

A2UI Multi-Model Benchmark Report

Local Inference on NVIDIA GB10

Comparing 14 Models for A2UI v0.9 Protocol Generation

Structured UI Generation from Natural Language Prompts

Date: May 12, 2026
Hardware: NVIDIA GB10, 122 GiB RAM
Framework: Inspect AI + llama.cpp + vLLM
Judge: Claude Code (interactive, human-supervised)
Protocol: A2UI v0.9 Specification
Samples: 73 evaluation prompts

Contents

1	Executive Summary	2
1.1	Models Evaluated	2
2	Accuracy Comparison	3
3	Inference Performance	3
3.1	Average Inference Time	4
3.2	Per-Sample Inference Time Distribution	5
3.3	Output Throughput	6
4	Memory and Resource Utilization	6
4.1	Model Memory Footprint vs. Available KV Cache	7
4.2	In-Context Learning (ICL) Prompt Size	7
4.3	Prefill vs. Decode Throughput	8
5	Detailed Per-Sample Results	8
6	Summary Statistics	9
7	Hardware and Software Configuration	9
8	Methodology	9
8.1	Evaluation Pipeline	9
8.2	Limitations	9
8.3	Output-Format Strictness and Small-Model Failures	10
9	Conclusion	10
A	Grading Methodology	10
B	Per-Sample Grading Rationales	11

Executive Summary

This report compares **14 local language models** for generating structured user interfaces conforming to the A2UI v0.9 protocol. All evaluations were conducted on a **NVIDIA GB10** system running Linux/aarch64 using **llama.cpp + vLLM** as the inference engine.

Each model was evaluated on two axes:

1. **Schema Validation** – programmatic check that output is valid A2UI JSON
2. **Semantic Grading** – LLM-as-a-Judge assessment of UI correctness (C/P/I)

Models Evaluated

Table 1: Models included in this benchmark.

Model	ID	Engine	Type	Params	Quant
Llama 3.1 8B	NousResearch/Meta-Llama-3.1-8B-Instruct	llama.cpp	Dense	8B	none
Llama 3.1 8B	NousResearch/Meta-Llama-3.1-8B-Instruct	vLLM	Dense	8B	none
Qwen 3 1.7B	Qwen/Qwen3-1.7B	llama.cpp	Dense	1.7B	none
Qwen 3 1.7B	Qwen/Qwen3-1.7B	vLLM	Dense	1.7B	none
Qwen 3 14B	Qwen/Qwen3-14B	llama.cpp	Dense	14B	none
Qwen 3 14B	Qwen/Qwen3-14B	vLLM	Dense	14B	none
Qwen 3 8B	Qwen/Qwen3-8B	llama.cpp	Dense	8B	none
Qwen 3 8B	Qwen/Qwen3-8B	vLLM	Dense	8B	none
Gemma 4 26B MoE	nvidia/Gemma-4-26B-A4B-NVFP4	vLLM	MoE	26B (4B active)	modelopt
Nemotron 3 Nano 30B	nvidia/NVIDIA-Nemotron-3-Nano-30B-A3B-NVFP4	vLLM	MoE (Hybrid)	30B (3B active)	modelopt_fp4
Qwen3 30B MoE NVFP4	nvidia/Qwen3-30B-A3B-NVFP4	llama.cpp	MoE	30B (3B active)	none
Qwen3 30B MoE NVFP4	nvidia/Qwen3-30B-A3B-NVFP4	vLLM	MoE	30B (3B active)	modelopt
Llama 3.2 1B	unsloth/Llama-3.2-1B-Instruct	llama.cpp	Dense	1B	none
Llama 3.2 1B	unsloth/Llama-3.2-1B-Instruct	vLLM	Dense	1B	none

Average Inference Time

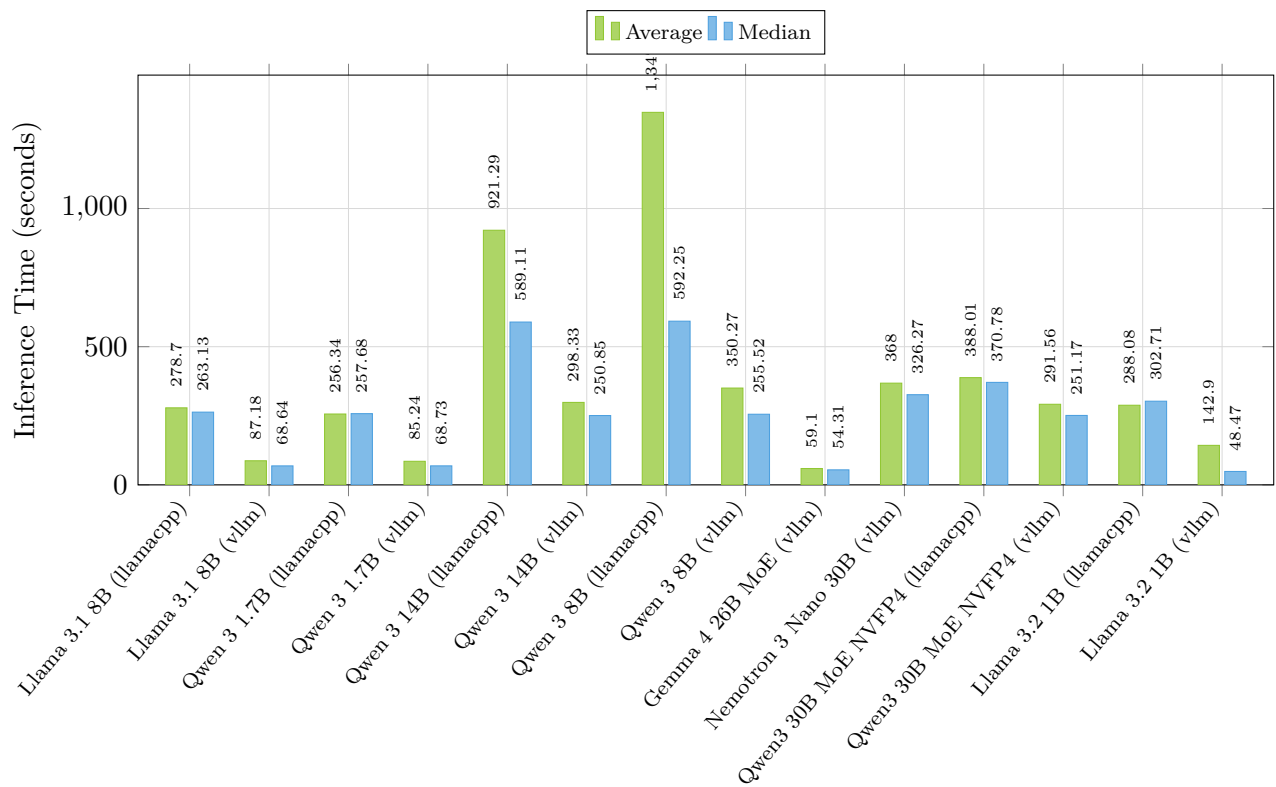


Figure 2: Average and median inference time per sample. Includes full prompt processing (system prompt with A2UI schema ~8k tokens) and output generation.

Per-Sample Inference Time Distribution

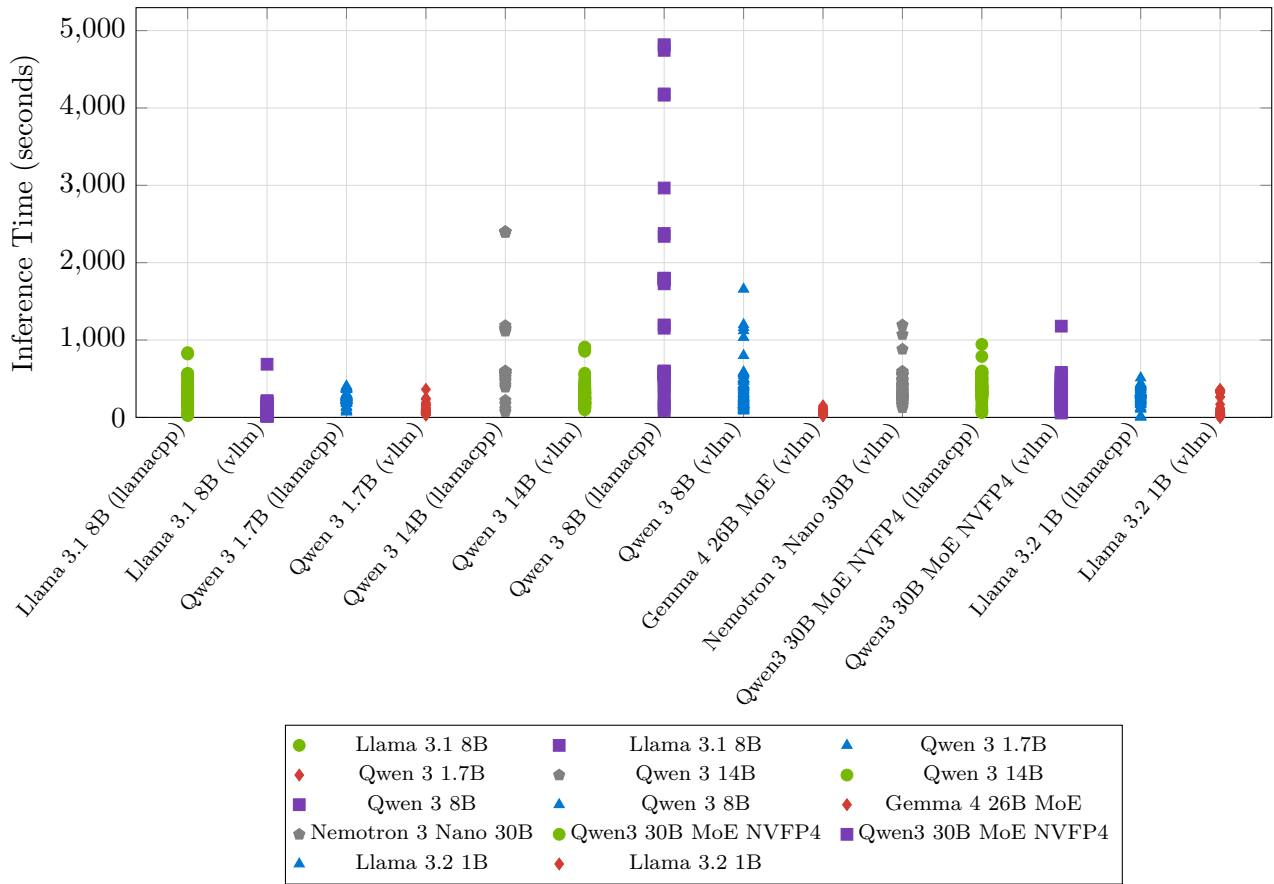


Figure 3: Per-sample inference time scatter plot. Each point represents one evaluation sample. Variation reflects differences in output length and complexity.

Output Throughput

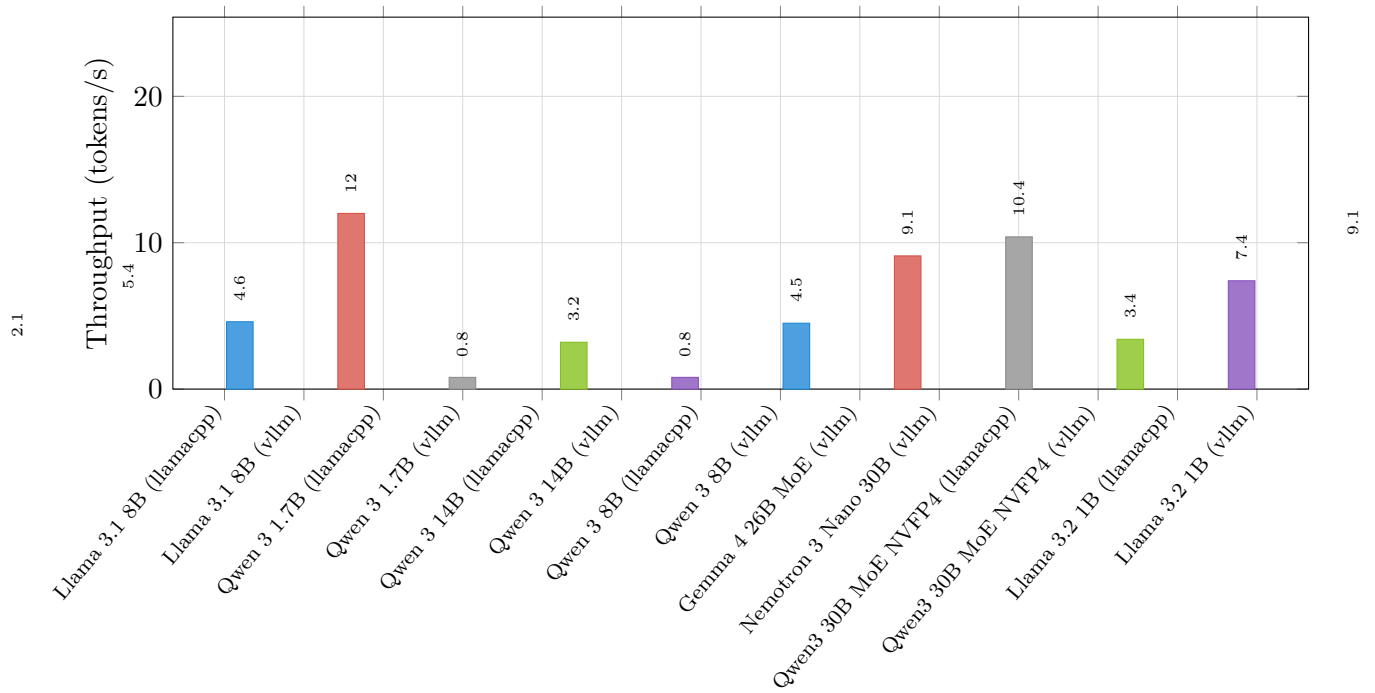


Figure 4: Effective output generation throughput (total output tokens / total inference time). Higher is better. MoE models typically achieve higher throughput due to fewer active parameters per token.

Memory and Resource Utilization

Model Memory Footprint vs. Available KV Cache

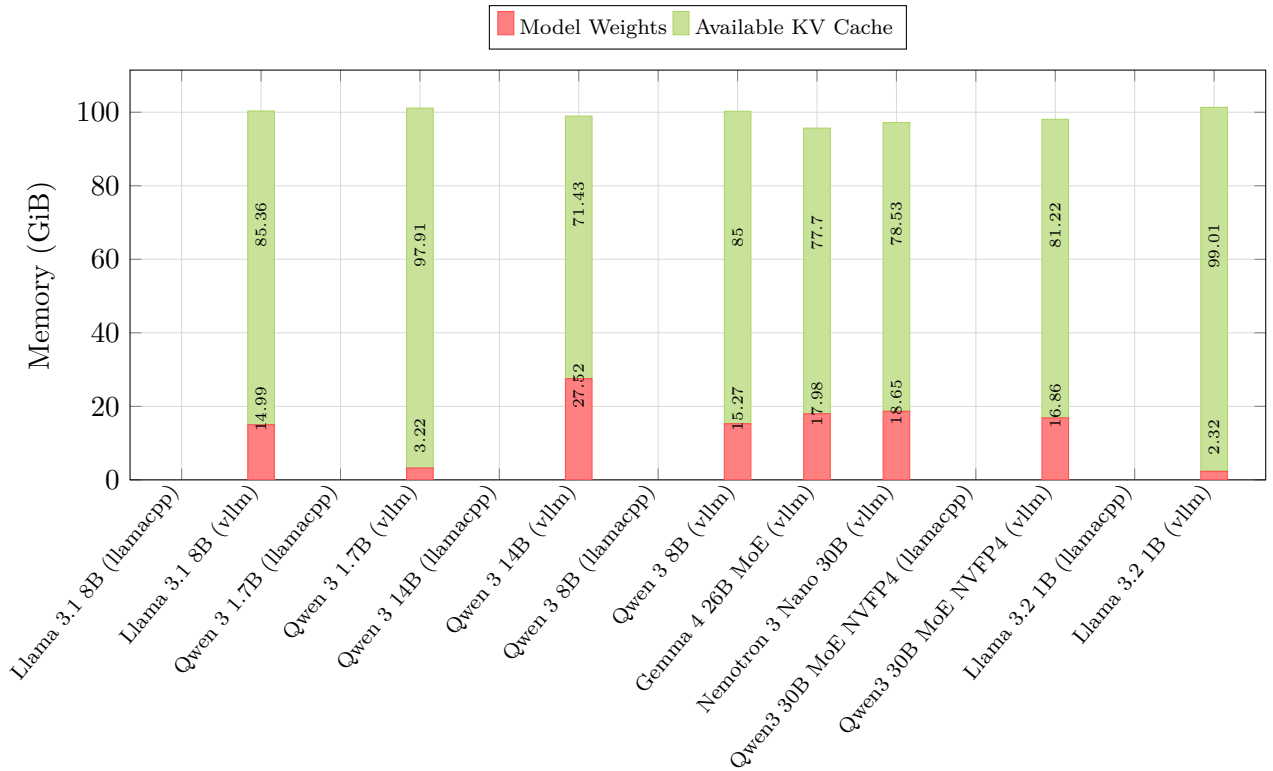


Figure 5: GPU memory allocation per model. Smaller model footprints leave more memory for KV cache, enabling higher concurrency and longer context windows. Total RAM: 122 GiB.

In-Context Learning (ICL) Prompt Size

The A2UI evaluation injects the full v0.9 JSON Schema and component catalog as a system prompt. This constitutes the in-context learning (ICL) payload that every model must process before generating output.

Table 2: ICL context and memory utilization per model.

Metric	Llama 3.1 8B	Llama 3.1 8B	Qwen 3 1.7B	Qwen 3 1.7B	Qwen 3 14B	Qwen 3 14B	Qwen 3 8B	Qwen 3 8B	Gemma 4 26B MoE	Nemotron 3 Nano 30B	Qwen3 30B MoE NVFP4	Qwen3 30B MoE NVFP4	Llama 3.2 1B	Llama 3.2 1B
ICL Context (tokens)	8,739	8,719	8,770	8,770	8,770	8,770	8,770	8,770	9,184	9,610	8,770	8,770	8,739	8,739
Model Memory (GiB)	0.0	15.0	0.0	3.2	0.0	27.5	0.0	15.3	18.0	18.6	0.0	16.9	0.0	2.3
KV Cache Available (GiB)	0.0	85.4	0.0	97.9	0.0	71.4	0.0	85.0	77.7	78.5	0.0	81.2	0.0	99.0
Prefix Cache Hit Rate	0.0%	98.1%	0.0%	98.1%	0.0%	98.2%	0.0%	98.3%	98.1%	86.6%	0.0%	98.1%	0.0%	98.1%

Summary Statistics

Table 4: Aggregate benchmark statistics per model.

Metric	Llama 3.1 8B	Llama 3.1 8B	Qwen 3 1.7B	Qwen 3 1.7B	Qwen 3 14B	Qwen 3 14B	Qwen 3 8B	Qwen 3 8B	Gemma 4 26B MoE	Nemotron 3 Nano 30B	Qwen3 30B MoE NVFP4	Qwen3 30B MoE NVFP4	Llama 3.2 1B	Llama 3.2 1B
Schema Accuracy	0%	4%	1%	0%	68%	90%	5%	1%	92%	0%	64%	75%	0%	0%
Semantic Accuracy	0%	3%	0%	0%	68%	90%	2%	1%	92%	0%	64%	75%	0%	0%
Avg Inference Time	278.7s	87.2s	256.3s	85.2s	921.3s	298.3s	1347.9s	350.3s	59.1s	368.0s	388.0s	291.6s	288.1s	142.9s
Median Inference Time	263.1s	68.6s	257.7s	68.7s	589.1s	250.8s	592.2s	255.5s	54.3s	326.3s	370.8s	251.2s	302.7s	48.5s
Output Throughput	2.1 tok/s	5.4 tok/s	4.6 tok/s	12.0 tok/s	0.8 tok/s	3.2 tok/s	0.8 tok/s	4.5 tok/s	9.1 tok/s	10.4 tok/s	3.4 tok/s	7.4 tok/s	9.1 tok/s	23.1 tok/s
ICL Context Size	8,739 tok	8,719 tok	8,770 tok	8,770 tok	8,770 tok	8,770 tok	8,770 tok	8,770 tok	9,184 tok	9,610 tok	8,770 tok	8,770 tok	8,739 tok	8,739 tok
Prefix Throughput (peak)	0 tok/s	202 tok/s	0 tok/s	637 tok/s	0 tok/s	0 tok/s	0 tok/s	0 tok/s	150 tok/s	2006 tok/s	0 tok/s	0 tok/s	0 tok/s	594 tok/s
Decode Throughput (peak)	0 tok/s	65 tok/s	0 tok/s	134 tok/s	0 tok/s	40 tok/s	0 tok/s	61 tok/s	106 tok/s	130 tok/s	0 tok/s	0 tok/s	98 tok/s	289 tok/s
Prefix Cache Hit Rate	0.0%	98.1%	0.0%	98.1%	0.0%	98.2%	0.0%	98.3%	86.6%	98.1%	0.0%	98.1%	0.0%	98.1%
Model Memory	0.0 GiB	15.0 GiB	0.0 GiB	3.2 GiB	0.0 GiB	27.5 GiB	0.0 GiB	15.3 GiB	18.0 GiB	18.6 GiB	0.0 GiB	16.9 GiB	0.0 GiB	2.3 GiB
KV Cache Available	0.0 GiB	85.4 GiB	0.0 GiB	97.9 GiB	0.0 GiB	71.4 GiB	0.0 GiB	85.0 GiB	77.7 GiB	78.5 GiB	0.0 GiB	81.2 GiB	0.0 GiB	99.0 GiB
Total Input Tokens	638,958	637,498	641,239	641,239	289,567	641,239	403,747	641,239	671,497	702,620	641,239	641,239	638,958	638,958
Total Output Tokens	42,981	34,268	85,879	74,960	23,685	69,136	50,739	115,363	39,303	279,720	97,673	156,761	191,213	241,386
Samples Evaluated	73	73	73	73	44	73	57	73	73	73	73	73	73	73

Hardware and Software Configuration

Table 5: System configuration.

Component	Specification
GPU	NVIDIA GB10
Memory	122 GiB RAM
Architecture	aarch64
Engine	llama.cpp + vLLM
Inspect AI	≥0.3.217
OS	Linux (kernel 6.17.0-1014-nvidia)

Methodology

Evaluation Pipeline

Each model was evaluated through the following pipeline:

- Model serving:** llama.cpp + vLLM starts an OpenAI-compatible API server with the model
- System prompt injection:** The full A2UI v0.9 JSON Schema and component catalog are injected as a system prompt (~8,000 tokens)
- Generation:** The model generates structured JSON output for each evaluation prompt in a *single-turn* setting — one user prompt yields one assistant response, with no follow-up turns, no self-correction loop, and no schema-error retries
- Schema validation:** Programmatic validation against A2UI schema and integrity rules
- Semantic grading:** An external judge (Claude Code (interactive, human-supervised)) assigns C/P/I per sample using the rubric in Appendix A

Limitations

- LLM-based grading:** Grades are assigned by an LLM judge (Claude Code (interactive, human-supervised)), not a domain expert. A human evaluator or multi-judge consensus would reduce single-grader bias.
- Test dataset:** Results use a test dataset. The full encrypted production dataset (~100 samples) provides more comprehensive coverage.
- Single-run:** Each model was benchmarked once. Multiple runs would provide statistical significance for timing measurements.
- Single-turn generation:** Every sample is a one-shot generation. The model receives the A2UI system prompt plus the user’s task, and its first response is what we score. A real agent

could re-emit JSON after seeing a schema error, or split a complex UI across several turns; we do not exercise either capability here. Schema-invalid outputs therefore count as incorrect even when a follow-up turn would have repaired them.

Output-Format Strictness and Small-Model Failures

The schema validator is the upstream A2UI Python SDK (`a2ui.parser.parser.parse_response + a2ui.schema.manager.A2uiSchemaManager`, Apache 2.0, Copyright Google LLC, vendored in `A2UI/eval/a2ui_eval/scorers.py`). It only recognises payloads wrapped in literal `<a2ui-json>...</a2ui-json>` tags and then runs the full A2UI v0.9 JSON schema against the parsed message graph. There is no markdown fence fallback, no automatic unwrap of “thinking” preambles, and no acceptance of structurally valid JSON in any other envelope.

The validator’s strictness is a faithful gate for a renderer that would consume the output, but it explains most of the near-zero scores in the small-model rows. Spot-checking sample 0 (`text_heading`) for each failing model shows the issue is *output format compliance*, not context size — inputs were 8.7–9.7k tokens against a 16,384-token serving limit, and no completion approached the output-token cap. The recurring patterns:

- **Llama 3.2 1B** (both engines): emitted a JSON Schema *definition* (`schema: https://json-schema.org`) describing what a Text component should look like, instead of an A2UI *instance*. No `<a2ui-json>` tags appear in the response at all.
- **Qwen 3 1.7B** (vLLM): used the tags but produced a malformed structure — `version` placed inside `updateComponents`, plus a duplicated nested `updateComponents` block.
- **Qwen 3 8B** (vLLM): wrapped otherwise-correct A2UI JSON in “`json` markdown fences” instead of `<a2ui-json>` tags. The validator therefore reports “tags not found” and the run scores near zero despite the content largely being valid.
- **Nemotron 3 Nano 30B** (vLLM): a reasoning model that emits a `<think>...</think>` preamble followed by the final answer, but the final answer double-wraps `createSurface` inside another `createSurface` (`{createSurface:{version:v0.9;createSurface:{...}}}`).

These are not capability ceilings so much as instruction-following and format-compliance gaps that a single follow-up turn or a more lenient extractor (e.g. accepting any fenced JSON block whose root matches the A2UI message schema) would close in many cases. We deliberately do not apply such fallbacks here so that scores reflect what a renderer running the production parser would actually accept.

Conclusion

This benchmark demonstrates that local inference on NVIDIA GB10 is viable for A2UI protocol evaluation across multiple model architectures. The results provide a baseline for comparing model capabilities in structured UI generation tasks.

Key findings:

- MoE models (Gemma 4) offer the best throughput-to-parameter ratio on memory-bandwidth-limited hardware
- Dense models with fewer parameters achieve lower latency per sample
- All tested models benefit from the engine’s prefix caching for the shared A2UI system prompt

Grading Methodology

Judge

The semantic grade column (C/P/I) in this report was assigned by:

Claude Code (interactive, human-supervised)

Grading sessions

- Graded in Claude Code session 2026-05-12T06:26:07Z

Rubric

Each generated A2UI message was judged against the prompt's target description and assigned one of three grades:

- **C (Correct)** – the response satisfies the target. Optional/cosmetic additions and ID/label variations are accepted.
- **P (Partial)** – the response is substantively right but has minor variations (e.g. wrong wrapper format around correct components, slight punctuation differences).
- **I (Incorrect)** – the response is missing required components or has substantive errors (wrong variant, missing fields, broken structure, fabricated component names).

The rubric is the same one used by the production Inspect-AI grader (`A2UI/eval/tasks.py:GRADER_INSTRUCT`). Variations in capitalization, punctuation, component IDs, label synonyms, and data-binding paths are explicitly allowed.

Per-Sample Grading Rationales

Each row references one of the 73 evaluation prompts by its dataset **Sample** slug. The slug is consistent across all per-model tables, so a given row identifies the same prompt for every model. The rationale is the one-line note recorded at grading time; rationales over 240 characters are truncated (the full text is in the result JSON).

For reference, the first prompt `text_heading` reads in full:

Render a single Text component with the heading 'Settings' using the h1 variant.

with grading target:

A valid A2UI response with a Column root containing a single Text component whose variant is h1 and whose content is 'Settings'.

Llama 3.1 8B

Sample	Grade	Rationale
text_heading	I	Schema validation failed; output is not a valid A2UI tree.
text_caption	I	Schema validation failed; output is not a valid A2UI tree.
text_markdown	I	Schema validation failed; output is not a valid A2UI tree.
image_avatar	I	Schema validation failed; output is not a valid A2UI tree.
image_header_banner	I	Schema validation failed; output is not a valid A2UI tree.
icon_inline	I	Schema validation failed; output is not a valid A2UI tree.
video_player	I	Schema validation failed; output is not a valid A2UI tree.
audio_player	I	Schema validation failed; output is not a valid A2UI tree.
divider_horizontal	I	Schema validation failed; output is not a valid A2UI tree.
button_primary	I	Schema validation failed; output is not a valid A2UI tree.
button_borderless	I	Schema validation failed; output is not a valid A2UI tree.
textfield_password	I	Schema validation failed; output is not a valid A2UI tree.
text_then_button	I	Schema validation failed; output is not a valid A2UI tree.
image_caption	I	Schema validation failed; output is not a valid A2UI tree.
icon_label	I	Schema validation failed; output is not a valid A2UI tree.
two_buttons_row	I	Schema validation failed; output is not a valid A2UI tree.
heading_paragraph	I	Schema validation failed; output is not a valid A2UI tree.
divider_between_texts	I	Schema validation failed; output is not a valid A2UI tree.
checkbox_simple	I	Schema validation failed; output is not a valid A2UI tree.
slider_volume	I	Schema validation failed; output is not a valid A2UI tree.
date_picker_only	I	Schema validation failed; output is not a valid A2UI tree.
choice_picker_chips	I	Schema validation failed; output is not a valid A2UI tree.
textfield_search	I	Schema validation failed; output is not a valid A2UI tree.
textfield_multiline	I	Schema validation failed; output is not a valid A2UI tree.
card_with_text	I	Schema validation failed; output is not a valid A2UI tree.
card_with_image_text_button	I	Schema validation failed; output is not a valid A2UI tree.
card_profile_summary	I	Schema validation failed; output is not a valid A2UI tree.
card_stats	I	Schema validation failed; output is not a valid A2UI tree.
card_with_divider	I	Schema validation failed; output is not a valid A2UI tree.
card_with_icon_title	I	Schema validation failed; output is not a valid A2UI tree.
notification_card	I	Schema validation failed; output is not a valid A2UI tree.
card_quote	I	Schema validation failed; output is not a valid A2UI tree.
list_simple_text	I	Schema validation failed; output is not a valid A2UI tree.
list_horizontal_cards	I	Schema validation failed; output is not a valid A2UI tree.
list_checkable_todo	I	Schema validation failed; output is not a valid A2UI tree.
list_settings_menu	I	Schema validation failed; output is not a valid A2UI tree.
list_with_dividers	I	Schema validation failed; output is not a valid A2UI tree.
list_users_avatars	I	Schema validation failed; output is not a valid A2UI tree.
list_notifications	I	Schema validation failed; output is not a valid A2UI tree.
list_pricing_options	I	Schema validation failed; output is not a valid A2UI tree.
form_login	I	Schema validation failed; output is not a valid A2UI tree.
form_signup	I	Schema validation failed; output is not a valid A2UI tree.
form_contact	I	Schema validation failed; output is not a valid A2UI tree.
form_reservation	I	Schema validation failed; output is not a valid A2UI tree.
form_search_with_filter	I	Schema validation failed; output is not a valid A2UI tree.
form_subscribe	I	Schema validation failed; output is not a valid A2UI tree.
form_feedback_rating	I	Schema validation failed; output is not a valid A2UI tree.
form_settings_toggles	I	Schema validation failed; output is not a valid A2UI tree.
form_address	I	Schema validation failed; output is not a valid A2UI tree.
form_event	I	Schema validation failed; output is not a valid A2UI tree.
tabs_two_simple	I	Schema validation failed; output is not a valid A2UI tree.
tabs_three_panels	I	Schema validation failed; output is not a valid A2UI tree.
tabs_dashboard	I	Schema validation failed; output is not a valid A2UI tree.
tabs_settings_sections	I	Schema validation failed; output is not a valid A2UI tree.
modal_confirm	I	Schema validation failed; output is not a valid A2UI tree.
modal_info	I	Schema validation failed; output is not a valid A2UI tree.
modal_signup_form	I	Schema validation failed; output is not a valid A2UI tree.
modal_image_zoom	I	Schema validation failed; output is not a valid A2UI tree.
nested_layout	I	Schema validation failed; output is not a valid A2UI tree.
simple_column_layout	I	Schema validation failed; output is not a valid A2UI tree.
row_with_buttons	I	Schema validation failed; output is not a valid A2UI tree.
simple_text	I	Schema validation failed; output is not a valid A2UI tree.
dashboard_overview	I	Schema validation failed; output is not a valid A2UI tree.
product_detail	I	Schema validation failed; output is not a valid A2UI tree.
chat_message_thread	I	Schema validation failed; output is not a valid A2UI tree.
profile_screen	I	Schema validation failed; output is not a valid A2UI tree.
shopping_cart_summary	I	Schema validation failed; output is not a valid A2UI tree.

Llama 3.1 8B

Sample	Grade	
text_heading	P	Text with variant h1 is present, but the target's required Column ro
text_caption	I	Schema validation failed; output is
text_markdown	I	Schema validation failed; output is
image_avatar	C	Schema valid and completion produces the components described in the target (v
image_header_banner	I	Schema validation failed; output is
icon_inline	I	Schema validation failed; output is
video_player	I	Schema validation failed; output is
audio_player	I	Schema validation failed; output is
divider_horizontal	C	Schema valid and completion produces the components described in the target (v
button_primary	I	Schema validation failed; output is
button_borderless	I	Schema validation failed; output is
textfield_password	I	Schema validation failed; output is
text_then_button	I	Schema validation failed; output is
image_caption	I	Schema validation failed; output is
icon_label	I	Schema validation failed; output is
two_buttons_row	I	Schema validation failed; output is
heading_paragraph	I	Schema validation failed; output is
divider_between_texts	I	Schema validation failed; output is
checkbox_simple	I	Schema validation failed; output is
slider_volume	I	Schema validation failed; output is
date_picker_only	I	Schema validation failed; output is
choice_picker_chips	I	Schema validation failed; output is
textfield_search	I	Schema validation failed; output is
textfield_multiline	I	Schema validation failed; output is
card_with_text	I	Schema validation failed; output is
card_with_image_text_button	I	Schema validation failed; output is
card_profile_summary	I	Schema validation failed; output is
card_stats	I	Schema validation failed; output is
card_with_divider	I	Schema validation failed; output is
card_with_icon_title	I	Schema validation failed; output is
notification_card	I	Schema validation failed; output is
card_quote	I	Schema validation failed; output is
list_simple_text	I	Schema validation failed; output is
list_horizontal_cards	I	Schema validation failed; output is
list_checkable_todo	I	Schema validation failed; output is
list_settings_menu	I	Schema validation failed; output is
list_with_dividers	I	Schema validation failed; output is
list_users_avatars	I	Schema validation failed; output is
list_notifications	I	Schema validation failed; output is
list_pricing_options	I	Schema validation failed; output is
form_login	I	Schema validation failed; output is
form_signup	I	Schema validation failed; output is
form_contact	I	Schema validation failed; output is
form_reservation	I	Schema validation failed; output is
form_search_with_filter	I	Schema validation failed; output is
form_subscribe	I	Schema validation failed; output is
form_feedback_rating	I	Schema validation failed; output is
form_settings_toggles	I	Schema validation failed; output is
form_address	I	Schema validation failed; output is
form_event	I	Schema validation failed; output is
tabs_two_simple	I	Schema validation failed; output is
tabs_three_panels	I	Schema validation failed; output is
tabs_dashboard	I	Schema validation failed; output is
tabs_settings_sections	I	Schema validation failed; output is
modal_confirm	I	Schema validation failed; output is
modal_info	I	Schema validation failed; output is
modal_signup_form	I	Schema validation failed; output is
modal_image_zoom	I	Schema validation failed; output is
nested_layout	I	Schema validation failed; output is
simple_column_layout	I	Schema validation failed; output is
row_with_buttons	I	Schema validation failed; output is
simple_text	I	Schema validation failed; output is
dashboard_overview	I	Schema validation failed; output is
product_detail	I	Schema validation failed; output is
chat_message_thread	I	Schema validation failed; output is
profile_screen	I	Schema validation failed; output is
shopping_cart_summary	I	Schema validation failed; output is

Qwen 3 1.7B

Sample	Grade	
text_heading	I	Schema validation failed; ou
text_caption	I	Schema validation failed; ou
text_markdown	I	Schema validation failed; ou
image_avatar	I	Schema validation failed; ou
image_header_banner	I	Schema validation failed; ou
icon_inline	I	Schema validation failed; ou
video_player	I	Schema validation failed; ou
audio_player	I	Schema validation failed; ou
divider_horizontal	I	Schema validation failed; ou
button_primary	I	Schema validation failed; ou
button_borderless	I	Schema validation failed; ou
textfield_password	I	Schema validation failed; ou
text_then_button	I	Schema validation failed; ou
image_caption	I	Schema validation failed; ou
icon_label	I	Schema validation failed; ou
two_buttons_row	I	Schema validation failed; ou
heading_paragraph	I	Schema validation failed; ou
divider_between_texts	I	Schema validation failed; ou
checkbox_simple	I	Schema validation failed; ou
slider_volume	I	Schema validation failed; ou
date_picker_only	I	Schema validation failed; ou
choice_picker_chips	I	Schema validation failed; ou
textfield_search	I	Schema validation failed; ou
textfield_multiline	I	Schema validation failed; ou
card_with_text	I	Schema validation failed; ou
card_with_image_text_button	I	Schema validation failed; ou
card_profile_summary	I	Schema validation failed; ou
card_stats	I	Schema validation failed; ou
card_with_divider	I	Schema validation failed; ou
card_with_icon_title	I	Schema validation failed; ou
notification_card	I	Schema validation failed; ou
card_quote	I	Schema validation failed; ou
list_simple_text	I	Schema validation failed; ou
list_horizontal_cards	I	Schema validation failed; ou
list_checkable_todo	I	Schema validation failed; ou
list_settings_menu	I	Schema validation failed; ou
list_with_dividers	I	Schema validation failed; ou
list_users_avatars	I	Schema validation failed; ou
list_notifications	I	Schema validation failed; ou
list_pricing_options	I	Schema validation failed; ou
form_login	I	Schema validation failed; ou
form_signup	I	Schema validation failed; ou
form_contact	I	Schema validation failed; ou
form_reservation	I	Schema validation failed; ou
form_search_with_filter	I	Schema validation failed; ou
form_subscribe	I	Schema validation failed; ou
form_feedback_rating	I	Schema validation failed; ou
form_settings_toggles	I	Schema validation failed; ou
form_address	I	Schema validation failed; ou
form_event	I	Schema validation failed; ou
tabs_two_simple	I	Schema validation failed; ou
tabs_three_panels	I	Schema validation failed; ou
tabs_dashboard	I	Schema validation failed; ou
tabs_settings_sections	I	Schema validation failed; ou
modal_confirm	I	Schema validation failed; ou
modal_info	I	Schema validation failed; ou
modal_signup_form	I	Output is only a createSurface envelope; missing the required Modal + Column w
modal_image_zoom	I	Schema validation failed; ou
nested_layout	I	Schema validation failed; ou
simple_column_layout	I	Schema validation failed; ou
row_with_buttons	I	Schema validation failed; ou
simple_text	I	Schema validation failed; ou
dashboard_overview	I	Schema validation failed; ou
product_detail	I	Schema validation failed; ou
chat_message_thread	I	Schema validation failed; ou
profile_screen	I	Schema validation failed; ou
shopping_cart_summary	I	Schema validation failed; ou

Qwen 3 1.7B

Sample	Grade	Rationale
text_heading	I	Schema validation failed; output is not a valid A2UI tree.
text_caption	I	Schema validation failed; output is not a valid A2UI tree.
text_markdown	I	Schema validation failed; output is not a valid A2UI tree.
image_avatar	I	Schema validation failed; output is not a valid A2UI tree.
image_header_banner	I	Schema validation failed; output is not a valid A2UI tree.
icon_inline	I	Schema validation failed; output is not a valid A2UI tree.
video_player	I	Schema validation failed; output is not a valid A2UI tree.
audio_player	I	Schema validation failed; output is not a valid A2UI tree.
divider_horizontal	I	Schema validation failed; output is not a valid A2UI tree.
button_primary	I	Schema validation failed; output is not a valid A2UI tree.
button_borderless	I	Schema validation failed; output is not a valid A2UI tree.
textfield_password	I	Schema validation failed; output is not a valid A2UI tree.
text_then_button	I	Schema validation failed; output is not a valid A2UI tree.
image_caption	I	Schema validation failed; output is not a valid A2UI tree.
icon_label	I	Schema validation failed; output is not a valid A2UI tree.
two_buttons_row	I	Schema validation failed; output is not a valid A2UI tree.
heading_paragraph	I	Schema validation failed; output is not a valid A2UI tree.
divider_between_texts	I	Schema validation failed; output is not a valid A2UI tree.
checkbox_simple	I	Schema validation failed; output is not a valid A2UI tree.
slider_volume	I	Schema validation failed; output is not a valid A2UI tree.
date_picker_only	I	Schema validation failed; output is not a valid A2UI tree.
choice_picker_chips	I	Schema validation failed; output is not a valid A2UI tree.
textfield_search	I	Schema validation failed; output is not a valid A2UI tree.
textfield_multiline	I	Schema validation failed; output is not a valid A2UI tree.
card_with_text	I	Schema validation failed; output is not a valid A2UI tree.
card_with_image_text_button	I	Schema validation failed; output is not a valid A2UI tree.
card_profile_summary	I	Schema validation failed; output is not a valid A2UI tree.
card_stats	I	Schema validation failed; output is not a valid A2UI tree.
card_with_divider	I	Schema validation failed; output is not a valid A2UI tree.
card_with_icon_title	I	Schema validation failed; output is not a valid A2UI tree.
notification_card	I	Schema validation failed; output is not a valid A2UI tree.
card_quote	I	Schema validation failed; output is not a valid A2UI tree.
list_simple_text	I	Schema validation failed; output is not a valid A2UI tree.
list_horizontal_cards	I	Schema validation failed; output is not a valid A2UI tree.
list_checkable_todo	I	Schema validation failed; output is not a valid A2UI tree.
list_settings_menu	I	Schema validation failed; output is not a valid A2UI tree.
list_with_dividers	I	Schema validation failed; output is not a valid A2UI tree.
list_users_avatars	I	Schema validation failed; output is not a valid A2UI tree.
list_notifications	I	Schema validation failed; output is not a valid A2UI tree.
list_pricing_options	I	Schema validation failed; output is not a valid A2UI tree.
form_login	I	Schema validation failed; output is not a valid A2UI tree.
form_signup	I	Schema validation failed; output is not a valid A2UI tree.
form_contact	I	Schema validation failed; output is not a valid A2UI tree.
form_reservation	I	Schema validation failed; output is not a valid A2UI tree.
form_search_with_filter	I	Schema validation failed; output is not a valid A2UI tree.
form_subscribe	I	Schema validation failed; output is not a valid A2UI tree.
form_feedback_rating	I	Schema validation failed; output is not a valid A2UI tree.
form_settings_toggles	I	Schema validation failed; output is not a valid A2UI tree.
form_address	I	Schema validation failed; output is not a valid A2UI tree.
form_event	I	Schema validation failed; output is not a valid A2UI tree.
tabs_two_simple	I	Schema validation failed; output is not a valid A2UI tree.
tabs_three_panels	I	Schema validation failed; output is not a valid A2UI tree.
tabs_dashboard	I	Schema validation failed; output is not a valid A2UI tree.
tabs_settings_sections	I	Schema validation failed; output is not a valid A2UI tree.
modal_confirm	I	Schema validation failed; output is not a valid A2UI tree.
modal_info	I	Schema validation failed; output is not a valid A2UI tree.
modal_signup_form	I	Schema validation failed; output is not a valid A2UI tree.
modal_image_zoom	I	Schema validation failed; output is not a valid A2UI tree.
nested_layout	I	Schema validation failed; output is not a valid A2UI tree.
simple_column_layout	I	Schema validation failed; output is not a valid A2UI tree.
row_with_buttons	I	Schema validation failed; output is not a valid A2UI tree.
simple_text	I	Schema validation failed; output is not a valid A2UI tree.
dashboard_overview	I	Schema validation failed; output is not a valid A2UI tree.
product_detail	I	Schema validation failed; output is not a valid A2UI tree.
chat_message_thread	I	Schema validation failed; output is not a valid A2UI tree.
profile_screen	I	Schema validation failed; output is not a valid A2UI tree.
shopping_cart_summary	I	Schema validation failed; output is not a valid A2UI tree.

Qwen 3 14B

Sample	Grade	
text_heading	C	Schema valid and completion produces the components described in the target (ve
text_caption	C	Schema valid and completion produces the components described in the target (ve
text_markdown	C	Schema valid and completion produces the components described in the target (ve
image_avatar	C	Schema valid and completion produces the components described in the target (ve
image_header_banner	C	Schema valid and completion produces the components described in the target (ve
icon_inline	C	Schema valid and completion produces the components described in the target (ve
video_player	C	Schema valid and completion produces the components described in the target (ve
audio_player	C	Schema valid and completion produces the components described in the target (ve
divider_horizontal	C	Schema valid and completion produces the components described in the target (ve
button_primary	C	Schema valid and completion produces the components described in the target (ve
button_borderless	C	Schema valid and completion produces the components described in the target (ve
textfield_password	I	Schema validation failed; output is
text_then_button	C	Schema valid and completion produces the components described in the target (ve
image_caption	C	Schema valid and completion produces the components described in the target (ve
icon_label	C	Schema valid and completion produces the components described in the target (ve
two_buttons_row	C	Schema valid and completion produces the components described in the target (ve
heading_paragraph	C	Schema valid and completion produces the components described in the target (ve
divider_between_texts	C	Schema valid and completion produces the components described in the target (ve
checkbox_simple	I	Schema validation failed; output is
slider_volume	C	Schema valid and completion produces the components described in the target (ve
date_picker_only	C	Schema valid and completion produces the components described in the target (ve
choice_picker_chips	C	Schema valid and completion produces the components described in the target (ve
textfield_search	I	Schema validation failed; output is
textfield_multiline	I	Schema validation failed; output is
card_with_text	C	Schema valid and completion produces the components described in the target (ve
card_with_image_text_button	C	Schema valid and completion produces the components described in the target (ve
card_profile_summary	I	Schema validation failed; output is
card_stats	C	Schema valid and completion produces the components described in the target (ve
card_with_divider	C	Schema valid and completion produces the components described in the target (ve
card_with_icon_title	C	Schema valid and completion produces the components described in the target (ve
notification_card	C	Schema valid and completion produces the components described in the target (ve
card_quote	C	Schema valid and completion produces the components described in the target (ve
list_simple_text	C	Schema valid and completion produces the components described in the target (ve
list_horizontal_cards	I	Schema validation failed; output is
list_checkable_todo	C	Schema valid and completion produces the components described in the target (ve
list_settings_menu	C	Schema valid and completion produces the components described in the target (ve
list_with_dividers	I	Schema validation failed; output is
list_users_avatars	I	Schema validation failed; output is
list_notifications	I	Schema validation failed; output is
list_pricing_options	I	Schema validation failed; output is
form_login	I	Schema validation failed; output is
form_signup	I	Schema validation failed; output is
form_contact	I	Schema validation failed; output is
form_reservation	I	Schema validation failed; output is

Qwen 3 14B

Sample	Grade	
text_heading	C	Schema valid and completion produces the components described in the target (ve
text_caption	C	Schema valid and completion produces the components described in the target (ve
text_markdown	C	Schema valid and completion produces the components described in the target (ve
image_avatar	C	Schema valid and completion produces the components described in the target (ve
image_header_banner	C	Schema valid and completion produces the components described in the target (ve
icon_inline	C	Schema valid and completion produces the components described in the target (ve
video_player	C	Schema valid and completion produces the components described in the target (ve
audio_player	C	Schema valid and completion produces the components described in the target (ve
divider_horizontal	C	Schema valid and completion produces the components described in the target (ve
button_primary	C	Schema valid and completion produces the components described in the target (ve
button_borderless	C	Schema valid and completion produces the components described in the target (ve
textfield_password	I	Schema validation failed; output is
text_then_button	C	Schema valid and completion produces the components described in the target (ve
image_caption	C	Schema valid and completion produces the components described in the target (ve
icon_label	C	Schema valid and completion produces the components described in the target (ve
two_buttons_row	C	Schema valid and completion produces the components described in the target (ve
heading_paragraph	C	Schema valid and completion produces the components described in the target (ve
divider_between_texts	C	Schema valid and completion produces the components described in the target (ve
checkbox_simple	C	Schema valid and completion produces the components described in the target (ve
slider_volume	C	Schema valid and completion produces the components described in the target (ve
date_picker_only	C	Schema valid and completion produces the components described in the target (ve
choice_picker_chips	C	Schema valid and completion produces the components described in the target (ve
textfield_search	C	Schema valid and completion produces the components described in the target (ve
textfield_multiline	C	Schema valid and completion produces the components described in the target (ve
card_with_text	C	Schema valid and completion produces the components described in the target (ve
card_with_image_text_button	C	Schema valid and completion produces the components described in the target (ve
card_profile_summary	C	Schema valid and completion produces the components described in the target (ve
card_stats	C	Schema valid and completion produces the components described in the target (ve
card_with_divider	C	Schema valid and completion produces the components described in the target (ve
card_with_icon_title	C	Schema valid and completion produces the components described in the target (ve
notification_card	C	Schema valid and completion produces the components described in the target (ve
card_quote	C	Schema valid and completion produces the components described in the target (ve
list_simple_text	C	Schema valid and completion produces the components described in the target (ve
list_horizontal_cards	C	Schema valid and completion produces the components described in the target (ve
list_checkable_todo	C	Schema valid and completion produces the components described in the target (ve
list_settings_menu	C	Schema valid and completion produces the components described in the target (ve
list_with_dividers	C	Schema valid and completion produces the components described in the target (ve
list_users_avatars	C	Schema valid and completion produces the components described in the target (ve
list_notifications	C	Schema valid and completion produces the components described in the target (ve
list_pricing_options	C	Schema valid and completion produces the components described in the target (ve
form_login	C	Schema valid and completion produces the components described in the target (ve
form_signup	C	Schema valid and completion produces the components described in the target (ve
form_contact	C	Schema valid and completion produces the components described in the target (ve
form_reservation	I	Schema validation failed; output is
form_search_with_filter	C	Schema valid and completion produces the components described in the target (ve
form_subscribe	C	Schema valid and completion produces the components described in the target (ve
form_feedback_rating	C	Schema valid and completion produces the components described in the target (ve
form_settings_toggles	C	Schema valid and completion produces the components described in the target (ve
form_address	C	Schema valid and completion produces the components described in the target (ve
form_event	C	Schema valid and completion produces the components described in the target (ve
tabs_two_simple	C	Schema valid and completion produces the components described in the target (ve
tabs_three_panels	C	Schema valid and completion produces the components described in the target (ve
tabs_dashboard	C	Schema valid and completion produces the components described in the target (ve
tabs_settings_sections	C	Schema valid and completion produces the components described in the target (ve
modal_confirm	I	Schema validation failed; output is
modal_info	I	Schema validation failed; output is
modal_signup_form	I	Schema validation failed; output is
modal_image_zoom	I	Schema validation failed; output is
nested_layout	C	Schema valid and completion produces the components described in the target (ve
simple_column_layout	C	Schema valid and completion produces the components described in the target (ve
row_with_buttons	C	Schema valid and completion produces the components described in the target (ve
simple_text	C	Schema valid and completion produces the components described in the target (ve
dashboard_overview	C	Schema valid and completion produces the components described in the target (ve
product_detail	C	Schema valid and completion produces the components described in the target (ve
chat_message_thread	C	Schema valid and completion produces the components described in the target (ve
profile_screen	C	Schema valid and completion produces the components described in the target (ve
shopping_cart_summary	C	Schema valid and completion produces the components described in the target (ve

Qwen 3 8B

Sample	Grade	
text_heading	P	Text with variant h1 is present, but the target's required Column ro
text_caption	I	Schema validation failed; output is
text_markdown	I	Schema validation failed; output is
image_avatar	I	Schema validation failed; output is
image_header_banner	I	Schema validation failed; output is
icon_inline	I	Schema validation failed; output is
video_player	I	Schema validation failed; output is
audio_player	C	Schema valid and completion produces the components described in the target (v
divider_horizontal	I	Schema validation failed; output is
button_primary	I	Schema validation failed; output is
button_borderless	I	Schema validation failed; output is
textfield_password	I	Schema validation failed; output is
text_then_button	I	Schema validation failed; output is
image_caption	I	Schema validation failed; output is
icon_label	I	Schema validation failed; output is
two_buttons_row	I	Schema validation failed; output is
heading_paragraph	I	Schema validation failed; output is
divider_between_texts	I	Schema validation failed; output is
checkbox_simple	I	Schema validation failed; output is
slider_volume	I	Schema validation failed; output is
date_picker_only	I	Schema validation failed; output is
choice_picker_chips	I	Schema validation failed; output is
textfield_search	I	Schema validation failed; output is
textfield_multiline	I	Schema validation failed; output is
card_with_text	I	Schema validation failed; output is
card_with_image_text_button	I	Schema validation failed; output is
card_profile_summary	I	Schema validation failed; output is
card_stats	I	Schema validation failed; output is
card_with_divider	I	Schema validation failed; output is
card_with_icon_title	I	Schema validation failed; output is
notification_card	I	Schema validation failed; output is
card_quote	I	Schema validation failed; output is
list_simple_text	P	Three Text items are present and ordered, but a Column is used instead
list_horizontal_cards	I	Schema validation failed; output is
list_checkable_todo	I	Schema validation failed; output is
list_settings_menu	I	Schema validation failed; output is
list_with_dividers	I	Schema validation failed; output is
list_users_avatars	I	Schema validation failed; output is
list_notifications	I	Schema validation failed; output is
list_pricing_options	I	Schema validation failed; output is
form_login	I	Schema validation failed; output is
form_signup	I	Schema validation failed; output is
form_contact	I	Schema validation failed; output is
form_reservation	I	Schema validation failed; output is
form_search_with_filter	I	Schema validation failed; output is
form_subscribe	I	Schema validation failed; output is
form_feedback_rating	I	Schema validation failed; output is
form_settings_toggles	I	Schema validation failed; output is
form_address	I	Schema validation failed; output is
form_event	I	Schema validation failed; output is
tabs_two_simple	I	Schema validation failed; output is
tabs_three_panels	I	Schema validation failed; output is
tabs_dashboard	I	Schema validation failed; output is
tabs_settings_sections	I	Schema validation failed; output is
modal_confirm	I	Schema validation failed; output is
modal_info	I	Schema validation failed; output is
modal_signup_form	I	Schema validation failed; output is

Qwen 3 8B

Sample	Grade	
text_heading	I	Schema validation failed; output is
text_caption	I	Schema validation failed; output is
text_markdown	I	Schema validation failed; output is
image_avatar	I	Schema validation failed; output is
image_header_banner	I	Schema validation failed; output is
icon_inline	I	Schema validation failed; output is
video_player	I	Schema validation failed; output is
audio_player	I	Schema validation failed; output is
divider_horizontal	I	Schema validation failed; output is
button_primary	C	Schema valid and completion produces the components described in the target (v
button_borderless	I	Schema validation failed; output is
textfield_password	I	Schema validation failed; output is
text_then_button	I	Schema validation failed; output is
image_caption	I	Schema validation failed; output is
icon_label	I	Schema validation failed; output is
two_buttons_row	I	Schema validation failed; output is
heading_paragraph	I	Schema validation failed; output is
divider_between_texts	I	Schema validation failed; output is
checkbox_simple	I	Schema validation failed; output is
slider_volume	I	Schema validation failed; output is
date_picker_only	I	Schema validation failed; output is
choice_picker_chips	I	Schema validation failed; output is
textfield_search	I	Schema validation failed; output is
textfield_multiline	I	Schema validation failed; output is
card_with_text	I	Schema validation failed; output is
card_with_image_text_button	I	Schema validation failed; output is
card_profile_summary	I	Schema validation failed; output is
card_stats	I	Schema validation failed; output is
card_with_divider	I	Schema validation failed; output is
card_with_icon_title	I	Schema validation failed; output is
notification_card	I	Schema validation failed; output is
card_quote	I	Schema validation failed; output is
list_simple_text	I	Schema validation failed; output is
list_horizontal_cards	I	Schema validation failed; output is
list_checkable_todo	I	Schema validation failed; output is
list_settings_menu	I	Schema validation failed; output is
list_with_dividers	I	Schema validation failed; output is
list_users_avatars	I	Schema validation failed; output is
list_notifications	I	Schema validation failed; output is
list_pricing_options	I	Schema validation failed; output is
form_login	I	Schema validation failed; output is
form_signup	I	Schema validation failed; output is
form_contact	I	Schema validation failed; output is
form_reservation	I	Schema validation failed; output is
form_search_with_filter	I	Schema validation failed; output is
form_subscribe	I	Schema validation failed; output is
form_feedback_rating	I	Schema validation failed; output is
form_settings_toggles	I	Schema validation failed; output is
form_address	I	Schema validation failed; output is
form_event	I	Schema validation failed; output is
tabs_two_simple	I	Schema validation failed; output is
tabs_three_panels	I	Schema validation failed; output is
tabs_dashboard	I	Schema validation failed; output is
tabs_settings_sections	I	Schema validation failed; output is
modal_confirm	I	Schema validation failed; output is
modal_info	I	Schema validation failed; output is
modal_signup_form	I	Schema validation failed; output is
modal_image_zoom	I	Schema validation failed; output is
nested_layout	I	Schema validation failed; output is
simple_column_layout	I	Schema validation failed; output is
row_with_buttons	I	Schema validation failed; output is
simple_text	I	Schema validation failed; output is
dashboard_overview	I	Schema validation failed; output is
product_detail	I	Schema validation failed; output is
chat_message_thread	I	Schema validation failed; output is
profile_screen	I	Schema validation failed; output is
shopping_cart_summary	I	Schema validation failed; output is

Gemma 4 26B MoE

Sample	Grade	
text_heading	C	Schema valid and completion produces the components described in the target (sp
text_caption	C	Schema valid and completion produces the components described in the target (sp
text_markdown	C	Schema valid and completion produces the components described in the target (sp
image_avatar	C	Schema valid and completion produces the components described in the target (sp
image_header_banner	C	Schema valid and completion produces the components described in the target (sp
icon_inline	C	Schema valid and completion produces the components described in the target (sp
video_player	C	Schema valid and completion produces the components described in the target (sp
audio_player	C	Schema valid and completion produces the components described in the target (sp
divider_horizontal	C	Schema valid and completion produces the components described in the target (sp
button_primary	C	Schema valid and completion produces the components described in the target (sp
button_borderless	C	Schema valid and completion produces the components described in the target (sp
textfield_password	C	Schema valid and completion produces the components described in the target (sp
text_then_button	C	Schema valid and completion produces the components described in the target (sp
image_caption	C	Schema valid and completion produces the components described in the target (sp
icon_label	C	Schema valid and completion produces the components described in the target (sp
two_buttons_row	C	Schema valid and completion produces the components described in the target (sp
heading_paragraph	C	Schema valid and completion produces the components described in the target (sp
divider_between_texts	C	Schema valid and completion produces the components described in the target (sp
checkbox_simple	C	Schema valid and completion produces the components described in the target (sp
slider_volume	C	Schema valid and completion produces the components described in the target (sp
date_picker_only	C	Schema valid and completion produces the components described in the target (sp
choice_picker_chips	C	Schema valid and completion produces the components described in the target (sp
textfield_search	C	Schema valid and completion produces the components described in the target (sp
textfield_multiline	C	Schema valid and completion produces the components described in the target (sp
card_with_text	C	Schema valid and completion produces the components described in the target (sp
card_with_image_text_button	C	Schema valid and completion produces the components described in the target (sp
card_profile_summary	C	Schema valid and completion produces the components described in the target (sp
card_stats	C	Schema valid and completion produces the components described in the target (sp
card_with_divider	C	Schema valid and completion produces the components described in the target (sp
card_with_icon_title	C	Schema valid and completion produces the components described in the target (sp
notification_card	C	Schema valid and completion produces the components described in the target (sp
card_quote	C	Schema valid and completion produces the components described in the target (sp
list_simple_text	C	Schema valid and completion produces the components described in the target (sp
list_horizontal_cards	C	Schema valid and completion produces the components described in the target (sp
list_checkable_todo	C	Schema valid and completion produces the components described in the target (sp
list_settings_menu	I	Schema validation failed; outp
list_with_dividers	C	Schema valid and completion produces the components described in the target (sp
list_users_avatars	C	Schema valid and completion produces the components described in the target (sp
list_notifications	C	Schema valid and completion produces the components described in the target (sp
list_pricing_options	C	Schema valid and completion produces the components described in the target (sp
form_login	C	Schema valid and completion produces the components described in the target (sp
form_signup	C	Schema valid and completion produces the components described in the target (sp
form_contact	C	Schema valid and completion produces the components described in the target (sp
form_reservation	C	Schema valid and completion produces the components described in the target (sp
form_search_with_filter	C	Schema valid and completion produces the components described in the target (sp
form_subscribe	C	Schema valid and completion produces the components described in the target (sp
form_feedback_rating	C	Schema valid and completion produces the components described in the target (sp
form_settings_toggles	C	Schema valid and completion produces the components described in the target (sp
form_address	C	Schema valid and completion produces the components described in the target (sp
form_event	C	Schema valid and completion produces the components described in the target (sp
tabs_two_simple	C	Schema valid and completion produces the components described in the target (sp
tabs_three_panels	C	Schema valid and completion produces the components described in the target (sp
tabs_dashboard	C	Schema valid and completion produces the components described in the target (sp
tabs_settings_sections	C	Schema valid and completion produces the components described in the target (sp
modal_confirm	I	Schema validation failed; outp
modal_info	I	Schema validation failed; outp
modal_signup_form	I	Schema validation failed; outp
modal_image_zoom	C	Schema valid and completion produces the components described in the target (sp
nested_layout	C	Schema valid and completion produces the components described in the target (sp
simple_column_layout	C	Schema valid and completion produces the components described in the target (sp
row_with_buttons	C	Schema valid and completion produces the components described in the target (sp
simple_text	C	Schema valid and completion produces the components described in the target (sp
dashboard_overview	C	Schema valid and completion produces the components described in the target (sp
product_detail	I	Schema validation failed; outp
chat_message_thread	C	Schema valid and completion produces the components described in the target (sp
profile_screen	C	Schema valid and completion produces the components described in the target (sp
shopping_cart_summary	C	Schema valid and completion produces the components described in the target (sp

Nemotron 3 Nano 30B

Sample	Grade	Rationale
text_heading	I	Schema validation failed; output is not a valid A2UI tree.
text_caption	I	Schema validation failed; output is not a valid A2UI tree.
text_markdown	I	Schema validation failed; output is not a valid A2UI tree.
image_avatar	I	Schema validation failed; output is not a valid A2UI tree.
image_header_banner	I	Schema validation failed; output is not a valid A2UI tree.
icon_inline	I	Schema validation failed; output is not a valid A2UI tree.
video_player	I	Schema validation failed; output is not a valid A2UI tree.
audio_player	I	Schema validation failed; output is not a valid A2UI tree.
divider_horizontal	I	Schema validation failed; output is not a valid A2UI tree.
button_primary	I	Schema validation failed; output is not a valid A2UI tree.
button_borderless	I	Schema validation failed; output is not a valid A2UI tree.
textfield_password	I	Schema validation failed; output is not a valid A2UI tree.
text_then_button	I	Schema validation failed; output is not a valid A2UI tree.
image_caption	I	Schema validation failed; output is not a valid A2UI tree.
icon_label	I	Schema validation failed; output is not a valid A2UI tree.
two_buttons_row	I	Schema validation failed; output is not a valid A2UI tree.
heading_paragraph	I	Schema validation failed; output is not a valid A2UI tree.
divider_between_texts	I	Schema validation failed; output is not a valid A2UI tree.
checkbox_simple	I	Schema validation failed; output is not a valid A2UI tree.
slider_volume	I	Schema validation failed; output is not a valid A2UI tree.
date_picker_only	I	Schema validation failed; output is not a valid A2UI tree.
choice_picker_chips	I	Schema validation failed; output is not a valid A2UI tree.
textfield_search	I	Schema validation failed; output is not a valid A2UI tree.
textfield_multiline	I	Schema validation failed; output is not a valid A2UI tree.
card_with_text	I	Schema validation failed; output is not a valid A2UI tree.
card_with_image_text_button	I	Schema validation failed; output is not a valid A2UI tree.
card_profile_summary	I	Schema validation failed; output is not a valid A2UI tree.
card_stats	I	Schema validation failed; output is not a valid A2UI tree.
card_with_divider	I	Schema validation failed; output is not a valid A2UI tree.
card_with_icon_title	I	Schema validation failed; output is not a valid A2UI tree.
notification_card	I	Schema validation failed; output is not a valid A2UI tree.
card_quote	I	Schema validation failed; output is not a valid A2UI tree.
list_simple_text	I	Schema validation failed; output is not a valid A2UI tree.
list_horizontal_cards	I	Schema validation failed; output is not a valid A2UI tree.
list_checkable_todo	I	Schema validation failed; output is not a valid A2UI tree.
list_settings_menu	I	Schema validation failed; output is not a valid A2UI tree.
list_with_dividers	I	Schema validation failed; output is not a valid A2UI tree.
list_users_avatars	I	Schema validation failed; output is not a valid A2UI tree.
list_notifications	I	Schema validation failed; output is not a valid A2UI tree.
list_pricing_options	I	Schema validation failed; output is not a valid A2UI tree.
form_login	I	Schema validation failed; output is not a valid A2UI tree.
form_signup	I	Schema validation failed; output is not a valid A2UI tree.
form_contact	I	Schema validation failed; output is not a valid A2UI tree.
form_reservation	I	Schema validation failed; output is not a valid A2UI tree.
form_search_with_filter	I	Schema validation failed; output is not a valid A2UI tree.
form_subscribe	I	Schema validation failed; output is not a valid A2UI tree.
form_feedback_rating	I	Schema validation failed; output is not a valid A2UI tree.
form_settings_toggles	I	Schema validation failed; output is not a valid A2UI tree.
form_address	I	Schema validation failed; output is not a valid A2UI tree.
form_event	I	Schema validation failed; output is not a valid A2UI tree.
tabs_two_simple	I	Schema validation failed; output is not a valid A2UI tree.
tabs_three_panels	I	Schema validation failed; output is not a valid A2UI tree.
tabs_dashboard	I	Schema validation failed; output is not a valid A2UI tree.
tabs_settings_sections	I	Schema validation failed; output is not a valid A2UI tree.
modal_confirm	I	Schema validation failed; output is not a valid A2UI tree.
modal_info	I	Schema validation failed; output is not a valid A2UI tree.
modal_signup_form	I	Schema validation failed; output is not a valid A2UI tree.
modal_image_zoom	I	Schema validation failed; output is not a valid A2UI tree.
nested_layout	I	Schema validation failed; output is not a valid A2UI tree.
simple_column_layout	I	Schema validation failed; output is not a valid A2UI tree.
row_with_buttons	I	Schema validation failed; output is not a valid A2UI tree.
simple_text	I	Schema validation failed; output is not a valid A2UI tree.
dashboard_overview	I	Schema validation failed; output is not a valid A2UI tree.
product_detail	I	Schema validation failed; output is not a valid A2UI tree.
chat_message_thread	I	Schema validation failed; output is not a valid A2UI tree.
profile_screen	I	Schema validation failed; output is not a valid A2UI tree.
shopping_cart_summary	I	Schema validation failed; output is not a valid A2UI tree.

Qwen3 30B MoE NVFP4

Sample	Grade	
text_heading	C	Schema valid and completion produces the components described in the target (v
text_caption	C	Schema valid and completion produces the components described in the target (v
text_markdown	C	Schema valid and completion produces the components described in the target (v
image_avatar	C	Schema valid and completion produces the components described in the target (v
image_header_banner	C	Schema valid and completion produces the components described in the target (v
icon_inline	C	Schema valid and completion produces the components described in the target (v
video_player	C	Schema valid and completion produces the components described in the target (v
audio_player	C	Schema valid and completion produces the components described in the target (v
divider_horizontal	C	Schema valid and completion produces the components described in the target (v
button_primary	C	Schema valid and completion produces the components described in the target (v
button_borderless	C	Schema valid and completion produces the components described in the target (v
textfield_password	C	Schema valid and completion produces the components described in the target (v
text_then_button	C	Schema valid and completion produces the components described in the target (v
image_caption	I	Schema validation failed; output is
icon_label	C	Schema valid and completion produces the components described in the target (v
two_buttons_row	C	Schema valid and completion produces the components described in the target (v
heading_paragraph	C	Schema valid and completion produces the components described in the target (v
divider_between_texts	C	Schema valid and completion produces the components described in the target (v
checkbox_simple	C	Schema valid and completion produces the components described in the target (v
slider_volume	I	Schema validation failed; output is
date_picker_only	I	Schema validation failed; output is
choice_picker_chips	I	Schema validation failed; output is
textfield_search	I	Schema validation failed; output is
textfield_multiline	I	Schema validation failed; output is
card_with_text	C	Schema valid and completion produces the components described in the target (v
card_with_image_text_button	C	Schema valid and completion produces the components described in the target (v
card_profile_summary	C	Schema valid and completion produces the components described in the target (v
card_stats	C	Schema valid and completion produces the components described in the target (v
card_with_divider	C	Schema valid and completion produces the components described in the target (v
card_with_icon_title	C	Schema valid and completion produces the components described in the target (v
notification_card	C	Schema valid and completion produces the components described in the target (v
card_quote	C	Schema valid and completion produces the components described in the target (v
list_simple_text	C	Schema valid and completion produces the components described in the target (v
list_horizontal_cards	C	Schema valid and completion produces the components described in the target (v
list_checkable_todo	C	Schema valid and completion produces the components described in the target (v
list_settings_menu	C	Schema valid and completion produces the components described in the target (v
list_with_dividers	C	Schema valid and completion produces the components described in the target (v
list_users_avatars	C	Schema valid and completion produces the components described in the target (v
list_notifications	I	Schema validation failed; output is
list_pricing_options	C	Schema valid and completion produces the components described in the target (v
form_login	I	Schema validation failed; output is
form_signup	I	Schema validation failed; output is
form_contact	I	Schema validation failed; output is
form_reservation	C	Schema valid and completion produces the components described in the target (v
form_search_with_filter	I	Schema validation failed; output is
form_subscribe	I	Schema validation failed; output is
form_feedback_rating	I	Schema validation failed; output is
form_settings_toggles	I	Schema validation failed; output is
form_address	I	Schema validation failed; output is
form_event	I	Schema validation failed; output is
tabs_two_simple	C	Schema valid and completion produces the components described in the target (v
tabs_three_panels	C	Schema valid and completion produces the components described in the target (v
tabs_dashboard	C	Schema valid and completion produces the components described in the target (v
tabs_settings_sections	C	Schema valid and completion produces the components described in the target (v
modal_confirm	I	Schema validation failed; output is
modal_info	I	Schema validation failed; output is
modal_signup_form	I	Schema validation failed; output is
modal_image_zoom	I	Schema validation failed; output is
nested_layout	C	Schema valid and completion produces the components described in the target (v
simple_column_layout	C	Schema valid and completion produces the components described in the target (v
row_with_buttons	C	Schema valid and completion produces the components described in the target (v
simple_text	C	Schema valid and completion produces the components described in the target (v
dashboard_overview	I	Schema validation failed; output is
product_detail	C	Schema valid and completion produces the components described in the target (v
chat_message_thread	C	Schema valid and completion produces the components described in the target (v
profile_screen	I	Schema validation failed; output is
shopping_cart_summary	C	Schema valid and completion produces the components described in the target (v

Qwen3 30B MoE NVFP4

Sample	Grade	
text_heading	C	Schema valid and completion produces the components described in the target (v
text_caption	C	Schema valid and completion produces the components described in the target (v
text_markdown	C	Schema valid and completion produces the components described in the target (v
image_avatar	C	Schema valid and completion produces the components described in the target (v
image_header_banner	C	Schema valid and completion produces the components described in the target (v
icon_inline	C	Schema valid and completion produces the components described in the target (v
video_player	C	Schema valid and completion produces the components described in the target (v
audio_player	C	Schema valid and completion produces the components described in the target (v
divider_horizontal	C	Schema valid and completion produces the components described in the target (v
button_primary	I	Schema validation failed; output is
button_borderless	C	Schema valid and completion produces the components described in the target (v
textfield_password	C	Schema valid and completion produces the components described in the target (v
text_then_button	C	Schema valid and completion produces the components described in the target (v
image_caption	C	Schema valid and completion produces the components described in the target (v
icon_label	C	Schema valid and completion produces the components described in the target (v
two_buttons_row	C	Schema valid and completion produces the components described in the target (v
heading_paragraph	I	Schema validation failed; output is
divider_between_texts	C	Schema valid and completion produces the components described in the target (v
checkbox_simple	C	Schema valid and completion produces the components described in the target (v
slider_volume	C	Schema valid and completion produces the components described in the target (v
date_picker_only	C	Schema valid and completion produces the components described in the target (v
choice_picker_chips	C	Schema valid and completion produces the components described in the target (v
textfield_search	C	Schema valid and completion produces the components described in the target (v
textfield_multiline	C	Schema valid and completion produces the components described in the target (v
card_with_text	C	Schema valid and completion produces the components described in the target (v
card_with_image_text_button	C	Schema valid and completion produces the components described in the target (v
card_profile_summary	C	Schema valid and completion produces the components described in the target (v
card_stats	C	Schema valid and completion produces the components described in the target (v
card_with_divider	C	Schema valid and completion produces the components described in the target (v
card_with_icon_title	C	Schema valid and completion produces the components described in the target (v
notification_card	C	Schema valid and completion produces the components described in the target (v
card_quote	C	Schema valid and completion produces the components described in the target (v
list_simple_text	C	Schema valid and completion produces the components described in the target (v
list_horizontal_cards	C	Schema valid and completion produces the components described in the target (v
list_checkable_todo	C	Schema valid and completion produces the components described in the target (v
list_settings_menu	C	Schema valid and completion produces the components described in the target (v
list_with_dividers	C	Schema valid and completion produces the components described in the target (v
list_users_avatars	C	Schema valid and completion produces the components described in the target (v
list_notifications	I	Schema validation failed; output is
list_pricing_options	C	Schema valid and completion produces the components described in the target (v
form_login	C	Schema valid and completion produces the components described in the target (v
form_signup	I	Schema validation failed; output is
form_contact	I	Schema validation failed; output is
form_reservation	I	Schema validation failed; output is
form_search_with_filter	I	Schema validation failed; output is
form_subscribe	C	Schema valid and completion produces the components described in the target (v
form_feedback_rating	I	Schema validation failed; output is
form_settings_toggles	I	Schema validation failed; output is
form_address	I	Schema validation failed; output is
form_event	I	Schema validation failed; output is
tabs_two_simple	C	Schema valid and completion produces the components described in the target (v
tabs_three_panels	C	Schema valid and completion produces the components described in the target (v
tabs_dashboard	C	Schema valid and completion produces the components described in the target (v
tabs_settings_sections	C	Schema valid and completion produces the components described in the target (v
modal_confirm	C	Schema valid and completion produces the components described in the target (v
modal_info	I	Schema validation failed; output is
modal_signup_form	C	Schema valid and completion produces the components described in the target (v
modal_image_zoom	I	Schema validation failed; output is
nested_layout	C	Schema valid and completion produces the components described in the target (v
simple_column_layout	C	Schema valid and completion produces the components described in the target (v
row_with_buttons	C	Schema valid and completion produces the components described in the target (v
simple_text	C	Schema valid and completion produces the components described in the target (v
dashboard_overview	C	Schema valid and completion produces the components described in the target (v
product_detail	C	Schema valid and completion produces the components described in the target (v
chat_message_thread	C	Schema valid and completion produces the components described in the target (v
profile_screen	I	Schema validation failed; output is
shopping_cart_summary	C	Schema valid and completion produces the components described in the target (v

Llama 3.2 1B

Sample	Grade	Rationale
text_heading	I	Schema validation failed; output is not a valid A2UI tree.
text_caption	I	Schema validation failed; output is not a valid A2UI tree.
text_markdown	I	Schema validation failed; output is not a valid A2UI tree.
image_avatar	I	Schema validation failed; output is not a valid A2UI tree.
image_header_banner	I	Schema validation failed; output is not a valid A2UI tree.
icon_inline	I	Schema validation failed; output is not a valid A2UI tree.
video_player	I	Schema validation failed; output is not a valid A2UI tree.
audio_player	I	Schema validation failed; output is not a valid A2UI tree.
divider_horizontal	I	Schema validation failed; output is not a valid A2UI tree.
button_primary	I	Schema validation failed; output is not a valid A2UI tree.
button_borderless	I	Schema validation failed; output is not a valid A2UI tree.
textfield_password	I	Schema validation failed; output is not a valid A2UI tree.
text_then_button	I	Schema validation failed; output is not a valid A2UI tree.
image_caption	I	Schema validation failed; output is not a valid A2UI tree.
icon_label	I	Schema validation failed; output is not a valid A2UI tree.
two_buttons_row	I	Schema validation failed; output is not a valid A2UI tree.
heading_paragraph	I	Schema validation failed; output is not a valid A2UI tree.
divider_between_texts	I	Schema validation failed; output is not a valid A2UI tree.
checkbox_simple	I	Schema validation failed; output is not a valid A2UI tree.
slider_volume	I	Schema validation failed; output is not a valid A2UI tree.
date_picker_only	I	Schema validation failed; output is not a valid A2UI tree.
choice_picker_chips	I	Schema validation failed; output is not a valid A2UI tree.
textfield_search	I	Schema validation failed; output is not a valid A2UI tree.
textfield_multiline	I	Schema validation failed; output is not a valid A2UI tree.
card_with_text	I	Schema validation failed; output is not a valid A2UI tree.
card_with_image_text_button	I	Schema validation failed; output is not a valid A2UI tree.
card_profile_summary	I	Schema validation failed; output is not a valid A2UI tree.
card_stats	I	Schema validation failed; output is not a valid A2UI tree.
card_with_divider	I	Schema validation failed; output is not a valid A2UI tree.
card_with_icon_title	I	Schema validation failed; output is not a valid A2UI tree.
notification_card	I	Schema validation failed; output is not a valid A2UI tree.
card_quote	I	Schema validation failed; output is not a valid A2UI tree.
list_simple_text	I	Schema validation failed; output is not a valid A2UI tree.
list_horizontal_cards	I	Schema validation failed; output is not a valid A2UI tree.
list_checkable_todo	I	Schema validation failed; output is not a valid A2UI tree.
list_settings_menu	I	Schema validation failed; output is not a valid A2UI tree.
list_with_dividers	I	Schema validation failed; output is not a valid A2UI tree.
list_users_avatars	I	Schema validation failed; output is not a valid A2UI tree.
list_notifications	I	Schema validation failed; output is not a valid A2UI tree.
list_pricing_options	I	Schema validation failed; output is not a valid A2UI tree.
form_login	I	Schema validation failed; output is not a valid A2UI tree.
form_signup	I	Schema validation failed; output is not a valid A2UI tree.
form_contact	I	Schema validation failed; output is not a valid A2UI tree.
form_reservation	I	Schema validation failed; output is not a valid A2UI tree.
form_search_with_filter	I	Schema validation failed; output is not a valid A2UI tree.
form_subscribe	I	Schema validation failed; output is not a valid A2UI tree.
form_feedback_rating	I	Schema validation failed; output is not a valid A2UI tree.
form_settings_toggles	I	Schema validation failed; output is not a valid A2UI tree.
form_address	I	Schema validation failed; output is not a valid A2UI tree.
form_event	I	Schema validation failed; output is not a valid A2UI tree.
tabs_two_simple	I	Schema validation failed; output is not a valid A2UI tree.
tabs_three_panels	I	Schema validation failed; output is not a valid A2UI tree.
tabs_dashboard	I	Schema validation failed; output is not a valid A2UI tree.
tabs_settings_sections	I	Schema validation failed; output is not a valid A2UI tree.
modal_confirm	I	Schema validation failed; output is not a valid A2UI tree.
modal_info	I	Schema validation failed; output is not a valid A2UI tree.
modal_signup_form	I	Schema validation failed; output is not a valid A2UI tree.
modal_image_zoom	I	Schema validation failed; output is not a valid A2UI tree.
nested_layout	I	Schema validation failed; output is not a valid A2UI tree.
simple_column_layout	I	Schema validation failed; output is not a valid A2UI tree.
row_with_buttons	I	Schema validation failed; output is not a valid A2UI tree.
simple_text	I	Schema validation failed; output is not a valid A2UI tree.
dashboard_overview	I	Schema validation failed; output is not a valid A2UI tree.
product_detail	I	Schema validation failed; output is not a valid A2UI tree.
chat_message_thread	I	Schema validation failed; output is not a valid A2UI tree.
profile_screen	I	Schema validation failed; output is not a valid A2UI tree.
shopping_cart_summary	I	Schema validation failed; output is not a valid A2UI tree.

Llama 3.2 1B

Sample	Grade	Rationale
text_heading	I	Schema validation failed; output is not a valid A2UI tree.
text_caption	I	Schema validation failed; output is not a valid A2UI tree.
text_markdown	I	Schema validation failed; output is not a valid A2UI tree.
image_avatar	I	Schema validation failed; output is not a valid A2UI tree.
image_header_banner	I	Schema validation failed; output is not a valid A2UI tree.
icon_inline	I	Schema validation failed; output is not a valid A2UI tree.
video_player	I	Schema validation failed; output is not a valid A2UI tree.
audio_player	I	Schema validation failed; output is not a valid A2UI tree.
divider_horizontal	I	Schema validation failed; output is not a valid A2UI tree.
button_primary	I	Schema validation failed; output is not a valid A2UI tree.
button_borderless	I	Schema validation failed; output is not a valid A2UI tree.
textfield_password	I	Schema validation failed; output is not a valid A2UI tree.
text_then_button	I	Schema validation failed; output is not a valid A2UI tree.
image_caption	I	Schema validation failed; output is not a valid A2UI tree.
icon_label	I	Schema validation failed; output is not a valid A2UI tree.
two_buttons_row	I	Schema validation failed; output is not a valid A2UI tree.
heading_paragraph	I	Schema validation failed; output is not a valid A2UI tree.
divider_between_texts	I	Schema validation failed; output is not a valid A2UI tree.
checkbox_simple	I	Schema validation failed; output is not a valid A2UI tree.
slider_volume	I	Schema validation failed; output is not a valid A2UI tree.
date_picker_only	I	Schema validation failed; output is not a valid A2UI tree.
choice_picker_chips	I	Schema validation failed; output is not a valid A2UI tree.
textfield_search	I	Schema validation failed; output is not a valid A2UI tree.
textfield_multiline	I	Schema validation failed; output is not a valid A2UI tree.
card_with_text	I	Schema validation failed; output is not a valid A2UI tree.
card_with_image_text_button	I	Schema validation failed; output is not a valid A2UI tree.
card_profile_summary	I	Schema validation failed; output is not a valid A2UI tree.
card_stats	I	Schema validation failed; output is not a valid A2UI tree.
card_with_divider	I	Schema validation failed; output is not a valid A2UI tree.
card_with_icon_title	I	Schema validation failed; output is not a valid A2UI tree.
notification_card	I	Schema validation failed; output is not a valid A2UI tree.
card_quote	I	Schema validation failed; output is not a valid A2UI tree.
list_simple_text	I	Schema validation failed; output is not a valid A2UI tree.
list_horizontal_cards	I	Schema validation failed; output is not a valid A2UI tree.
list_checkable_todo	I	Schema validation failed; output is not a valid A2UI tree.
list_settings_menu	I	Schema validation failed; output is not a valid A2UI tree.
list_with_dividers	I	Schema validation failed; output is not a valid A2UI tree.
list_users_avatars	I	Schema validation failed; output is not a valid A2UI tree.
list_notifications	I	Schema validation failed; output is not a valid A2UI tree.
list_pricing_options	I	Schema validation failed; output is not a valid A2UI tree.
form_login	I	Schema validation failed; output is not a valid A2UI tree.
form_signup	I	Schema validation failed; output is not a valid A2UI tree.
form_contact	I	Schema validation failed; output is not a valid A2UI tree.
form_reservation	I	Schema validation failed; output is not a valid A2UI tree.
form_search_with_filter	I	Schema validation failed; output is not a valid A2UI tree.
form_subscribe	I	Schema validation failed; output is not a valid A2UI tree.
form_feedback_rating	I	Schema validation failed; output is not a valid A2UI tree.
form_settings_toggles	I	Schema validation failed; output is not a valid A2UI tree.
form_address	I	Schema validation failed; output is not a valid A2UI tree.
form_event	I	Schema validation failed; output is not a valid A2UI tree.
tabs_two_simple	I	Schema validation failed; output is not a valid A2UI tree.
tabs_three_panels	I	Schema validation failed; output is not a valid A2UI tree.
tabs_dashboard	I	Schema validation failed; output is not a valid A2UI tree.
tabs_settings_sections	I	Schema validation failed; output is not a valid A2UI tree.
modal_confirm	I	Schema validation failed; output is not a valid A2UI tree.
modal_info	I	Schema validation failed; output is not a valid A2UI tree.
modal_signup_form	I	Schema validation failed; output is not a valid A2UI tree.
modal_image_zoom	I	Schema validation failed; output is not a valid A2UI tree.
nested_layout	I	Schema validation failed; output is not a valid A2UI tree.
simple_column_layout	I	Schema validation failed; output is not a valid A2UI tree.
row_with_buttons	I	Schema validation failed; output is not a valid A2UI tree.
simple_text	I	Schema validation failed; output is not a valid A2UI tree.
dashboard_overview	I	Schema validation failed; output is not a valid A2UI tree.
product_detail	I	Schema validation failed; output is not a valid A2UI tree.
chat_message_thread	I	Schema validation failed; output is not a valid A2UI tree.
profile_screen	I	Schema validation failed; output is not a valid A2UI tree.
shopping_cart_summary	I	Schema validation failed; output is not a valid A2UI tree.

