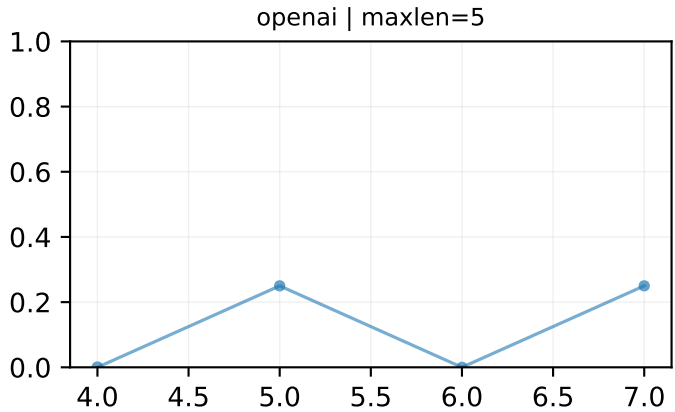
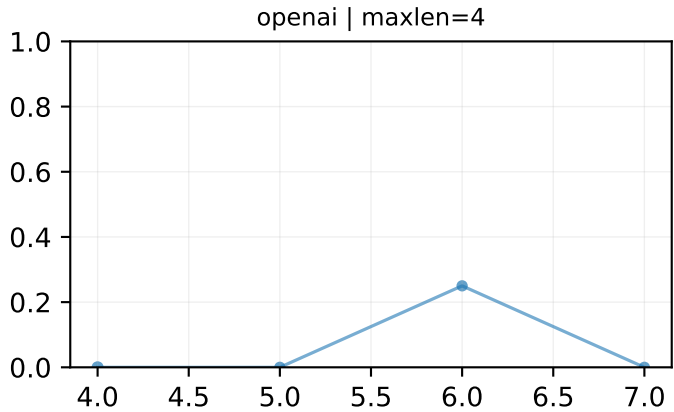
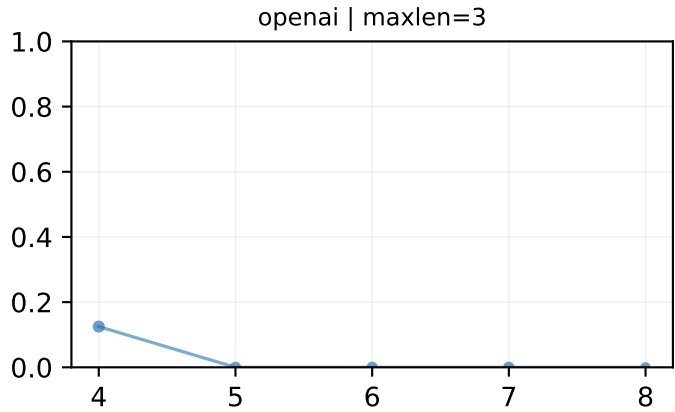
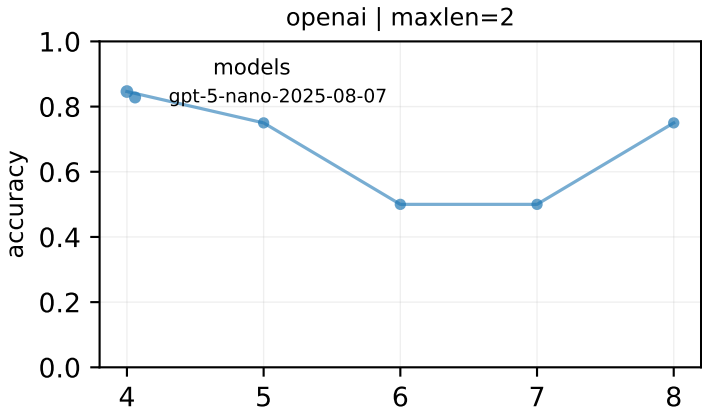
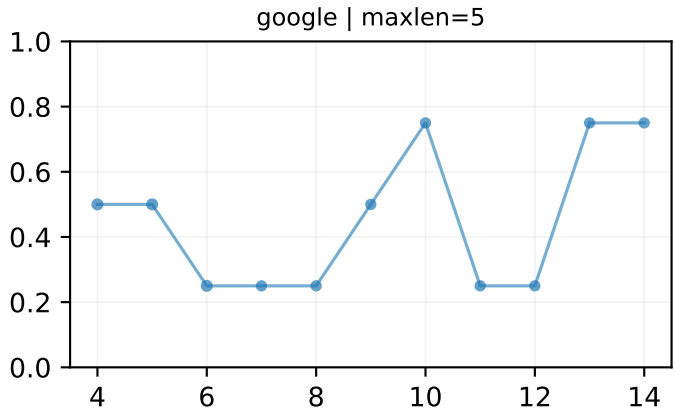
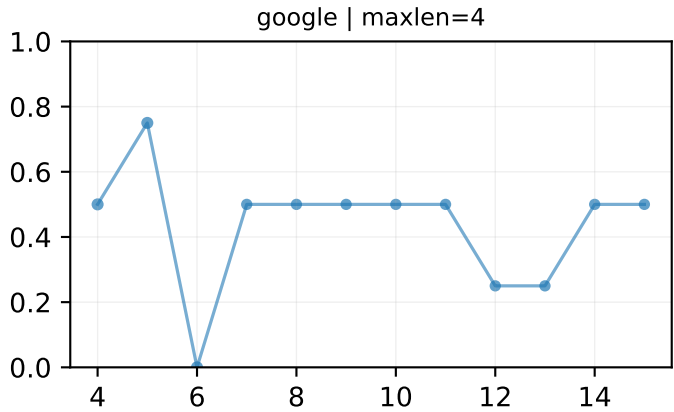
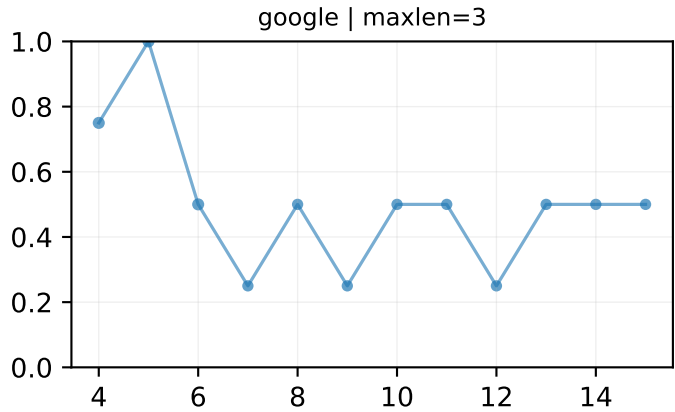
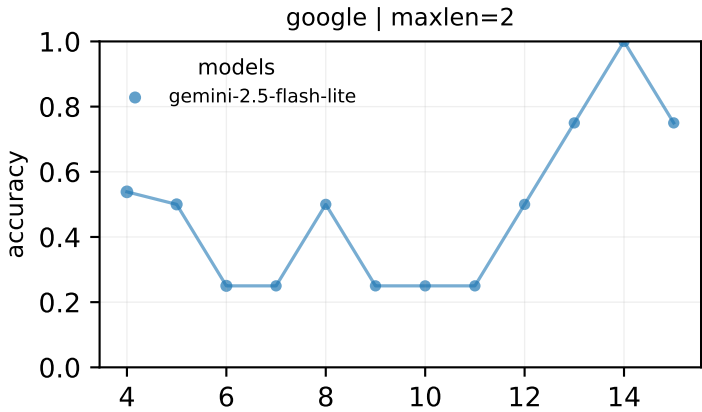
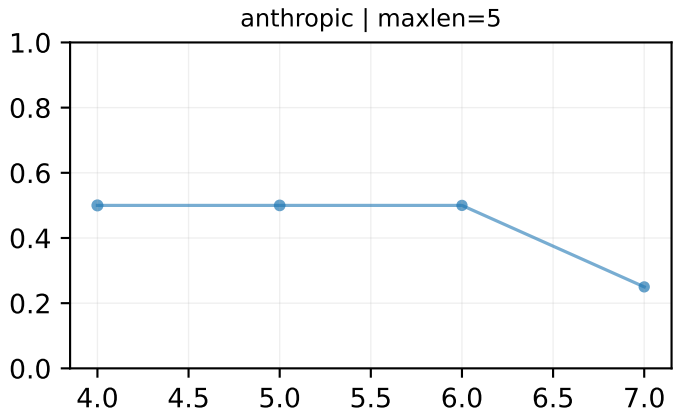
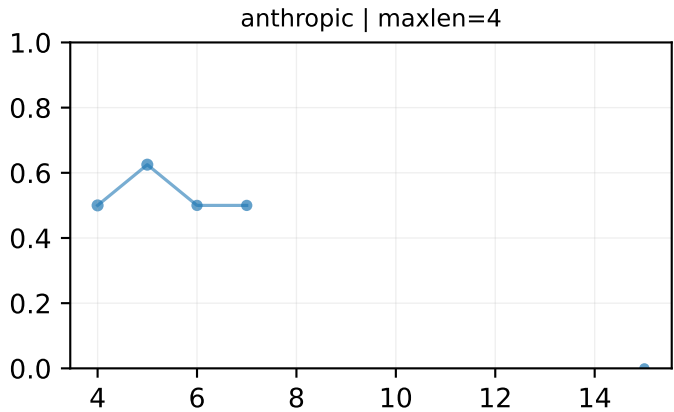
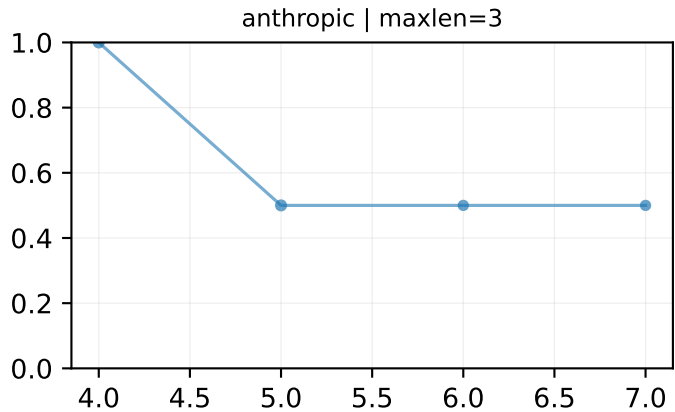
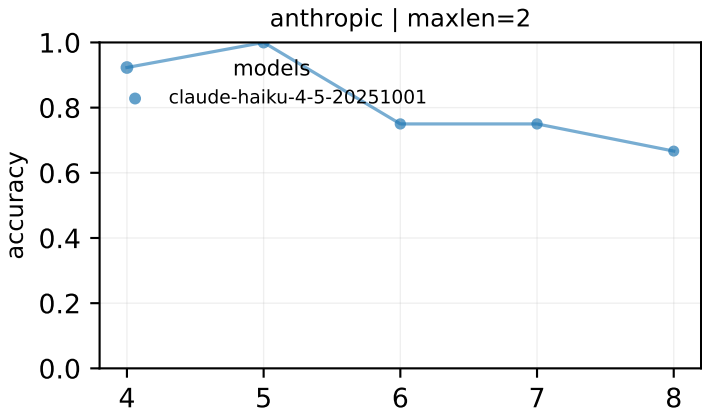


**Condensed: family=horn\_unified\_answer | thinking=nothink | horn=0 | metric=accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

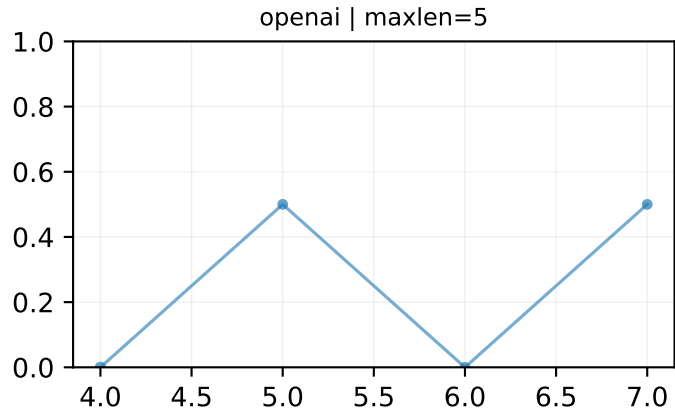
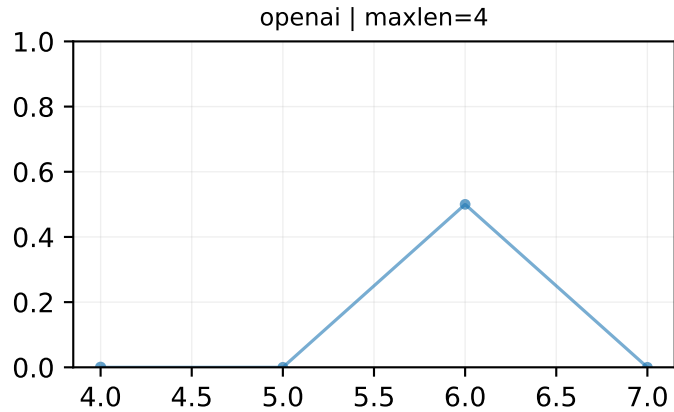
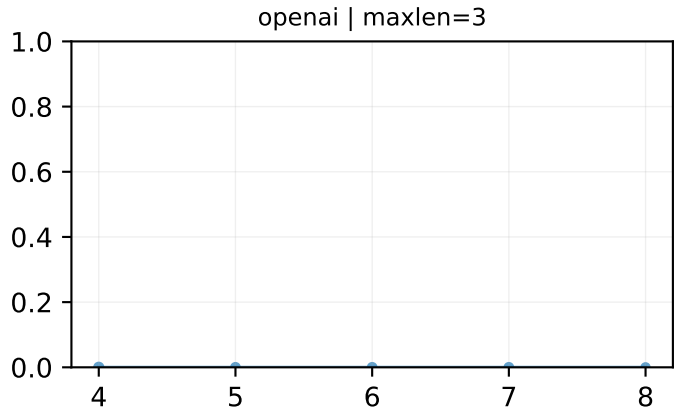
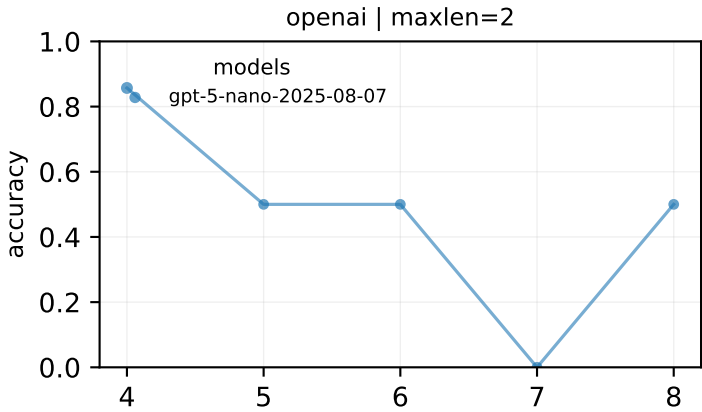
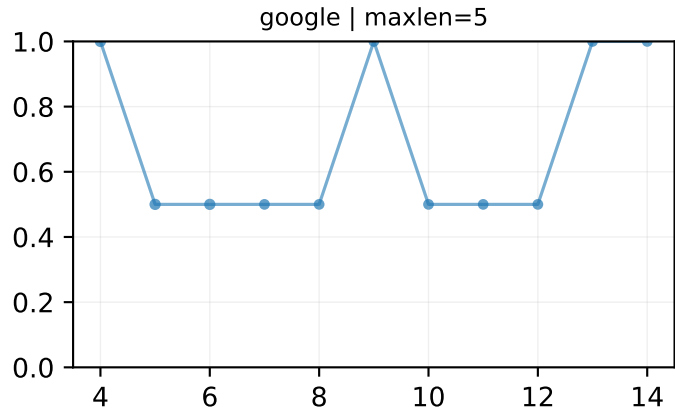
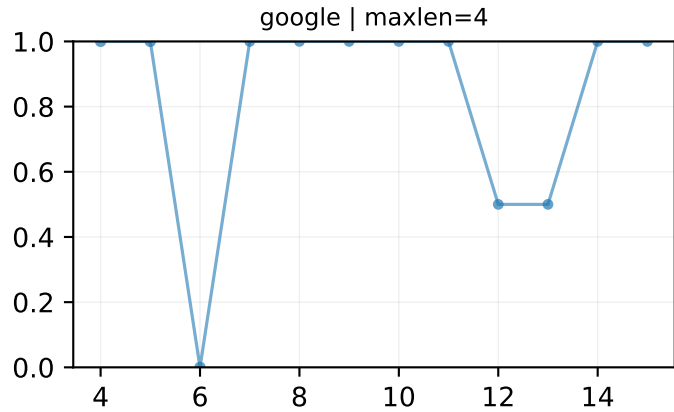
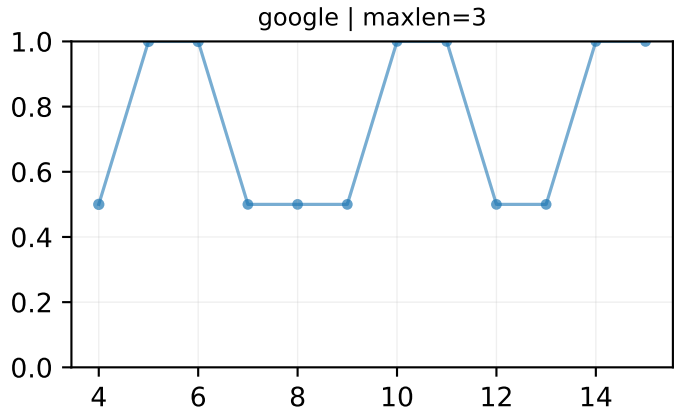
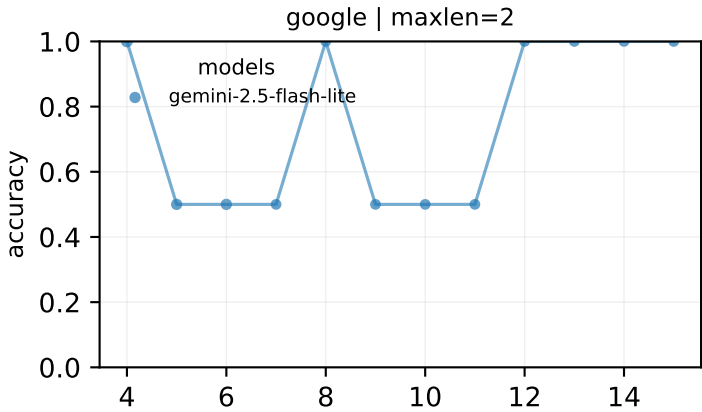
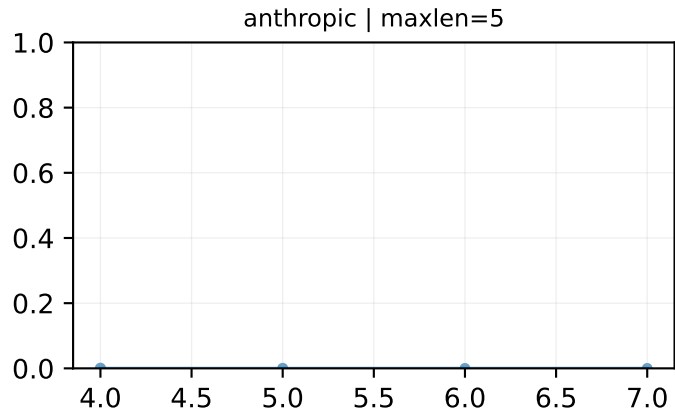
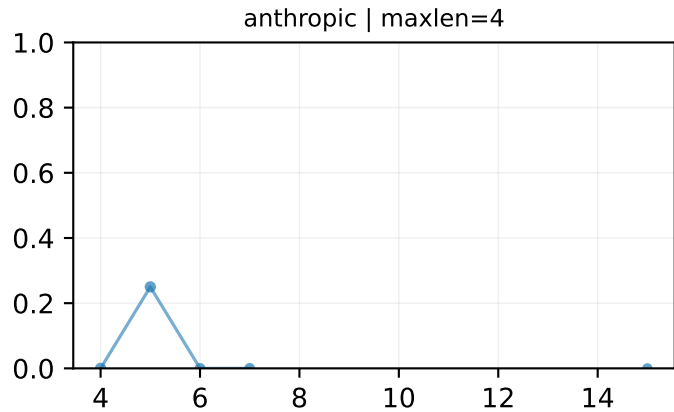
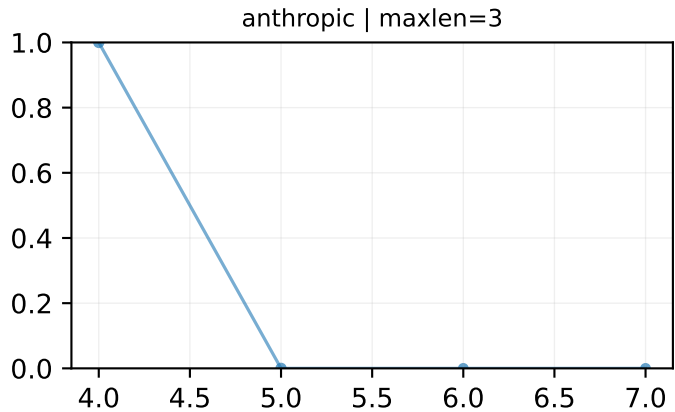
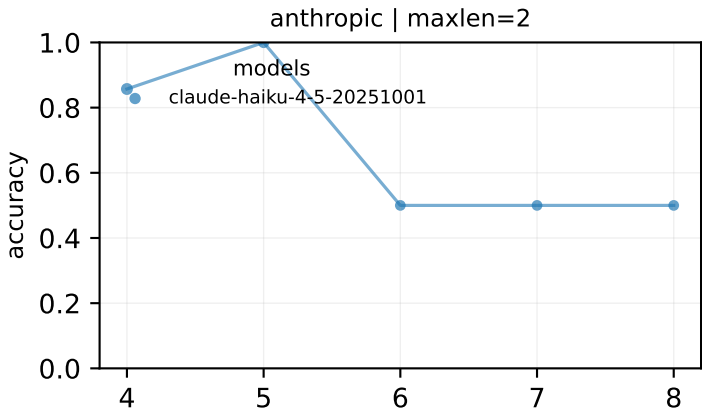


**Condensed: family=horn\_unified\_answer | thinking=nothink | horn=0 | metric=sat\_accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

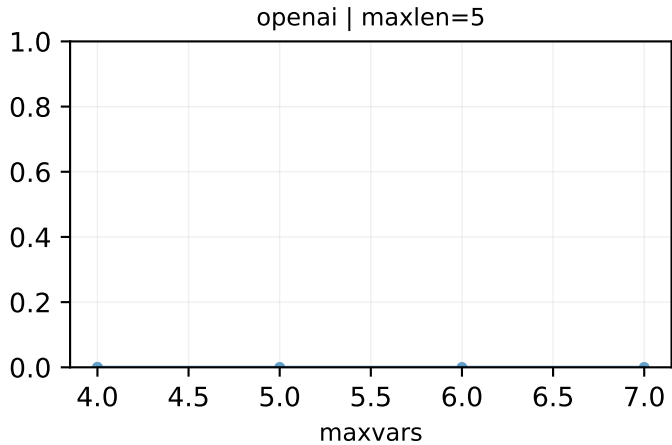
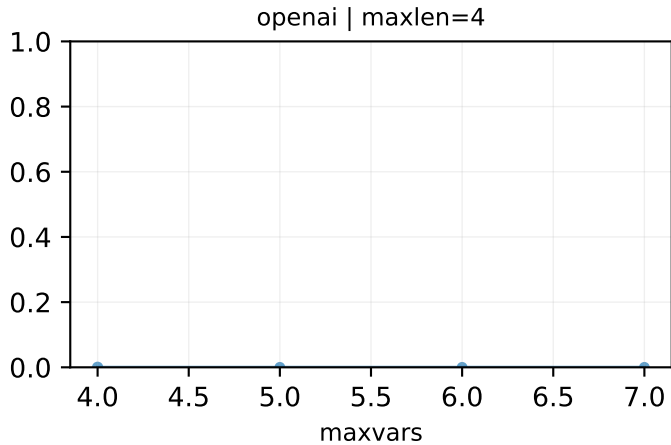
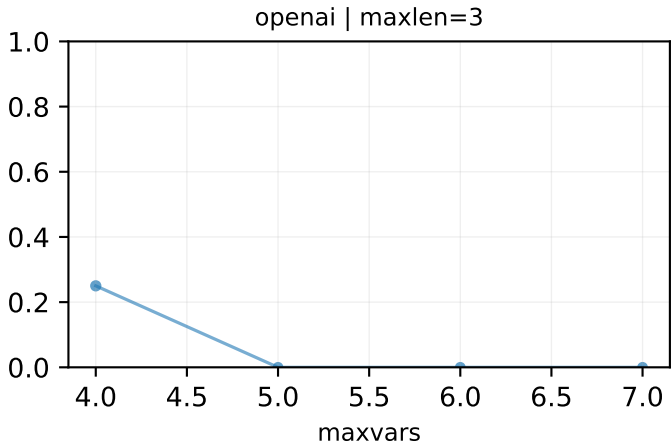
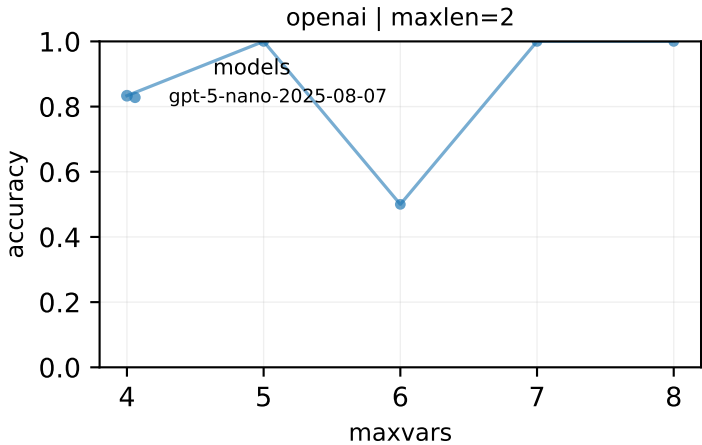
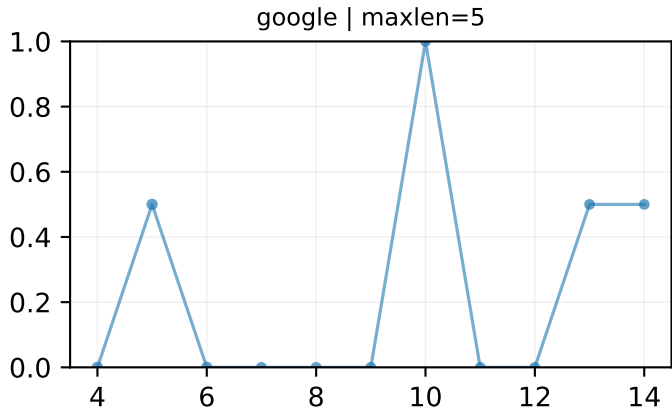
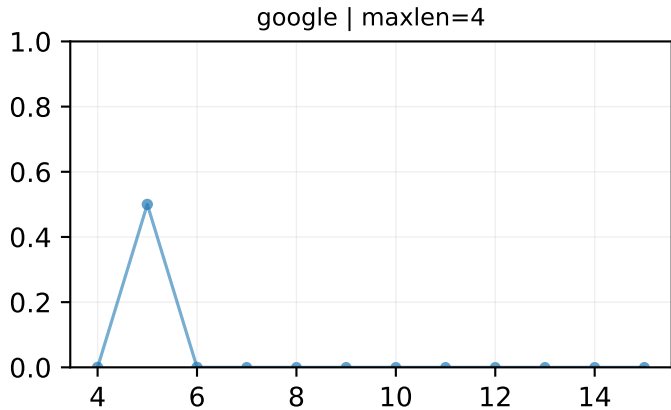
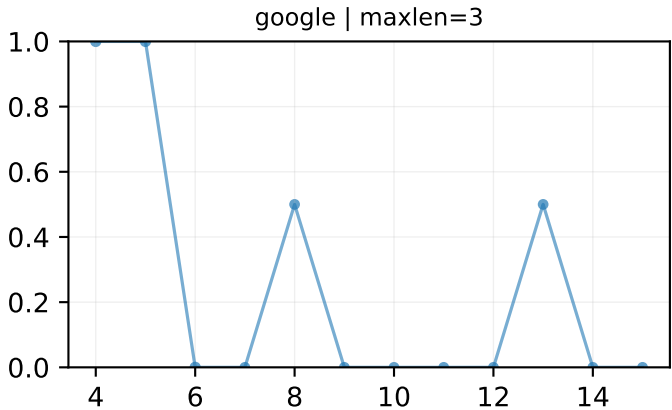
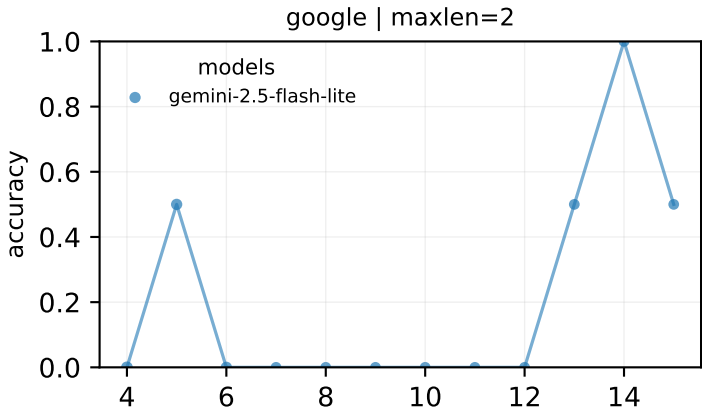
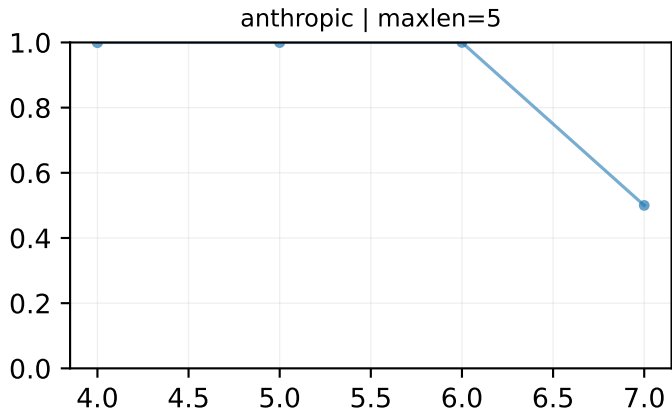
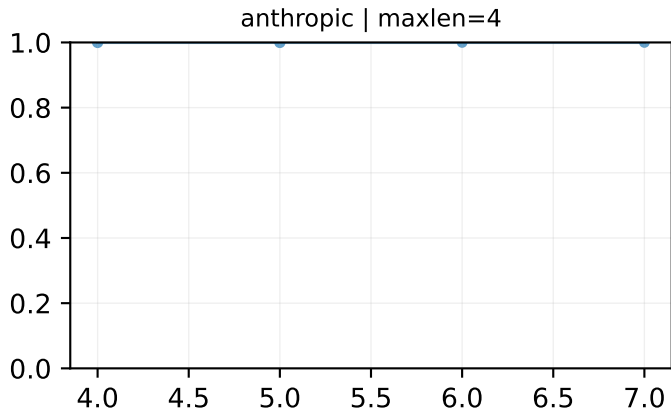
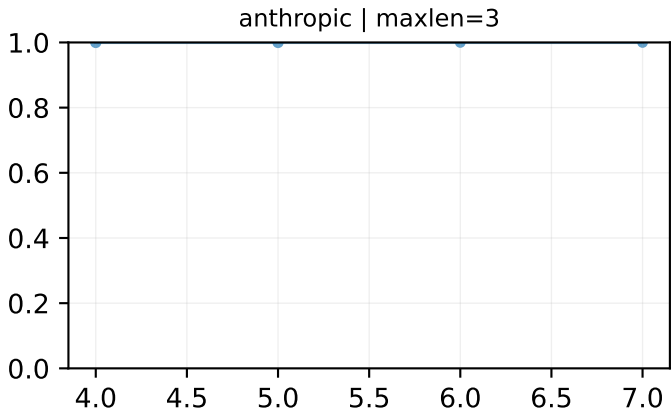
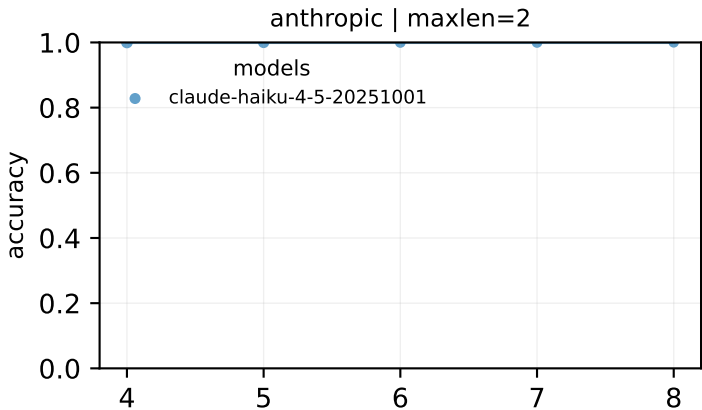


**Condensed: family=horn\_unified\_answer | thinking=nothink | horn=0 | metric=unsat\_accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

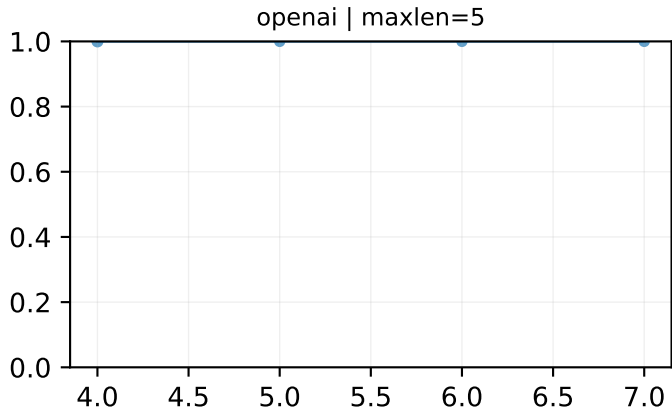
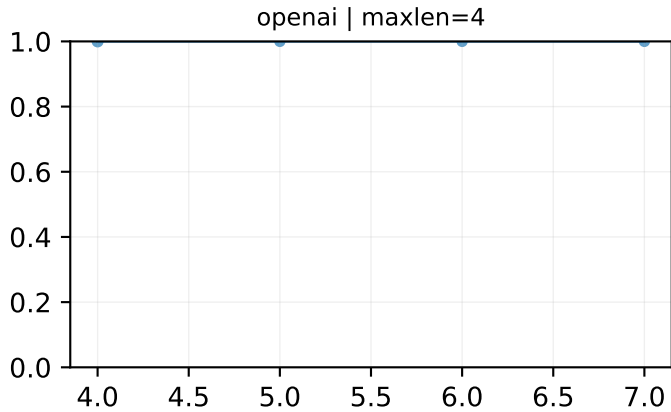
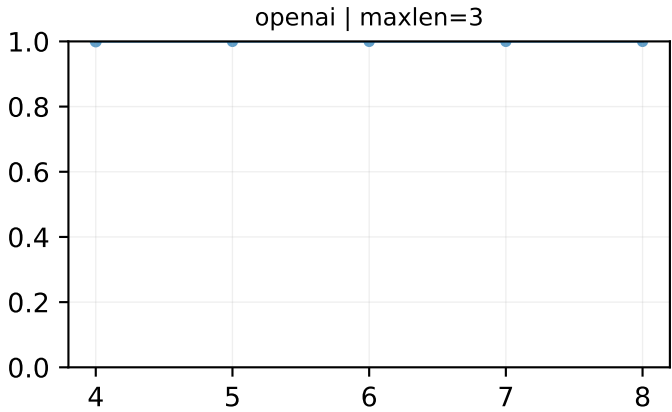
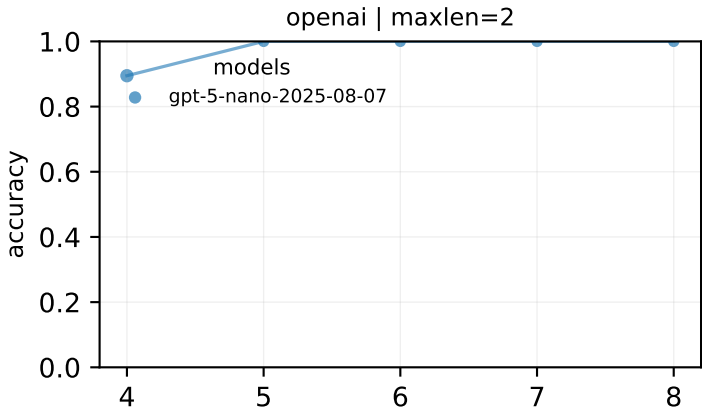
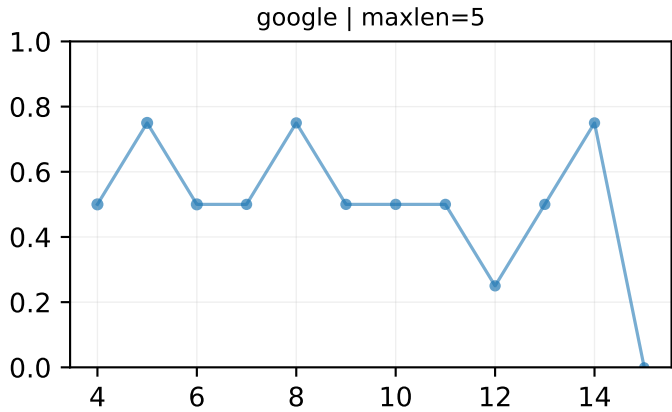
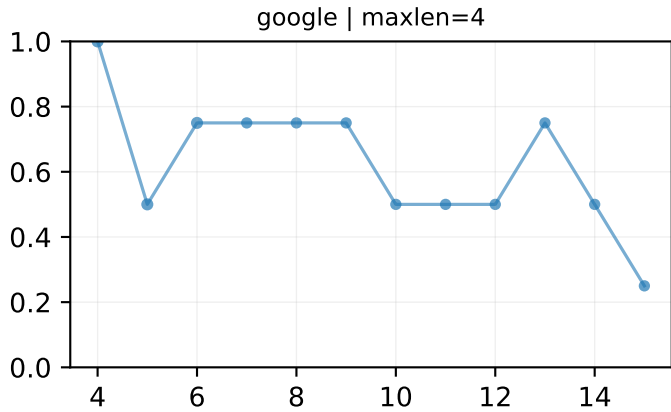
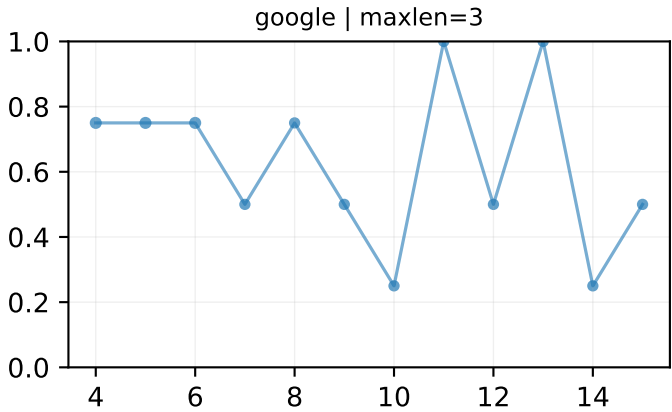
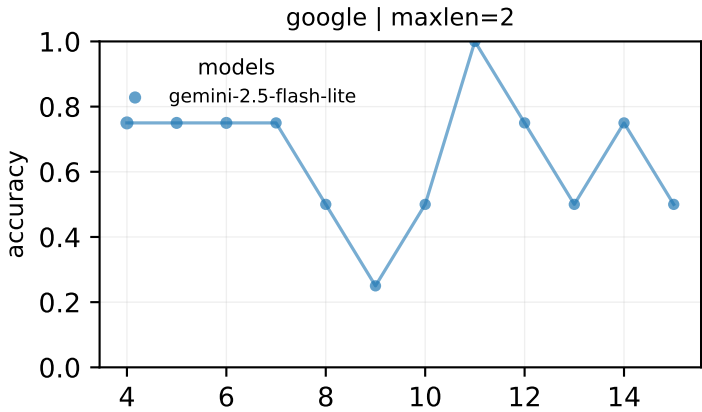
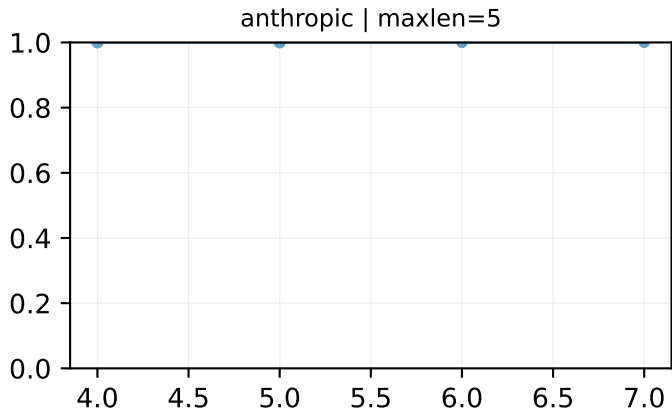
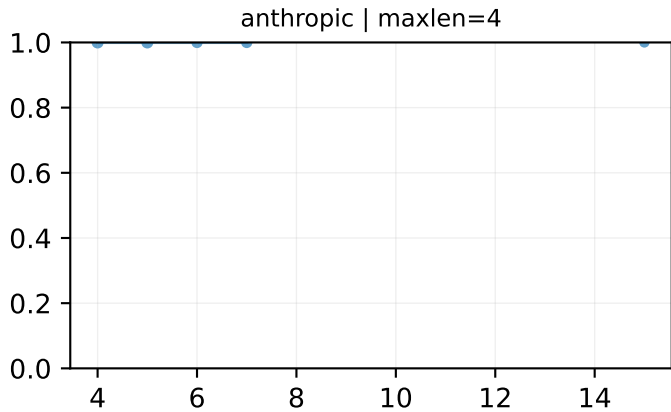
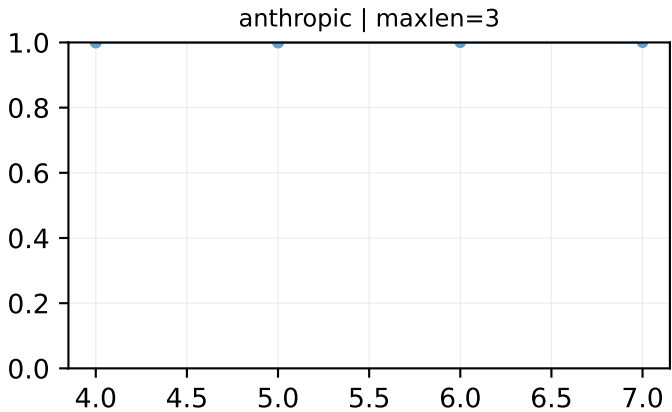
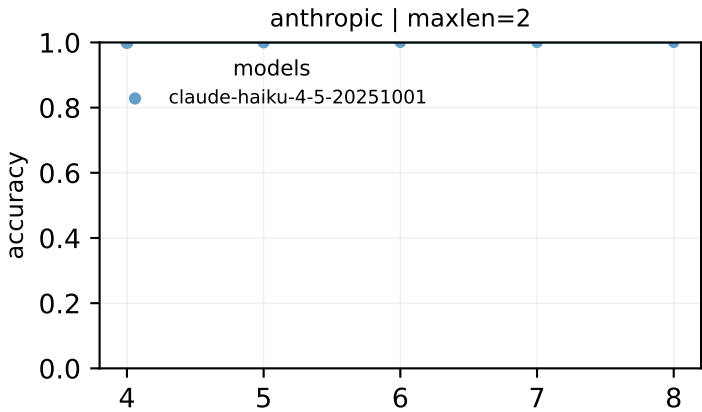


**Condensed: family=horn\_unified\_answer | thinking=nothink | horn=1 | metric=accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

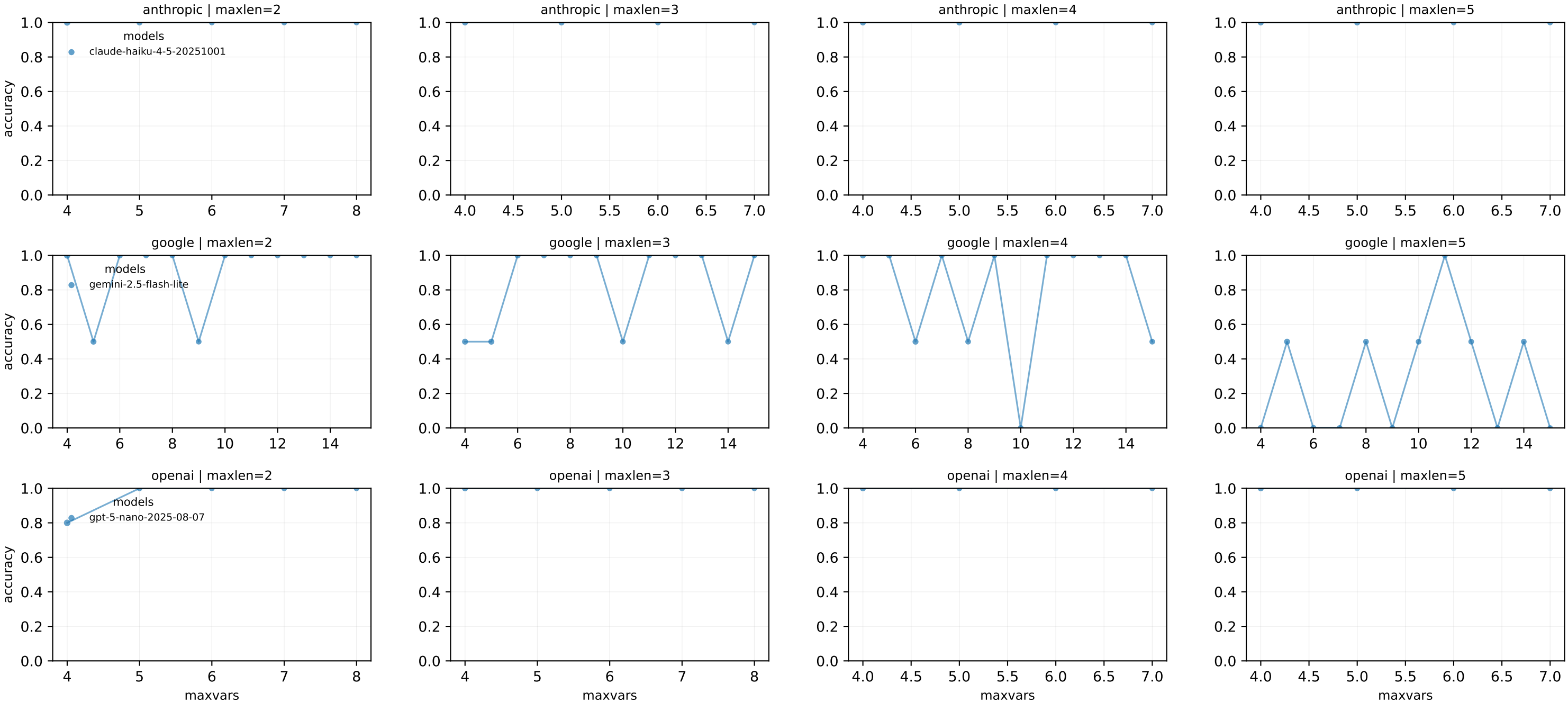


**Condensed: family=horn\_unified\_answer | thinking=nothink | horn=1 | metric=sat\_accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

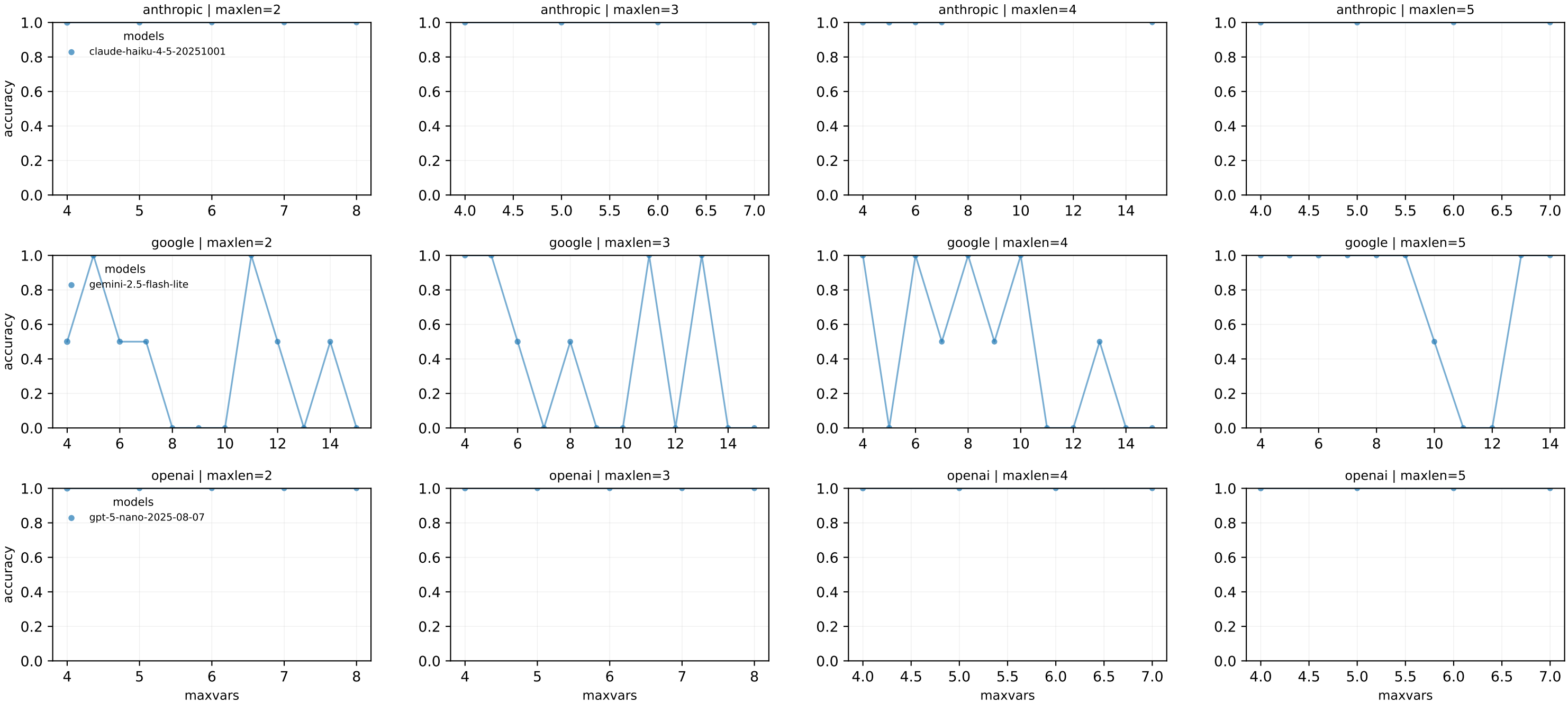


**Condensed: family=horn\_unified\_answer | thinking=nothink | horn=1 | metric=unsat\_accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

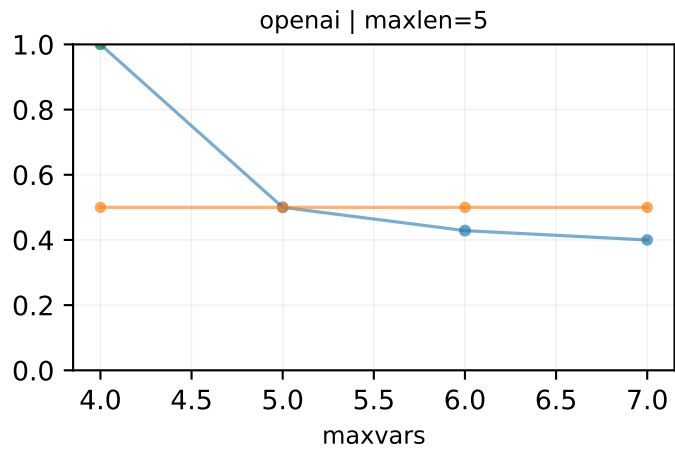
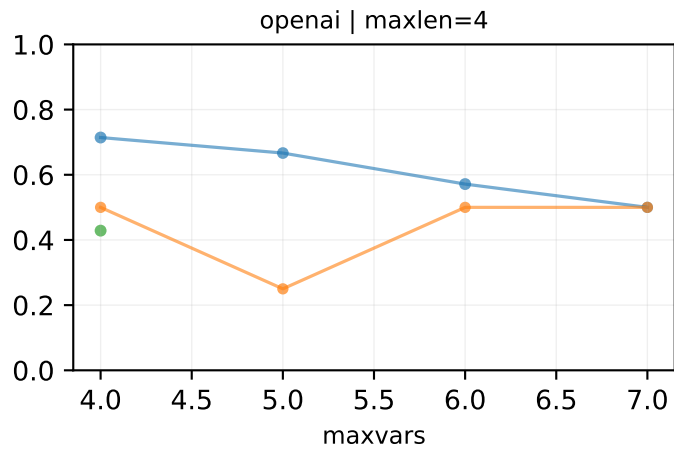
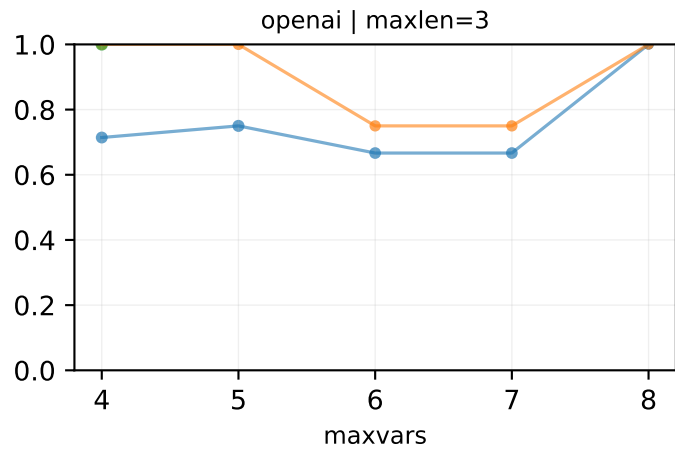
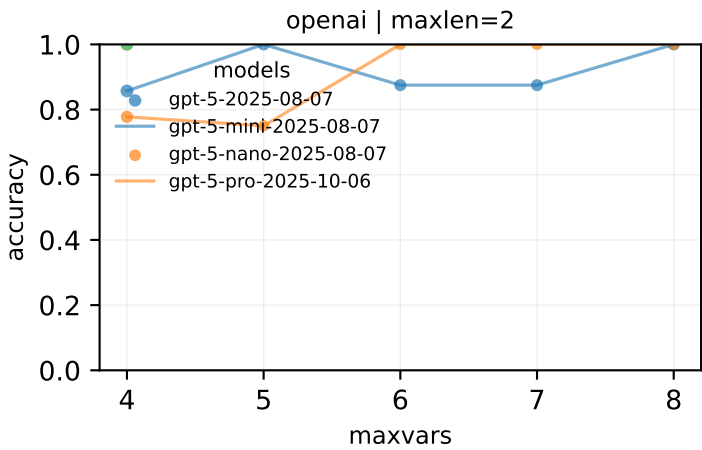
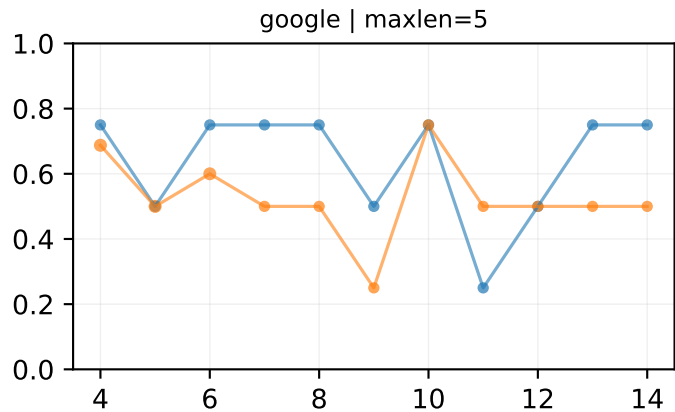
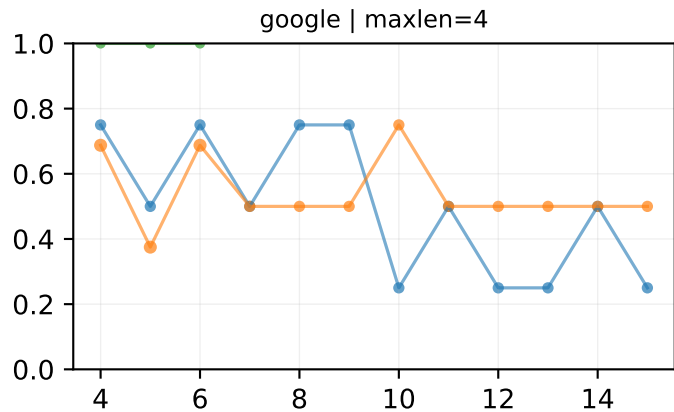
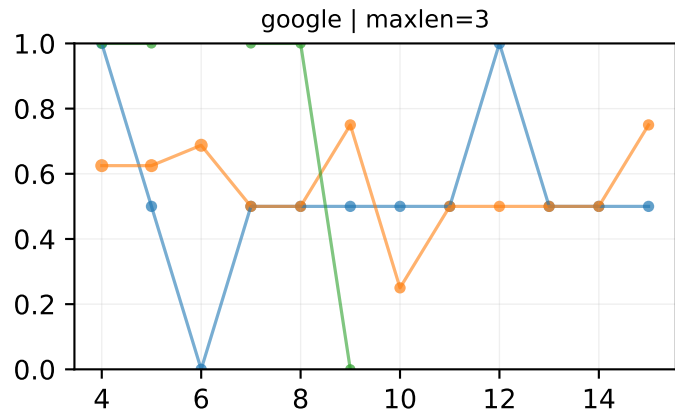
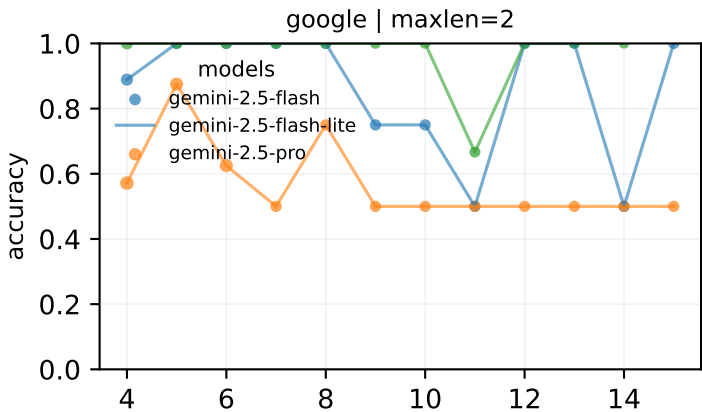
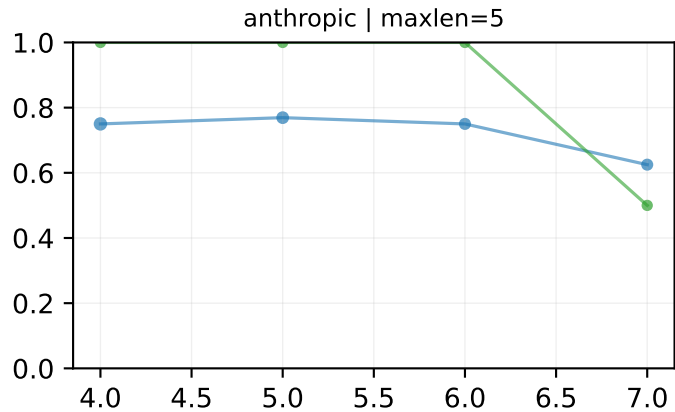
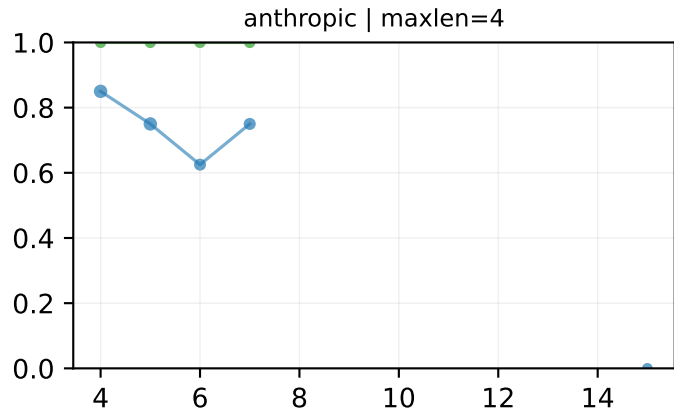
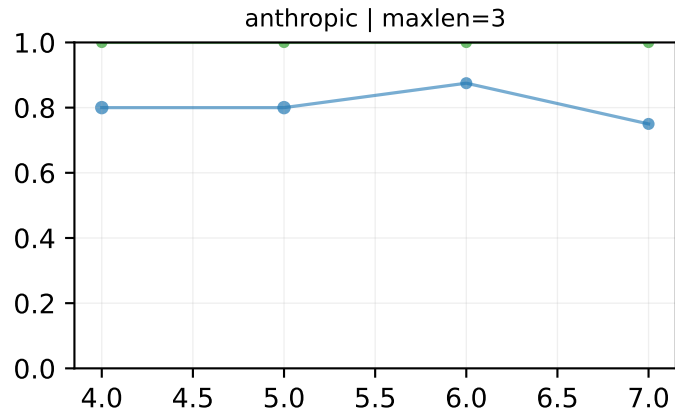
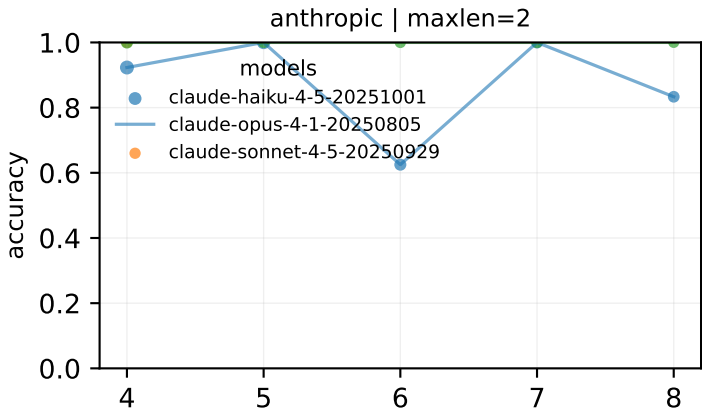


**Condensed: family=horn\_unified\_answer | thinking=thinking | horn=0 | metric=accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

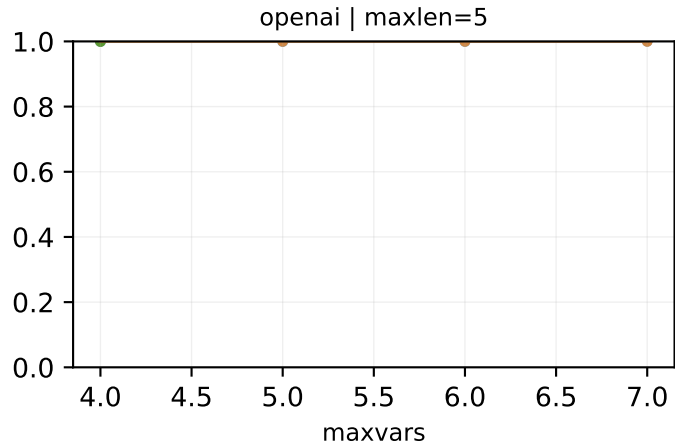
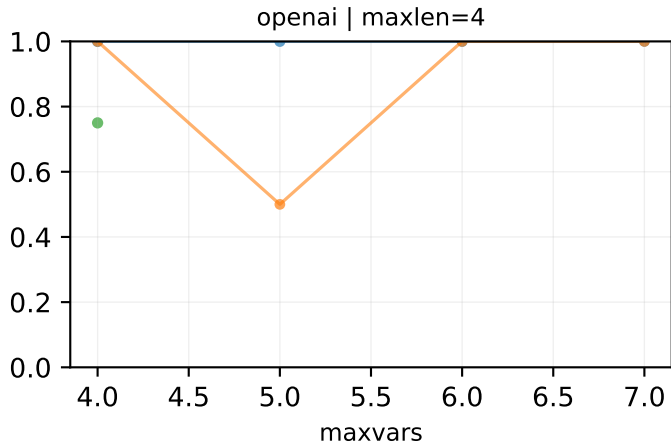
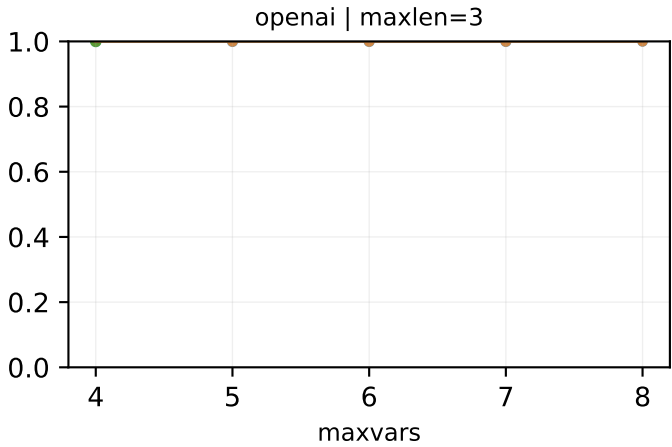
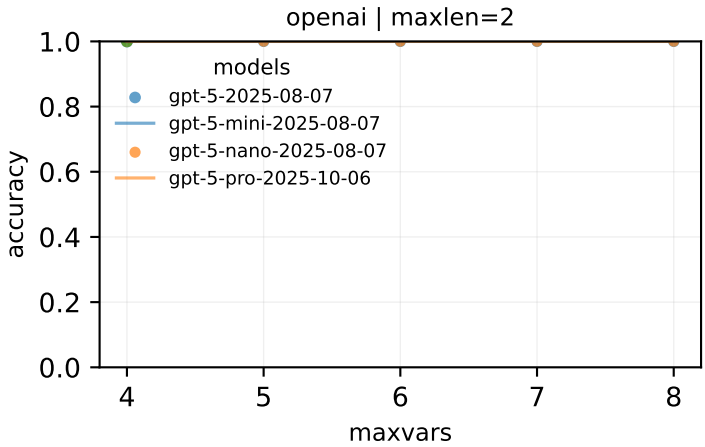
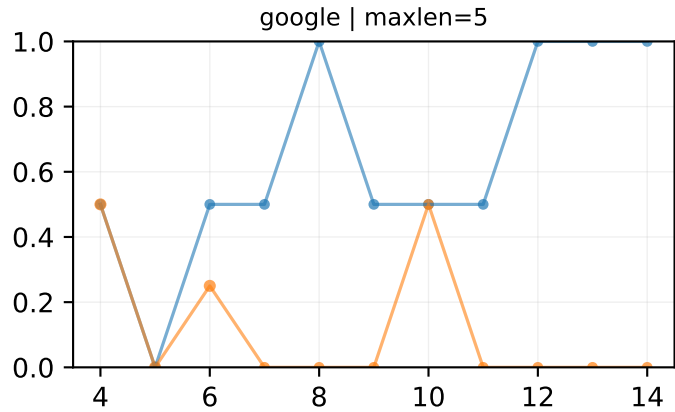
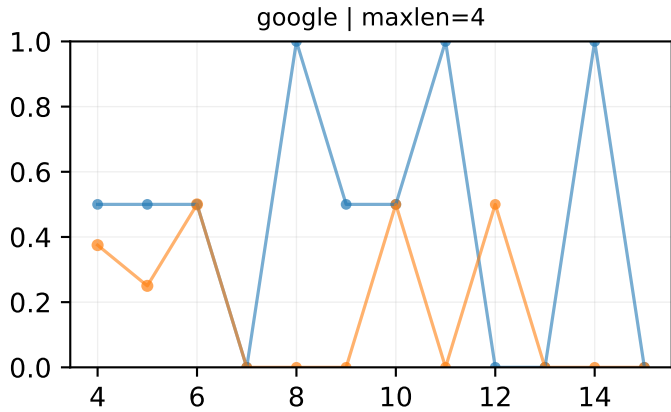
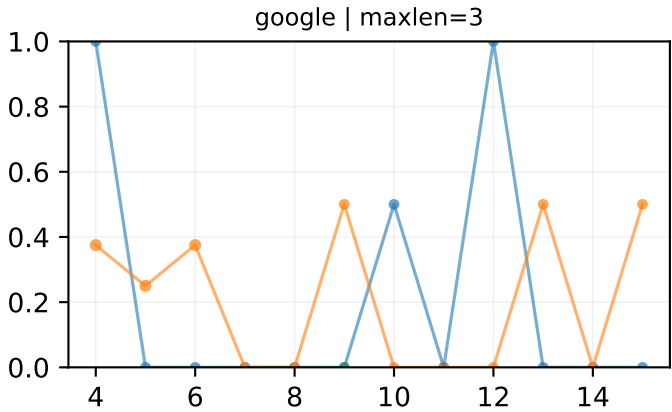
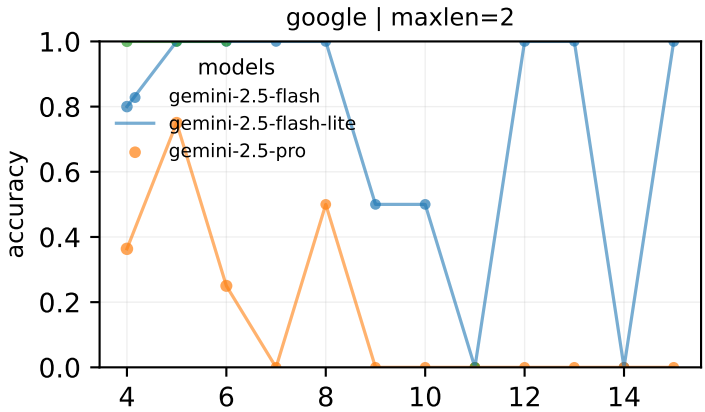
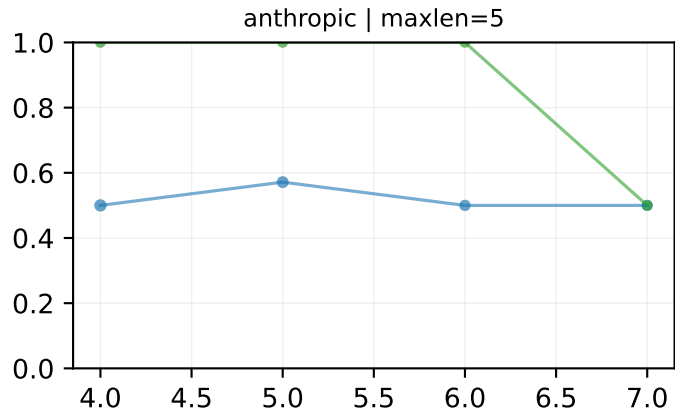
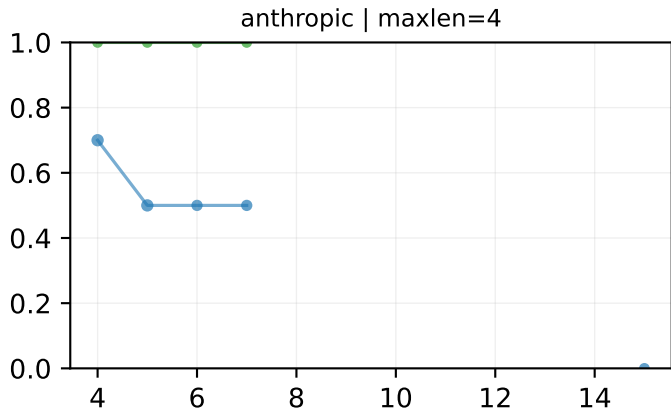
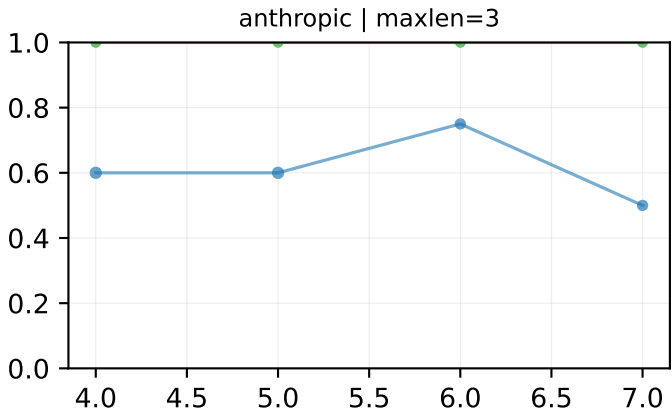
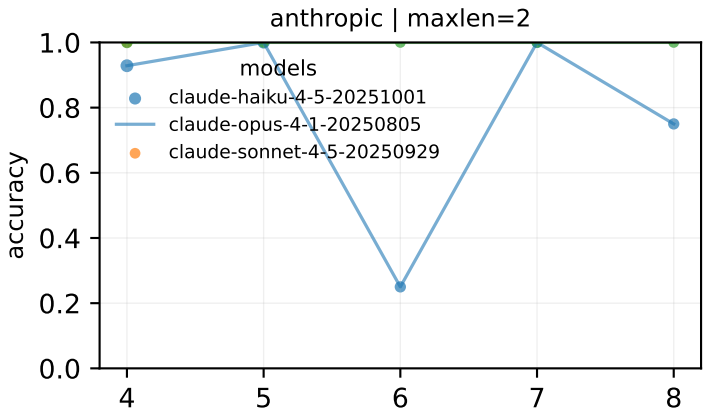


**Condensed: family=horn\_unified\_answer | thinking=thinking | horn=0 | metric=sat\_accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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



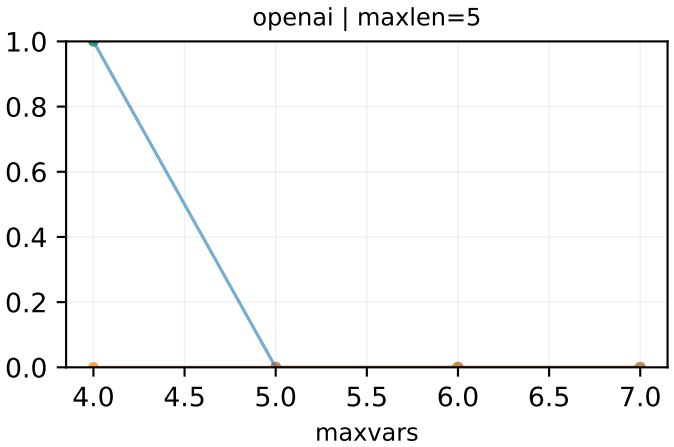
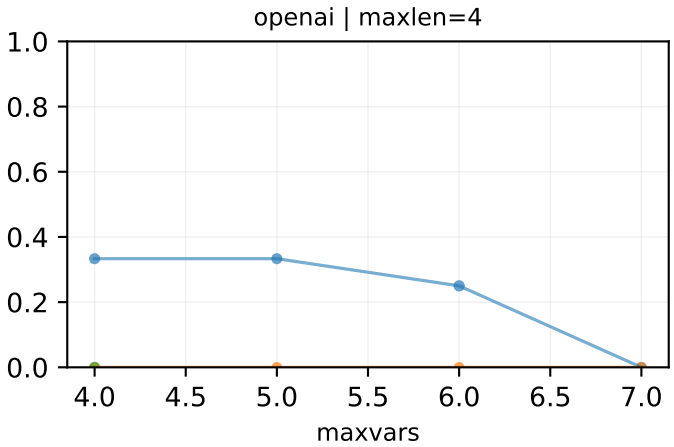
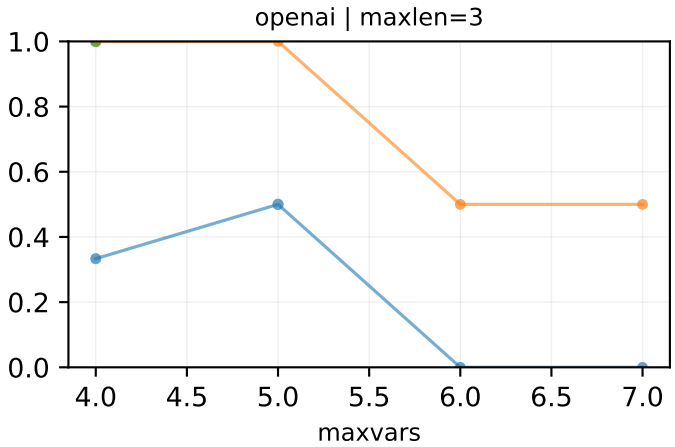
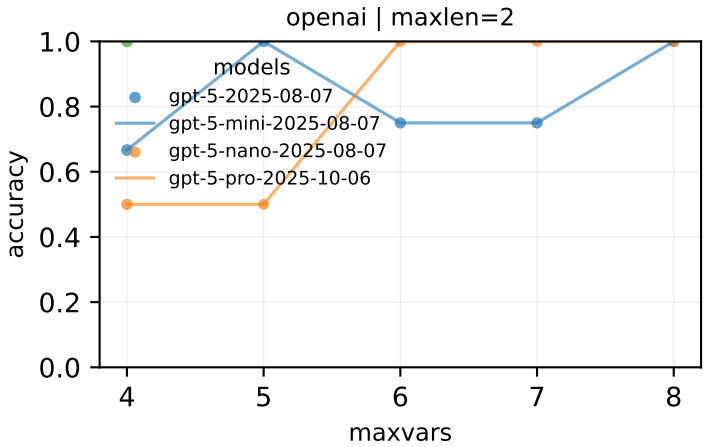
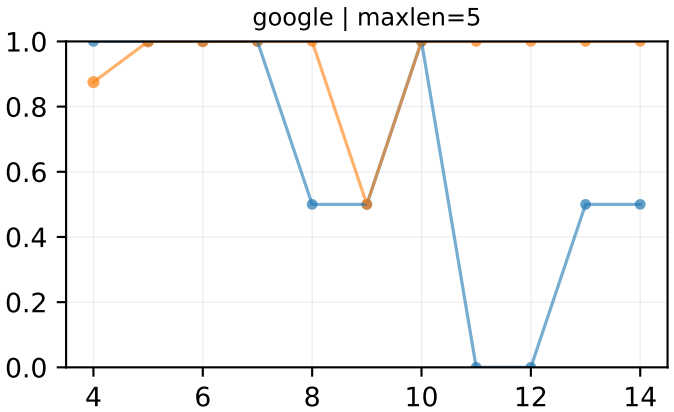
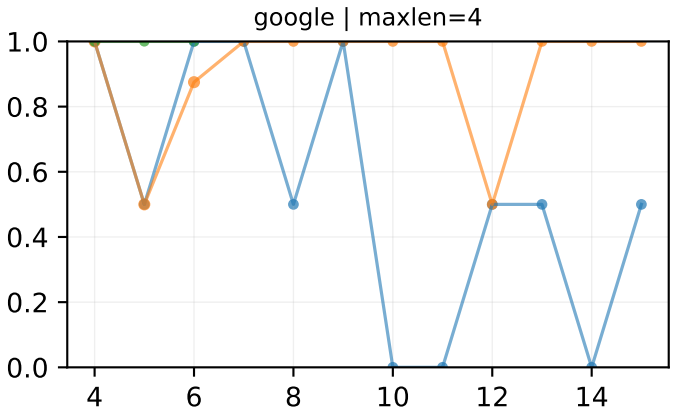
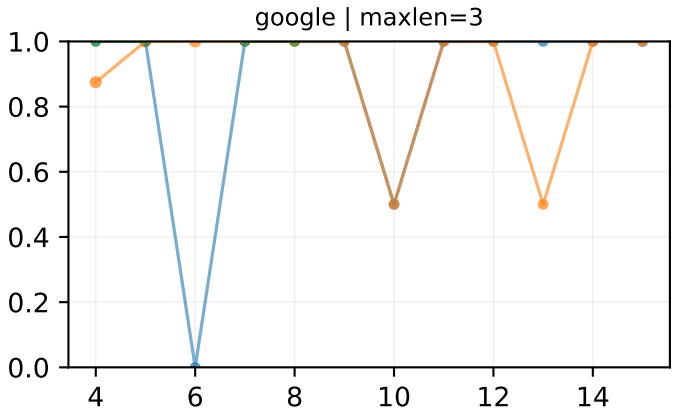
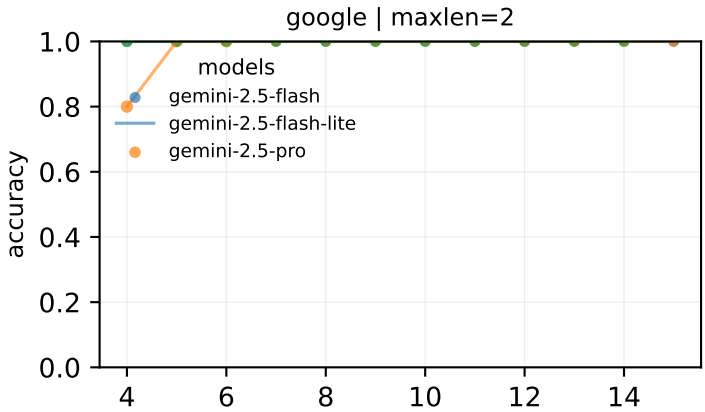
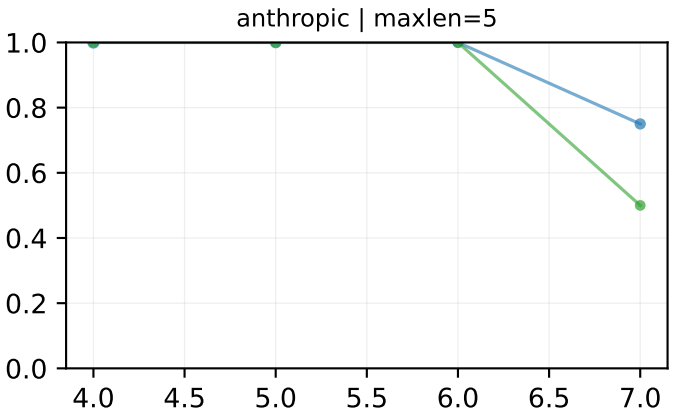
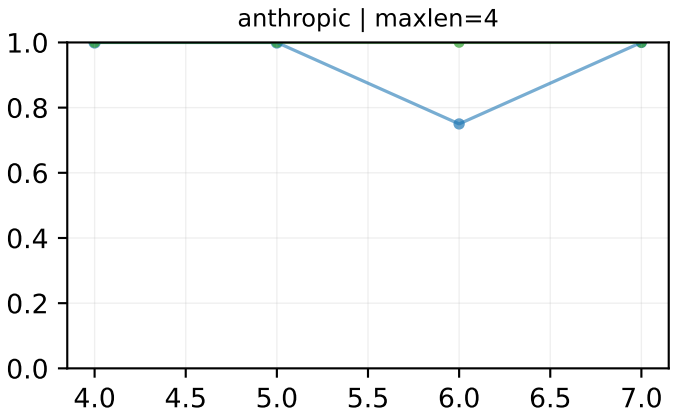
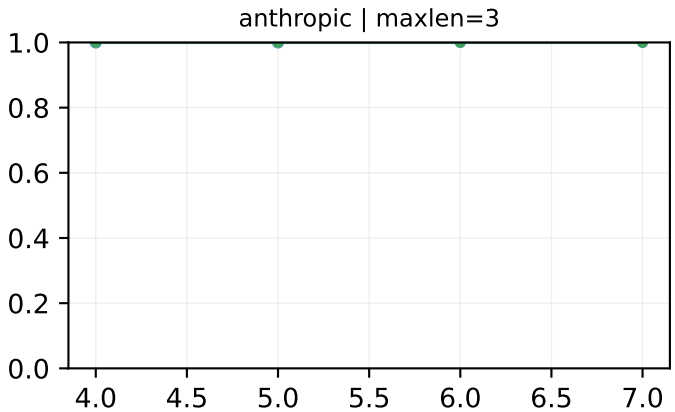
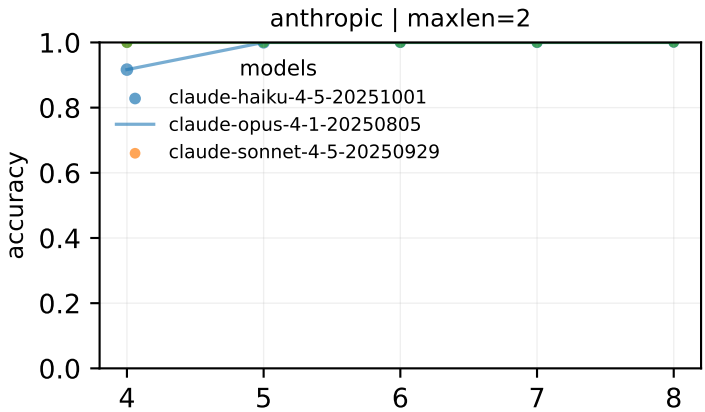


**Condensed: family=horn\_unified\_answer | thinking=thinking | horn=0 | metric=unsat\_accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

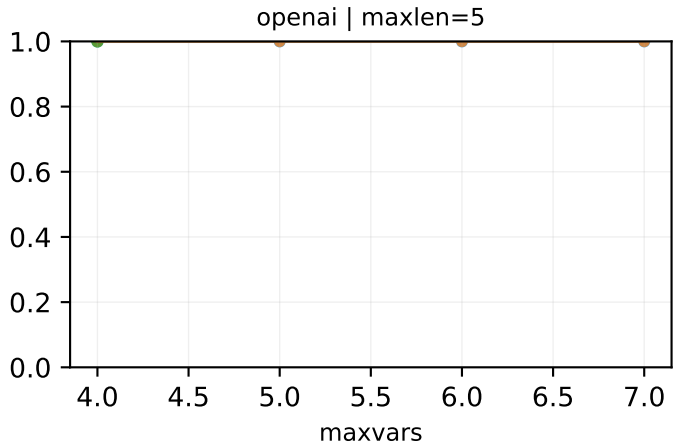
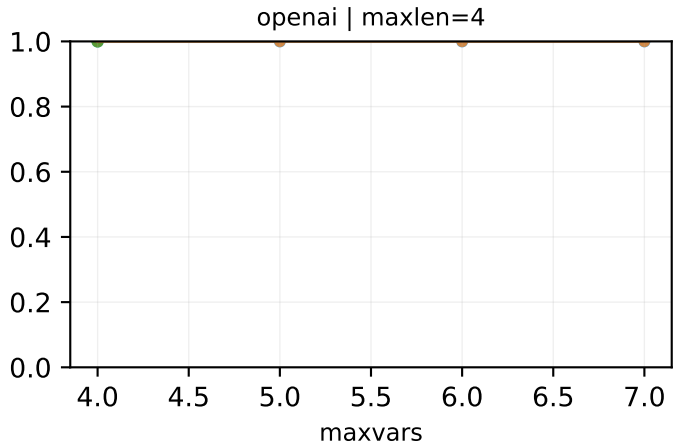
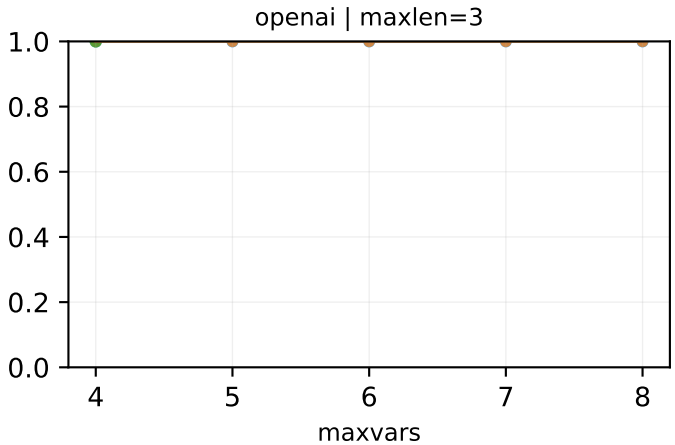
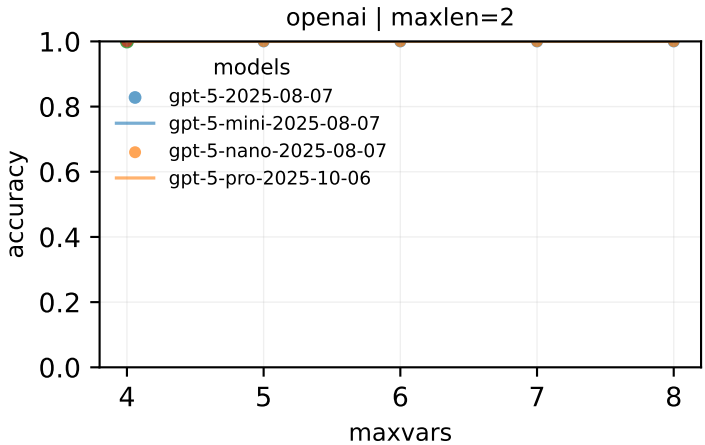
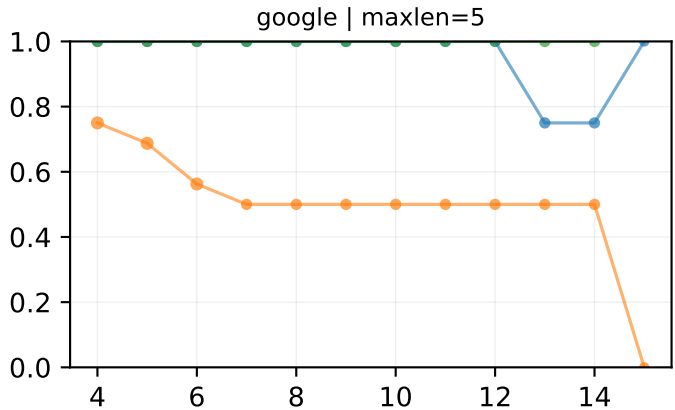
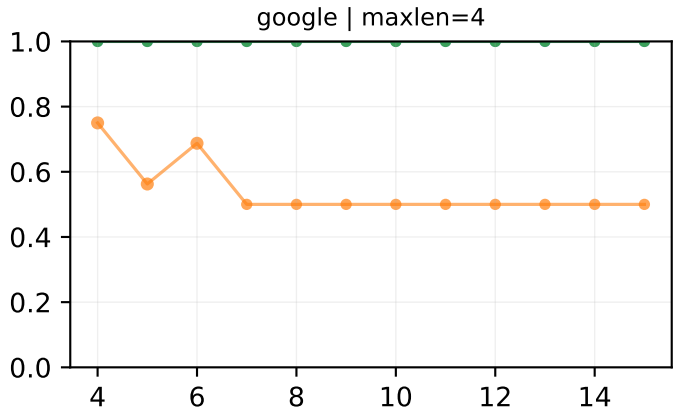
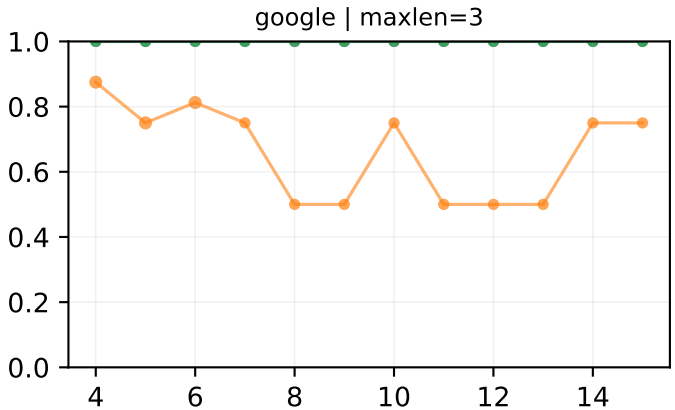
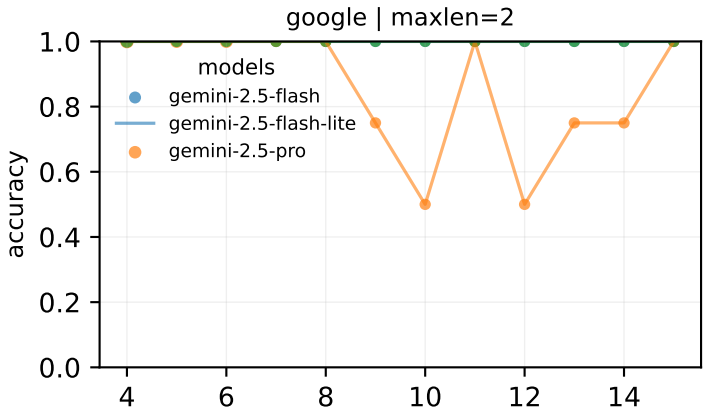
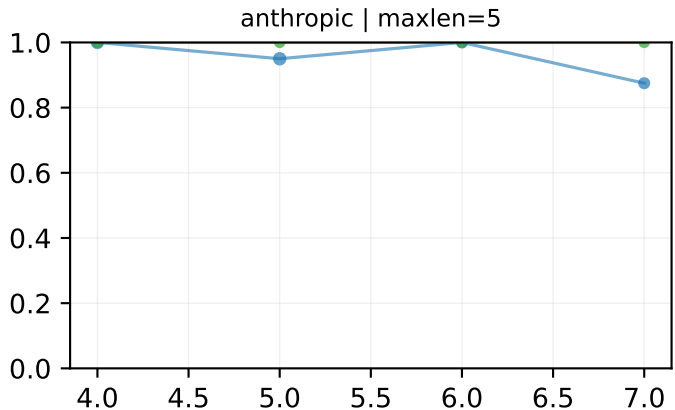
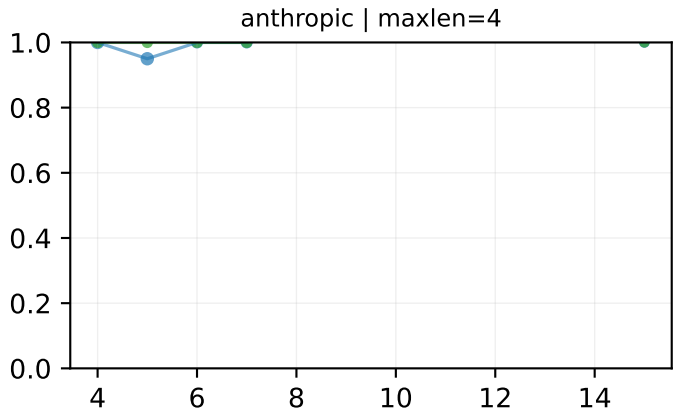
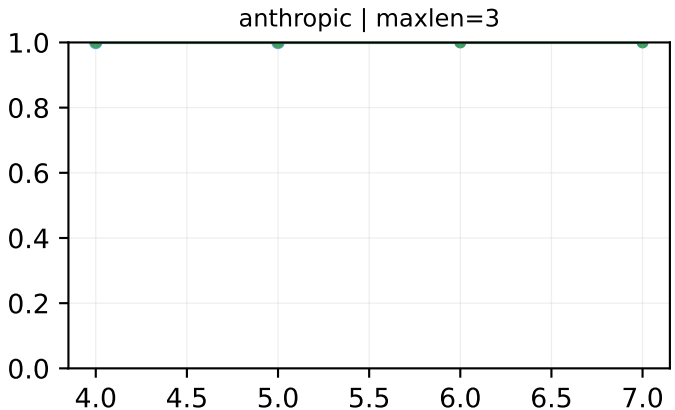
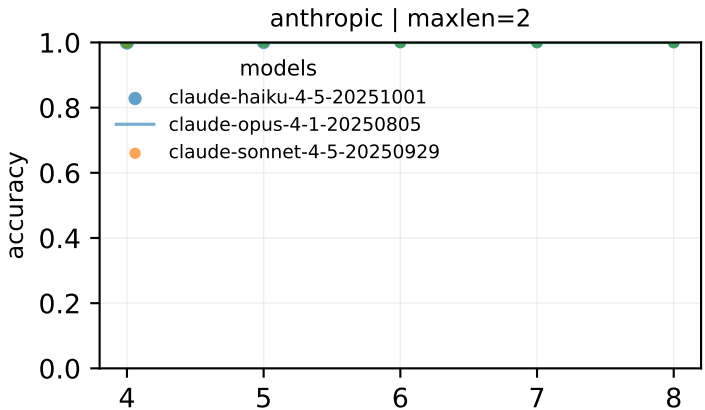


**Condensed: family=horn\_unified\_answer | thinking=thinking | horn=1 | metric=accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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

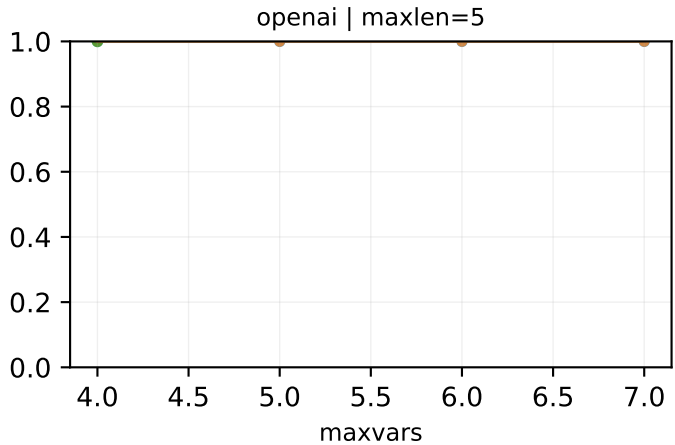
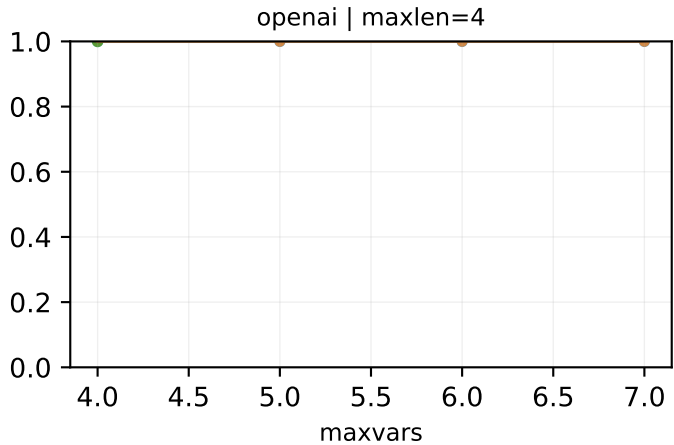
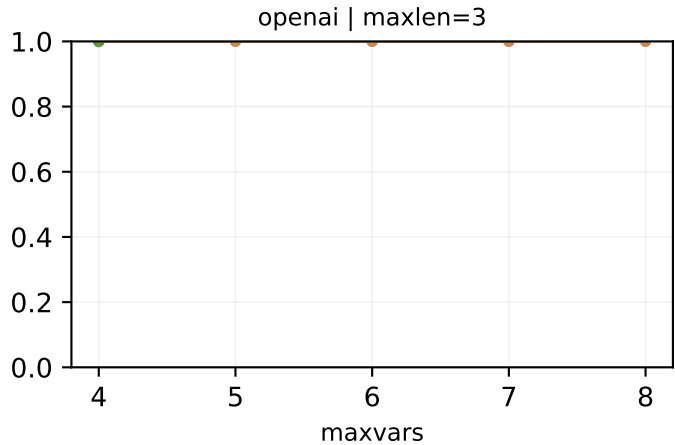
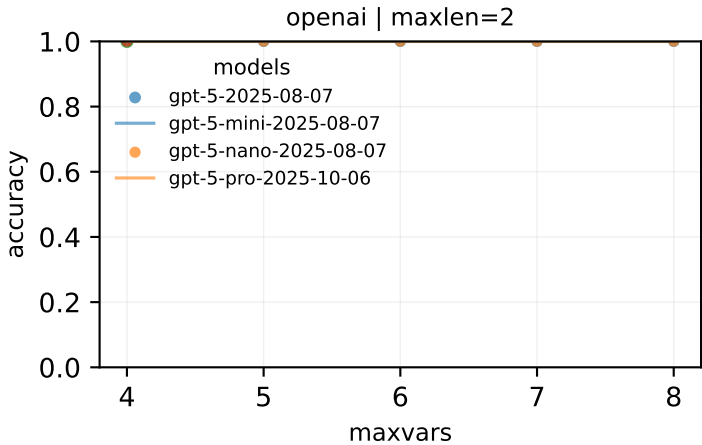
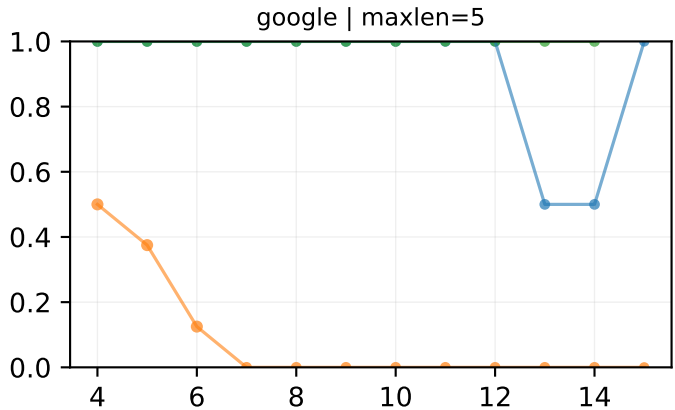
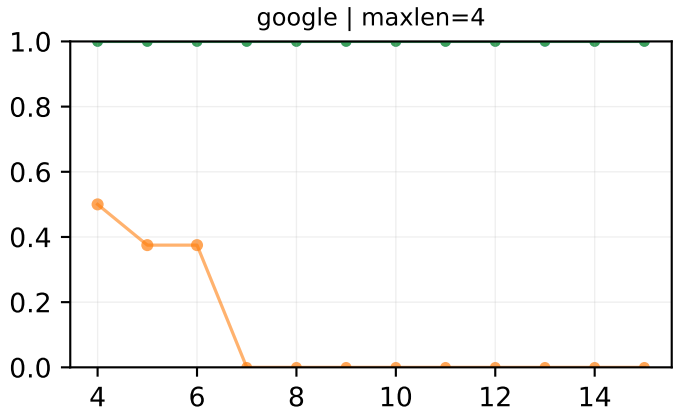
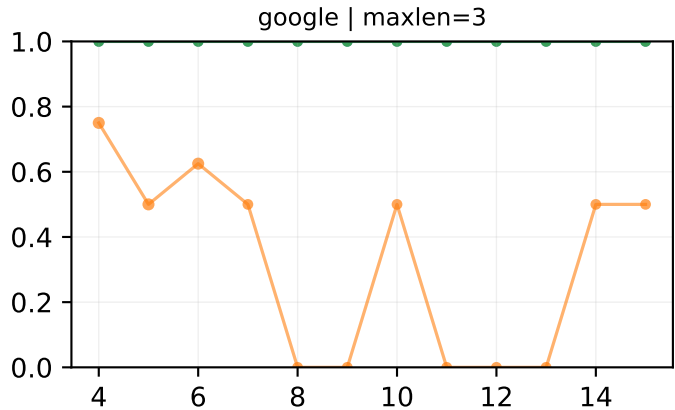
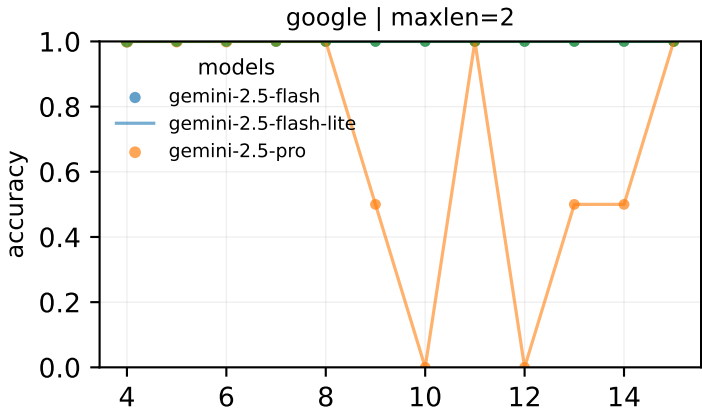
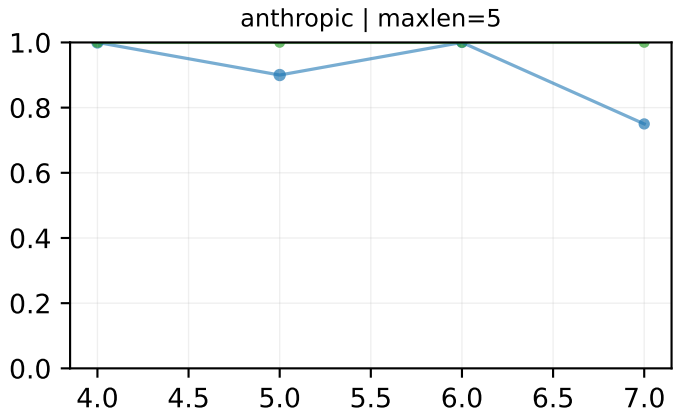
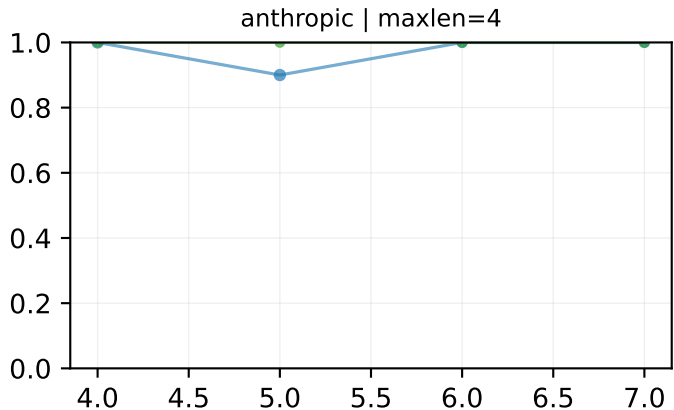
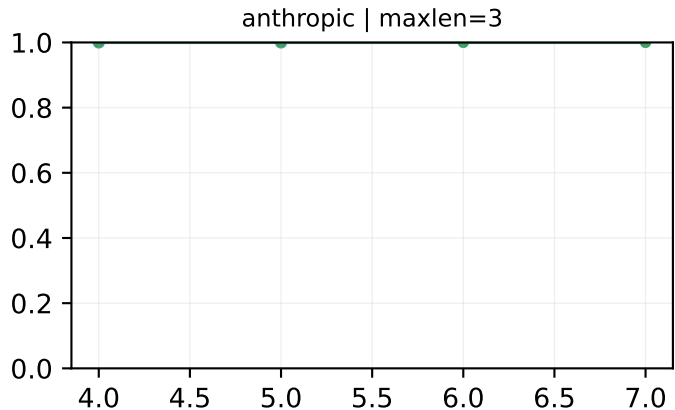
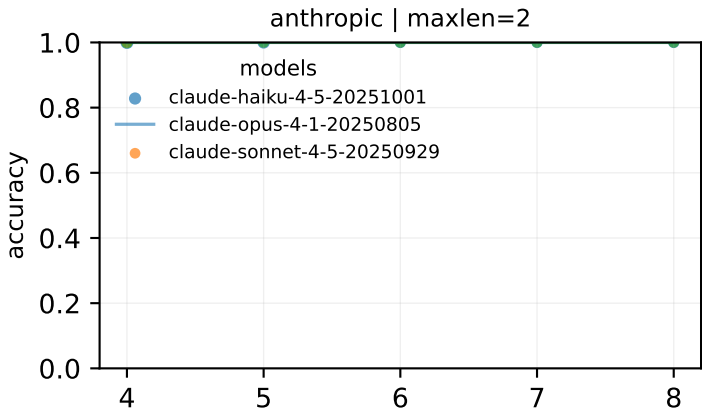


**Condensed: family=horn\_unified\_answer | thinking=thinking | horn=1 | metric=sat\_accuracy**

Line plots: x=maxvars, y=accuracy. Marker size ~ 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 ~ 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

