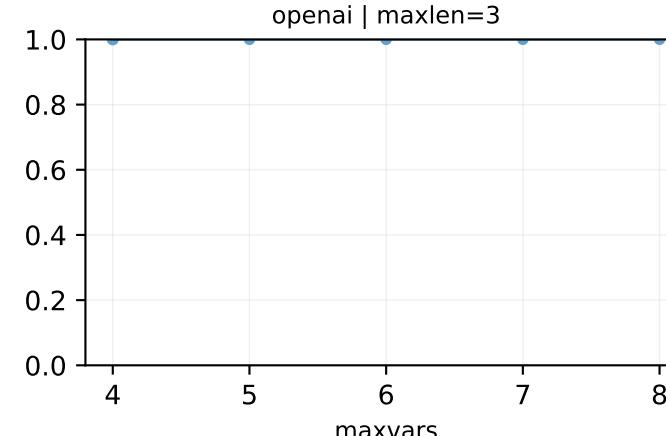
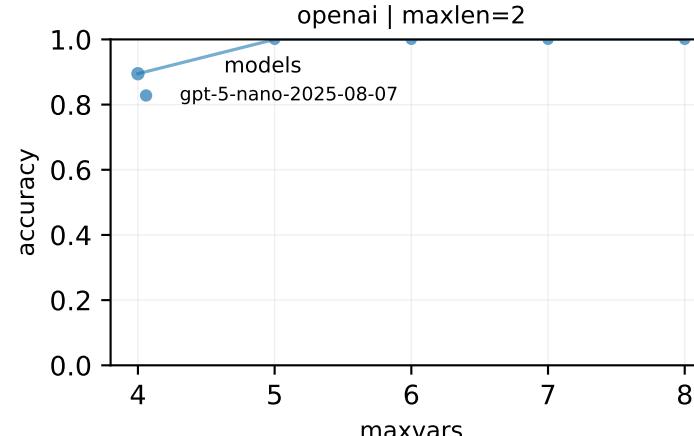
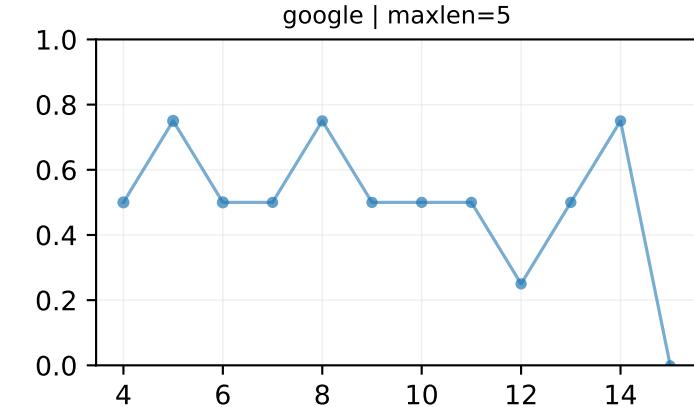
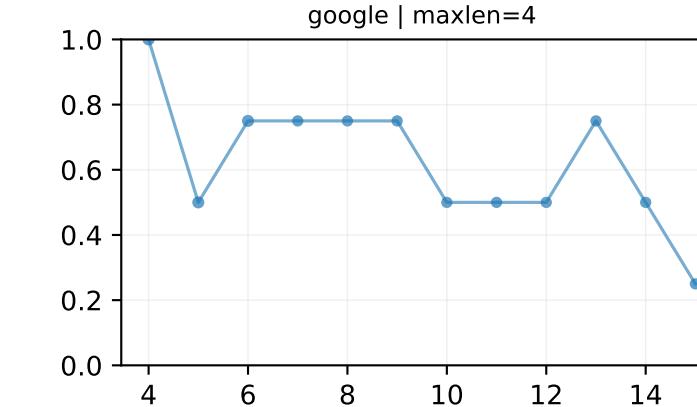
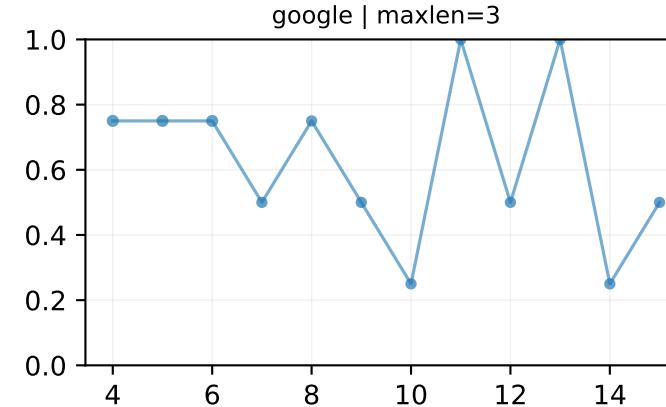
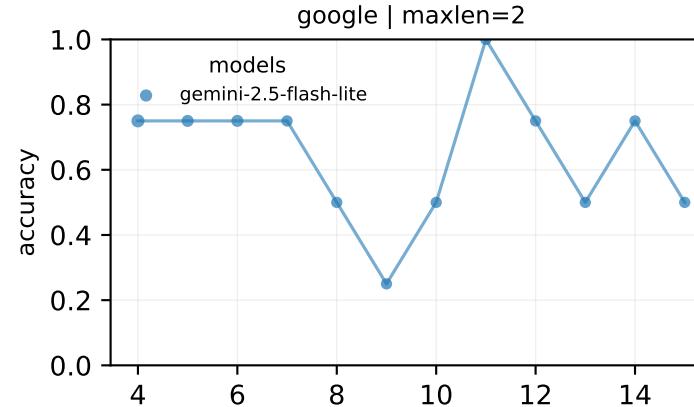
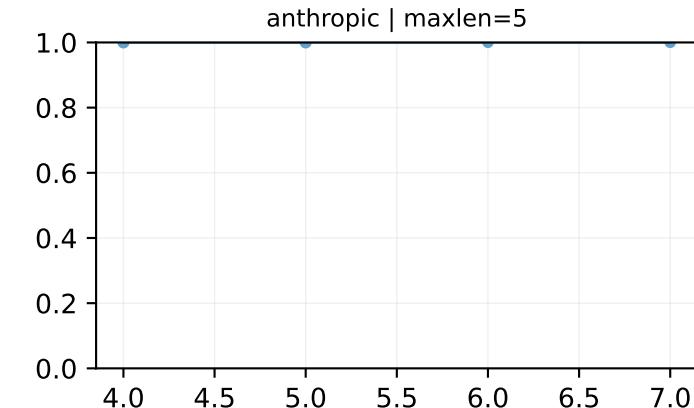
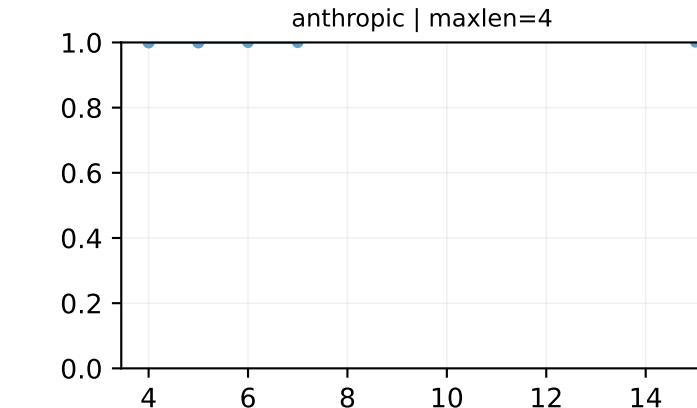
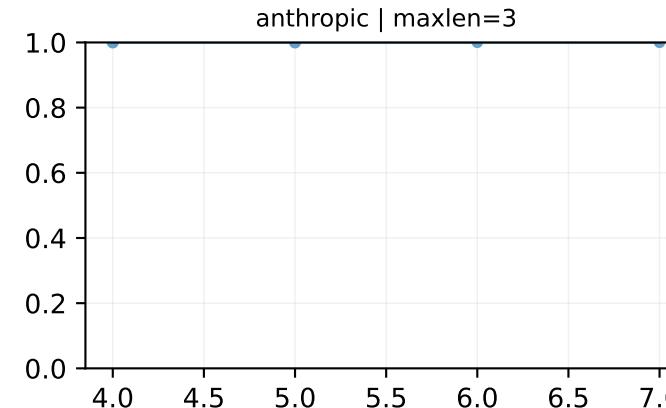
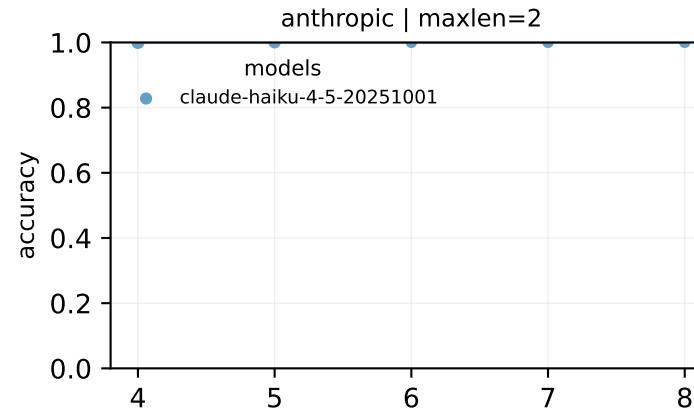


Condensed: family=horn_unified_answer | thinking=nothink | horn=1 | metric=accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \text{sqrt}(n)$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

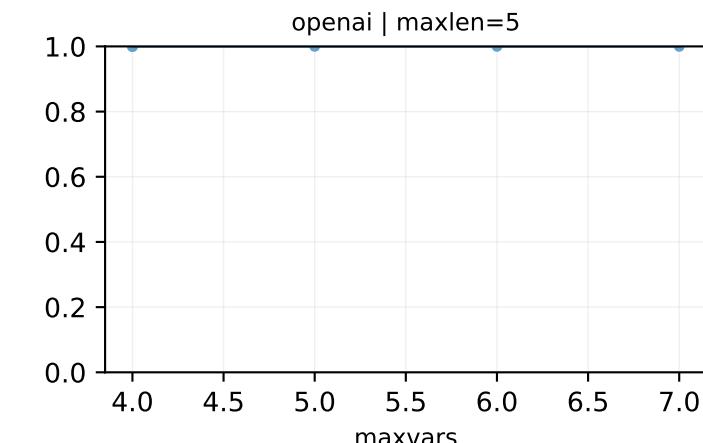
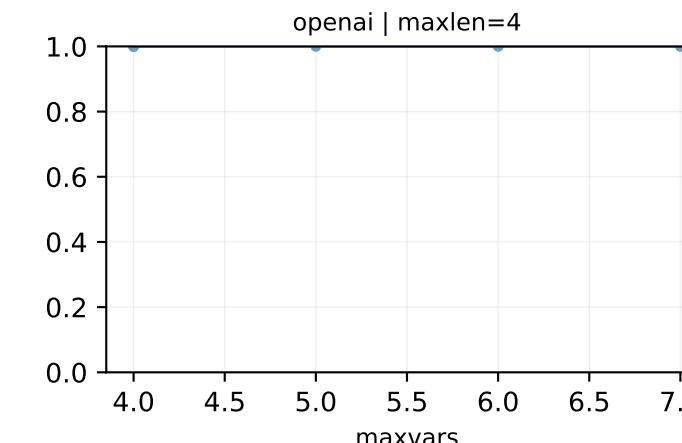
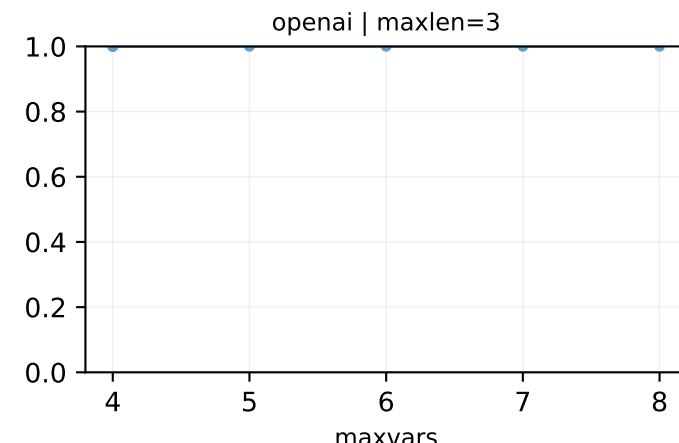
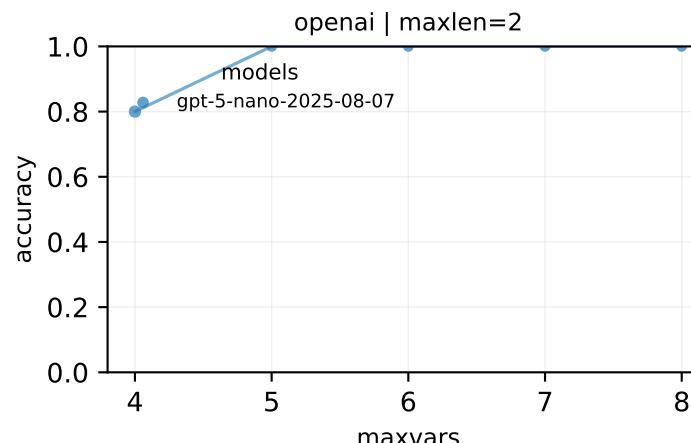
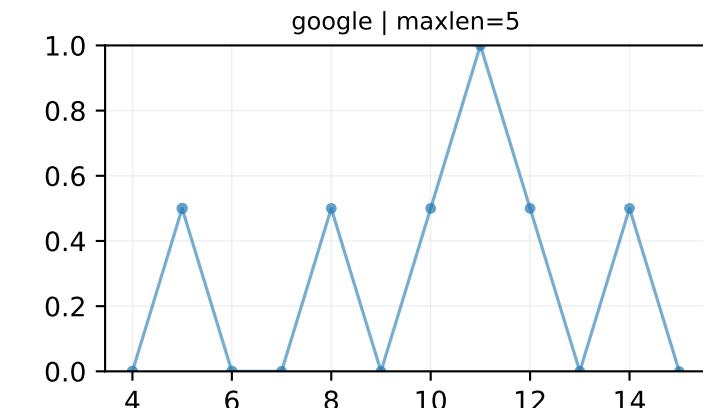
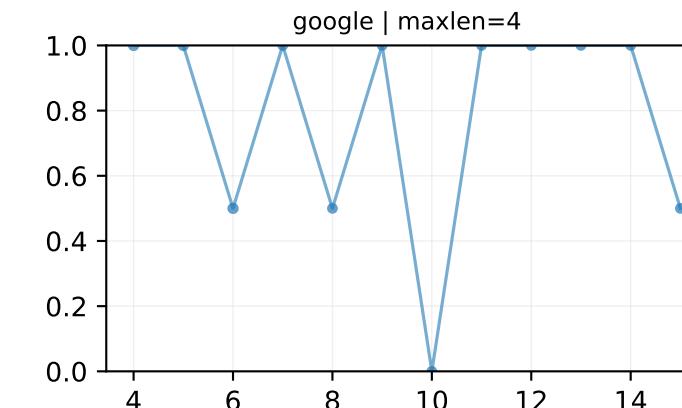
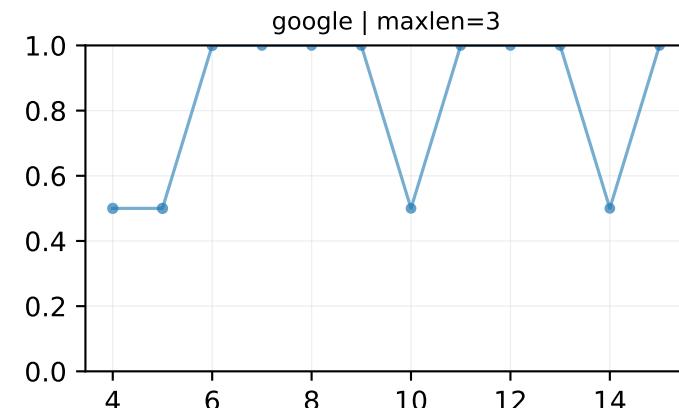
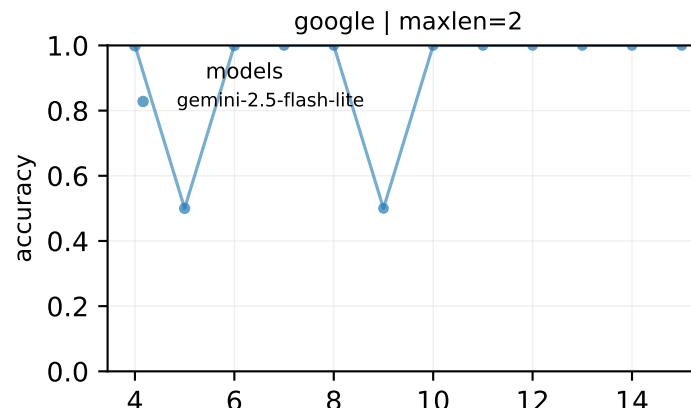
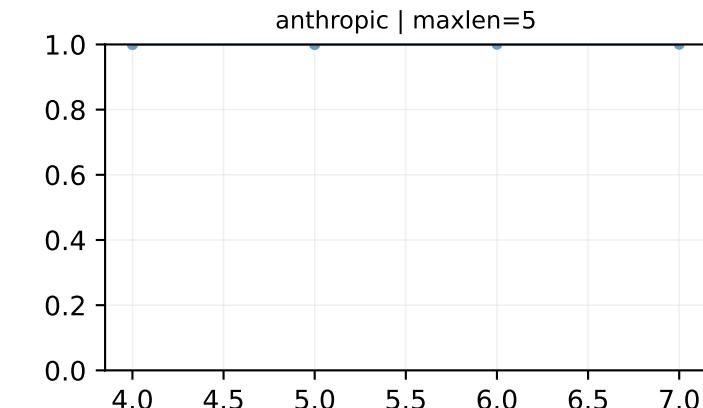
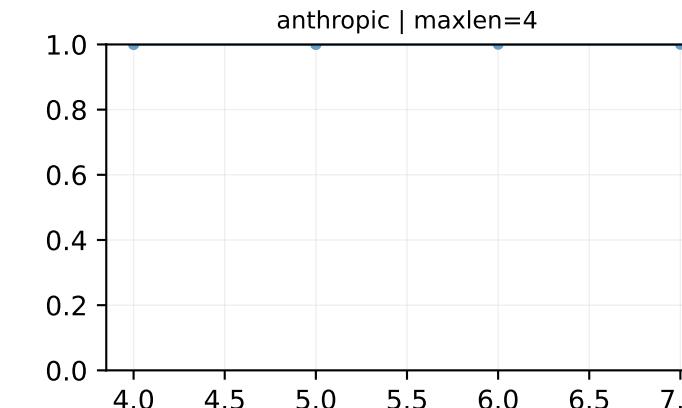
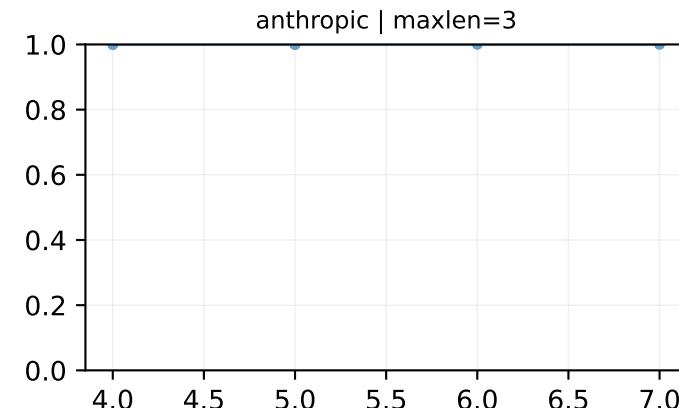
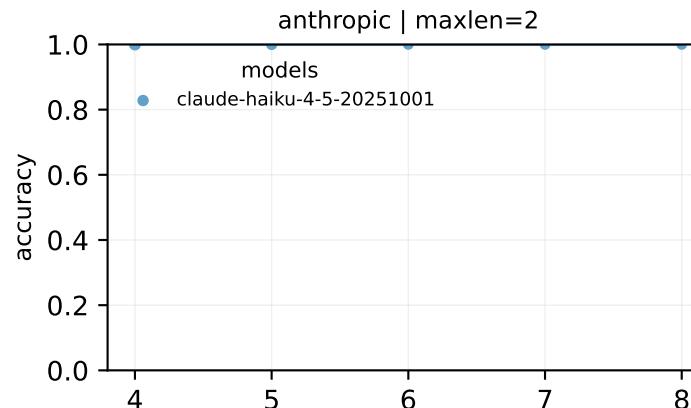


Condensed: family=horn_unified_answer | thinking=nothink | horn=1 | metric=sat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \sqrt{n}$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

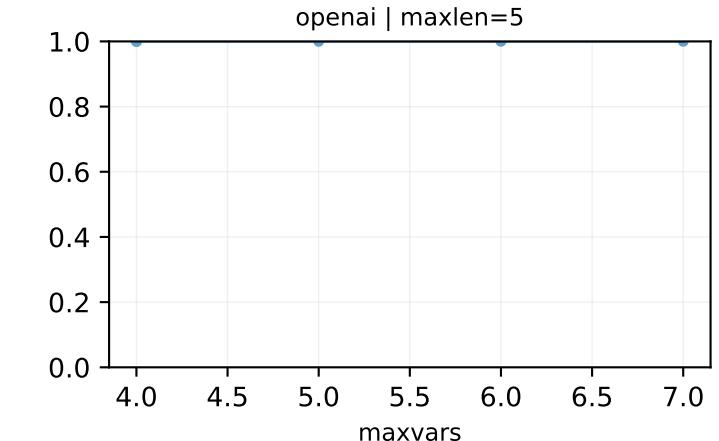
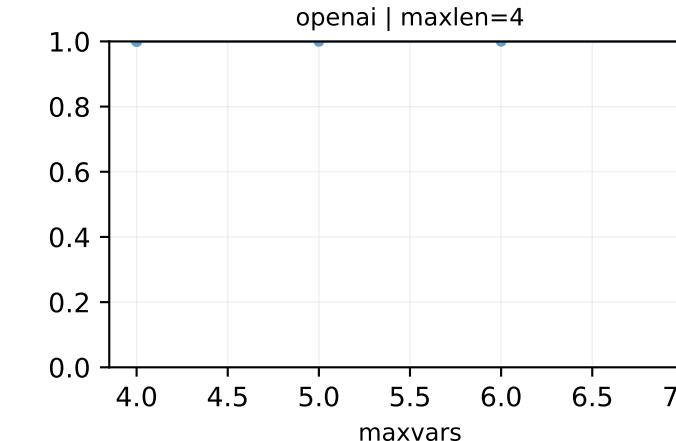
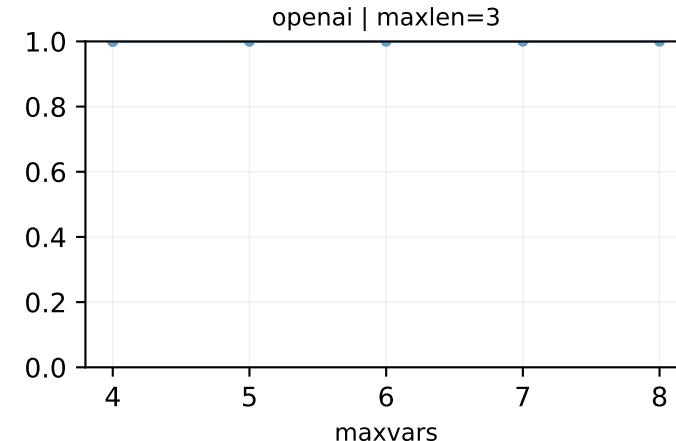
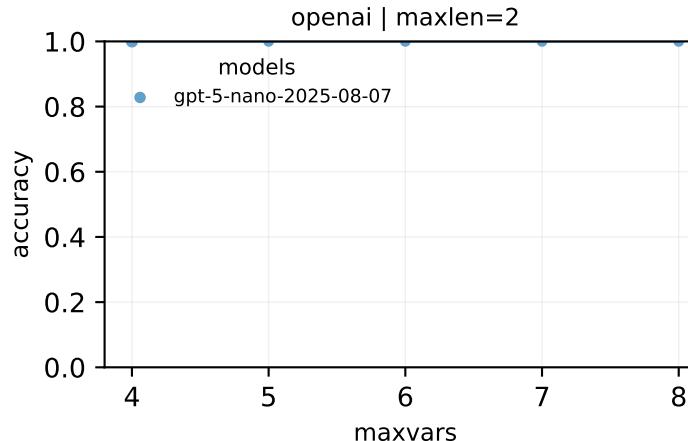
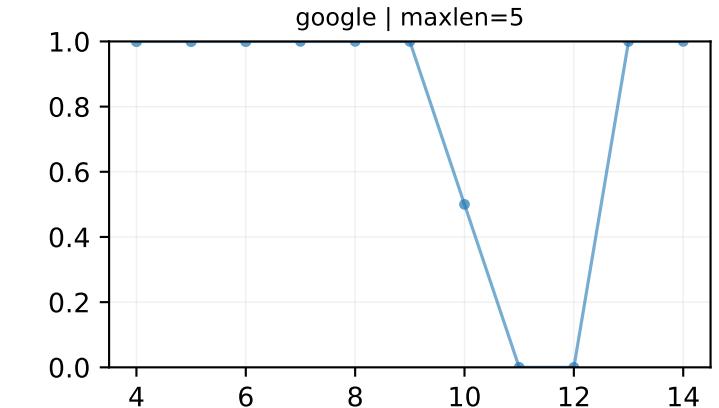
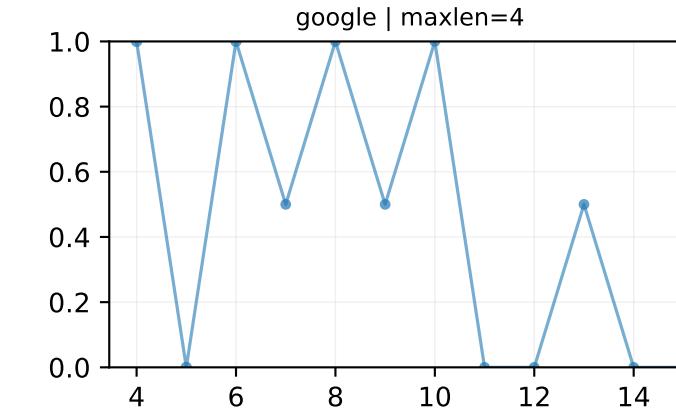
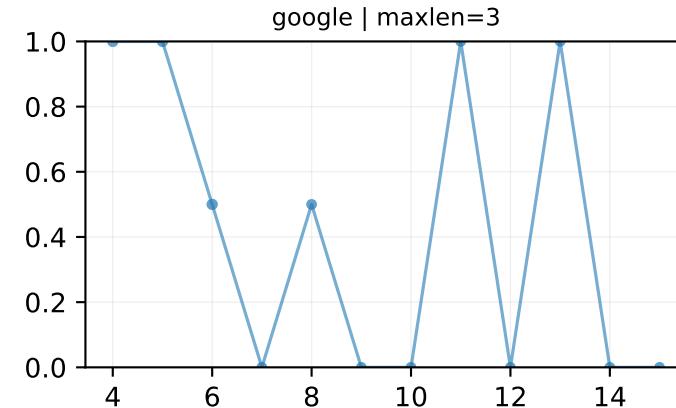
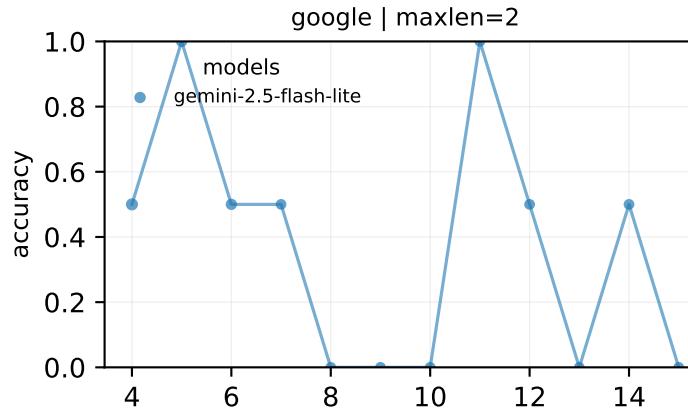
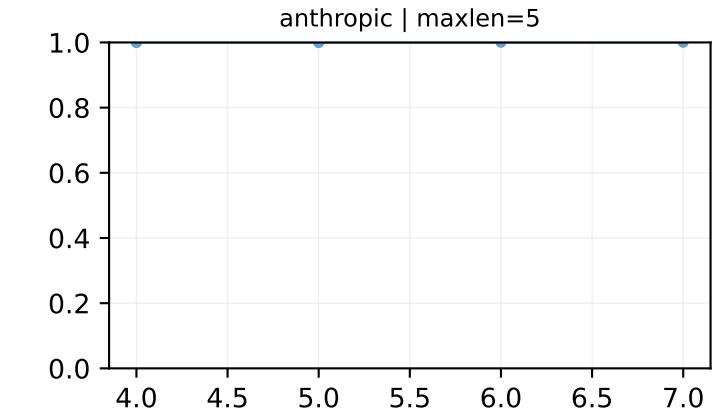
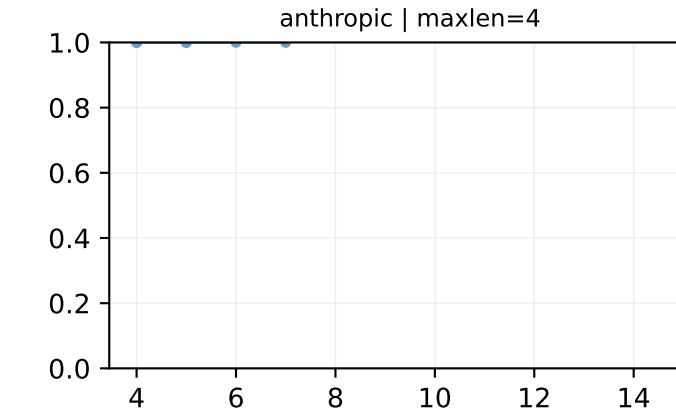
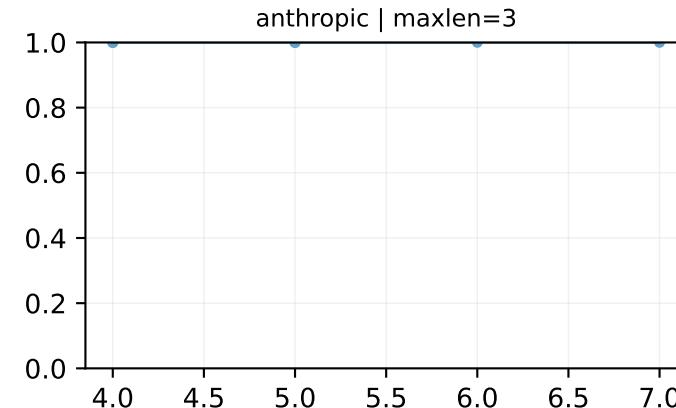
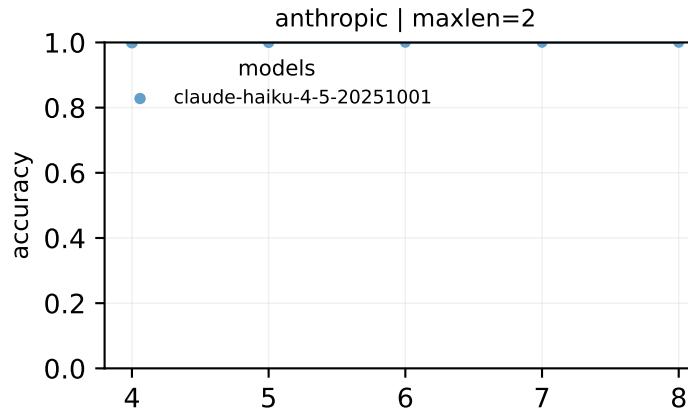


Condensed: family=horn_unified_answer | thinking=nothink | horn=1 | metric=unsat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \sqrt{n}$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

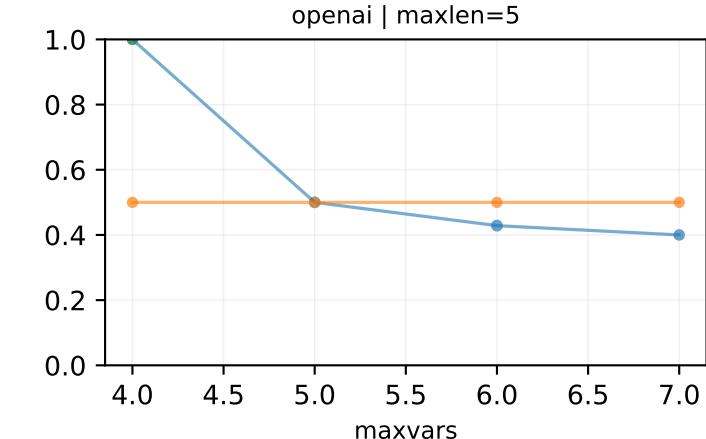
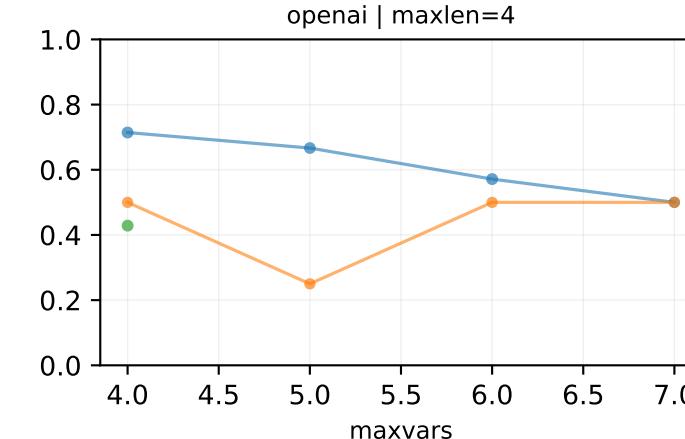
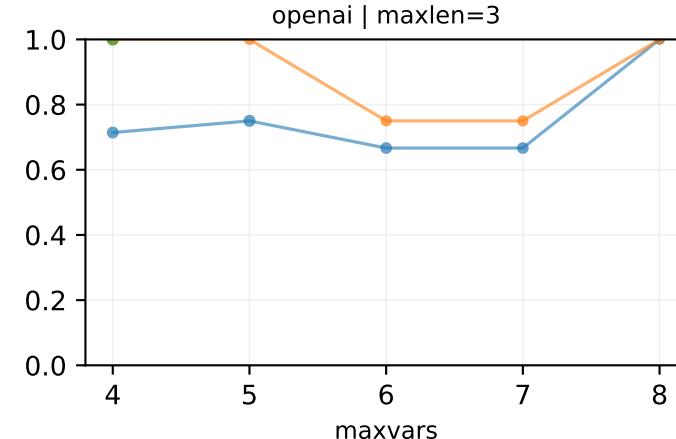
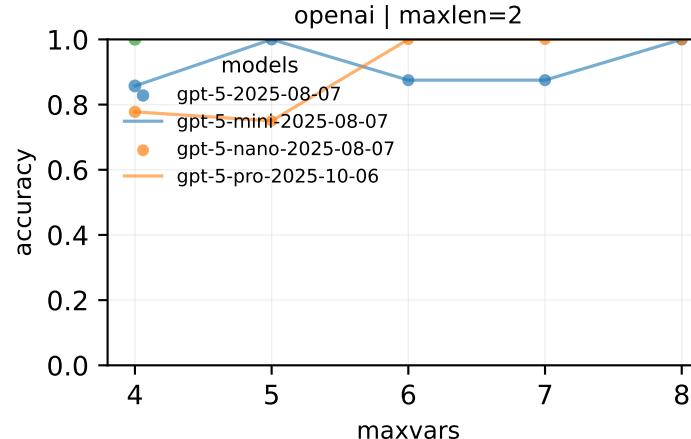
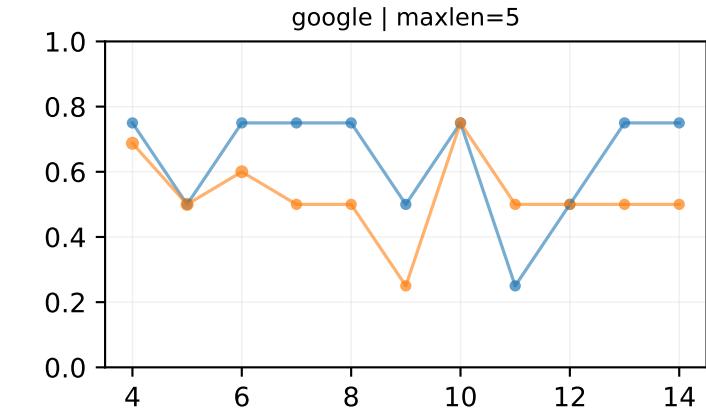
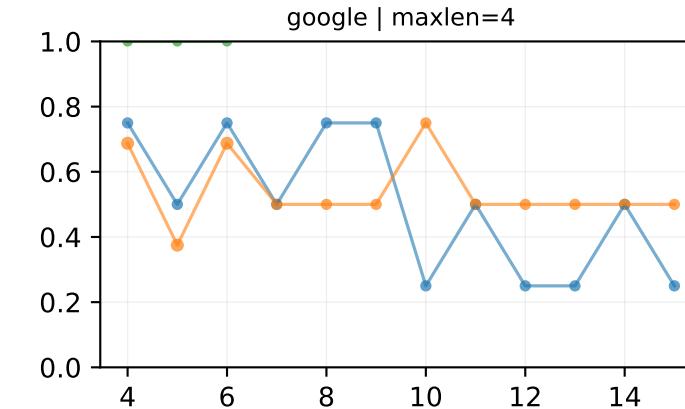
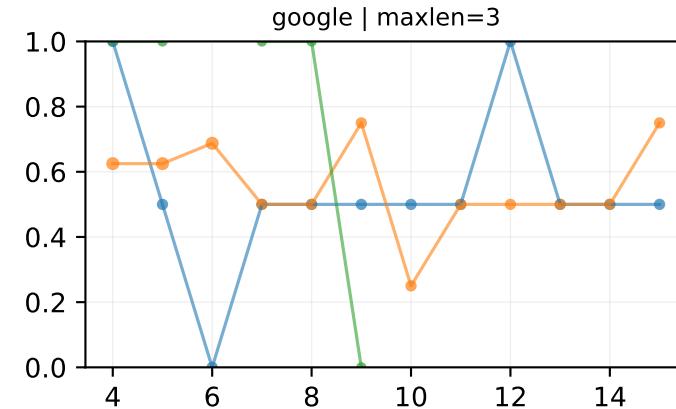
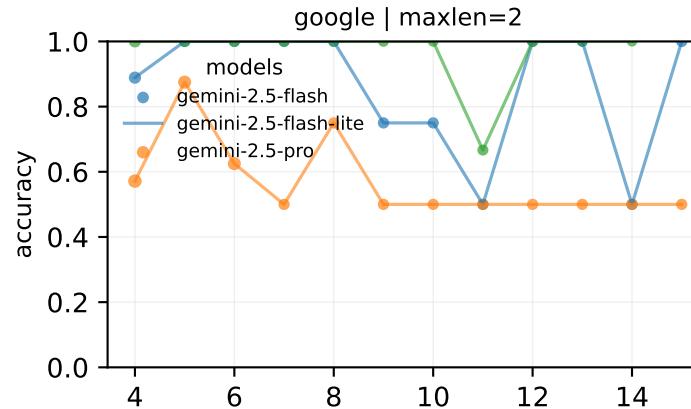
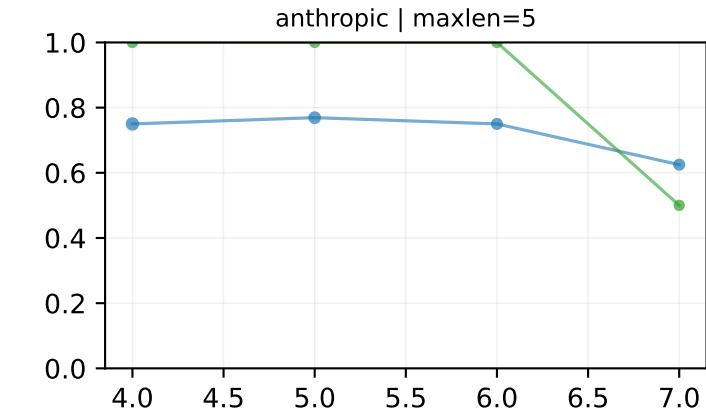
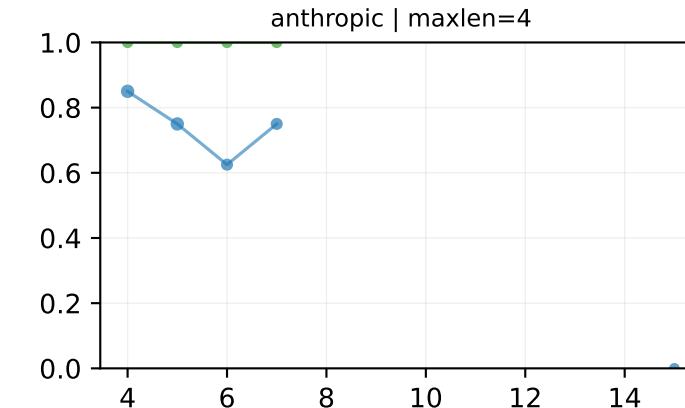
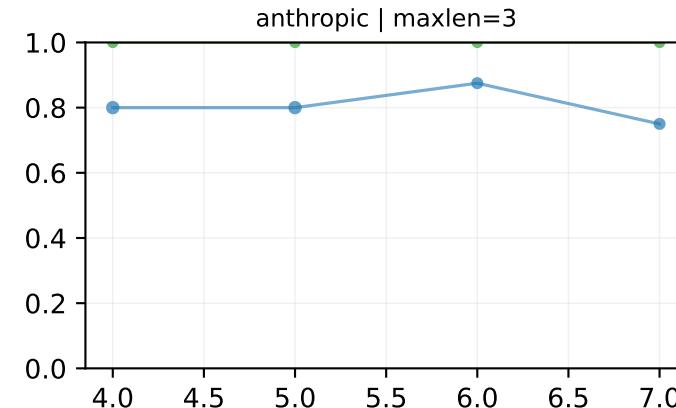
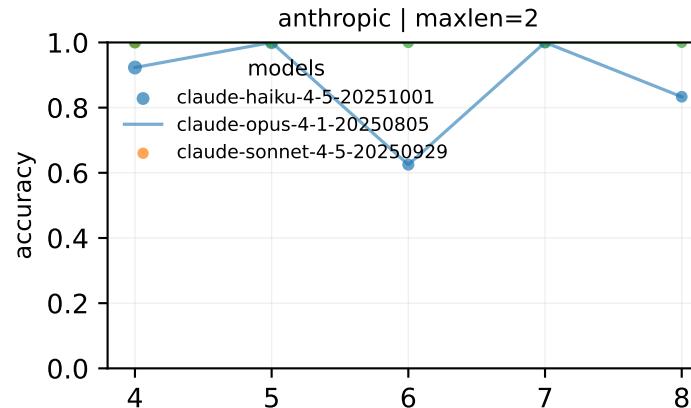


Condensed: family=horn_unified_answer | thinking=thinking | horn=0 | metric=accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \text{sqrt}(n)$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

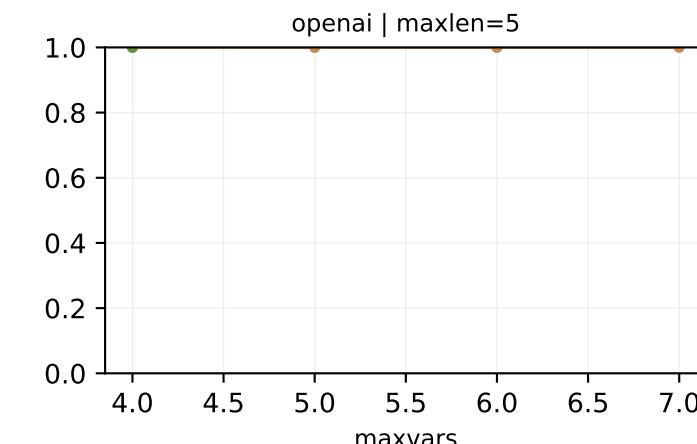
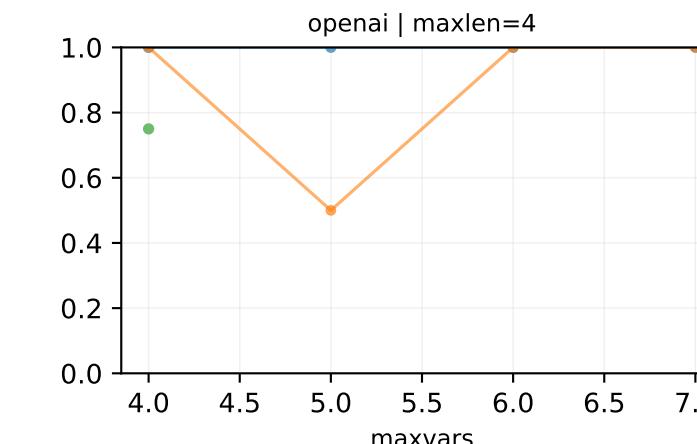
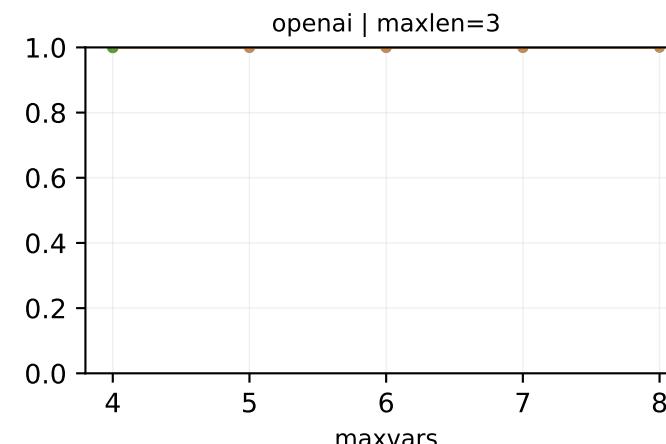
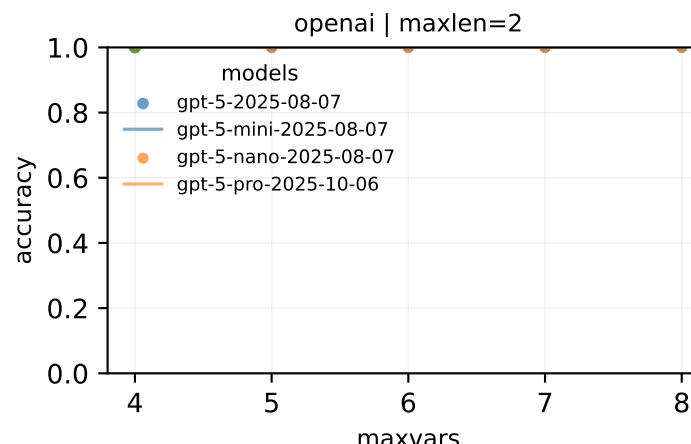
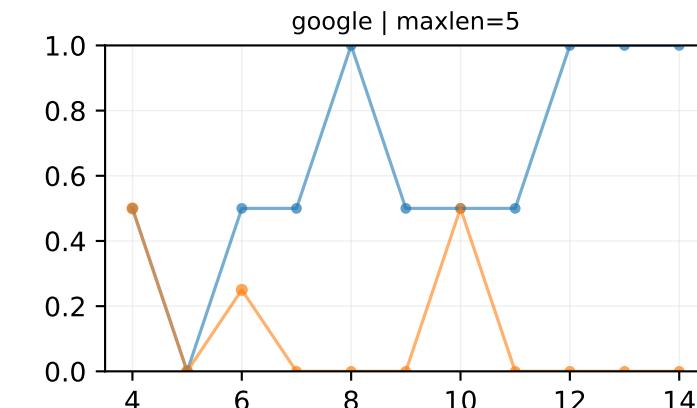
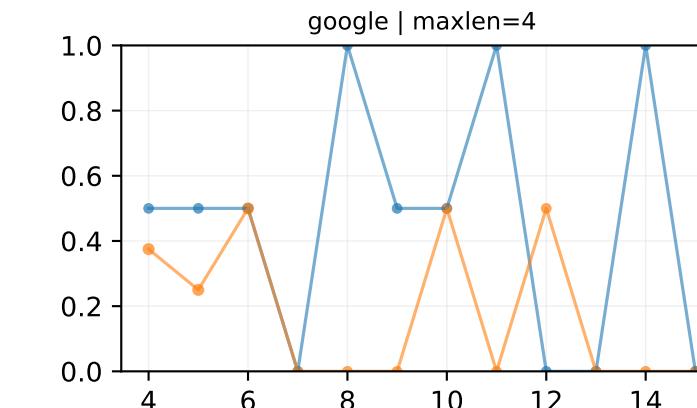
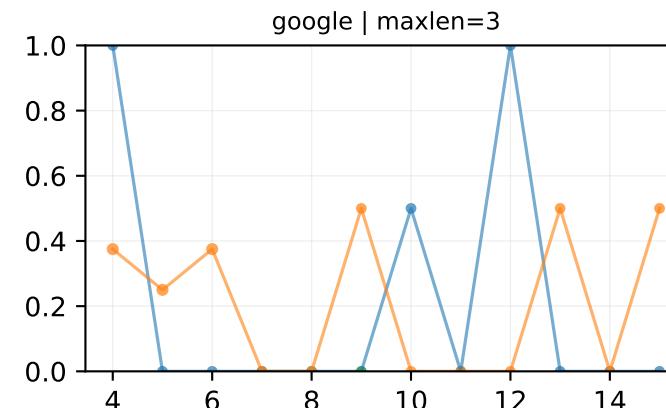
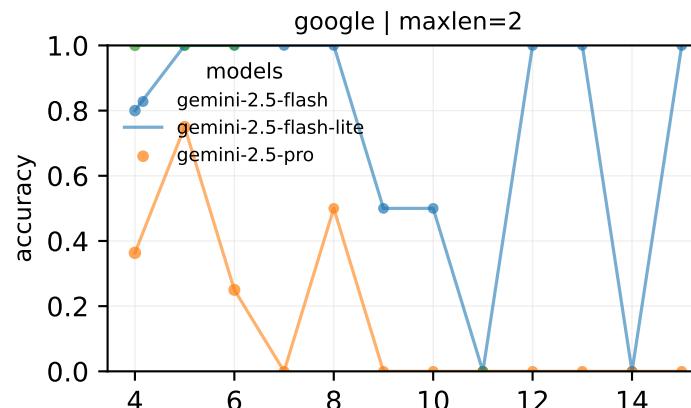
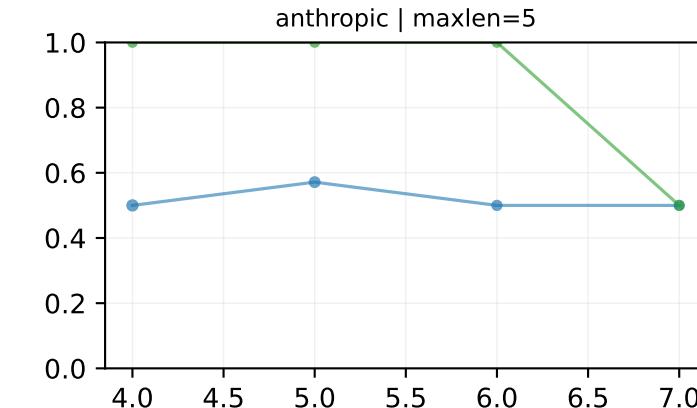
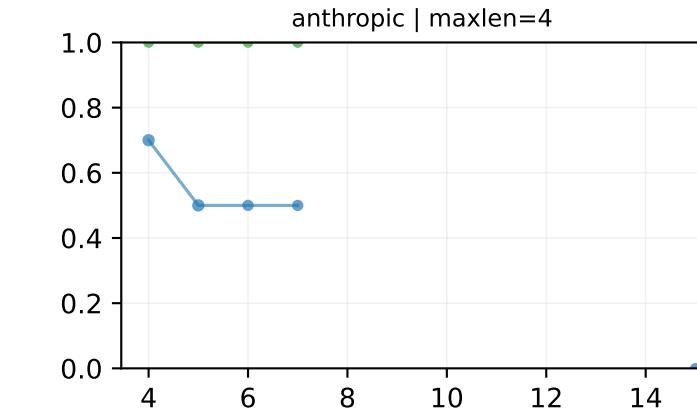
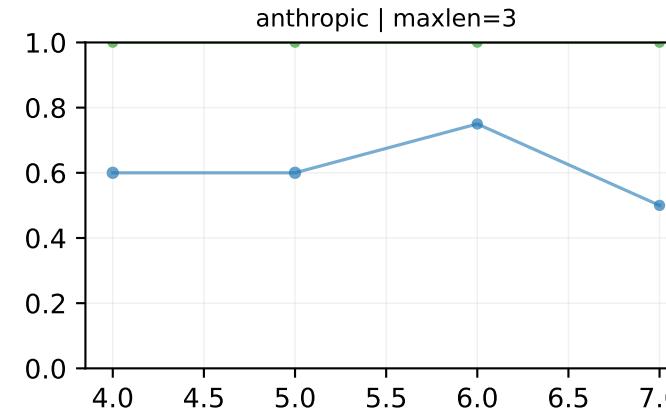
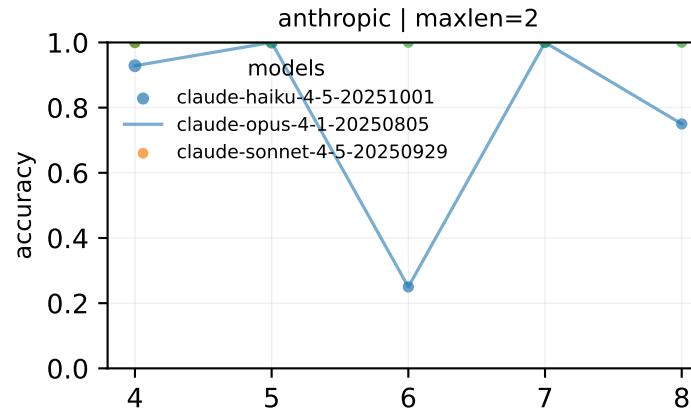


Condensed: family=horn_unified_answer | thinking=thinking | horn=0 | metric=sat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \text{sqrt}(n)$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

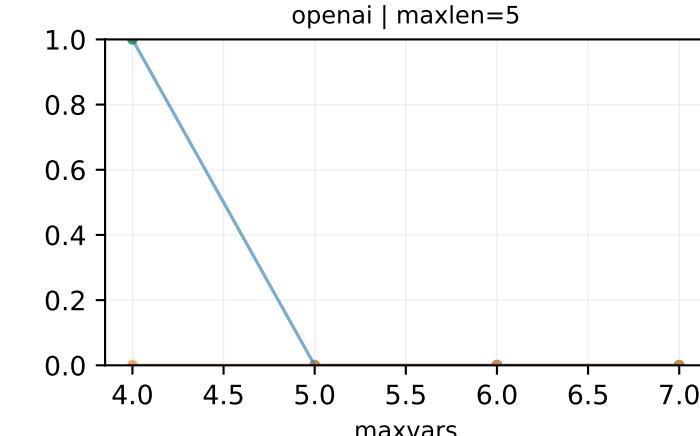
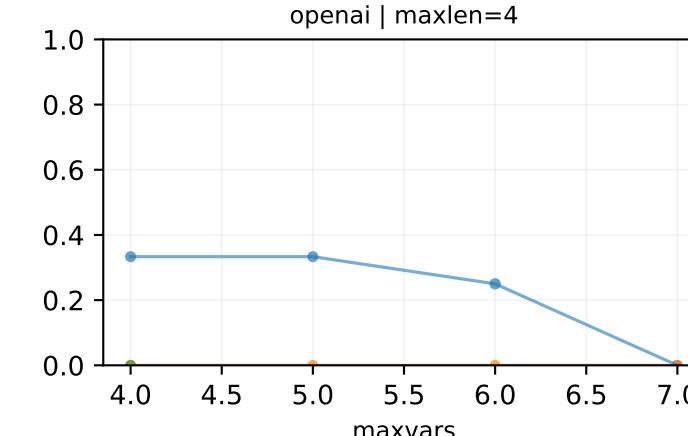
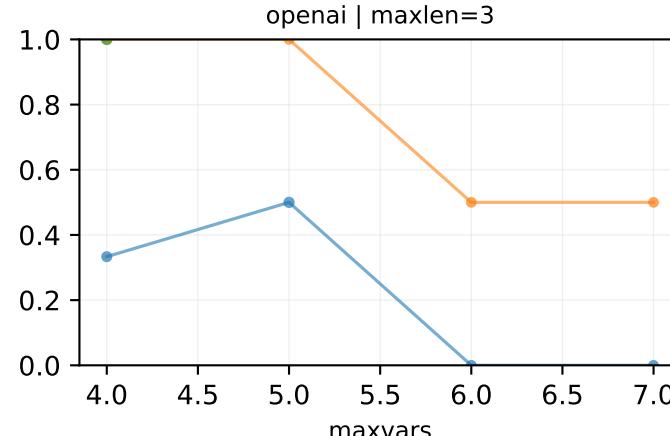
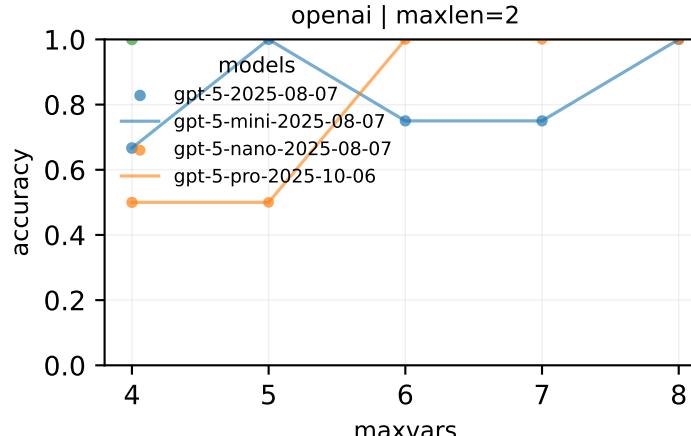
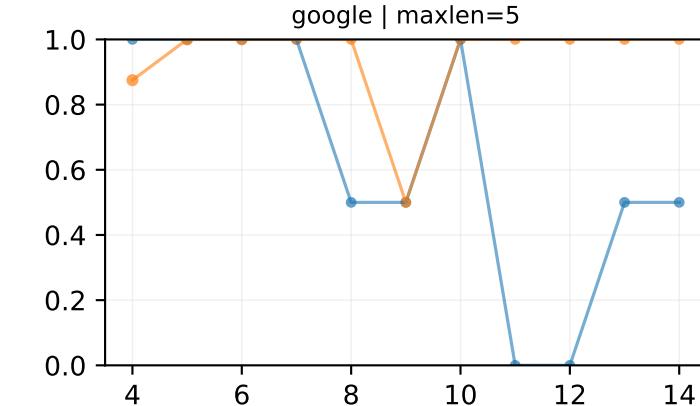
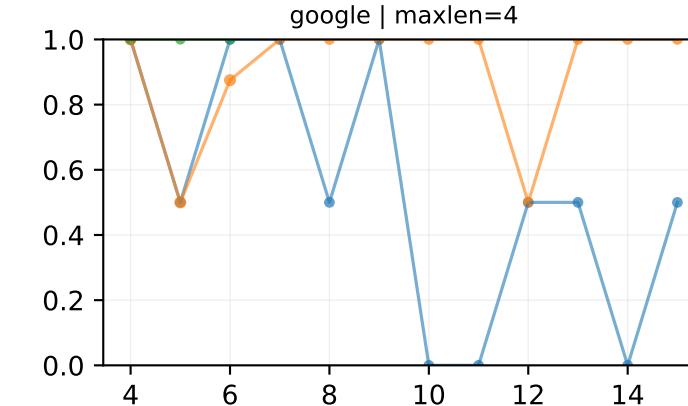
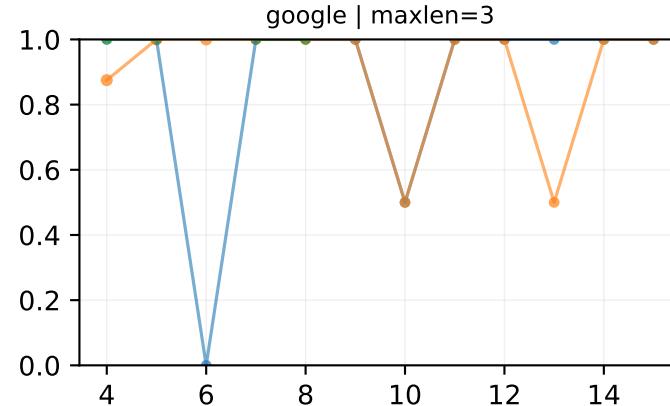
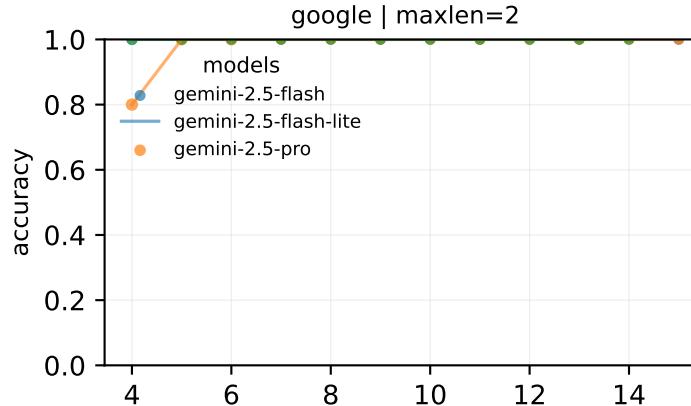
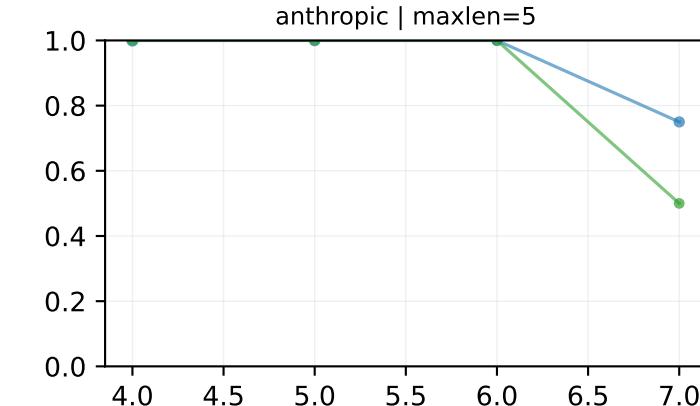
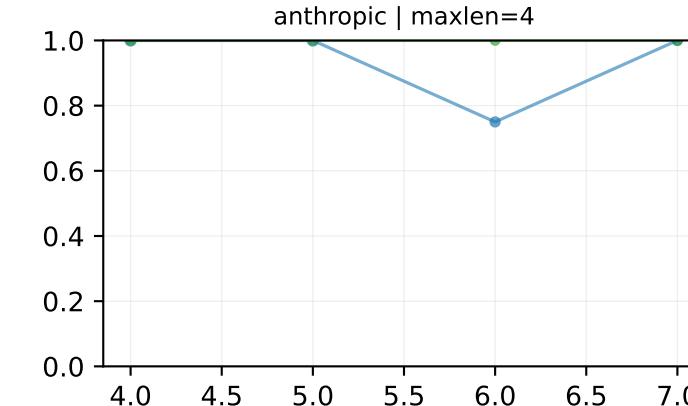
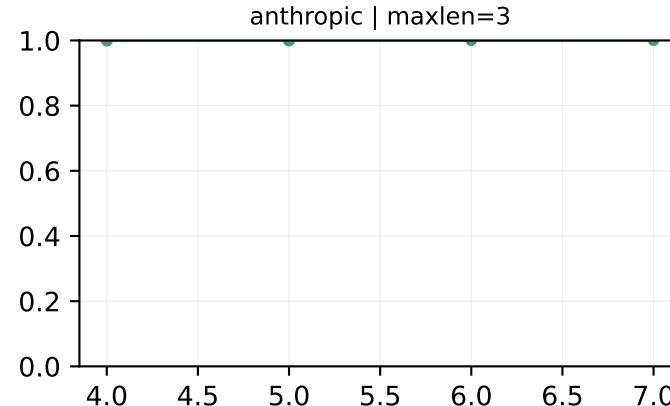
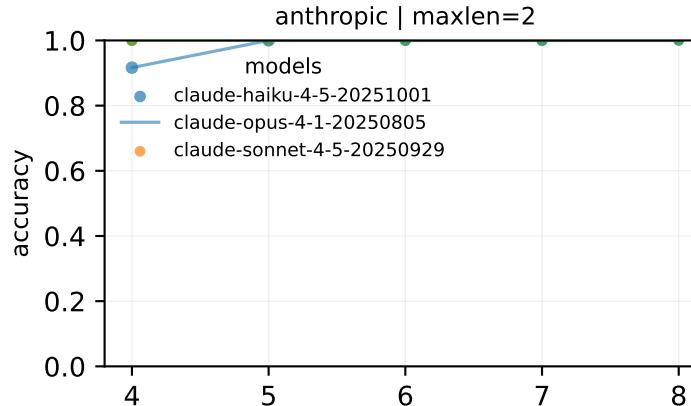


Condensed: family=horn_unified_answer | thinking=thinking | horn=0 | metric=unsat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \text{sqrt}(n)$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

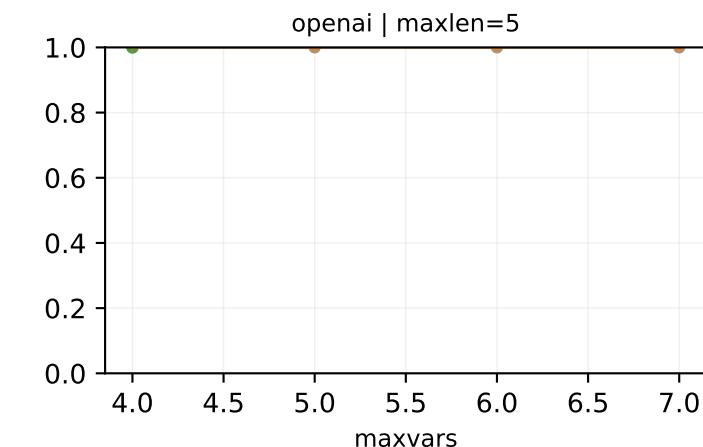
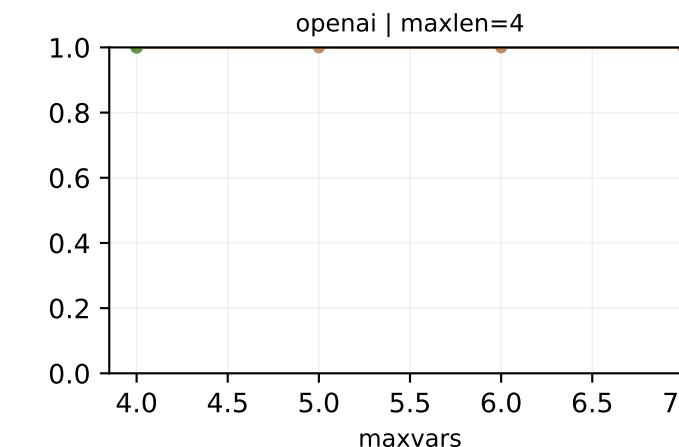
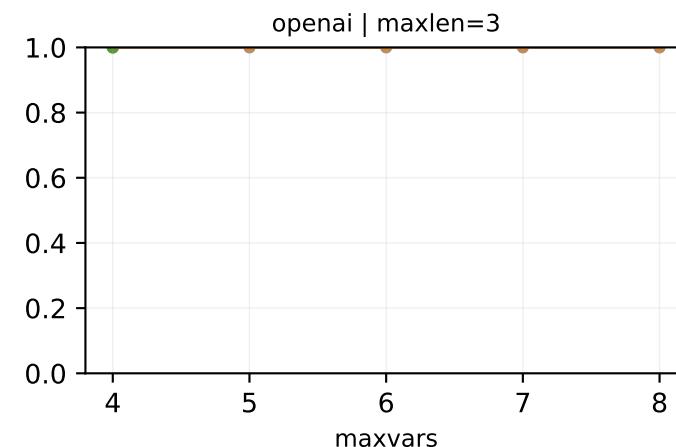
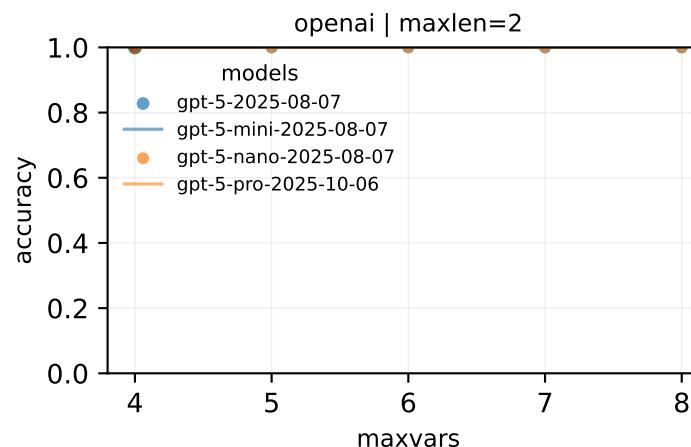
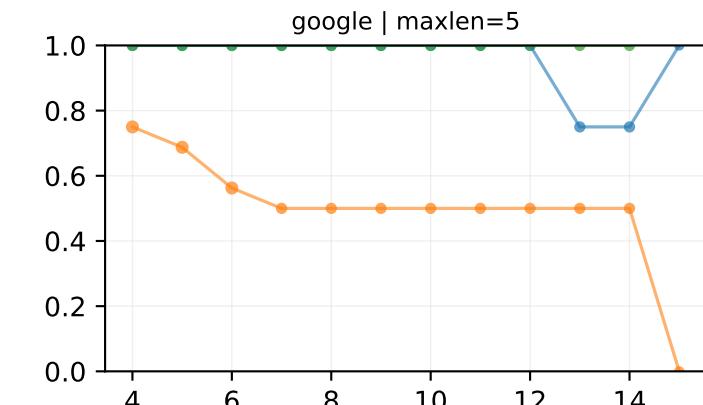
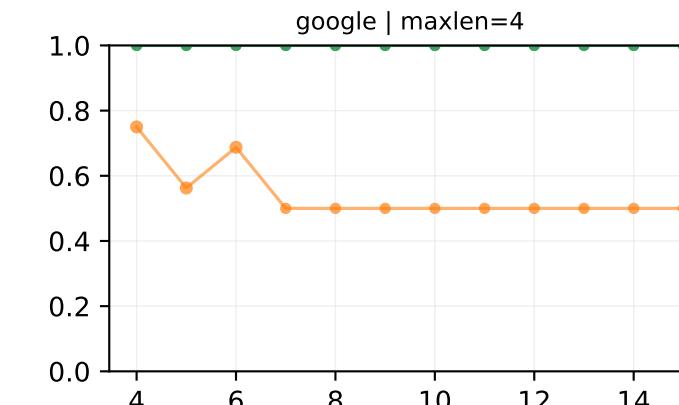
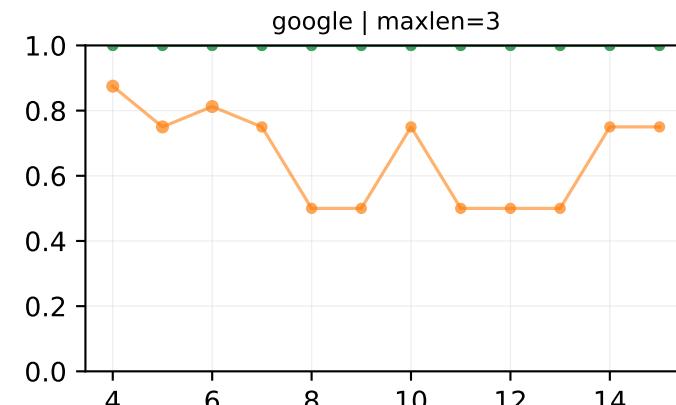
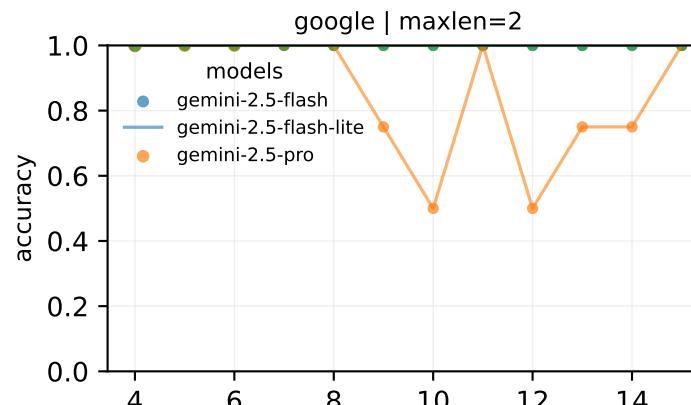
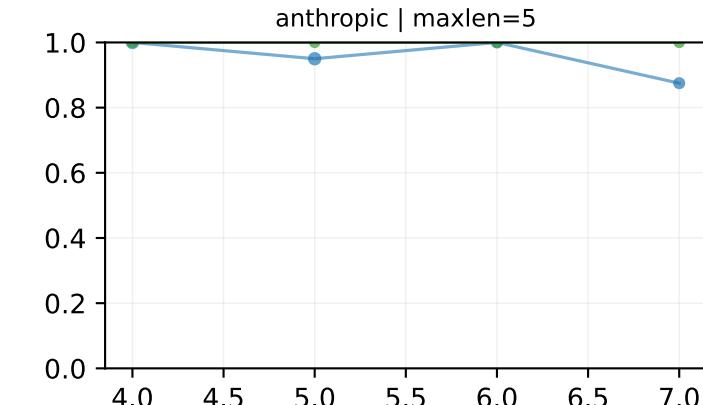
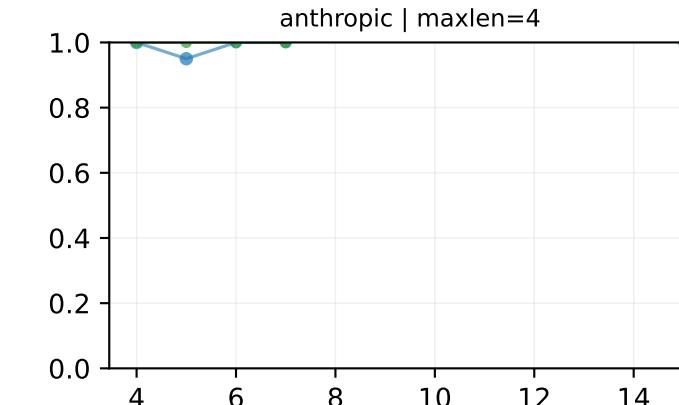
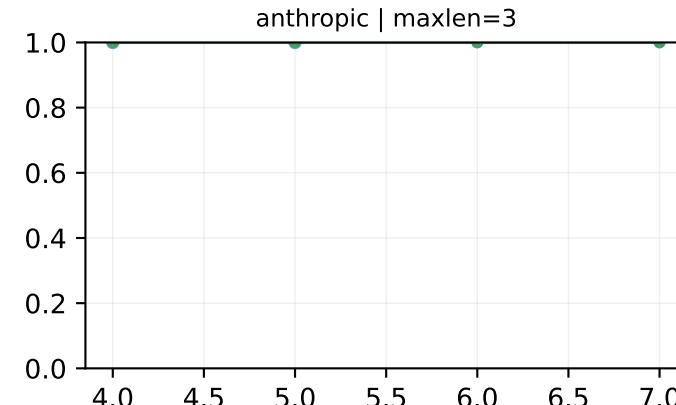
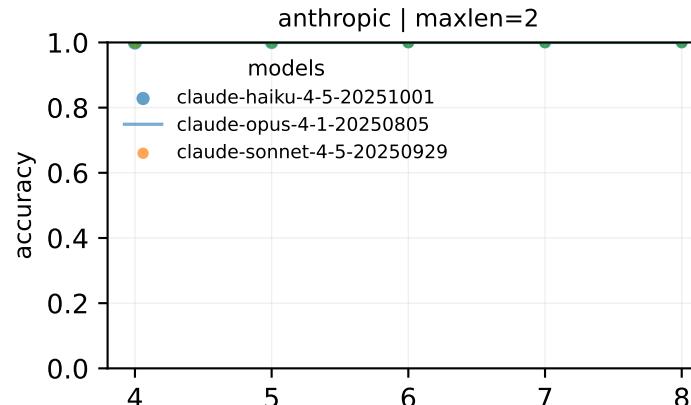


Condensed: family=horn_unified_answer | thinking=thinking | horn=1 | metric=accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \text{sqrt}(n)$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

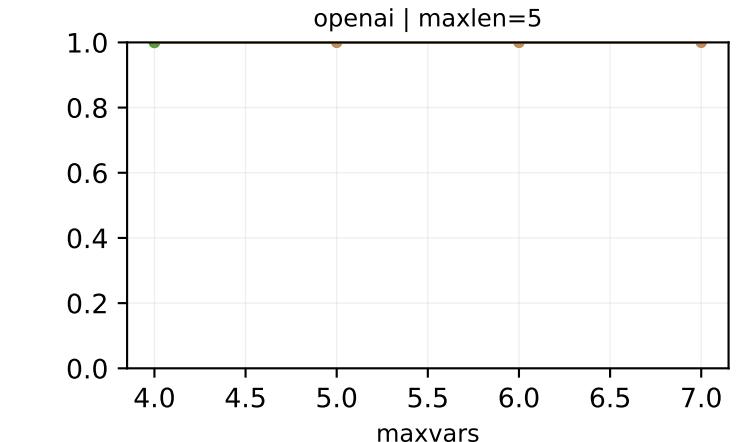
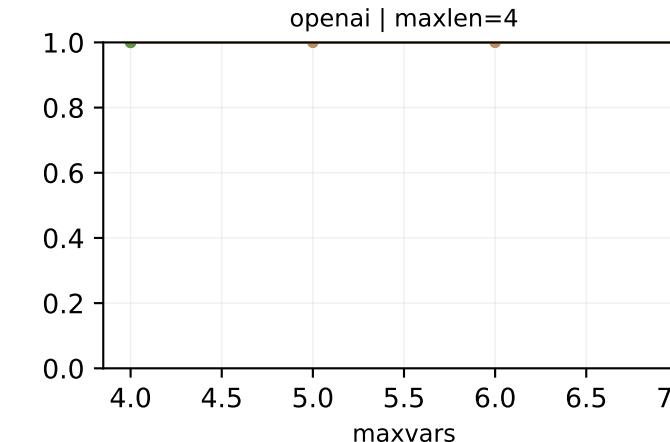
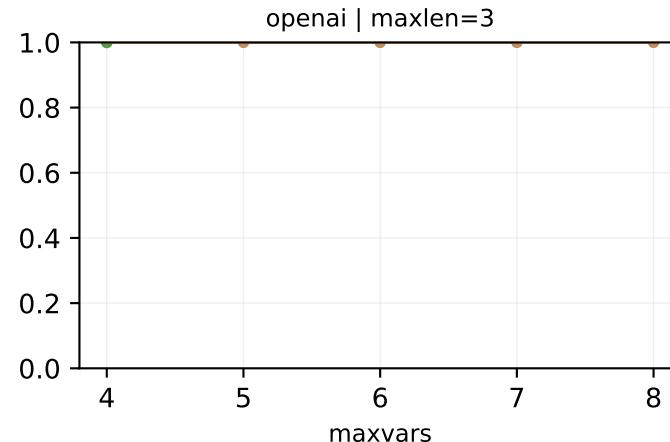
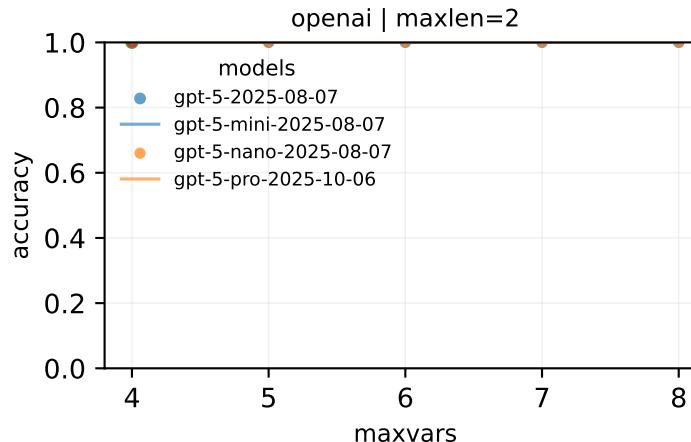
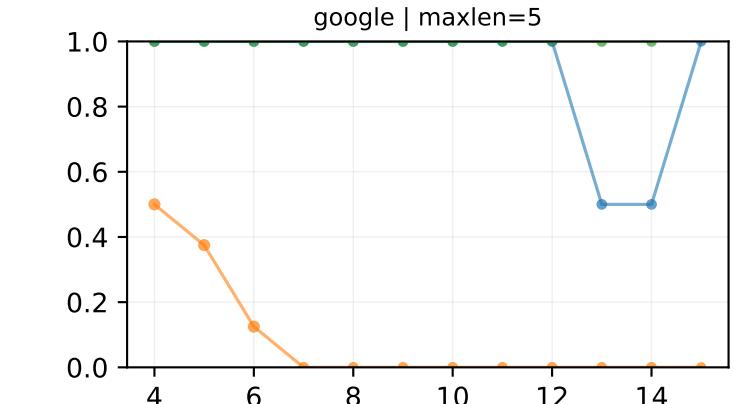
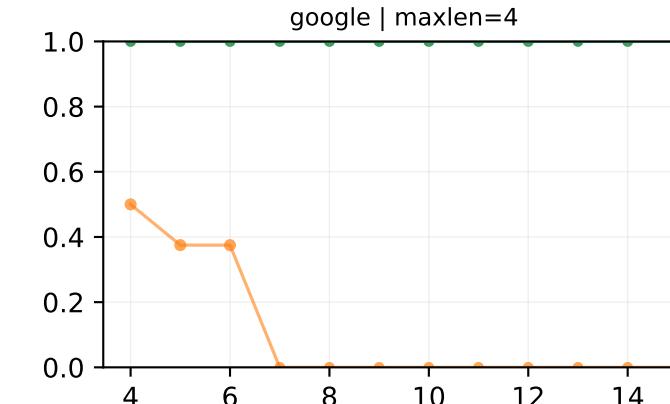
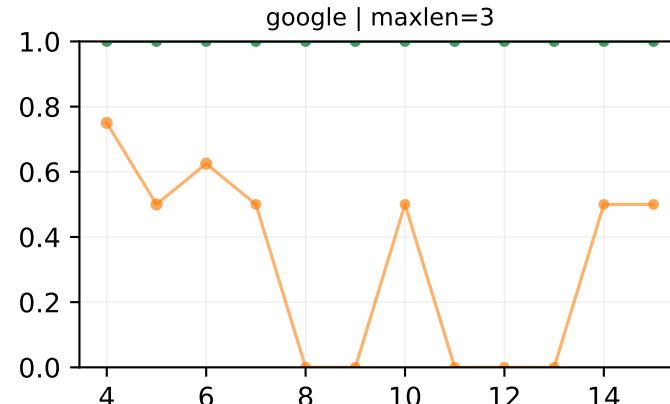
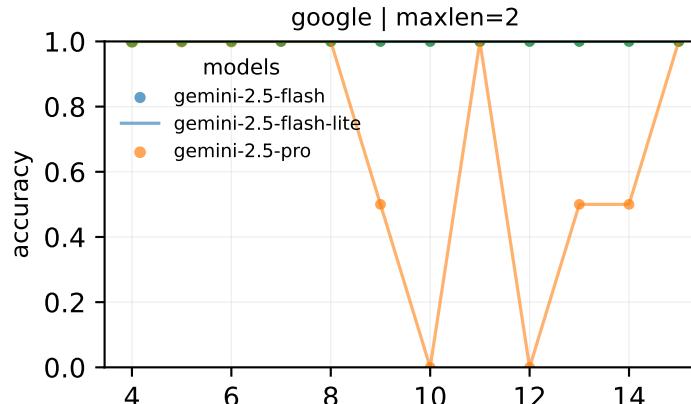
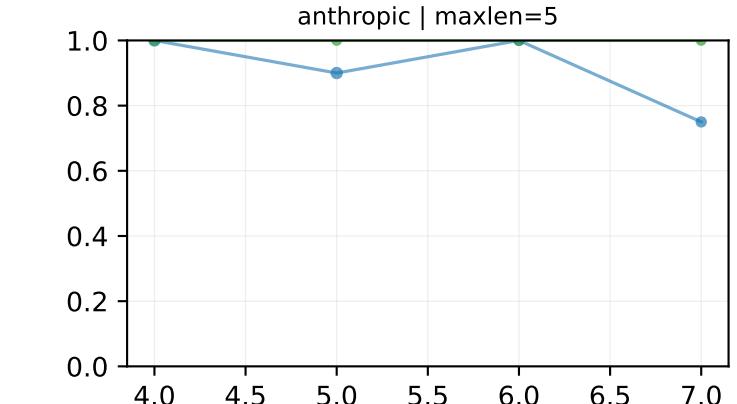
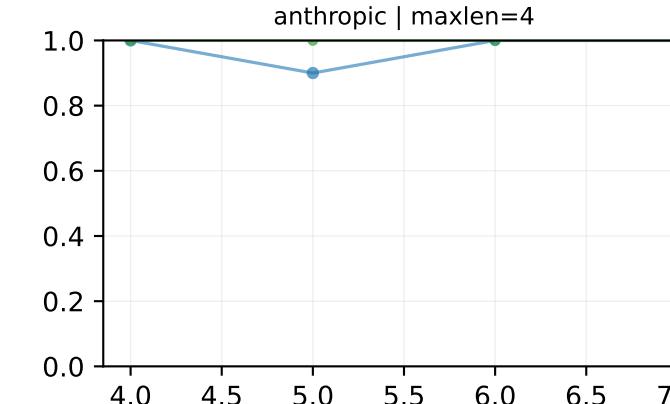
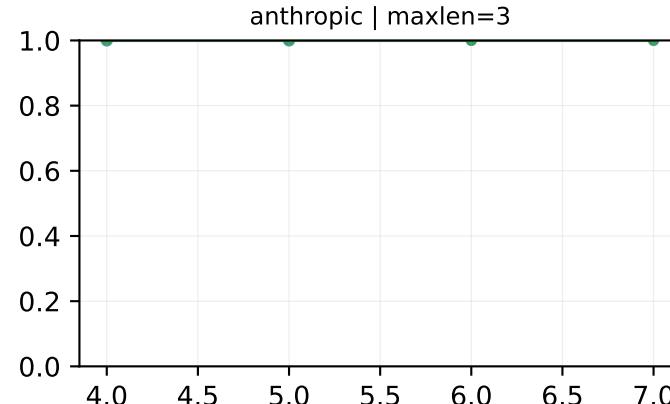
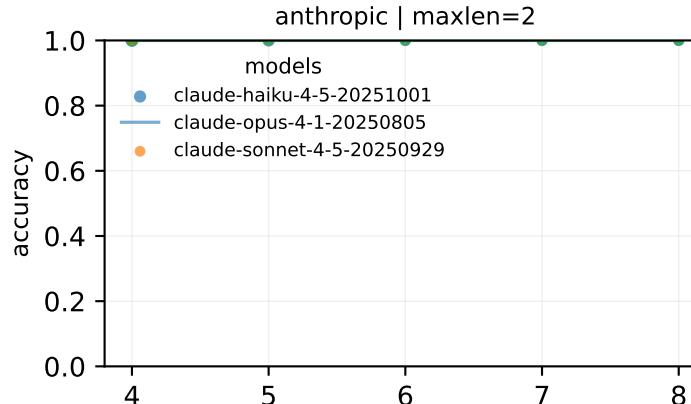


Condensed: family=horn_unified_answer | thinking=thinking | horn=1 | metric=sat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \text{sqrt}(n)$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md



Condensed: family=horn_unified_answer | thinking=thinking | horn=1 | metric=unsat_accuracy

Line plots: x=maxvars, y=accuracy. Marker size $\sim \text{sqrt}(n)$ for the metric denominator. Line segments connect only consecutive maxvars values.

Aggregated across all runs (provider-error rows excluded). Prompt IDs included: prompt_2376d1fca7, prompt_62ba908560

Exact prompts: experiments/paper_outputs/prompts/prompt_catalog.md

