

Moonshot releases Kimi K2.6, scaling its agent swarm to 300 sub-agents

Source copy kept by wheresthe.ai for the story "Moonshot releases Kimi K2.6, scaling its agent swarm to 300 sub-agents" (<https://wheresthe.ai/d/cn-moonshot-kimi-k2-6-release-2026-04/>).

Source: <https://huggingface.co/moonshotai/Kimi-K2.6>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 1d3003220c0d20f5cda3613472e70c8b253bda3b9216827b3d945b5be1ba3329

```
image<|media_content|><|media_pad|><|media_end|>\n {% elif content['type'] == 'video' or content['type']==
'video_url'-%}\n <|kimi_k25_video_placeholder|>\n {% else -%}\n {{ content['text'] }}\n {%- endif -%}\n {%- endfor -
%}\n {%- endif -%}\n{%- endmacro -%}\n\n{% macro set_roles(message) -%}\n {%- set role_name =
message.get('name') or message['role'] -%}\n {%- if message['role'] == 'user' -%}\n <|im_user|>{{role_name}}<|
im_middle|>\n {%- elif message['role'] == 'assistant' -%}\n <|im_assistant|>{{role_name}}<|im_middle|>\n {%- else -
%}\n <|im_system|>{{role_name}}<|im_middle|>\n {%- endif -%}\n{%- endmacro -%}\n\n{% macro
render_toolcalls(message) -%}\n <|tool_calls_section_begin|>\n {%- for tool_call in message['tool_calls'] -%}\n {%-
set formatted_id = tool_call['id'] -%}\n <|tool_call_begin|>{{ formatted_id }}<|tool_call_argument_begin|>{% if
tool_call['function']['arguments'] is string %}{{ tool_call['function']['arguments'] }}{% else %}{{ tool_call['function']
['arguments'] | tojson }}{% endif %}<|tool_call_end|>\n {%- endfor -%}\n <|tool_calls_section_end|>\n{%- endmacro
-%}\n\n{%- set preserve_thinking = preserve_thinking | default(false) -%}\n\n# Find last non-tool-call assistant
message. If preserve_thinking, keep -1 so hist is empty and all msgs use suffix (retain reasoning). #}\n{%- set ns
= namespace(last_non_tool_call_assistant_msg=-1) -%}\n{%- if not preserve_thinking -%}\n{%- for idx in
range(messages|length-1, -1, -1) -%}\n {%- if messages[idx]['role'] == 'assistant' and not
messages[idx].get('tool_calls') -%}\n {%- set ns.last_non_tool_call_assistant_msg = idx -%}\n {%- break -%}\n {%-
endif -%}\n{%- endfor -%}\n{%- endif -%}\n\n# split all messages into history & suffix, reasoning_content in suffix
should be reserved.#}\n{%- set hist_msgs = messages[:ns.last_non_tool_call_assistant_msg+1] -%}\n{%- set
suffix_msgs = messages[ns.last_non_tool_call_assistant_msg+1:] -%}\n\n{%- if tools -%}\n {%- if tools_ts_str -
%}\n <|im_system|>tool_declare<|im_middle|>{{ tools_ts_str }}<|im_end|>\n {%- else -%}\n <|im_system|
>tool_declare<|im_middle|>{{ tools | tojson(separators=(',', ':')) }}<|im_end|>\n {%- endif -%}\n{%- endif -%}\n\n
\n{%- for message in hist_msgs -%}\n {{set_roles(message)}}\n {%- if message['role'] == 'assistant' -%}\n
<think></think>{{render_content(message)}}\n {%- if message.get('tool_calls') -%}\n
{{render_toolcalls(message)}}\n {%- endif -%}\n {%- elif message['role'] == 'tool' -%}\n {%- set tool_call_id =
message.tool_call_id -%}\n ## Return of {{ tool_call_id }}\n{{render_content(message)}}\n {%- elif
message['content'] is not none -%}\n {{render_content(message)}}\n {%- endif -%}\n <|im_end|>\n{%- endfor -
%}\n\n{%- for message in suffix_msgs -%}\n {{set_roles(message)}}\n {%- if message['role'] == 'assistant' -%}\n
{%- if thinking is defined and thinking is false and preserve_thinking is false -%}\n <think></
think>{{render_content(message)}}\n {%- else -%}\n {%- set rc = message.get('reasoning',
message.get('reasoning_content', '')) -%}\n <think>{{rc}}</think>{{render_content(message)}}\n {%- endif -%}\n
{%- if message.get('tool_calls') -%}\n {{render_toolcalls(message)}}\n {%- endif -%}\n {%- elif message['role'] ==
'tool' -%}\n {%- set tool_call_id = message.tool_call_id -%}\n ## Return of {{ tool_call_id
}}\n{{render_content(message)}}\n {%- elif message['content'] is not none -%}\n {{render_content(message)}}\n
{%- endif -%}\n <|im_end|>\n{%- endfor -%}\n\n{%- if add_generation_prompt -%}\n <|im_assistant|>assistant<|
im_middle|>\n {%- if thinking is defined and thinking is false -%}\n <think></think>\n {%- else -%}\n <think>\n {%-
endif -%}\n{%- endif -
%}\n", "createdAt": "2026-04-14T04:23:36.000Z", "discussionsDisabled": false, "discussionsSorting": "recently-
created", "downloads": 382421, "downloadsAllTime": 8561780, "id": "moonshotai/Kimi-
K2.6", "isLikedByUser": false, "availableInferenceProviders":
[{"provider": "novita", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/kimi-
k2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features":
{"structuredOutput": false, "toolCalling": true}, "isCheapestPricingOutput": true, "isFastestThroughput": true, "isModelAuthc
{"provider": "featherless-ai", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-
of-000064.safetensors", "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuthor": false},
{"provider": "deepinfra", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
```

K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features": {"structuredOutput": true, "toolCalling": true}, "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuth": {"provider": "baseten", "modelStatus": "error", "providerStatus": "live", "providerId": "moonshotai/Kimi-K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features": {"structuredOutput": true, "toolCalling": true}, "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuth": {"text-to-text": true, "library_name": "transformers", "librariesOther": [], "trackDownloads": true, "model-index": null, "evalResults": [{"dataset": {"id": "SWE-bench/SWE-bench_Verified", "isBenchmark": true, "task_id": "swe_bench_%_resolved"}, "value": 80.2, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "swe_bench_verified.yaml", "verified": false, "rank": 8, "label": "Swe Bench Resolved"}, {"dataset": {"id": "Idavidrein/gpqa", "isBenchmark": true, "task_id": "diamond"}, "value": 90.5, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "gpqa_diamond.yaml", "verified": false, "rank": 12, "label": "Diamond"}, {"dataset": {"id": "ScaleAI/SWE-bench_Pro", "isBenchmark": true, "task_id": "SWE_Bench_Pro"}, "value": 58.6, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "swe_bench_pro.yaml", "verified": false, "rank": 15, "label": "SWE Bench Pro"}, {"dataset": {"id": "MathArena/aime_2026", "isBenchmark": true, "task_id": "MathArena/aime_2026"}, "value": 96.4, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "aime_2026.yaml", "verified": false, "rank": 6, "label": "MathArena Aime 2026"}, {"dataset": {"id": "MMMU/MMMU_Pro", "isBenchmark": true, "task_id": "mmmu_pro_vision"}, "value": 79.4, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "filename": ".eval_results/mmmu_pro.yaml", "verified": false, "rank": 2, "label": "Mmmu Pro Vision"}, {"dataset": {"id": "harborframework/terminal-bench-2.0", "isBenchmark": true, "task_id": "terminalbench_2"}, "value": 66.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "terminal_bench_2.yaml", "verified": false, "rank": 7, "label": "Terminalbench 2"}, {"dataset": {"id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb_solved"}, "value": 0, "source": {"url": "https://zli12321.github.io/LHTB/leaderboard.html", "name": "LHTB leaderboard", "isExternal": true, "filename": ".eval_results/lhtb.yaml", "verified": false, "pullRequest": 55, "rank": 9, "label": "Lhtb Solved", "notes": "0/46 tasks solved at reward >= 0.95 (pass@0.95); mean reward x100 = 25.5; official LHTB Harbor harness"}, {"dataset": {"id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb2_solved"}, "value": 2, "source": {"url": "https://github.com/zli12321/LHTB/tree/lhtb-2.0#lhtb-20-results-october-2026", "name": "LHTB 2.0 results", "isExternal": true, "filename": ".eval_results/lhtb2_solved.yaml", "verified": false, "pullRequest": 55, "rank": 8, "label": "Lhtb2 Solved", "notes": "2/77 tasks solved at reward >= 0.95 (pass@0.95); mean reward x100 = 42.0; LHTB 2.0 (77 tasks), terminus-2, 3-hour budget per task, official LHTB Harbor harness"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "chi_bench"}, "value": 15.6, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Chi Bench", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "prior_authorization"}, "value": 18.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true, "filename": ".eval_results/chi-

bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Prior Authorization", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)", {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "utilization_management"}, "value": 21.3, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 3, "label": "Utilization Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)", {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "care_management"}, "value": 6.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 1, "label": "Care Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)", {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle"}, "value": 34.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwitt", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle.yaml", "verified": false, "label": "Hle"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle"}, "value": 54, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwitt", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle_with_tools.yaml", "verified": false, "rank": 9, "label": "Hle", "notes": "With tools"}, {"dataset": {"id": "MathArena/hmmt_feb_2026", "isBenchmark": true, "task_id": "MathArena/hmmt_feb_2026"}, "value": 92.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwitt", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hmmt_feb_2026.yaml", "verified": false, "rank": 2, "label": "MathArena Hmmt Feb 2026"}, {"dataset": {"id": "joelniklaus/LEXam-hard", "isBenchmark": true, "task_id": "lexam_hard"}, "value": 36.18, "source": {"url": "https://huggingface.co/buckets/joelniklaus/SwissLegalEvals", "name": "SwissLegalEvals per-sample details (lighteval)", "isExternal": true, "author": {"_id": "5fae5f68b8423e1d80b8a988", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1656078368144-5fae5f68b8423e1d80b8a988.jpeg", "fullName": "Joel Niklaus", "name": "joelniklaus", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 17, "lexam-hard.yaml", "verified": false, "pullRequest": 58, "rank": 9, "label": "Lexam Hard", "notes": "lighteval, LEXam paper prompts, one response per question, no tools; DeepSeek-R1-0528 judge; mean of the German and English means over the 518 questions, 0-100"}, {"dataset": {"id": "SWE-bench/SWE-bench_Multilingual", "isBenchmark": true, "task_id": "swe_bench_multilingual_%_resolved"}, "value": 76.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false}, {"filename": ".eval_results/swe-bench_multilingual.yaml", "verified": false, "pullRequest": 57, "rank": 7, "label": "Swe Bench Multilingual Resolved"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "overall"}, "value": 89.76, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 6, "label": "Overall"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "scanning"}, "value": 90.08, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Scanning"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "warping"}, "value": 89.62, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench

Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Warping"}, {"dataset": {"id": "PaddlePaddle/Real5-OmnidocBench", "isBenchmark": true, "task_id": "screen_photography", "value": 89.58, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmnidocBench", "name": "Real5-OmnidocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 6, "label": "Screen Photography"}, {"dataset": {"id": "PaddlePaddle/Real5-OmnidocBench", "isBenchmark": true, "task_id": "illumination", "value": 89.91, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmnidocBench", "name": "Real5-OmnidocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Illumination"}, {"dataset": {"id": "PaddlePaddle/Real5-OmnidocBench", "isBenchmark": true, "task_id": "skew", "value": 89.61, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmnidocBench", "name": "Real5-OmnidocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 5, "label": "Skew"}, {"dataset": {"id": "benchflow/skillsbench", "isBenchmark": true, "task_id": "skillsbench_v1_1", "value": 54, "source": {"url": "https://huggingface.co/datasets/benchflow/skillsbench-leaderboard/raw/main/leaderboard/skillsbench/v1.1/official.json", "name": "SkillsBench v1.1 official leaderboard", "isExternal": true, "author": {"_id": "69657f24ec1d157f1590a81d", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/663fe2d26304d377fc253322/L6ik2VpL5iizADEXK4W-A.png", "fullName": "BenchFlow", "name": "benchflow", "type": "org", "isHf": false, "isHfAdmin": false, "isMod": false, "plan": "te", "skillsbench.yaml", "verified": false, "pullRequest": 53, "rank": 5, "label": "Skillsbench V1 1", "notes": "with-skills; BenchFlow harness; OpenHands agent; 87 tasks x 3 trials; full 261/261 coverage"}, {"dataset": {"id": "InternScience/ResearchClawBench", "isBenchmark": false, "task_id": "overall", "value": 18, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false}, "filename": ".eval_results/researchclawbench.yaml", "verified": false, "pullRequest": 42, "rank": 3, "label": "Overall", "notes": "ResearchHarness evaluation with tools enabled, code execution, and a file-system workspace; completed 38/40 ResearchClawBench tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "general", "value": 61.5, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 15, "label": "General", "notes": "Pass^{3%} | N=3 | 161 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multi_turn", "value": 65.8, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 1, "label": "Multi Turn", "notes": "Pass^{3%} | N=3 | 38 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multimodal", "value": 18.8, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 5, "label": "Multimodal", "notes": "Pass^{3%} | N=3 | 101 tasks"}, {"dataset": {"id": "collinear-ai/yc-bench", "isBenchmark": true, "task_id": "medium", "value": 511137, "source": {"url": "https://github.com/collinear-ai/yc-bench", "name": "YC-Bench

```
eval", "isExternal": true, "filename": ".eval_results/yc-
bench.yaml", "verified": false, "pullRequest": 29, "rank": 5, "label": "Medium", "notes": "medium preset, avg final funds
(USD) across seeds 1,2,3"}], "private": false, "repoType": "model", "gated": false, "tags":
["transformers", "safetensors", "kimi_k25", "feature-extraction", "compressed-tensors", "image-text-to-
text", "conversational", "custom_code", "arxiv:2602.02276", "license:other", "eval-results", "region:us"], "tag_objs":
[{"id": "image-text-to-text", "label": "Image-Text-to-Text", "type": "pipeline_tag", "subType": "multimodal"},
{"id": "transformers", "label": "Transformers", "type": "library"}, {"id": "safetensors", "label": "Safetensors", "type": "library"},
{"id": "kimi_k25", "label": "kimi_k25", "type": "other", "clickable": true}, {"id": "feature-extraction", "label": "feature-
extraction", "type": "other", "clickable": true}, {"id": "compressed-tensors", "label": "compressed-
tensors", "type": "other", "clickable": true},
{"id": "conversational", "label": "conversational", "type": "other", "clickable": true},
{"id": "custom_code", "label": "custom_code", "type": "other", "clickable": true}, {"id": "eval-results", "label": "Eval
Results", "type": "other", "clickable": true}, {"id": "arxiv:2602.02276", "label": "arxiv:2602.02276", "type": "arxiv", "extra":
{"paperTitle": "Kimi K2.5: Visual Agentic Intelligence"}], {"id": "license:other", "label": "other", "type": "license"},
{"type": "region", "label": "🇺🇸 Region: US", "id": "region:us"}], "transformersInfo":
{"auto_model": "AutoModel", "custom_class": "modeling_kimi_k25.KimiK25ForConditionalGeneration", "pipeline_tag": "fe
ature-extraction", "processor": "AutoProcessor"}, "widgetData": [{"text": "Hi, what can you help me with?"}, {"text": "What is
84 * 3 / 2?"}, {"text": "Tell me an interesting fact about the universe!"}, {"text": "Explain quantum computing in simple
terms."}], "safetensors": {"parameters":
{"F32": 23040, "I32": 1014686023680, "BF16": 12193329648}, "total": 1026879376368, "sharded": true, "totalFileSize": 5952
{"closed": 13, "open": 42, "total": 55}, "query": {}, "inferenceContextData": {"billableEntities": [], "entityName2Providers":
{}}, "defaultProviders":
[{"isOriginalProvider": false, "name": "novita", "enabled": true, "position": 1, "isReleased": true, "accuratePricing": true},
{"isOriginalProvider": false, "name": "deepinfra", "enabled": true, "position": 15, "isReleased": true, "accuratePricing": true},
{"isOriginalProvider": false, "name": "featherless-
ai", "enabled": true, "position": 7, "isReleased": true, "accuratePricing": true},
{"isOriginalProvider": false, "name": "baseten", "enabled": true, "position": 13, "isReleased": true, "accuratePricing": true}], "h
[]}]>
```

huggingchat |  Tech Blog

1. Model Introduction

Kimi K2.6 is an open-source, native multimodal agentic model that advances practical capabilities in long-horizon coding, coding-driven design, proactive autonomous execution, and swarm-based task orchestration.

Key Features

Long-Horizon Coding : K2.6 achieves significant improvements on complex, end-to-end coding tasks, generalizing robustly across programming languages (Rust, Go, Python) and domains spanning front-end, DevOps, and performance optimization.

Coding-Driven Design : K2.6 is capable of transforming simple prompts and visual inputs into production-ready interfaces and lightweight full-stack workflows, generating structured layouts, interactive elements, and rich animations with deliberate aesthetic precision.

Elevated Agent Swarm : Scaling horizontally to 300 sub-agents executing 4,000 coordinated steps, K2.6 can dynamically decompose tasks into parallel, domain-specialized subtasks, delivering end-to-end outputs from documents to websites to spreadsheets in a single autonomous run.

Proactive & Open Orchestration : For autonomous tasks, K2.6 demonstrates strong performance in powering persistent, 24/7 background agents that proactively manage schedules, execute code, and orchestrate cross-platform operations without human oversight.

2. Model Summary

Architecture Mixture-of-Experts (MoE)

Total Parameters 1T

Activated Parameters 32B

Number of Layers (Dense layer included) 61

Number of Dense Layers 1

Attention Hidden Dimension 7168

MoE Hidden Dimension (per Expert) 2048

Number of Attention Heads 64

Number of Experts 384

Selected Experts per Token 8

Number of Shared Experts 1

Vocabulary Size 160K

Context Length 256K

Attention Mechanism MLA

Activation Function SwiGLU

Vision Encoder MoonViT

Parameters of Vision Encoder 400M

3. Evaluation Results

Benchmark Kimi K2.6 GPT-5.4 (xhigh) Claude Opus 4.6 (max effort) Gemini 3.1 Pro (thinking high) Kimi K2.5 Agentic

HLE-Full (w/ tools) 54.0 52.1 53.0 51.4 50.2

BrowseComp 83.2 82.7 83.7 85.9 74.9

BrowseComp (Agent Swarm) 86.3 78.4

DeepSearchQA (f1-score) 92.5 78.6 91.3 81.9 89.0

DeepSearchQA (accuracy) 83.0 63.7 80.6 60.2 77.1

WideSearch (item-f1) 80.8 - - - 72.7

Toolathlon 50.0 54.6 47.2 48.8 27.8

MCPMark 55.9 62.5* 56.7* 55.9* 29.5

Claw Eval (pass^3) 62.3 60.3 70.4 57.8 52.3

Claw Eval (pass@3) 80.9 78.4 82.4 82.9 75.4

APEX-Agents 27.9 33.3 33.0 32.0 11.5

OSWorld-Verified 73.1 75.0 72.7 - 63.3

Coding

Terminal-Bench 2.0 (Terminus-2) 66.7 65.4* 65.4 68.5 50.8

SWE-Bench Pro 58.6 57.7 53.4 54.2 50.7

SWE-Bench Multilingual 76.7 - 77.8 76.9* 73.0

SWE-Bench Verified 80.2 - 80.8 80.6 76.8

SciCode 52.2 56.6 51.9 58.9 48.7

OJBench (python) 60.6 - 60.3 70.7 54.7

LiveCodeBench (v6) 89.6 - 88.8 91.7 85.0

Reasoning & Knowledge

HLE-Full 34.7 39.8 40.0 44.4 30.1

AIME 2026 96.4 99.2 96.7 98.3 95.8

HMMT 2026 (Feb) 92.7 97.7 96.2 94.7 87.1

IMO-AnswerBench 86.0 91.4 75.3 91.0* 81.8

GPQA-Diamond 90.5 92.8 91.3 94.3 87.6

Vision

MMMU-Pro 79.4 81.2 73.9 83.0* 78.5

MMMU-Pro (w/ python) 80.1 82.1 77.3 85.3* 77.7

CharXiv (RQ) 80.4 82.8* 69.1 80.2* 77.5

CharXiv (RQ) (w/ python) 86.7 90.0* 84.7 89.9* 78.7

MathVision 87.4 92.0* 71.2* 89.8* 84.2

MathVision (w/ python) 93.2 96.1* 84.6* 95.7* 85.0

BabyVision 39.8 49.7 14.8 51.6 36.5

BabyVision (w/ python) 68.5 80.2* 38.4* 68.3* 40.5

V* (w/ python) 96.9 98.4* 86.4* 96.9* 86.9

Footnotes

General Testing Details We report results for Kimi K2.6 and Kimi K2.5 with thinking mode enabled, Claude Opus 4.6 with max effort, GPT-5.4 with xhigh reasoning effort, and Gemini 3.1 Pro with a high thinking level.

Unless otherwise specified, all Kimi K2.6 experiments were conducted with temperature = 1.0, top-p = 1.0, and a context length of 262,144 tokens.

Benchmarks without publicly available scores were re-evaluated under the same conditions used for Kimi K2.6 and are marked with an asterisk (*). Except where noted with an asterisk, all other results are cited from official reports.

Reasoning Benchmarks IMO-AnswerBench scores for GPT-5.4 and Claude 4.6 were obtained from z.ai/blog/glm-5.1 .

Humanity's Last Exam (HLE) and other reasoning tasks were evaluated with a maximum generation length of 98,304 tokens. By default, we report results on the HLE full set. For the text-only subset, Kimi K2.6 achieves 36.4% accuracy without tools and 55.5% with tools.

Tool-Augmented / Agentic Tasks Kimi K2.6 was equipped with search, code-interpreter, and web-browsing tools for HLE with tools, BrowseComp, DeepSearchQA, and WideSearch.

For HLE-Full with tools, the maximum generation length is 262,144 tokens with a per-step limit of 49,152 tokens. We employ a simple context management strategy: once the context window exceeds the threshold, only the most recent round of tool-related messages is retained.

For BrowseComp, we report scores obtained with context management using the same discard-all strategy as Kimi K2.5 and DeepSeek-V3.2.

For DeepSearchQA, no context management was applied to Kimi K2.6 tests, and tasks exceeding the supported context length were directly counted as failed. Scores for Claude Opus 4.6, GPT-5.4, and Gemini 3.1 Pro on DeepSearchQA are cited from the Claude Opus 4.7 System Card .

For WideSearch, we report results under the "hide tool result" context management setting. Once the context window exceeds the threshold, only the most recent round of tool-related messages is retained.

The test system prompts are identical to those used in the Kimi K2.5 technical report .

Claw Eval was conducted using version 1.1 with max-tokens-per-step = 16384.

For APEX-Agents, we evaluate 452 tasks from the public 480-task release, as done by Artificial Analysis (excluding Investment Banking Worlds 244 and 246, which have external runtime dependencies)

Coding Tasks Terminal-Bench 2.0 scores were obtained with the default agent framework (Terminus-2) and the provided JSON parser, operating in preserve thinking mode.

For the SWE-Bench series of evaluations (including Verified, Multilingual, and Pro), we used an in-house evaluation framework adapted from SWE-agent. This framework includes a minimal set of tools—bash tool, createfile tool, insert tool, view tool, strreplace tool, and submit tool.

All reported scores for coding tasks are averaged over 10 independent runs.

Vision Benchmarks Max-tokens = 98,304, averaged over three runs (avg@3).

Settings with Python tool use max-tokens-per-step = 65,536 and max-steps = 50 for multi-step reasoning.

MMMU-Pro follows the official protocol, preserving input order and prepending images.

4. Native INT4 Quantization

Kimi-K2.6 adopts the same native int4 quantization method as Kimi-K2-Thinking .

5. Deployment

You can access Kimi-K2.6's API on <https://platform.moonshot.ai> and we provide OpenAI/Anthropic-compatible API for you. To verify the deployment is correct, we also provide the Kimi Vendor Verifier . Currently, Kimi-K2.6 is recommended to run on the following inference engines:

vLLM

SGLang

KTransformers

Kimi-K2.6 has the same architecture as Kimi-K2.5, and the deployment method can be directly reused.

The version requirement for transformers is $\geq 4.57.1$, $< 5.0.0$.

Deployment examples can be found in the Model Deployment Guide .

6. Model Usage

The usage demos below demonstrate how to call our official API.

For third-party APIs deployed with vLLM or SGLang, please note that:

Chat with video content is an experimental feature and is only supported in our official API for now.

The recommended temperature will be 1.0 for Thinking mode and 0.6 for Instant mode.

The recommended top_p is 0.95 .

To use instant mode, you need to pass {'chat_template_kwargs': {"thinking": False}} in extra_body .

Chat Completion

This is a simple chat completion script which shows how to call K2.6 API in Thinking and Instant modes.

```
import openai
import base64
import requests

def simple_chat ( client: openai.OpenAI, model_name: str ):
    messages = [ { 'role' : 'system' , 'content' : 'You are Kimi, an AI assistant created by Moonshot AI.' }, { 'role' : 'user' , 'content' : [ { 'type' : 'text' , 'text' : 'which one is bigger, 9.11 or 9.9? think carefully.' } ] , } ]
    response = client.chat.completions.create( model=model_name, messages=messages, stream= False , max_tokens= 4096 )
    print ( '==== Below is reasoning content in Thinking Mode =====' )
    print ( f'reasoning content: {response.choices[ 0 ].message.reasoning} ' )
    print ( '==== Below is response in Thinking Mode =====' )
    print ( f'response: {response.choices[ 0 ].message.content} ' )

# To use instant mode, pass {"thinking" = {"type": "disabled"}}
response = client.chat.completions.create(
```

```
model=model_name, messages=messages, stream= False , max_tokens= 4096 , extra_body={ 'thinking' : { 'type' : 'disabled' } }, # this is for official API # extra_body= { 'chat_template_kwargs': { "thinking": False } } # this is for vLLM/ SGLang ) print ( '==== Below is response in Instant Mode =====' ) print ( f'response: {response.choices[ 0 ].message.content} ' )
```

Chat Completion with visual content

K2.6 supports Image and Video input.

The following example demonstrates how to call K2.6 API with image input:

```
import openai import base64 import requests
```

```
def chat_with_image ( client: openai.OpenAI, model_name: str ): url = 'https://huggingface.co/moonshotai/Kimi-K2.6/resolve/main/figures/kimi-logo.png' image_base64 = base64.b64encode(requests.get(url).content).decode() messages = [ { 'role' : 'user' , 'content' : [ { 'type' : 'text' , 'text' : 'Describe this image in detail.' }, { 'type' : 'image_url' , 'image_url' : { 'url' : f'data:image/png;base64, {image_base64} ' }, }, ], } ]
```

```
response = client.chat.completions.create( model=model_name, messages=messages, stream= False , max_tokens= 8192 ) print ( '==== Below is reasoning content in Thinking Mode =====' ) print ( f'reasoning content: {response.choices[ 0 ].message.reasoning} ' ) print ( '==== Below is response in Thinking Mode =====' ) print ( f'response: {response.choices[ 0 ].message.content} ' )
```

```
# Also support instant mode if you pass {"thinking" = {"type":"disabled"}} response = client.chat.completions.create( model=model_name, messages=messages, stream= False , max_tokens= 4096 , extra_body={ 'thinking' : { 'type' : 'disabled' } }, # this is for official API # extra_body= { 'chat_template_kwargs': { "thinking": False } } # this is for vLLM/SGLang ) print ( '==== Below is response in Instant Mode =====' ) print ( f'response: {response.choices[ 0 ].message.content} ' )
```

```
return response.choices[ 0 ].message.content
```

The following example demonstrates how to call K2.6 API with video input:

```
import openai import base64 import requests
```

```
def chat_with_video ( client: openai.OpenAI, model_name: str ): url = 'https://huggingface.co/moonshotai/Kimi-K2.6/resolve/main/figures/demo_video.mp4' video_base64 = base64.b64encode(requests.get(url).content).decode() messages = [ { "role" : "user" , "content" : [ { "type" : "text" , "text" : "Describe the video in detail." }, { "type" : "video_url" , "video_url" : { "url" : f'data:video/mp4;base64, {video_base64} " }, }, ], } ]
```

```
response = client.chat.completions.create(model=model_name, messages=messages) print ( '==== Below is reasoning content in Thinking Mode =====' ) print ( f'reasoning content: {response.choices[ 0 ].message.reasoning} ' ) print ( '==== Below is response in Thinking Mode =====' ) print ( f'response: {response.choices[ 0 ].message.content} ' )
```

```
# Also support instant mode if pass {"thinking" = {"type":"disabled"}} response = client.chat.completions.create( model=model_name, messages=messages, stream= False , max_tokens= 4096 , extra_body={ 'thinking' : { 'type' : 'disabled' } }, # this is for official API # extra_body= { 'chat_template_kwargs': { "thinking": False } } # this is for vLLM/ SGLang ) print ( '==== Below is response in Instant Mode =====' ) print ( f'response: {response.choices[ 0 ].message.content} ' ) return response.choices[ 0 ].message.content
```

Preserve Thinking

Kimi K2.6 supports preserve_thinking mode, which retains full reasoning content across multi-turn interactions and enhances performance in coding agent scenarios.

This feature is disabled by default. The following example demonstrates how to call K2.6 API in preserve_thinking mode:

```
def chat_with_preserve_thinking ( client: openai.OpenAI, model_name: str ): messages = [ { "role" : "user" , "content" : "Tell me three random numbers." }, { "role" : "assistant" , "reasoning_content" : "I'll start by listing five numbers: 473, 921, 235, 215, 222, and I'll tell you the first three." , "content" : "473, 921, 235" }, { "role" : "user" , "content" : "What are the other two numbers you have in mind?" } ]
```

```
response = client.chat.completions.create( model=model_name, messages=messages, stream= False ,
max_tokens= 4096 , extra_body={ 'thinking' : { 'type' : 'enabled' , 'keep' : 'all' }}, # this is for official API #
extra_body={"chat_template_kwargs": {"thinking":True, "preserve_thinking": True}}, # this is for vLLM/SGLang #
We recommend enabling preserve_thinking only in think mode. ) # the assistant should mention 215 and 222 that
appear in the prior reasoning content print ( f"response: {response.choices[ 0 ].message.reasoning} " ) return
response.choices[ 0 ].message.content
```

Interleaved Thinking and Multi-Step Tool Call

K2.6 shares the same design of Interleaved Thinking and Multi-Step Tool Call as K2 Thinking. For usage example, please refer to the K2 Thinking documentation .

Coding Agent Framework

Kimi K2.6 works best with Kimi Code CLI as its agent framework — give it a try at <https://www.kimi.com/code> .

7. License

Both the code repository and the model weights are released under the Modified MIT License .

8. Third Party Notices

See THIRD PARTY NOTICES

9. Contact Us

If you have any questions, please reach out at support@moonshot.ai .

Downloads last month 382,421

```
image<|media_content|><|media_pad|><|media_end|>\n {% elif content['type'] == 'video' or content['type']==
'video_url'-%}\n <|kimi_k25_video_placeholder|>\n {% else -%}\n {{ content['text'] }}\n {%- endif -%}\n {%- endfor -
%}\n {%- endif -%}\n{%- endmacro -%}\n\n{% macro set_roles(message) -%}\n {%- set role_name =
message.get('name') or message['role'] -%}\n {%- if message['role'] == 'user' -%}\n <|im_user|>{{role_name}}<|
im_middle|>\n {%- elif message['role'] == 'assistant' -%}\n <|im_assistant|>{{role_name}}<|im_middle|>\n {%- else -
%}\n <|im_system|>{{role_name}}<|im_middle|>\n {%- endif -%}\n{%- endmacro -%}\n\n{% macro
render_toolcalls(message) -%}\n <|tool_calls_section_begin|>\n {%- for tool_call in message['tool_calls'] -%}\n {%-
set formatted_id = tool_call['id'] -%}\n <|tool_call_begin|>{{ formatted_id }}<|tool_call_argument_begin|>{% if
tool_call['function']['arguments'] is string %}{{ tool_call['function']['arguments'] }}{% else %}{{ tool_call['function']
['arguments'] | tojson }}{% endif %}<|tool_call_end|>\n {%- endfor -%}\n <|tool_calls_section_end|>\n{%- endmacro
-%}\n\n{%- set preserve_thinking = preserve_thinking | default(false) -%}\n\n# Find last non-tool-call assistant
message. If preserve_thinking, keep -1 so hist is empty and all msgs use suffix (retain reasoning). #}\n{%- set ns
= namespace(last_non_tool_call_assistant_msg=-1) -%}\n{%- if not preserve_thinking -%}\n{%- for idx in
range(messages|length-1, -1, -1) -%}\n {%- if messages[idx]['role'] == 'assistant' and not
messages[idx].get('tool_calls') -%}\n {%- set ns.last_non_tool_call_assistant_msg = idx -%}\n {%- break -%}\n {%-
endif -%}\n{%- endfor -%}\n{%- endif -%}\n\n# split all messages into history & suffix, reasoning_content in suffix
should be reserved.#}\n{%- set hist_msgs = messages[:ns.last_non_tool_call_assistant_msg+1] -%}\n{%- set
suffix_msgs = messages[ns.last_non_tool_call_assistant_msg+1:] -%}\n\n{%- if tools -%}\n {%- if tools_ts_str -
%}\n <|im_system|>tool_declare<|im_middle|>{{ tools_ts_str }}<|im_end|>\n {%- else -%}\n <|im_system|
>tool_declare<|im_middle|>{{ tools | tojson(separators=(',', ':')) }}<|im_end|>\n {%- endif -%}\n{%- endif -%}\n\n
\n{%- for message in hist_msgs -%}\n {{set_roles(message)}}\n {%- if message['role'] == 'assistant' -%}\n
<think></think>{{render_content(message)}}\n {%- if message.get('tool_calls') -%}\n
{{render_toolcalls(message)}}\n {%- endif -%}\n {%- elif message['role'] == 'tool' -%}\n {%- set tool_call_id =
message.tool_call_id -%}\n ## Return of {{ tool_call_id }}\n{{render_content(message)}}\n {%- elif
message['content'] is not none -%}\n {{render_content(message)}}\n {%- endif -%}\n <|im_end|>\n{%- endfor -
%}\n\n{%- for message in suffix_msgs -%}\n {{set_roles(message)}}\n {%- if message['role'] == 'assistant' -%}\n
{%- if thinking is defined and thinking is false and preserve_thinking is false -%}\n <think></
think>{{render_content(message)}}\n {%- else -%}\n {%- set rc = message.get('reasoning',
message.get('reasoning_content', '')) -%}\n <think>{{rc}}</think>{{render_content(message)}}\n {%- endif -%}\n
{%- if message.get('tool_calls') -%}\n {{render_toolcalls(message)}}\n {%- endif -%}\n {%- elif message['role'] ==
'tool' -%}\n {%- set tool_call_id = message.tool_call_id -%}\n ## Return of {{ tool_call_id
}}\n{{render_content(message)}}\n {%- elif message['content'] is not none -%}\n {{render_content(message)}}\n
```

```

{% endif %}
<|im_end|>
{% endfor %}
{% if add_generation_prompt %}
<|im_assistant|>assistant<|
im_middle|>
{% if thinking is defined and thinking is false %}
<think></think>
{% else %}
<think>
{% endif %}
{% endif %}
}, "createdAt": "2026-04-14T04:23:36.000Z", "discussionsDisabled": false, "discussionsSorting": "recently-created", "downloads": 382421, "downloadsAllTime": 8561780, "id": "moonshotai/Kimi-K2.6", "isLikedByUser": false, "availableInferenceProviders": [
  {
    "provider": "novita", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/kimi-k2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features": {
      "structuredOutput": false, "toolCalling": true, "isCheapestPricingOutput": true, "isFastestThroughput": true, "isModelAuth": true
    }, "provider": "featherless-ai", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuthor": false,
    {
      "provider": "deepinfra", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features": {
        "structuredOutput": true, "toolCalling": true, "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuth": true
      }, "provider": "baseten", "modelStatus": "error", "providerStatus": "live", "providerId": "moonshotai/Kimi-K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features": {
        "structuredOutput": true, "toolCalling": true, "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuth": true
      }, "text-to-text", "library_name": "transformers", "librariesOther": [], "trackDownloads": true, "model-index": null, "evalResults": {
        "dataset": {
          "id": "SWE-bench/SWE-bench_Verified", "isBenchmark": true, "task_id": "swe_bench_%_resolved", "value": 80.2, "source": {
            "url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {
              "id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,
              swe_bench_verified.yaml", "verified": false, "rank": 8, "label": "Swe Bench Resolved", "dataset": {
                "id": "Idavidrein/gpqa", "isBenchmark": true, "task_id": "diamond", "value": 90.5, "source": {
                  "url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {
                    "id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,
                    gpqa_diamond.yaml", "verified": false, "rank": 12, "label": "Diamond", "dataset": {
                      "id": "ScaleAI/SWE-bench_Pro", "isBenchmark": true, "task_id": "SWE_Bench_Pro", "value": 58.6, "source": {
                        "url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {
                          "id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,
                          swe_bench_pro.yaml", "verified": false, "rank": 15, "label": "SWE Bench Pro", "dataset": {
                            "id": "MathArena/aime_2026", "isBenchmark": true, "task_id": "MathArena/aime_2026", "value": 96.4, "source": {
                              "url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {
                                "id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,
                                aime_2026.yaml", "verified": false, "rank": 6, "label": "MathArena Aime 2026", "dataset": {
                                  "id": "MMMU/MMMU_Pro", "isBenchmark": true, "task_id": "mmmu_pro_vision", "value": 79.4, "source": {
                                    "url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "filename": ".eval_results/mmmu_pro.yaml", "verified": false, "rank": 2, "label": "Mmmu Pro Vision", "dataset": {
                                      "id": "harborframework/terminal-bench-2.0", "isBenchmark": true, "task_id": "terminalbench_2", "value": 66.7, "source": {
                                        "url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {
                                          "id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,
                                          terminal_bench_2.yaml", "verified": false, "rank": 7, "label": "Terminalbench 2", "dataset": {
                                            "id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb_solved", "value": 0, "source": {
                                              "url": "https://zli12321.github.io/LHTB/leaderboard.html", "name": "LHTB leaderboard", "isExternal": true, "filename": ".eval_results/lhtb.yaml", "verified": false, "pullRequest": 55, "rank": 9, "label": "Lhtb Solved", "notes": "0/46 tasks solved at reward >=

```

0.95 (pass@0.95); mean reward x100 = 25.5; official LHTB Harbor harness"}, {"dataset": {"id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb2_solved"}, "value": 2, "source": {"url": "https://github.com/zli12321/LHTB/tree/lhtb-2.0#lhtb-20-results-october-2026", "name": "LHTB 2.0 results", "isExternal": true}, "filename": ".eval_results/lhtb2_solved.yaml", "verified": false, "pullRequest": 55, "rank": 8, "label": "Lhtb2 Solved", "notes": "2/77 tasks solved at reward >= 0.95 (pass@0.95); mean reward x100 = 42.0; LHTB 2.0 (77 tasks), terminus-2, 3-hour budget per task, official LHTB Harbor harness"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "chi_bench"}, "value": 15.6, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Chi Bench", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "prior_authorization"}, "value": 18.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Prior Authorization", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "utilization_management"}, "value": 21.3, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 3, "label": "Utilization Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "care_management"}, "value": 6.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 1, "label": "Care Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle"}, "value": 34.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwitt", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle.yaml", "verified": false, "label": "Hle"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle"}, "value": 54, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwitt", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle_with_tools.yaml", "verified": false, "rank": 9, "label": "Hle", "notes": "With tools"}, {"dataset": {"id": "MathArena/hmmt_feb_2026", "isBenchmark": true, "task_id": "MathArena/hmmt_feb_2026"}, "value": 92.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwitt", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hmmt_feb_2026.yaml", "verified": false, "rank": 2, "label": "MathArena Hmmt Feb 2026"}, {"dataset": {"id": "joelniklaus/LEXam-hard", "isBenchmark": true, "task_id": "lexam_hard"}, "value": 36.18, "source": {"url": "https://huggingface.co/buckets/joelniklaus/SwissLegalEvals", "name": "SwissLegalEvals per-sample details (lighteval)", "isExternal": true, "author": {"id": "5fae5f68b8423e1d80b8a988", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1656078368144-5fae5f68b8423e1d80b8a988.jpeg", "fullName": "Joel Niklaus", "name": "joelniklaus", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 17, "lexam-hard.yaml", "verified": false, "pullRequest": 58, "rank": 9, "label": "Lexam Hard", "notes": "lighteval, LEXam paper prompts, one response per question, no tools; DeepSeek-R1-0528 judge; mean of the German and English means over the 518 questions, 0-100"}, {"dataset": {"id": "SWE-bench/SWE-bench_Multilingual", "isBenchmark": true, "task_id": "swe_bench_multilingual_%_resolved"}, "value": 76.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false}, "filename": ".eval_results/swe-bench_multilingual.yaml", "verified": false, "pullRequest": 57, "rank": 7, "label": "Swe Bench Multilingual Resolved"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "overall"}, "value": 89.76, "source": {"url": "https://huggingface.co/

datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench
 Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type":
 real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 6, "label": "Overall"}, {"dataset":
 {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "scanning", "value": 90.08, "source":
 {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench
 Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type":
 real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Scanning"}, {"dataset":
 {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "warping", "value": 89.62, "source":
 {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench
 Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type":
 real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Warping"}, {"dataset":
 {"id": "PaddlePaddle/Real5-
 OmniDocBench", "isBenchmark": true, "task_id": "screen_photography", "value": 89.58, "source": {"url": "https://
 huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench
 Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type":
 real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 6, "label": "Screen Photography"}, {"dataset":
 {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "illumination", "value": 89.91, "source":
 {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench
 Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type":
 real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Illumination"}, {"dataset":
 {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "skew", "value": 89.61, "source":
 {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench
 Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type":
 real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 5, "label": "Skew"}, {"dataset": {"id": "benchflow/
 skillsbench", "isBenchmark": true, "task_id": "skillsbench_v1_1", "value": 54, "source": {"url": "https://huggingface.co/
 datasets/benchflow/skillsbench-leaderboard/raw/main/leaderboard/skillsbench/v1.1/
 official.json", "name": "SkillsBench v1.1 official leaderboard", "isExternal": true, "author":
 {"_id": "69657f24ec1d157f1590a81d", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/
 uploads/663fe2d26304d377fc253322/L6ik2VpL5iizADEXK4W-
 A.png", "fullName": "BenchFlow", "name": "benchflow", "type": "org", "isHf": false, "isHfAdmin": false, "isMod": false, "plan": "te:
 skillsbench.yaml", "verified": false, "pullRequest": 53, "rank": 5, "label": "Skillsbench V1 1", "notes": "with-skills;
 BenchFlow harness; OpenHands agent; 87 tasks x 3 trials; full 261/261 coverage"}, {"dataset": {"id": "InternScience/
 ResearchClawBench", "isBenchmark": false, "task_id": "overall", "value": 18, "source": {"url": "https://huggingface.co/
 moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "filename": ".eval_results/
 researchclawbench.yaml", "verified": false, "pullRequest": 42, "rank": 3, "label": "Overall", "notes": "ResearchHarness
 evaluation with tools enabled, code execution, and a file-system workspace; completed 38/40
 ResearchClawBench tasks."}, {"dataset": {"id": "claw-eval/Claw-
 Eval", "isBenchmark": true, "task_id": "general", "value": 61.5, "source": {"url": "https://claw-
 eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author":
 {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/
 uploads/6038d6d0612f5eef3cc05ea9/ryhVAX5djQpD5OrllZQ1f.jpeg", "fullName": "Lei
 Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse
 claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 15, "label": "General", "notes": "Pass% | N=3 | 161 tasks"},
 {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multi_turn", "value": 65.8, "source":

```

{"url":"https://claw-eval.github.io","name":"Claw-Eval Leaderboard","isExternal":true,"author":
{"_id":"6038d6d0612f5eef3cc05ea9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/
uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg","fullName":"Lei
Li","name":"tobiaslee","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":30,"isUse
claw_eval.yaml","verified":false,"pullRequest":30,"rank":1,"label":"Multi Turn","notes":"Pass³% | N=3 | 38 tasks"},
{"dataset":{"id":"claw-eval/Claw-Eval","isBenchmark":true,"task_id":"multimodal"},"value":18.8,"source":
{"url":"https://claw-eval.github.io","name":"Claw-Eval Leaderboard","isExternal":true,"author":
{"_id":"6038d6d0612f5eef3cc05ea9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/
uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg","fullName":"Lei
Li","name":"tobiaslee","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":30,"isUse
claw_eval.yaml","verified":false,"pullRequest":30,"rank":5,"label":"Multimodal","notes":"Pass³% | N=3 | 101 tasks"},
{"dataset":{"id":"collinear-ai/yc-bench","isBenchmark":true,"task_id":"medium"},"value":511137,"source":
{"url":"https://github.com/collinear-ai/yc-bench","name":"YC-Bench
eval","isExternal":true,"filename":".eval_results/yc-
bench.yaml","verified":false,"pullRequest":29,"rank":5,"label":"Medium","notes":"medium preset, avg final funds
(USD) across seeds 1,2,3"},"private":false,"repoType":"model","gated":false,"tags":
["transformers","safetensors","kimi_k25","feature-extraction","compressed-tensors","image-text-to-
text","conversational","custom_code","arxiv:2602.02276","license:other","eval-results","region:us"],"tag_objs":
[{"id":"image-text-to-text","label":"Image-Text-to-Text","type":"pipeline_tag","subType":"multimodal"},
{"id":"transformers","label":"Transformers","type":"library"},{"id":"safetensors","label":"Safetensors","type":"library"},
{"id":"kimi_k25","label":"kimi_k25","type":"other","clickable":true},{"id":"feature-extraction","label":"feature-
extraction","type":"other","clickable":true},{"id":"compressed-tensors","label":"compressed-
tensors","type":"other","clickable":true},
{"id":"conversational","label":"conversational","type":"other","clickable":true},
{"id":"custom_code","label":"custom_code","type":"other","clickable":true},{"id":"eval-results","label":"Eval
Results","type":"other","clickable":true},{"id":"arxiv:2602.02276","label":"arxiv:2602.02276","type":"arxiv","extra":
{"paperTitle":"Kimi K2.5: Visual Agentic Intelligence"}},{"id":"license:other","label":"other","type":"license"},
{"type":"region","label":"🇺🇸 Region: US","id":"region:us"},"transformersInfo":
{"auto_model":"AutoModel","custom_class":"modeling_kimi_k25.KimiK25ForConditionalGeneration","pipeline_tag":"ft
extraction","processor":"AutoProcessor"},"widgetData":[{"text":"Hi, what can you help me with?"},{"text":"What is
84 * 3 / 2?"},{"text":"Tell me an interesting fact about the universe!"},{"text":"Explain quantum computing in simple
terms."}],safetensors":{"parameters":
{"F32":23040,"I32":1014686023680,"BF16":12193329648},"total":1026879376368,"sharded":true,"totalFileSize":5952
{"billableEntities":[],"entityName2Providers":{},"defaultProviders":
[{"isOriginalProvider":false,"name":"novita","enabled":true,"position":1,"isReleased":true,"accuratePricing":true},
{"isOriginalProvider":false,"name":"deepinfra","enabled":true,"position":15,"isReleased":true,"accuratePricing":true},
{"isOriginalProvider":false,"name":"featherless-
ai","enabled":true,"position":7,"isReleased":true,"accuratePricing":true},
{"isOriginalProvider":false,"name":"baseten","enabled":true,"position":13,"isReleased":true,"accuratePricing":true}]],"c
Safetensors Model size 1T params

```

Tensor type F32 · I32 · BF16 ·

Chat template Files info

```

image<|media_content|><|media_pad|><|media_end|>\n {% elif content['type'] == 'video' or content['type']==
'video_url'-%>\n <|kimi_k25_video_placeholder|>\n {% else -%>\n {{ content['text'] }}\n {%- endif -%>\n {%- endfor -
%>\n {%- endif -%>\n{%- endmacro -%>\n\n{% macro set_roles(message) -%>\n {%- set role_name =
message.get('name') or message['role'] -%>\n {%- if message['role'] == 'user' -%>\n <|im_user|>{{role_name}}<|
im_middle|>\n {%- elif message['role'] == 'assistant' -%>\n <|im_assistant|>{{role_name}}<|im_middle|>\n {%- else -
%>\n <|im_system|>{{role_name}}<|im_middle|>\n {%- endif -%>\n{%- endmacro -%>\n\n{% macro
render_toolcalls(message) -%>\n <|tool_calls_section_begin|>\n {%- for tool_call in message['tool_calls'] -%>\n {%-
set formatted_id = tool_call['id'] -%>\n <|tool_call_begin|>{{ formatted_id }}<|tool_call_argument_begin|>{% if
tool_call['function']['arguments'] is string %>{{ tool_call['function']['arguments'] }}{% else %>{{ tool_call['function']
['arguments'] | tojson }}{% endif %><|tool_call_end|>\n {%- endfor -%>\n <|tool_calls_section_end|>\n{%- endmacro
-%>\n\n{%- set preserve_thinking = preserve_thinking | default(false) -%>\n# Find last non-tool-call assistant
message. If preserve_thinking, keep -1 so hist is empty and all msgs use suffix (retain reasoning). #>\n{%- set ns
= namespace(last_non_tool_call_assistant_msg=-1) -%>\n{%- if not preserve_thinking -%>\n{%- for idx in

```

```

range(messages|length-1, -1, -1) -%}\n %- if messages[idx]['role'] == 'assistant' and not
messages[idx].get('tool_calls') -%}\n %- set ns.last_non_tool_call_assistant_msg = idx -%}\n %- break -%}\n %-
endif -%}\n %- endfor -%}\n %- endif -%}\n\n# split all messages into history & suffix, reasoning_content in suffix
should be reserved.#}\n %- set hist_msgs = messages[:ns.last_non_tool_call_assistant_msg+1] -%}\n %- set
suffix_msgs = messages[ns.last_non_tool_call_assistant_msg+1:] -%}\n\n %- if tools -%}\n %- if tools_ts_str -
%}\n <|im_system|>tool_declare<|im_middle|>{{ tools_ts_str }}<|im_end|>\n %- else -%}\n %- <|im_system|
>tool_declare<|im_middle|>{{ tools | tojson(separators=(',', ':')) }}<|im_end|>\n %- endif -%}\n %- endif -%}\n\n
\n %- for message in hist_msgs -%}\n {{set_roles(message)}}\n %- if message['role'] == 'assistant' -%}\n
<think></think>{{render_content(message)}}\n %- if message.get('tool_calls') -%}\n
{{render_toolcalls(message)}}\n %- endif -%}\n %- elif message['role'] == 'tool' -%}\n %- set tool_call_id =
message.tool_call_id -%}\n ## Return of {{ tool_call_id }}\n {{render_content(message)}}\n %- elif
message['content'] is not none -%}\n {{render_content(message)}}\n %- endif -%}\n %- <|im_end|>\n %- endfor -
%}\n\n %- for message in suffix_msgs -%}\n {{set_roles(message)}}\n %- if message['role'] == 'assistant' -%}\n
 %- if thinking is defined and thinking is false and preserve_thinking is false -%}\n <think></
think>{{render_content(message)}}\n %- else -%}\n %- set rc = message.get('reasoning',
message.get('reasoning_content', '')) -%}\n <think>{{rc}}</think>{{render_content(message)}}\n %- endif -%}\n
 %- if message.get('tool_calls') -%}\n {{render_toolcalls(message)}}\n %- endif -%}\n %- elif message['role'] ==
'tool' -%}\n %- set tool_call_id = message.tool_call_id -%}\n ## Return of {{ tool_call_id
}}\n {{render_content(message)}}\n %- elif message['content'] is not none -%}\n {{render_content(message)}}\n
 %- endif -%}\n %- <|im_end|>\n %- endfor -%}\n\n %- if add_generation_prompt -%}\n %- <|im_assistant|>assistant<|
im_middle|>\n %- if thinking is defined and thinking is false -%}\n <think></think>\n %- else -%}\n <think>\n %-
endif -%}\n %- endif -
%}\n", "createdAt": "2026-04-14T04:23:36.000Z", "discussionsDisabled": false, "discussionsSorting": "recently-
created", "downloads": 382421, "downloadsAllTime": 8561780, "id": "moonshotai/Kimi-
K2.6", "isLikedByUser": false, "availableInferenceProviders":
[{"provider": "novita", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/kimi-
k2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features":
{"structuredOutput": false, "toolCalling": true}, "isCheapestPricingOutput": true, "isFastestThroughput": true, "isModelAuthc
{"provider": "featherless-ai", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-
of-000064.safetensors", "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuthor": false},
{"provider": "deepinfra", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features":
{"structuredOutput": true, "toolCalling": true}, "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuth
{"provider": "baseten", "modelStatus": "error", "providerStatus": "live", "providerId": "moonshotai/Kimi-
K2.6", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features":
{"structuredOutput": true, "toolCalling": true}, "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuth
text-to-text", "library_name": "transformers", "librariesOther": [], "trackDownloads": true, "model-
index": null, "evalResults": [{"dataset": {"id": "SWE-bench/SWE-
bench_Verified", "isBenchmark": true, "task_id": "swe_bench_%_resolved"}, "value": 80.2, "source": {"url": "https://
huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author":
{"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/
uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan
Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,
swe_bench_verified.yaml", "verified": false, "rank": 8, "label": "Swe Bench Resolved"}, {"dataset": {"id": "Idavidrein/
gpqa", "isBenchmark": true, "task_id": "diamond"}, "value": 90.5, "source": {"url": "https://huggingface.co/moonshotai/
Kimi-K2.6", "name": "Model Card", "isExternal": false, "author":
{"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/
uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan
Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,
gpqa_diamond.yaml", "verified": false, "rank": 12, "label": "Diamond"}, {"dataset": {"id": "ScaleAI/SWE-
bench_Pro", "isBenchmark": true, "task_id": "SWE_Bench_Pro"}, "value": 58.6, "source": {"url": "https://huggingface.co/
moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author":
{"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/
uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan
Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490,

```

swe_bench_pro.yaml", "verified": false, "rank": 15, "label": "SWE Bench Pro"}, {"dataset": {"id": "MathArena/aime_2026", "isBenchmark": true, "task_id": "MathArena/aime_2026"}, {"value": 96.4, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "aime_2026.yaml", "verified": false, "rank": 6, "label": "MathArena Aime 2026"}, {"dataset": {"id": "MMMU/MMMU_Pro", "isBenchmark": true, "task_id": "mmmu_pro_vision", "value": 79.4, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "filename": ".eval_results/mmmu_pro.yaml", "verified": false, "rank": 2, "label": "Mmmu Pro Vision"}, {"dataset": {"id": "harborframework/terminal-bench-2.0", "isBenchmark": true, "task_id": "terminalbench_2", "value": 66.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "terminal_bench_2.yaml", "verified": false, "rank": 7, "label": "Terminalbench 2"}, {"dataset": {"id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb_solved", "value": 0, "source": {"url": "https://zli12321.github.io/LHTB/leaderboard.html", "name": "LHTB leaderboard", "isExternal": true, "filename": ".eval_results/lhtb.yaml", "verified": false, "pullRequest": 55, "rank": 9, "label": "Lhtb Solved", "notes": "0/46 tasks solved at reward >= 0.95 (pass@0.95); mean reward x100 = 25.5; official LHTB Harbor harness"}, {"dataset": {"id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb2_solved", "value": 2, "source": {"url": "https://github.com/zli12321/LHTB/tree/lhtb-2.0#lhtb-20-results-october-2026", "name": "LHTB 2.0 results", "isExternal": true, "filename": ".eval_results/lhtb2_solved.yaml", "verified": false, "pullRequest": 55, "rank": 8, "label": "Lhtb2 Solved", "notes": "2/77 tasks solved at reward >= 0.95 (pass@0.95); mean reward x100 = 42.0; LHTB 2.0 (77 tasks), terminus-2, 3-hour budget per task, official LHTB Harbor harness"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "chi_bench", "value": 15.6, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Chi Bench", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "prior_authorization", "value": 18.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Prior Authorization", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "utilization_management", "value": 21.3, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 3, "label": "Utilization Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "care_management", "value": 6.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 1, "label": "Care Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle", "value": 34.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle.yaml", "verified": false, "label": "Hle"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle", "value": 54, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle_with_tools.yaml", "verified": false, "rank": 9, "label": "Hle", "notes": "With tools"}, {"dataset": {"id": "MathArena/hmmt_feb_2026", "isBenchmark": true, "task_id": "MathArena/hmmt_feb_2026", "value": 92.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author":

{"_id":"63e0eea7af523c37e5a77966","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg","fullName":"Nathan Habib","name":"SaylorTwitt","type":"user","isPro":true,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":490,"hmmmt_feb_2026.yaml","verified":false,"rank":2,"label":"MathArena Hmmt Feb 2026"},{"dataset":{"id":"joelniklaus/LEXam-hard","isBenchmark":true,"task_id":"lexam_hard"},"value":36.18,"source":{"url":"https://huggingface.co/buckets/joelniklaus/SwissLegalEvals","name":"SwissLegalEvals per-sample details (lighteval)","isExternal":true,"author":{"_id":"5fae5f68b8423e1d80b8a988","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1656078368144-5fae5f68b8423e1d80b8a988.jpeg","fullName":"Joel Niklaus","name":"joelniklaus","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":17,"lexam-hard.yaml","verified":false,"pullRequest":58,"rank":9,"label":"Lexam Hard","notes":"lighteval, LEXam paper prompts, one response per question, no tools; DeepSeek-R1-0528 judge; mean of the German and English means over the 518 questions, 0-100"},{"dataset":{"id":"SWE-bench/SWE-bench_Multilingual","isBenchmark":true,"task_id":"swe_bench_multilingual_%_resolved"},"value":76.7,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.6","name":"Model Card","isExternal":false,"filename":"eval_results/swe-bench_multilingual.yaml","verified":false,"pullRequest":57,"rank":7,"label":"Swe Bench Multilingual Resolved"},{"dataset":{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"overall"},"value":89.76,"source":{"url":"https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench","name":"Real5-OmniDocBench Leaderboard","isExternal":true,"author":{"_id":"62067d5d3906f102bc9658bd","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png","fullName":"PaddlePaddle","name":"PaddlePaddle","type":"real5_omnidocbench.yaml","verified":false,"pullRequest":56,"rank":6,"label":"Overall"},{"dataset":{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"scanning"},"value":90.08,"source":{"url":"https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench","name":"Real5-OmniDocBench Leaderboard","isExternal":true,"author":{"_id":"62067d5d3906f102bc9658bd","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png","fullName":"PaddlePaddle","name":"PaddlePaddle","type":"real5_omnidocbench.yaml","verified":false,"pullRequest":56,"rank":7,"label":"Scanning"},{"dataset":{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"warping"},"value":89.62,"source":{"url":"https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench","name":"Real5-OmniDocBench Leaderboard","isExternal":true,"author":{"_id":"62067d5d3906f102bc9658bd","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png","fullName":"PaddlePaddle","name":"PaddlePaddle","type":"real5_omnidocbench.yaml","verified":false,"pullRequest":56,"rank":7,"label":"Warping"},{"dataset":{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"screen_photography"},"value":89.58,"source":{"url":"https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench","name":"Real5-OmniDocBench Leaderboard","isExternal":true,"author":{"_id":"62067d5d3906f102bc9658bd","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png","fullName":"PaddlePaddle","name":"PaddlePaddle","type":"real5_omnidocbench.yaml","verified":false,"pullRequest":56,"rank":6,"label":"Screen Photography"},{"dataset":{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"illumination"},"value":89.91,"source":{"url":"https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench","name":"Real5-OmniDocBench Leaderboard","isExternal":true,"author":{"_id":"62067d5d3906f102bc9658bd","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png","fullName":"PaddlePaddle","name":"PaddlePaddle","type":"real5_omnidocbench.yaml","verified":false,"pullRequest":56,"rank":7,"label":"Illumination"},{"dataset":{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"skew"},"value":89.61,"source":{"url":"https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench","name":"Real5-OmniDocBench Leaderboard","isExternal":true,"author":{"_id":"62067d5d3906f102bc9658bd","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png","fullName":"PaddlePaddle","name":"PaddlePaddle","type":"real5_omnidocbench.yaml","verified":false,"pullRequest":56,"rank":5,"label":"Skew"},{"dataset":{"id":"benchflow/skillsbench","isBenchmark":true,"task_id":"skillsbench_v1_1"},"value":54,"source":{"url":"https://huggingface.co/

datasets/benchflow/skillsbench-leaderboard/raw/main/leaderboard/skillsbench/v1.1/official.json", "name": "SkillsBench v1.1 official leaderboard", "isExternal": true, "author": {"_id": "69657f24ec1d157f1590a81d", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/663fe2d26304d377fc253322/L6ik2VpL5iizADEXK4W-A.png", "fullName": "BenchFlow", "name": "benchflow", "type": "org", "isHf": false, "isHfAdmin": false, "isMod": false, "plan": "te: skillsbench.yaml", "verified": false, "pullRequest": 53, "rank": 5, "label": "Skillsbench V1 1", "notes": "with-skills; BenchFlow harness; OpenHands agent; 87 tasks x 3 trials; full 261/261 coverage"}, {"dataset": {"id": "InternScience/ResearchClawBench", "isBenchmark": false, "task_id": "overall", "value": 18, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false}, "filename": ".eval_results/researchclawbench.yaml", "verified": false, "pullRequest": 42, "rank": 3, "label": "Overall", "notes": "ResearchHarness evaluation with tools enabled, code execution, and a file-system workspace; completed 38/40 ResearchClawBench tasks."}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "general", "value": 61.5, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 15, "label": "General", "notes": "Pass3% | N=3 | 161 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multi_turn", "value": 65.8, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 1, "label": "Multi Turn", "notes": "Pass3% | N=3 | 38 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multimodal", "value": 18.8, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrlIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 5, "label": "Multimodal", "notes": "Pass3% | N=3 | 101 tasks"}, {"dataset": {"id": "collinear-ai/yc-bench", "isBenchmark": true, "task_id": "medium", "value": 511137, "source": {"url": "https://github.com/collinear-ai/yc-bench", "name": "YC-Bench eval", "isExternal": true}, "filename": ".eval_results/yc-bench.yaml", "verified": false, "pullRequest": 29, "rank": 5, "label": "Medium", "notes": "medium preset, avg final funds (USD) across seeds 1,2,3"}], "private": false, "repoType": "model", "gated": false, "tags": ["transformers", "safetensors", "kimi_k25", "feature-extraction", "compressed-tensors", "image-text-to-text", "conversational", "custom_code", "arxiv:2602.02276", "license:other", "eval-results", "region:us"], "tag_objs": [{"id": "image-text-to-text", "label": "Image-Text-to-Text", "type": "pipeline_tag", "subType": "multimodal"}, {"id": "transformers", "label": "Transformers", "type": "library"}, {"id": "safetensors", "label": "Safetensors", "type": "library"}, {"id": "kimi_k25", "label": "kimi_k25", "type": "other", "clickable": true}, {"id": "feature-extraction", "label": "feature-extraction", "type": "other", "clickable": true}, {"id": "compressed-tensors", "label": "compressed-tensors", "type": "other", "clickable": true}, {"id": "conversational", "label": "conversational", "type": "other", "clickable": true}, {"id": "custom_code", "label": "custom_code", "type": "other", "clickable": true}, {"id": "eval-results", "label": "Eval Results", "type": "other", "clickable": true}, {"id": "arxiv:2602.02276", "label": "arxiv:2602.02276", "type": "arxiv", "extra": {"paperTitle": "Kimi K2.5: Visual Agentic Intelligence"}}, {"id": "license:other", "label": "other", "type": "license"}, {"type": "region", "label": "🇺🇸 Region: US", "id": "region:us"}], "transformersInfo": {"auto_model": "AutoModel", "custom_class": "modeling_kimi_k25.KimiK25ForConditionalGeneration", "pipeline_tag": "feature-extraction", "processor": "AutoProcessor"}, "widgetData": [{"text": "Hi, what can you help me with?"}, {"text": "What is 84 * 3 / 2?"}, {"text": "Tell me an interesting fact about the universe!"}, {"text": "Explain quantum computing in simple terms."}], "safetensors": {"parameters": {"F32": 23040, "I32": 1014686023680, "BF16": 12193329648}, "total": 1026879376368, "sharded": true, "totalFileSize": 5952}, {"inferenceContextData": {"billableEntities": [], "entityName2Providers": {}, "defaultProviders": [{"isOriginalProvider": false, "name": "novita", "enabled": true, "position": 1, "isReleased": true, "accuratePricing": true}, {"isOriginalProvider": false, "name": "deepinfra", "enabled": true, "position": 15, "isReleased": true, "accuratePricing": true}, {"isOriginalProvider": false, "name": "featherless-

```
ai","enabled":true,"position":7,"isReleased":true,"accuratePricing":true},
{"isOriginalProvider":false,"name":"baseten","enabled":true,"position":13,"isReleased":true,"accuratePricing":true}},{"text":"Hi, what can you help me with?","text":"What is  $84 \times 3 / 2$ ?","text":"Tell me an interesting fact about the universe!","text":"Explain quantum computing in simple terms."}],{"initialSelectedProvider":"novita","widgetType":"conversational","isMaximized":false,"isMac":true}}">
Inference Providers NEW Novita
```

+1

Image-Text-to-Text

Examples

Input a message to start chatting with moonshotai/Kimi-K2.6 .

Send

View Code Snippets

Compare providers

Model tree for moonshotai/Kimi-K2.6

Adapters

9 models

Finetunes

18 models

Merges

2 models

Quantizations

46 models

Spaces using moonshotai/Kimi-K2.6 100

Collection including moonshotai/Kimi-K2.6 Moonshot's large visual-language model • 4 items • Updated Jul 27 • 84

Paper for moonshotai/Kimi-K2.6 Paper • 2602.02276 • Published Feb 2 • 278

```
image<|media_content|><|media_pad|><|media_end|>\n {% elif content['type'] == 'video' or content['type']==
'video_url'-%}\n <|kimi_k25_video_placeholder|>\n {% else -%}\n {{ content['text'] }}\n {%- endif -%}\n {%- endfor -
%}\n {%- endif -%}\n{%- endmacro -%}\n\n{% macro set_roles(message) -%}\n {%- set role_name =
message.get('name') or message['role'] -%}\n {%- if message['role'] == 'user' -%}\n <|im_user|>{{role_name}}<|
im_middle|>\n {%- elif message['role'] == 'assistant' -%}\n <|im_assistant|>{{role_name}}<|im_middle|>\n {%- else -
%}\n <|im_system|>{{role_name}}<|im_middle|>\n {%- endif -%}\n{%- endmacro -%}\n\n{% macro
render_toolcalls(message) -%}\n <|tool_calls_section_begin|>\n {%- for tool_call in message['tool_calls'] -%}\n {%-
set formatted_id = tool_call['id'] -%}\n <|tool_call_begin|>{{ formatted_id }}<|tool_call_argument_begin|>{% if
tool_call['function']['arguments'] is string %}{{ tool_call['function']['arguments'] }}{% else %}{{ tool_call['function']
['arguments'] | tojson }}{% endif %}<|tool_call_end|>\n {%- endfor -%}\n <|tool_calls_section_end|>\n{%- endmacro
-%}\n\n{%- set preserve_thinking = preserve_thinking | default(false) -%}\n\n# Find last non-tool-call assistant
message. If preserve_thinking, keep -1 so hist is empty and all msgs use suffix (retain reasoning). #}\n{%- set ns
= namespace(last_non_tool_call_assistant_msg=-1) -%}\n{%- if not preserve_thinking -%}\n{%- for idx in
range(messages|length-1, -1, -1) -%}\n {%- if messages[idx]['role'] == 'assistant' and not
messages[idx].get('tool_calls') -%}\n {%- set ns.last_non_tool_call_assistant_msg = idx -%}\n {%- break -%}\n {%-
endif -%}\n{%- endfor -%}\n{%- endif -%}\n\n# split all messages into history & suffix, reasoning_content in suffix
should be reserved.#}\n{%- set hist_msgs = messages[:ns.last_non_tool_call_assistant_msg+1] -%}\n{%- set
suffix_msgs = messages[ns.last_non_tool_call_assistant_msg+1:] -%}\n\n{%- if tools -%}\n {%- if tools_ts_str -
%}\n <|im_system|>tool_declare<|im_middle|>{{ tools_ts_str }}<|im_end|>\n {%- else -%}\n <|im_system|
```

```
>tool_declare<  
tools |<br>tojson(separators=(',', ':')) }<br></lim_end|>\n {%- endif -%}<br>%\n\n%\n\n{%- for message in hist_msgs -%}\n {{set_roles(message)}}\n {%- if message['role'] == 'assistant' -%}\n\n<think></think>{{render_content(message)}}\n {%- if message.get('tool_calls') -%}\n  
{{render_toolcalls(message)}}\n {%- endif -%}\n {%- elif message['role'] == 'tool' -%}\n {%- set tool_call_id =  
message.tool_call_id -%}\n ## Return of {{ tool_call_id }}\n{{render_content(message)}}\n {%- elif  
message['content'] is not none -%}\n {{render_content(message)}}\n {%- endif -%}\n <|im_end|>\n {%- endfor -  
%}\n\n{%- for message in suffix_msgs -%}\n {{set_roles(message)}}\n {%- if message['role'] == 'assistant' -%}\n  
{%- if thinking is defined and thinking is false and preserve_thinking is false -%}\n <think></  
think>{{render_content(message)}}\n {%- else -%}\n {%- set rc = message.get('reasoning',  
message.get('reasoning_content', '')) -%}\n <think>{{rc}}</think>{{render_content(message)}}\n {%- endif -%}\n  
{%- if message.get('tool_calls') -%}\n {{render_toolcalls(message)}}\n {%- endif -%}\n {%- elif message['role'] ==  
'tool' -%}\n {%- set tool_call_id = message.tool_call_id -%}\n ## Return of {{ tool_call_id  
}}\n{{render_content(message)}}\n {%- elif message['content'] is not none -%}\n {{render_content(message)}}\n  
{%- endif -%}\n <|im_end|>\n {%- endfor -%}\n\n{%- if add_generation_prompt -%}\n <|im_assistant|>assistant<|  
im_middle|>\n {%- if thinking is defined and thinking is false -%}\n <think></think>\n {%- else -%}\n <think>\n {%-  
endif -%}\n {%- endif -  
%}"},"createdAt":"2026-04-14T04:23:36.000Z","discussionsDisabled":false,"discussionsSorting":"recently-  
created","downloads":382421,"downloadsAllTime":8561780,"id":"moonshotai/Kimi-K2.6","isLikedByUser":false,"availableInferenceProviders":  
[{"provider":"novita","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/kimi-k2.6","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","features":{"structuredOutput":false,"toolCalling":true},"isCheapestPricingOutput":true,"isFastestThroughput":true,"isModelAuthc {"provider":"featherless-ai","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/Kimi-K2.6","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","isCheapestPricingOutput":false,"isFastestThroughput":false,"isModelAuthor":false}, {"provider":"deepinfra","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/Kimi-K2.6","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","features":{"structuredOutput":true,"toolCalling":true},"isCheapestPricingOutput":false,"isFastestThroughput":false,"isModelAuth {"provider":"baseten","modelStatus":"error","providerStatus":"live","providerId":"moonshotai/Kimi-K2.6","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","features":{"structuredOutput":true,"toolCalling":true},"isCheapestPricingOutput":false,"isFastestThroughput":false,"isModelAuth text-to-text","library_name":"transformers","librariesOther":[],"trackDownloads":true,"model-index":null,"evalResults":[{"dataset":{"id":"SWE-bench/SWE-bench_Verified"},"isBenchmark":true,"task_id":"swe_bench_%_resolved"},"value":80.2,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.6","name":"Model Card","isExternal":false,"author":{"_id":"63e0eea7af523c37e5a77966","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg"},"fullName":"Nathan Habib","name":"SaylorTwift","type":"user","isPro":true,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":490,swe_bench_verified.yaml"},"verified":false,"rank":8,"label":"Swe Bench Resolved"}, {"dataset":{"id":"ldavidrein/gpqa"},"isBenchmark":true,"task_id":"diamond"},"value":90.5,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.6","name":"Model Card","isExternal":false,"author":{"_id":"63e0eea7af523c37e5a77966","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg"},"fullName":"Nathan Habib","name":"SaylorTwift","type":"user","isPro":true,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":490,gpqa_diamond.yaml"},"verified":false,"rank":12,"label":"Diamond"}, {"dataset":{"id":"ScaleAI/SWE-bench_Pro"},"isBenchmark":true,"task_id":"SWE_Bench_Pro"},"value":58.6,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.6","name":"Model Card","isExternal":false,"author":{"_id":"63e0eea7af523c37e5a77966","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg"},"fullName":"Nathan Habib","name":"SaylorTwift","type":"user","isPro":true,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":490,swe_bench_pro.yaml"},"verified":false,"rank":15,"label":"SWE Bench Pro"}, {"dataset":{"id":"MathArena/aime_2026"},"isBenchmark":true,"task_id":"MathArena/aime_2026"},"value":96.4,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.6","name":"Model Card","isExternal":false,"author":{"_id":"63e0eea7af523c37e5a77966","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg"},"fullName":"Nathan Habib","name":"SaylorTwift","type":"user","isPro":true,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":490,
```

aime_2026.yaml", "verified": false, "rank": 6, "label": "MathArena Aime 2026"}, {"dataset": {"id": "MMMU/MMMU_Pro", "isBenchmark": true, "task_id": "mmmu_pro_vision", "value": 79.4, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false}, "filename": ".eval_results/mmmu_pro.yaml", "verified": false, "rank": 2, "label": "Mmmu Pro Vision"}, {"dataset": {"id": "harborframework/terminal-bench-2.0", "isBenchmark": true, "task_id": "terminalbench_2", "value": 66.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "terminal_bench_2.yaml", "verified": false, "rank": 7, "label": "Terminalbench 2"}, {"dataset": {"id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb_solved", "value": 0, "source": {"url": "https://zli12321.github.io/LHTB/leaderboard.html", "name": "LHTB leaderboard", "isExternal": true}, "filename": ".eval_results/lhtb.yaml", "verified": false, "pullRequest": 55, "rank": 9, "label": "Lhtb Solved", "notes": "0/46 tasks solved at reward >= 0.95 (pass@0.95); mean reward x100 = 25.5; official LHTB Harbor harness"}, {"dataset": {"id": "IntelligenceLab/Long-Horizon-Terminal-Bench", "isBenchmark": true, "task_id": "lhtb2_solved", "value": 2, "source": {"url": "https://github.com/zli12321/LHTB/tree/lhtb-2.0#lhtb-20-results-october-2026", "name": "LHTB 2.0 results", "isExternal": true}, "filename": ".eval_results/lhtb2_solved.yaml", "verified": false, "pullRequest": 55, "rank": 8, "label": "Lhtb2 Solved", "notes": "2/77 tasks solved at reward >= 0.95 (pass@0.95); mean reward x100 = 42.0; LHTB 2.0 (77 tasks), terminus-2, 3-hour budget per task, official LHTB Harbor harness"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "chi_bench", "value": 15.6, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Chi Bench", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "prior_authorization", "value": 18.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 2, "label": "Prior Authorization", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "utilization_management", "value": 21.3, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 3, "label": "Utilization Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "actava/chi-bench", "isBenchmark": true, "task_id": "care_management", "value": 6.7, "source": {"url": "https://arxiv.org/abs/2605.16679", "name": "CHI-Bench", "isExternal": true}, "filename": ".eval_results/chi-bench.yaml", "verified": false, "pullRequest": 46, "rank": 1, "label": "Care Management", "notes": "Harness: Hermes; Protocol: 75 tasks x 3 trials; Metric: pass@1 (%)"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle", "value": 34.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle.yaml", "verified": false, "label": "Hle"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle", "value": 54, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hle_with_tools.yaml", "verified": false, "rank": 9, "label": "Hle", "notes": "With tools"}, {"dataset": {"id": "MathArena/hmmt_feb_2026", "isBenchmark": true, "task_id": "MathArena/hmmt_feb_2026", "value": 92.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "author": {"_id": "63e0eea7af523c37e5a77966", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1678663263366-63e0eea7af523c37e5a77966.jpeg", "fullName": "Nathan Habib", "name": "SaylorTwift", "type": "user", "isPro": true, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 490, "hmmt_feb_2026.yaml", "verified": false, "rank": 2, "label": "MathArena Hmmt Feb 2026"}, {"dataset": {"id": "joelniklaus/LEXam-hard", "isBenchmark": true, "task_id": "lexam_hard", "value": 36.18, "source": {"url": "https://huggingface.co/buckets/joelniklaus/SwissLegalEvals", "name": "SwissLegalEvals per-sample details

(lighteval)", "isExternal": true, "author": {"_id": "5fae5f68b8423e1d80b8a988", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1656078368144-5fae5f68b8423e1d80b8a988.jpeg", "fullName": "Joel Niklaus", "name": "joelniklaus", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 17, "lexam-hard.yaml", "verified": false, "pullRequest": 58, "rank": 9, "label": "Lexam Hard", "notes": "lighteval, LEXam paper prompts, one response per question, no tools; DeepSeek-R1-0528 judge; mean of the German and English means over the 518 questions, 0-100"}, {"dataset": {"id": "SWE-bench/SWE-bench_Multilingual", "isBenchmark": true, "task_id": "swe_bench_multilingual_%_resolved", "value": 76.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false, "filename": ".eval_results/swe-bench_multilingual.yaml", "verified": false, "pullRequest": 57, "rank": 7, "label": "Swe Bench Multilingual Resolved"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "overall", "value": 89.76, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 6, "label": "Overall"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "scanning", "value": 90.08, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Scanning"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "warping", "value": 89.62, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Warping"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "screen_photography", "value": 89.58, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 6, "label": "Screen Photography"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "illumination", "value": 89.91, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 7, "label": "Illumination"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "skew", "value": 89.61, "source": {"url": "https://huggingface.co/datasets/PaddlePaddle/Real5-OmniDocBench", "name": "Real5-OmniDocBench Leaderboard", "isExternal": true, "author": {"_id": "62067d5d3906f102bc9658bd", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1654942635336-5f3ff69679c1ba4c353d0c5a.png", "fullName": "PaddlePaddle", "name": "PaddlePaddle", "type": "real5_omnidocbench.yaml", "verified": false, "pullRequest": 56, "rank": 5, "label": "Skew"}, {"dataset": {"id": "benchflow/skillsbench", "isBenchmark": true, "task_id": "skillsbench_v1_1", "value": 54, "source": {"url": "https://huggingface.co/datasets/benchflow/skillsbench-leaderboard/raw/main/leaderboard/skillsbench/v1.1/official.json", "name": "SkillsBench v1.1 official leaderboard", "isExternal": true, "author": {"_id": "69657f24ec1d157f1590a81d", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/663fe2d26304d377fc253322/L6ik2VpL5iizADEXK4W-A.png", "fullName": "BenchFlow", "name": "benchflow", "type": "org", "isHf": false, "isHfAdmin": false, "isMod": false, "plan": "te:skillsbench.yaml", "verified": false, "pullRequest": 53, "rank": 5, "label": "Skillsbench V1 1", "notes": "with-skills;

BenchFlow harness; OpenHands agent; 87 tasks x 3 trials; full 261/261 coverage"}, {"dataset": {"id": "InternScience/ResearchClawBench", "isBenchmark": false, "task_id": "overall", "value": 18, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.6", "name": "Model Card", "isExternal": false}, "filename": ".eval_results/researchclawbench.yaml", "verified": false, "pullRequest": 42, "rank": 3, "label": "Overall", "notes": "ResearchHarness evaluation with tools enabled, code execution, and a file-system workspace; completed 38/40 ResearchClawBench tasks."}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "general", "value": 61.5, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrIIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse": "claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 15, "label": "General", "notes": "Pass³% | N=3 | 161 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multi_turn", "value": 65.8, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrIIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse": "claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 1, "label": "Multi Turn", "notes": "Pass³% | N=3 | 38 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multimodal", "value": 18.8, "source": {"url": "https://claw-eval.github.io", "name": "Claw-Eval Leaderboard", "isExternal": true, "author": {"_id": "6038d6d0612f5eef3cc05ea9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6038d6d0612f5eef3cc05ea9/ryhvAX5djQpD5OrIIZQ1f.jpeg", "fullName": "Lei Li", "name": "tobiaslee", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 30, "isUse": "claw_eval.yaml", "verified": false, "pullRequest": 30, "rank": 5, "label": "Multimodal", "notes": "Pass³% | N=3 | 101 tasks"}, {"dataset": {"id": "collinear-ai/yc-bench", "isBenchmark": true, "task_id": "medium", "value": 511137, "source": {"url": "https://github.com/collinear-ai/yc-bench", "name": "YC-Bench eval", "isExternal": true}, "filename": ".eval_results/yc-bench.yaml", "verified": false, "pullRequest": 29, "rank": 5, "label": "Medium", "notes": "medium preset, avg final funds (USD) across seeds 1,2,3"}, {"private": false, "repoType": "model", "gated": false, "tags": ["transformers", "safetensors", "kimi_k25", "feature-extraction", "compressed-tensors", "image-text-to-text", "conversational", "custom_code", "arxiv:2602.02276", "license:other", "eval-results", "region:us"], "tag_objs": [{"id": "image-text-to-text", "label": "Image-Text-to-Text", "type": "pipeline_tag", "subType": "multimodal"}, {"id": "transformers", "label": "Transformers", "type": "library"}, {"id": "safetensors", "label": "Safetensors", "type": "library"}, {"id": "kimi_k25", "label": "kimi_k25", "type": "other", "clickable": true}, {"id": "feature-extraction", "label": "feature-extraction", "type": "other", "clickable": true}, {"id": "compressed-tensors", "label": "compressed-tensors", "type": "other", "clickable": true}, {"id": "conversational", "label": "conversational", "type": "other", "clickable": true}, {"id": "custom_code", "label": "custom_code", "type": "other", "clickable": true}, {"id": "eval-results", "label": "Eval Results", "type": "other", "clickable": true}, {"id": "arxiv:2602.02276", "label": "arxiv:2602.02276", "type": "arxiv", "extra": {"paperTitle": "Kimi K2.5: Visual Agentic Intelligence"}}, {"id": "license:other", "label": "other", "type": "license"}, {"type": "region", "label": "🇺🇸 Region: US", "id": "region:us"}], "transformersInfo": {"auto_model": "AutoModel", "custom_class": "modeling_kimi_k25.KimiK25ForConditionalGeneration", "pipeline_tag": "feature-extraction", "processor": "AutoProcessor"}, "widgetData": [{"text": "Hi, what can you help me with?"}, {"text": "What is 84 * 3 / 2?"}, {"text": "Tell me an interesting fact about the universe!"}, {"text": "Explain quantum computing in simple terms."}], "safetensors": {"parameters": {"F32": 23040, "I32": 1014686023680, "BF16": 12193329648}, "total": 1026879376368, "sharded": true, "totalFileSize": 5952}

Evaluation results SWE-bench/SWE-bench_Verified · Swe Bench Resolved View evaluation results leaderboard 80.2

Idavidrein/gpqa · Diamond View evaluation results leaderboard 90.5

ScaleAI/SWE-bench_Pro · SWE Bench Pro View evaluation results leaderboard 58.6

MathArena/aime_2026 · MathArena Aime 2026 View evaluation results leaderboard 96.4

MMMU/MMMU_Pro · Mmmu Pro Vision View evaluation results leaderboard 79.4

harborframework/terminal-bench-2.0 · Terminalbench 2 View evaluation results leaderboard 66.7

IntelligenceLab/Long-Horizon-Terminal-Bench leaderboard

+2 more

Expand 10 benchmarks