

Moonshot releases Kimi K2.5, an open multimodal model with an agent swarm mode

Source copy kept by wheresthe.ai for the story "Moonshot releases Kimi K2.5, an open multimodal model with an agent swarm mode" (<https://wheresthe.ai/d/cn-moonshot-kimi-k2-5-release-2026-01/>).

Source: <https://huggingface.co/moonshotai/Kimi-K2.5>

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

SHA-256 of the saved text: 308d31309edf8ea45f227c654c6136680a09a6977e55289e4ece1998f9b8d95d

```
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\n# Find last non-tool-call assistant message #\n{%- set ns =
namespace(last_non_tool_call_assistant_msg=-1) -%}\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\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 -%}\n <think></think>{{render_content(message)}}\n {%- else -%}\n {%- set rc =
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\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", "created_at": "2026-01-01T06:06:03.000Z", "discussionsDisabled": false, "discussionsSorting": "recently-
created", "downloads": 327584, "downloadsAllTime": 16128649, "id": "moonshotai/Kimi-
K2.5", "isLikedByUser": false, "availableInferenceProviders":
[{"provider": "novita", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/kimi-
k2.5", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features":
{"structuredOutput": false, "toolCalling": true}, "isCheapestPricingOutput": false, "isFastestThroughput": true, "isModelAuth
{"provider": "featherless-ai", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
K2.5", "task": "conversational", "adapterWeightsPath": "model-00001-
of-000064.safetensors", "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuthor": false},
{"provider": "deepinfra", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
K2.5", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features":
{"structuredOutput": true, "toolCalling": true}, "isCheapestPricingOutput": true, "isFastestThroughput": false, "isModelAuthc
```

text-to-text", "library_name": "transformers", "librariesOther": [], "trackDownloads": true, "model-index": null, "evalResults": [{"dataset": {"id": "TIGER-Lab/MMLU-Pro", "isBenchmark": true, "task_id": "mmlu_pro"}, "value": 87.1, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullname": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isMmlu_pro.yaml", "verified": false, "pullRequest": 51, "rank": 6, "label": "Mmlu Pro"}, {"dataset": {"id": "SWE-bench/SWE-bench_Verified", "isBenchmark": true, "task_id": "swe_bench_%_resolved", "value": 70.8, "source": {"url": "https://www.swebench.com/", "name": "SWE-Bench official evaluation", "isExternal": true, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullname": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "is swe_bench_verified.yaml", "verified": false, "pullRequest": 97, "rank": 38, "label": "Swe Bench Resolved", "notes": "high reasoning"}, {"dataset": {"id": "Idavidrein/gpqa", "isBenchmark": true, "task_id": "diamond", "value": 87.3737373737, "source": {"url": "https://huggingface.co/datasets/evalEval/EEE_datastore/blob/192329fb7d6b15b7b0936a1a58ae862aa7e8ba24/flat/objects/5c/4b/5c4b9bd9-14db-4c11-bf99-9661461459f6.json", "name": "EvalEval", "isExternal": true}, "filename": ".eval_results/gpqa-diamond.yaml", "verified": false, "pullRequest": 126, "label": "Diamond", "notes": "GPQA Diamond"}, {"dataset": {"id": "ScaleAI/SWE-bench_Pro", "isBenchmark": true, "task_id": "SWE_Bench_Pro", "value": 50.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "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, "pullRequest": 101, "rank": 31, "label": "SWE Bench Pro"}, {"dataset": {"id": "Idavidrein/gpqa", "isBenchmark": true, "task_id": "diamond", "value": 87.6, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullname": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "is gpqa_diamond.yaml", "verified": false, "pullRequest": 50, "rank": 28, "label": "Diamond"}, {"dataset": {"id": "mercor/apex-agents", "isBenchmark": true, "task_id": "apex-agents", "value": 14.4, "source": {"url": "https://www.mercor.com/apex/apex-agents-leaderboard/", "name": "APEX-Agents Leaderboard", "isExternal": true, "author": {"_id": "6531a6d5c93c85e11c82d495", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/63b4a97834df1f9a90a0da76/HHWjFJmEitVTMv3Fk85wu.jpeg", "fullname": "Mercor", "name": "mercor", "type": "org", "isHf": false, "isHfAdmin": false, "is apex-agents.yaml", "verified": false, "pullRequest": 113, "rank": 8, "label": "Apex Agents", "notes": "harness: react_toolbelt_agent; reasoning_effort: thinking"}, {"dataset": {"id": "MMMU/MMMU_Pro", "isBenchmark": true, "task_id": "mmmu_pro_vision", "value": 78.5, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false}, "filename": ".eval_results/mmmu_pro.yaml", "verified": false, "pullRequest": 119, "rank": 3, "label": "Mmmu Pro Vision"}, {"dataset": {"id": "harborframework/terminal-bench-2.0", "isBenchmark": true, "task_id": "terminalbench_2", "value": 43.2, "source": {"url": "https://www.tbench.ai/leaderboard/terminal-bench/2.0", "name": "Terminal-Bench Leaderboard", "isExternal": true, "author": {"_id": "62d648291fa3e4e7ae3fa6e8", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/62d648291fa3e4e7ae3fa6e8/oatOwf8Xqe5eDbCSuYqCd.png", "fullname": "ben burtenshaw", "name": "burtenshaw", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCour terminal_bench.yaml", "verified": false, "pullRequest": 90, "rank": 19, "label": "Terminalbench 2", "notes": "agent: Terminus 2"}, {"dataset": {"id": "MathArena/aima_2026", "isBenchmark": true, "task_id": "MathArena/aima_2026", "value": 95.83, "source": {"url": "https://matharena.ai/?comp=aima--aima_2026", "name": "Official MathArena Evaluation", "isExternal": true}, "filename": ".eval_results/MathArena--aima_2026.yaml", "verified": false, "pullRequest": 88, "rank": 7, "label": "MathArena Aima 2026"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle", "value": 50.2, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/

uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullname": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isHle_with_tools.yaml", "verified": false, "pullRequest": 49, "rank": 17, "label": "Hle", "notes": "With tools"}, {"dataset": {"id": "SWE-bench/SWE-bench_Multilingual", "isBenchmark": true, "task_id": "swe_bench_multilingual_%_resolved"}, {"value": 73, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "filename": ".eval_results/swe-bench_multilingual.yaml", "verified": false, "pullRequest": 131, "rank": 14, "label": "Swe Bench Multilingual Resolved"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "overall"}, {"value": 89.09, "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": 130, "rank": 7, "label": "Overall"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "scanning"}, {"value": 89.67, "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": 130, "rank": 9, "label": "Scanning"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "warping"}, {"value": 88.86, "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": 130, "rank": 8, "label": "Warping"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "screen_photography"}, {"value": 88.39, "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": 130, "rank": 10, "label": "Screen Photography"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "illumination"}, {"value": 89.66, "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": 130, "rank": 8, "label": "Illumination"}, {"dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "skew"}, {"value": 88.86, "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": 130, "rank": 6, "label": "Skew"}, {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "overall"}, {"value": 77.5, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullname": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}}, {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "digital"}, {"value": 85, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullname": "Delores Lin", "name": "Delores-

```

Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 9, "label": "Digital"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "photographed"}, {"value": 75, "source": {"url": "https://huggingface.co/
datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 7, "label": "Photographed"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "latin"}, {"value": 81.6, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Latin"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "de"}, {"value": 85.9, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 3, "label": "De"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "en"}, {"value": 86.2, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 4, "label": "En"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "es"}, {"value": 72.7, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Es"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "fr"}, {"value": 71, "source": {"url": "https://huggingface.co/datasets/Delores-
Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Fr"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "id"}, {"value": 80.6, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 7, "label": "Id"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "it"}, {"value": 86.6, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "It"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "nl"}, {"value": 77.4, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 9, "label": "Nl"}, {"dataset": {"id": "Delores-Lin/

```

```
MDPBench","isBenchmark":true,"task_id":"pt"},"value":87.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":3,"label":"Pt"),{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"vi"},"value":86.2,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":1,"label":"Vi"),{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"non_latin"},"value":72.9,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":8,"label":"Non Latin"}, {"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"ar"},"value":75.8,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":6,"label":"Ar"}, {"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"hi"},"value":74.5,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Hi"}, {"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"jp"},"value":72.5,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":3,"label":"Jp"}, {"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"ko"},"value":70.9,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":8,"label":"Ko"}, {"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"ru"},"value":61.8,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":9,"label":"Ru"}, {"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"th"},"value":67,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfdr9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Th"}, {"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"zh"},"value":81.7,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
```

```

{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":4,"label":"Zh"),{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"zh_t"},"value":78.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":2,"label":"Zh T"},"value":81.2,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":4,"label":"Private"),{"dataset":{"id":"cais/hle"},"isBenchmark":false,"task_id":"default"},"value":24.37,"source":{"url":"https://huggingface.co/datasets/evalval/EEE_datastore/blob/192329fb7d6b15b7b0936a1a58ae862aa7e8ba24/flat/objects/aa/05/aa055fe5-5174-4517-8b10-19d46da94f3e.json"},"name":"EvalEval","isExternal":true},"filename":".eval_results/hle.yaml","verified":false,"pullRequest":126,"rank":1,"label":"Default"),{"dataset":{"id":"internlm/WildClawBench"},"isBenchmark":false,"task_id":"overall"},"value":30.8,"source":{"url":"https://internlm.github.io/WildClawBench"},"name":"WildClawBench","isExternal":true,"author":
{"_id":"64a2d5fa81252883206f24c9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png","fullname":"Intern Large Models","name":"internlm","type":"org","isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":1347,"isUserFollowwildclawbench.yaml","verified":false,"pullRequest":121,"rank":20,"label":"Overall"),{"dataset":{"id":"internlm/WildClawBench"},"isBenchmark":false,"task_id":"avg_time"},"value":406,"source":{"url":"https://internlm.github.io/WildClawBench"},"name":"WildClawBench","isExternal":true,"author":
{"_id":"64a2d5fa81252883206f24c9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png","fullname":"Intern Large Models","name":"internlm","type":"org","isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":1347,"isUserFollowwildclawbench.yaml","verified":false,"pullRequest":121,"rank":15,"label":"Avg Time"),{"dataset":{"id":"internlm/WildClawBench"},"isBenchmark":false,"task_id":"avg_cost"},"value":6.6,"source":{"url":"https://internlm.github.io/WildClawBench"},"name":"WildClawBench","isExternal":true,"author":
{"_id":"64a2d5fa81252883206f24c9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png","fullname":"Intern Large Models","name":"internlm","type":"org","isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":1347,"isUserFollowwildclawbench.yaml","verified":false,"pullRequest":121,"rank":11,"label":"Avg Cost"),{"dataset":{"id":"InternScience/ResearchClawBench"},"isBenchmark":false,"task_id":"overall"},"value":13.96,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.5"},"name":"Model Card","isExternal":false},"filename":".eval_results/researchclawbench.yaml","verified":false,"pullRequest":120,"rank":7,"label":"Overall","notes":"ResearchHarness evaluation with tools enabled, code execution, and a file-system workspace; completed 39/40 ResearchClawBench tasks."},"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"overall"},"value":0.3026,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Overall","notes":"agent: Claude Code"},"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"appworld"},"value":0.08,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Appworld","notes":"agent: Claude Code"},"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"browsecomp_plus"},"value":0.56,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"rank":3,"label":"Browsecomp Plus","notes":"agent: Claude Code"},"dataset":{"id":"open-agent-leaderboard/

```

```

results", "isBenchmark": false, "task_id": "swebench", "value": 0.5204, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent:
Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline", "value": 0.12, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Airline", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail", "value": 0.03, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom", "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Telecom", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall", "value": 0.2501, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld", "value": 0.08, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus", "value": 0.35, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench", "value": 0.567, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline", "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Airline", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail", "value": 0.01, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom", "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Telecom", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall", "value": 0.42, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent:
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld", "value": 0.11, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Appworld", "note
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus", "value": 0.33, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/

```

results", "isBenchmark": false, "task_id": "swebench", "value": 0.5761, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Swebench", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "taubench_airline", "value": 0.56, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench Airline", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "taubench_retail", "value": 0.7245, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taubench Retail", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "taubench_telecom", "value": 0.7071, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench Telecom", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "overall", "value": 0.4251, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "appworld", "value": 0.09, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "browsecomp_plus", "value": 0.34, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp Plus", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "swebench", "value": 0.5714, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "taubench_airline", "value": 0.62, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taubench Airline", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "taubench_retail", "value": 0.6465, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Taubench Retail", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "taubench_telecom", "value": 0.83, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "rank": 4, "label": "Taubench Telecom", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "overall", "value": 0.4276, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Over React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "appworld", "value": 0.1, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "browsecomp_plus", "value": 0.34, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp Plus", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/"

results", "isBenchmark": false, "task_id": "swebench", "value": 0.5714, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/results", "isBenchmark": false, "task_id": "taubench_airline", "value": 0.62, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taut Airline", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/results", "isBenchmark": false, "task_id": "taubench_retail", "value": 0.6465, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Taubench Retail", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/results", "isBenchmark": false, "task_id": "taubench_telecom", "value": 0.83, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 4, "label": "Taut Telecom", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "general", "value": 52.8, "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": 117, "rank": 20, "label": "General", "notes": "Pass3% | N=3 | 161 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multimodal", "value": 14.8, "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": 117, "rank": 6, "label": "Multimodal", "notes": "Pass3% | N=3 | 101 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multi_turn", "value": 50, "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": 117, "rank": 10, "label": "Multi Turn", "notes": "Pass3% | N=3 | 38 tasks"}, {"dataset": {"id": "MME-Benchmarks/Video-MME-v2", "isBenchmark": true, "task_id": "video-mme-v2", "value": 61.1, "source": {"url": "https://video-mme-v2.netlify.app/#leaderboard", "name": "VideoMMEv2 leaderboard", "isExternal": true, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isUse videomme_v2.yaml", "verified": false, "pullRequest": 116, "rank": 3, "label": "Video Mme V2", "notes": "512 frames"}, {"dataset": {"id": "MME-Benchmarks/Video-MME-v2", "isBenchmark": true, "task_id": "video-mme-v2", "value": 54.4, "source": {"url": "https://video-mme-v2.netlify.app/#leaderboard", "name": "VideoMMEv2 leaderboard", "isExternal": true, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isUse videomme_v2.yaml", "verified": false, "pullRequest": 116, "label": "Video Mme V2", "notes": "64 frames"}, {"dataset": {"id": "mercor/APEX-SWE", "isBenchmark": false, "task_id": "apex-swe", "value": 30.3, "source": {"url": "https://www.mercor.com/apex/apex-swe-leaderboard/", "name": "APEX-SWE Leaderboard", "isExternal": true, "author": {"_id": "6531a6d5c93c85e11c82d495", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/63b4a97834df1f9a90a0da76/HHWjFJmEitVTMv3Fk85wu.jpeg", "fullName": "Mercor", "name": "mercor", "type": "org", "isHf": false, "isHfAdmin": false, "isUse apex-swe.yaml", "verified": false, "pullRequest": 114, "rank": 1, "label": "Apex Swe", "notes": "harness: claude_code; reasoning_effort: thinking"}, {"dataset": {"id": "collinear-ai/yc-bench", "isBenchmark": true, "task_id": "medium", "value": 408822, "source": {"url": "https://github.com/collinear-ai/yc-

```

bench", "name": "YC-Bench eval", "isExternal": true, "filename": ".eval_results/yc-
bench.yaml", "verified": false, "pullRequest": 104, "rank": 6, "label": "Medium", "notes": "avg final funds (USD) across
seeds 1,2,3. Kimi K2.5 (via OpenRouter moonshotai/kimi-k2.5)", {"dataset": {"id": "MathArena/
hmmmt_feb_2026", "isBenchmark": true, "task_id": "MathArena/hmmmt_feb_2026"}, "value": 87.12, "source": {"url": "https://
matharena.ai/?comp=hmmmt--hmmmt_feb_2026", "name": "Official MathArena
Evaluation", "isExternal": true, "filename": ".eval_results/MathArena--
hmmmt_feb_2026.yaml", "verified": false, "pullRequest": 95, "rank": 6, "label": "MathArena Hmmt Feb
2026"}}, {"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", "deploy:sagemaker", "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"},
{"id": "deploy:sagemaker", "label": "sagemaker", "type": "deploy"}, {"type": "region", "label": "🌐 Region:
US", "id": "region:us"}], "transformersInfo":
{"auto_model": "AutoModel", "custom_class": "modeling_kimi_k25.KimiK25ForConditionalGeneration", "pipeline_tag": "fe
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": 40, "open": 87, "total": 127}, "query": {}, "inferenceContextData": {"billableEntities": [], "entityName2Providers":
{}}, "defaultProviders":
[{"isOriginalProvider": false, "name": "novita", "enabled": true, "position": 1, "isReleased": true, "accuratePricing": true},
{"isOriginalProvider": false, "name": "deepinfra", "enabled": true, "position": 16, "isReleased": true, "accuratePricing": true},
{"isOriginalProvider": false, "name": "featherless-
ai", "enabled": true, "position": 7, "isReleased": true, "accuratePricing": true}]], "hasQuantizations": true, "copyToBucketNam
[]]">

```

 Tech Blog |  Paper

0. Changelog

2026.1.29: The default system prompt might cause confusion to users and unexpected behaviours, so we remove it.

The token `<|media_start|>` is incorrect; it has been replaced with `<|media_begin|>` in the chat template.

1. Model Introduction

Kimi K2.5 is an open-source, native multimodal agentic model built through continual pretraining on approximately 15 trillion mixed visual and text tokens atop Kimi-K2-Base. It seamlessly integrates vision and language understanding with advanced agentic capabilities, instant and thinking modes, as well as conversational and agentic paradigms.

Key Features

Native Multimodality : Pre-trained on vision–language tokens, K2.5 excels in visual knowledge, cross-modal reasoning, and agentic tool use grounded in visual inputs.

Coding with Vision : K2.5 generates code from visual specifications (UI designs, video workflows) and autonomously orchestrates tools for visual data processing.

Agent Swarm : K2.5 transitions from single-agent scaling to a self-directed, coordinated swarm-like execution scheme. It decomposes complex tasks into parallel sub-tasks executed by dynamically instantiated, domain-

specific agents.

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 K2.5 (Thinking) GPT-5.2 (xhigh) Claude 4.5 Opus (Extended Thinking) Gemini 3 Pro (High Thinking Level) DeepSeek V3.2 (Thinking) Qwen3-VL- 235B-A22B- Thinking

Reasoning & Knowledge

HLE-Full 30.1 34.5 30.8 37.5 25.1 † -

HLE-Full (w/ tools) 50.2 45.5 43.2 45.8 40.8 † -

AIME 2025 96.1 100 92.8 95.0 93.1 -

HMMT 2025 (Feb) 95.4 99.4 92.9* 97.3* 92.5 -

IMO-AnswerBench 81.8 86.3 78.5* 83.1* 78.3 -

GPQA-Diamond 87.6 92.4 87.0 91.9 82.4 -

MMLU-Pro 87.1 86.7* 89.3* 90.1 85.0 -

Image & Video

MMMU-Pro 78.5 79.5* 74.0 81.0 - 69.3

CharXiv (RQ) 77.5 82.1 67.2* 81.4 - 66.1

MathVision 84.2 83.0 77.1* 86.1* - 74.6

MathVista (mini) 90.1 82.8* 80.2* 89.8* - 85.8

ZeroBench 9 9* 3* 8* - 4*

ZeroBench (w/ tools) 11 7* 9* 12* - 3*

OCRBench 92.3 80.7* 86.5* 90.3* - 87.5
 OmniDocBench 1.5 88.8 85.7 87.7* 88.5 - 82.0*
 InfoVQA (val) 92.6 84* 76.9* 57.2* - 89.5
 SimpleVQA 71.2 55.8* 69.7* 69.7* - 56.8*
 WorldVQA 46.3 28.0 36.8 47.4 - 23.5
 VideoMMMU 86.6 85.9 84.4* 87.6 - 80.0
 MMVU 80.4 80.8* 77.3 77.5 - 71.1
 MotionBench 70.4 64.8 60.3 70.3 - -
 VideoMME 87.4 86.0* - 88.4* - 79.0
 LongVideoBench 79.8 76.5* 67.2* 77.7* - 65.6*
 LVBench 75.9 - - 73.5* - 63.6
 Coding
 SWE-Bench Verified 76.8 80.0 80.9 76.2 73.1 -
 SWE-Bench Pro 50.7 55.6 55.4* - - -
 SWE-Bench Multilingual 73.0 72.0 77.5 65.0 70.2 -
 Terminal Bench 2.0 50.8 54.0 59.3 54.2 46.4 -
 PaperBench 63.5 63.7* 72.9* - 47.1 -
 CyberGym 41.3 - 50.6 39.9* 17.3* -
 SciCode 48.7 52.1 49.5 56.1 38.9 -
 OJBench (cpp) 57.4 - 54.6* 68.5* 54.7* -
 LiveCodeBench (v6) 85.0 - 82.2* 87.4* 83.3 -
 Long Context
 Longbench v2 61.0 54.5* 64.4* 68.2* 59.8* -
 AA-LCR 70.0 72.3* 71.3* 65.3* 64.3* -
 Agentic Search
 BrowseComp 60.6 65.8 37.0 37.8 51.4 -
 BrowseComp (w/ctx manage) 74.9 57.8 59.2 67.6 -
 BrowseComp (Agent Swarm) 78.4 - - - -
 WideSearch (item-f1) 72.7 - 76.2* 57.0 32.5* -
 WideSearch (item-f1 Agent Swarm) 79.0 - - - -
 DeepSearchQA 77.1 71.3* 76.1* 63.2* 60.9* -
 FinSearchCompT2&T3 67.8 - 66.2* 49.9 59.1* -
 Seal-0 57.4 45.0 47.7* 45.5* 49.5* -

Footnotes

General Testing Details We report results for Kimi K2.5 and DeepSeek-V3.2 with thinking mode enabled, Claude Opus 4.5 with extended thinking mode, GPT-5.2 with xhigh reasoning effort, and Gemini 3 Pro with a high thinking level. For vision benchmarks, we additionally report results for Qwen3-VL-235B-A22B-Thinking.

Unless otherwise specified, all Kimi K2.5 experiments were conducted with temperature = 1.0, top-p = 0.95, and a context length of 256k tokens.

Benchmarks without publicly available scores were re-evaluated under the same conditions used for Kimi K2.5 and are marked with an asterisk (*).

We could not evaluate GPT-5.2 xhigh on all benchmarks due to service stability issues. For benchmarks that were not tested, we mark them as "-".

Text and Reasoning HLE, AIME 2025, HMMT 2025 (Feb), and GPQA-Diamond were evaluated with a maximum completion budget of 96k tokens.

Results for AIME and HMMT are averaged over 32 runs (avg@32); GPQA-Diamond over 8 runs (avg@8).

For HLE, we report scores on the full set (text & image). Kimi K2.5 scores 31.5 (text) and 21.3 (image) without tools, and 51.8 (text) and 39.8 (image) with tools. The DeepSeek-V3.2 score corresponds to its text-only subset (marked with †) . Hugging Face access was blocked to prevent potential data leakage. HLE with tools uses simple context management: once the context exceeds a threshold, only the latest round of tool messages is retained.

Tool-Augmented / Agentic Search Kimi K2.5 was equipped with search, code-interpreter, and web-browsing tools for HLE with tools and all agentic search benchmarks.

Except for BrowseComp (where K2.5 and DeepSeek-V3.2 used the discard-all strategy), no context management was applied, and tasks exceeding the supported context length were directly counted as failed.

The test system prompts emphasize deep and proactive tool use, instructing models to reason carefully, leverage tools, and verify uncertain information. Full prompts will be provided in the technical report.

Results for Seal-0 and WideSearch are averaged over four runs (avg@4).

Vision Benchmarks Max-tokens = 64k, averaged over three runs (avg@3).

ZeroBench (w/ tools) uses max-tokens-per-step = 24k and max-steps = 30 for multi-step reasoning.

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

GPT-5.2-xhigh had ~10% failure rate (no output despite 3 retries), treated as incorrect; reported scores likely underestimate true performance.

WorldVQA, a benchmark designed to evaluate atomic vision-centric world knowledge. Access WorldVQA at <https://github.com/MoonshotAI/WorldVQA> .

OmniDocBench Score is computed as $(1 - \text{normalized Levenshtein distance}) \times 100$, where a higher score denotes superior accuracy.

Coding Tasks Terminal-Bench 2.0 scores were obtained with the default agent framework (Terminus-2) and the provided JSON parser. In our implementation, we evaluated Terminal-Bench 2.0 under non-thinking mode. This choice was made because our current context management strategy for the thinking mode is incompatible with Terminus-2.

For the SWE-Bench series of evaluations (including verified, multilingual, and pro), we used an internally developed evaluation framework. This framework includes a minimal set of tools—bash tool, createfile tool, insert tool, view tool, strreplace tool, and submit tool—along with tailored system prompts designed for the tasks. The highest scores were achieved under non-thinking mode.

The score of Claude Opus 4.5 on CyberGym is reported under the non-thinking setting.

All reported scores of coding tasks are averaged over 5 independent runs.

Long-Context Benchmarks AA-LCR: scores averaged over three runs (avg@3).

LongBench-V2: identical prompts and input contexts standardized to ~128k tokens.

Agent Swarm BrowseComp (Swarm Mode): main agent max 15 steps; sub-agents max 100 steps.

WideSearch (Swarm Mode): main and sub-agents max 100 steps.

4. Native INT4 Quantization

Kimi-K2.5 adopts the same native int4 quantization method as Kimi-K2-Thinking .

5. Deployment

You can access Kimi-K2.5'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-K2.5 is recommended to run on the following inference engines:

vLLM

SGLang

KTransformers

The minimum version requirement for transformers is 4.57.1 .

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 K2.5 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_content} ' )
    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

K2.5 supports Image and Video input.

The following example demonstrates how to call K2.5 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-K2.5/resolve/main/figures/kimi-logo.png'
    image_base64 = base64.b64encode(requests.get(url).content).decode()
    messages = [ { 'role' : 'user' , 'content' : [ { 'type' : 'image_url' , 'image_url' : { 'url' : f'data:image/png;base64,{image_base64} ' } }, { 'type' : 'text' , 'text' : 'Describe this image in detail.' }, ], }, ]

    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_content} ' )
    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 K2.5 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-
K2.5/resolve/main/figures/demo_video.mp4' video_base64 =
base64.b64encode(requests.get(url).content).decode() messages = [ { "role" : "user" , "content" : [ { "type" :
"video_url" , "video_url" : { "url" : f"data:video/mp4;base64, {video_base64} " }, }, { "type" : "text" , "text" : "Describe
the video in detail." }, ], } ]

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_content} ' ) 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

```

Interleaved Thinking and Multi-Step Tool Call

K2.5 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 K2.5 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.cn .

10. Reference

If you find K2.5 useful for your research, please kindly cite K2.5 technical report as follows:

@misc{kimiteam2026kimik25visualagentic, title={Kimi K2.5: Visual Agentic Intelligence}, author={Kimi Team and Tongtong Bai and Yifan Bai and Yiping Bao and S. H. Cai and Yuan Cao and Y. Charles and H. S. Che and Cheng Chen and Guanduo Chen and Huarong Chen and Jia Chen and Jiahao Chen and Jianlong Chen and Jun Chen and Kefan Chen and Liang Chen and Ruijie Chen and Xinhao Chen and Yanru Chen and Yanxu Chen and Yicun Chen and Yimin Chen and Yingjiang Chen and Yuankun Chen and Yujie Chen and Yutian Chen and Zhirong Chen and Ziwei Chen and Dazhi Cheng and Minghan Chu and Jialei Cui and Jiaqi Deng and Muxi Diao and Hao Ding and Mengfan Dong and Mengnan Dong and Yuxin Dong and Yuhao Dong and Angang Du and Chenzhuang Du and Dikang Du and Lingxiao Du and Yulun Du and Yu Fan and Shengjun Fang and Qiulin Feng and Yichen Feng and Garimugai Fu and Kelin Fu and Hongcheng Gao and Tong Gao and Yuyao Ge and Shangyi

Geng and Chengyang Gong and Xiaochen Gong and Zhuoma Gongque and Qizheng Gu and Xinran Gu and Yicheng Gu and Longyu Guan and Yuanying Guo and Xiaoru Hao and Weiran He and Wenyang He and Yunjia He and Chao Hong and Hao Hu and Jiayi Hu and Yangyang Hu and Zhenxing Hu and Ke Huang and Ruiyuan Huang and Weixiao Huang and Zhiqi Huang and Tao Jiang and Zhejun Jiang and Xinyi Jin and Yu Jing and Guokun Lai and Aidi Li and C. Li and Cheng Li and Fang Li and Guanghe Li and Guanyu Li and Haitao Li and Haoyang Li and Jia Li and Jingwei Li and Junxiong Li and Lincan Li and Mo Li and Weihong Li and Wentao Li and Xinhang Li and Xinhao Li and Yang Li and Yanhao Li and Yiwei Li and Yuxiao Li and Zhaowei Li and Zheming Li and Weilong Liao and Jiawei Lin and Xiaohan Lin and Zhishan Lin and Zichao Lin and Cheng Liu and Chenyu Liu and Hongzhang Liu and Liang Liu and Shaowei Liu and Shudong Liu and Shuran Liu and Tianwei Liu and Tianyu Liu and Weizhou Liu and Xiangyan Liu and Yangyang Liu and Yanming Liu and Yibo Liu and Yuanxin Liu and Yue Liu and Zhengying Liu and Zhongnuo Liu and Enzhe Lu and Haoyu Lu and Zhiyuan Lu and Junyu Luo and Tongxu Luo and Yashuo Luo and Long Ma and Yingwei Ma and Shaoguang Mao and Yuan Mei and Xin Men and Fanqing Meng and Zhiyong Meng and Yibo Miao and Mingqing Ni and Kun Ouyang and Siyuan Pan and Bo Pang and Yuchao Qian and Ruoyu Qin and Zeyu Qin and Jiezhong Qiu and Bowen Qu and Zeyu Shang and Youbo Shao and Tianxiao Shen and Zhennan Shen and Juanfeng Shi and Lidong Shi and Shengyuan Shi and Feifan Song and Pengwei Song and Tianhui Song and Xiaoxi Song and Hongjin Su and Jianlin Su and Zhaochen Su and Lin Sui and Jinsong Sun and Junyao Sun and Tongyu Sun and Flood Sung and Yunpeng Tai and Chuning Tang and Heyi Tang and Xiaojuan Tang and Zhengyang Tang and Jiawen Tao and Shiyuan Teng and Chaoran Tian and Pengfei Tian and Ao Wang and Bowen Wang and Chensi Wang and Chuang Wang and Congcong Wang and Dingkun Wang and Dinglu Wang and Dongliang Wang and Feng Wang and Hailong Wang and Haiming Wang and Hengzhi Wang and Huaqing Wang and Hui Wang and Jiahao Wang and Jinhong Wang and Jiuzheng Wang and Kaixin Wang and Linian Wang and Qibin Wang and Shengjie Wang and Shuyi Wang and Si Wang and Wei Wang and Xiaochen Wang and Xinyuan Wang and Yao Wang and Yejie Wang and Yipu Wang and Yiqin Wang and Yucheng Wang and Yuzhi Wang and Zhaoji Wang and Zhaowei Wang and Zhengtao Wang and Zhexu Wang and Zihan Wang and Zizhe Wang and Chu Wei and Ming Wei and Chuan Wen and Zichen Wen and Chengjie Wu and Haoning Wu and Junyan Wu and Rucong Wu and Wenhao Wu and Yuefeng Wu and Yuhao Wu and Yuxin Wu and Zijian Wu and Chenjun Xiao and Jin Xie and Xiaotong Xie and Yuchong Xie and Yifei Xin and Bowei Xing and Boyu Xu and Jianfan Xu and Jing Xu and Jinjing Xu and L. H. Xu and Lin Xu and Suting Xu and Weixin Xu and Xinbo Xu and Xinran Xu and Yangchuan Xu and Yichang Xu and Yuemeng Xu and Zelai Xu and Ziyao Xu and Junjie Yan and Yuzi Yan and Guangyao Yang and Hao Yang and Junwei Yang and Kai Yang and Ningyuan Yang and Ruihan Yang and Xiaofei Yang and Xinlong Yang and Ying Yang and Yi Yang and Yi Yang and Zhen Yang and Zhilin Yang and Zonghan Yang and Haotian Yao and Dan Ye and Wenjie Ye and Zhuorui Ye and Bohong Yin and Chengzhen Yu and Longhui Yu and Tao Yu and Tianxiang Yu and Enming Yuan and Mengjie Yuan and Xiaokun Yuan and Yang Yue and Weihao Zeng and Dunyuan Zha and Haobing Zhan and Dehao Zhang and Hao Zhang and Jin Zhang and Puqi Zhang and Qiao Zhang and Rui Zhang and Xiaobin Zhang and Y. Zhang and Yadong Zhang and Yangkun Zhang and Yichi Zhang and Yizhi Zhang and Yongting Zhang and Yu Zhang and Yushun Zhang and Yutao Zhang and Yutong Zhang and Zheng Zhang and Chenguang Zhao and Feifan Zhao and Jinxiang Zhao and Shuai Zhao and Xiangyu Zhao and Yikai Zhao and Zijia Zhao and Huabin Zheng and Ruihan Zheng and Shaojie Zheng and Tengyang Zheng and Junfeng Zhong and Longguang Zhong and Weiming Zhong and M. Zhou and Runjie Zhou and Xinyu Zhou and Zaida Zhou and Jinguo Zhu and Liya Zhu and Xinhao Zhu and Yuxuan Zhu and Zhen Zhu and Jingze Zhuang and Weiyu Zhuang and Ying Zou and Xinxing Zu}, year={2026}, eprint={2602.02276}, archivePrefix={arXiv}, primaryClass={cs.CL}, url={https://arxiv.org/abs/2602.02276}, }

Downloads last month 327,584

```
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\n# Find last non-tool-call assistant message #}\n{%- set ns =
```

```

namespace(last_non_tool_call_assistant_msg=-1) -%}\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\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 %- 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 -%}\n <think></think>{{render_content(message)}}\n %- else -%}\n %- set rc =
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\n%- endif -
%}\n"},"createdAt":"2026-01-01T06:06:03.000Z","discussionsDisabled":false,"discussionsSorting":"recently-
created","downloads":327584,"downloadsAllTime":16128649,"id":"moonshotai/Kimi-
K2.5","isLikedByUser":false,"availableInferenceProviders":
[{"provider":"novita","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/kimi-
k2.5","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","features":
{"structuredOutput":false,"toolCalling":true},"isCheapestPricingOutput":false,"isFastestThroughput":true,"isModelAuth
{"provider":"featherless-ai","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/Kimi-
K2.5","task":"conversational","adapterWeightsPath":"model-00001-
of-000064.safetensors","isCheapestPricingOutput":false,"isFastestThroughput":false,"isModelAuthor":false},
{"provider":"deepinfra","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/Kimi-
K2.5","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","features":
{"structuredOutput":true,"toolCalling":true},"isCheapestPricingOutput":true,"isFastestThroughput":false,"isModelAuthc
text-to-text","library_name":"transformers","librariesOther":[],"trackDownloads":true,"model-
index":null,"evalResults":[{"dataset":{"id":"TIGER-Lab/MMLU-
Pro"},"isBenchmark":true,"task_id":"mmlu_pro"},"value":87.1,"source":{"url":"https://huggingface.co/moonshotai/
Kimi-K2.5"},"name":"Model Card","isExternal":false,"author":
{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/
uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullName":"Niels
Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isl
mmlu_pro.yaml","verified":false,"pullRequest":51,"rank":6,"label":"Mmlu Pro"},"dataset":{"id":"SWE-bench/SWE-
bench_Verified"},"isBenchmark":true,"task_id":"swe_bench_%_resolved"},"value":70.8,"source":{"url":"https://
www.swebench.com/"},"name":"SWE-Bench official evaluation","isExternal":true,"author":
{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/
uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullName":"Niels
Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isl
swe_bench_verified.yaml","verified":false,"pullRequest":97,"rank":38,"label":"Swe Bench Resolved","notes":"high
reasoning"},"dataset":{"id":"Idavidrein/
gpqa"},"isBenchmark":true,"task_id":"diamond"},"value":87.3737373737,"source":{"url":"https://huggingface.co/
datasets/evaleval/EEE_datastore/blob/192329fb7d6b15b7b0936a1a58ae862aa7e8ba24/flat/
objects/5c/4b/5c4b9bd9-14db-4c11-
bf99-9661461459f6.json"},"name":"EvalEval","isExternal":true},"filename":"eval_results/gpqa-
diamond.yaml","verified":false,"pullRequest":126,"label":"Diamond","notes":"GPQA Diamond"},"dataset":
{"id":"ScaleAI/SWE-bench_Pro"},"isBenchmark":true,"task_id":"SWE_Bench_Pro"},"value":50.7,"source":
{"url":"https://huggingface.co/moonshotai/Kimi-K2.5"},"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,
swe_bench_pro.yaml","verified":false,"pullRequest":101,"rank":31,"label":"SWE Bench Pro"},{"dataset":
{"id":"ldavidrein/gpqa","isBenchmark":true,"task_id":"diamond"},"value":87.6,"source":{"url":"https://huggingface.co/
moonshotai/Kimi-K2.5","name":"Model Card","isExternal":false,"author":
{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/
uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullName":"Niels
Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isl
gpqa_diamond.yaml","verified":false,"pullRequest":50,"rank":28,"label":"Diamond"},{"dataset":{"id":"mercor/apex-
agents","isBenchmark":true,"task_id":"apex-agents"},"value":14.4,"source":{"url":"https://www.mercor.com/apex/
apex-agents-leaderboard/","name":"APEX-Agents Leaderboard","isExternal":true,"author":
{"_id":"6531a6d5c93c85e11c82d495","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/
uploads/63b4a97834df1f9a90a0da76/
HHWjFJmEitVTMv3Fk85wu.jpeg","fullName":"Mercor","name":"mercor","type":"org","isHf":false,"isHfAdmin":false,"isL
apex-agents.yaml","verified":false,"pullRequest":113,"rank":8,"label":"Apex Agents","notes":"harness:
react_toolbelt_agent; reasoning_effort: thinking"},{"dataset":{"id":"MMMU/
MMMU_Pro","isBenchmark":true,"task_id":"mmmu_pro_vision"},"value":78.5,"source":{"url":"https://
huggingface.co/moonshotai/Kimi-K2.5","name":"Model Card","isExternal":false},"filename":"eval_results/
mmmu_pro.yaml","verified":false,"pullRequest":119,"rank":3,"label":"Mmmu Pro Vision"},{"dataset":
{"id":"harborframework/terminal-bench-2.0","isBenchmark":true,"task_id":"terminalbench_2"},"value":43.2,"source":
{"url":"https://www.tbench.ai/leaderboard/terminal-bench/2.0","name":"Terminal-Bench
Leaderboard","isExternal":true,"author":{"_id":"62d648291fa3e4e7ae3fa6e8","avatarUrl":"https://cdn-
avatars.huggingface.co/v1/production/uploads/62d648291fa3e4e7ae3fa6e8/
oatOwf8Xqe5eDbCSuYqCd.png","fullName":"ben
burtenshaw","name":"burtenshaw","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCour
terminal_bench.yaml","verified":false,"pullRequest":90,"rank":19,"label":"Terminalbench 2","notes":"agent:
Terminus 2"},{"dataset":{"id":"MathArena/aime_2026","isBenchmark":true,"task_id":"MathArena/
aime_2026"},"value":95.83,"source":{"url":"https://matharena.ai/?comp=aime--aime_2026","name":"Official
MathArena Evaluation","isExternal":true},"filename":"eval_results/MathArena--
aime_2026.yaml","verified":false,"pullRequest":88,"rank":7,"label":"MathArena Aime 2026"},{"dataset":{"id":"cais/
hle","isBenchmark":false,"task_id":"hle"},"value":50.2,"source":{"url":"https://huggingface.co/moonshotai/Kimi-
K2.5","name":"Model Card","isExternal":false,"author":{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-
avatars.huggingface.co/v1/production/
uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullName":"Niels
Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isl
hle_with_tools.yaml","verified":false,"pullRequest":49,"rank":17,"label":"Hle","notes":"With tools"},{"dataset":
{"id":"SWE-bench/SWE-
bench_Multilingual","isBenchmark":true,"task_id":"swe_bench_multilingual_%_resolved"},"value":73,"source":
{"url":"https://huggingface.co/moonshotai/Kimi-K2.5","name":"Model
Card","isExternal":false},"filename":"eval_results/swe-
bench_multilingual.yaml","verified":false,"pullRequest":131,"rank":14,"label":"Swe Bench Multilingual Resolved"},
{"dataset":{"id":"PaddlePaddle/Real5-
OmniDocBench","isBenchmark":true,"task_id":"overall"},"value":89.09,"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":130,"rank":7,"label":"Overall"},{"dataset":
{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"scanning"},"value":89.67,"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":130,"rank":9,"label":"Scanning"},{"dataset":
{"id":"PaddlePaddle/Real5-OmniDocBench","isBenchmark":true,"task_id":"warping"},"value":88.86,"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":130, "rank":8, "label":"Warping"}, {"dataset":
{"id":"PaddlePaddle/Real5-
OmniDocBench", "isBenchmark":true, "task_id":"screen_photography", "value":88.39, "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":130, "rank":10, "label":"Screen Photography"}, {"dataset":
{"id":"PaddlePaddle/Real5-OmniDocBench", "isBenchmark":true, "task_id":"illumination", "value":89.66, "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":130, "rank":8, "label":"Illumination"}, {"dataset":
{"id":"PaddlePaddle/Real5-OmniDocBench", "isBenchmark":true, "task_id":"skew", "value":88.86, "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":130, "rank":6, "label":"Skew"}, {"dataset":{"id":"Delores-Lin/
MDPBench", "isBenchmark":true, "task_id":"overall", "value":77.5, "source":{"url":"https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name":"MDPBench leaderboard", "isExternal":true, "author":
{"_id":"687a14565d8f9617af6ab38b", "avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName":"Delores Lin", "name":"Delores-
Lin", "type":"user", "isPro":false, "isHf":false, "isHfAdmin":false, "isMod":false, "followerCount":2, "isUserFollowing":false}},
mdpbench.yaml", "verified":false, "pullRequest":129, "rank":7, "label":"Overall"}, {"dataset":{"id":"Delores-Lin/
MDPBench", "isBenchmark":true, "task_id":"digital", "value":85, "source":{"url":"https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name":"MDPBench leaderboard", "isExternal":true, "author":
{"_id":"687a14565d8f9617af6ab38b", "avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName":"Delores Lin", "name":"Delores-
Lin", "type":"user", "isPro":false, "isHf":false, "isHfAdmin":false, "isMod":false, "followerCount":2, "isUserFollowing":false}},
mdpbench.yaml", "verified":false, "pullRequest":129, "rank":9, "label":"Digital"}, {"dataset":{"id":"Delores-Lin/
MDPBench", "isBenchmark":true, "task_id":"photographed", "value":75, "source":{"url":"https://huggingface.co/
datasets/Delores-Lin/MDPBench", "name":"MDPBench leaderboard", "isExternal":true, "author":
{"_id":"687a14565d8f9617af6ab38b", "avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName":"Delores Lin", "name":"Delores-
Lin", "type":"user", "isPro":false, "isHf":false, "isHfAdmin":false, "isMod":false, "followerCount":2, "isUserFollowing":false}},
mdpbench.yaml", "verified":false, "pullRequest":129, "rank":7, "label":"Photographed"}, {"dataset":{"id":"Delores-Lin/
MDPBench", "isBenchmark":true, "task_id":"latin", "value":81.6, "source":{"url":"https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name":"MDPBench leaderboard", "isExternal":true, "author":
{"_id":"687a14565d8f9617af6ab38b", "avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName":"Delores Lin", "name":"Delores-
Lin", "type":"user", "isPro":false, "isHf":false, "isHfAdmin":false, "isMod":false, "followerCount":2, "isUserFollowing":false}},
mdpbench.yaml", "verified":false, "pullRequest":129, "rank":5, "label":"Latin"}, {"dataset":{"id":"Delores-Lin/
MDPBench", "isBenchmark":true, "task_id":"de", "value":85.9, "source":{"url":"https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name":"MDPBench leaderboard", "isExternal":true, "author":
{"_id":"687a14565d8f9617af6ab38b", "avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName":"Delores Lin", "name":"Delores-
Lin", "type":"user", "isPro":false, "isHf":false, "isHfAdmin":false, "isMod":false, "followerCount":2, "isUserFollowing":false}},
mdpbench.yaml", "verified":false, "pullRequest":129, "rank":3, "label":"De"}, {"dataset":{"id":"Delores-Lin/
MDPBench", "isBenchmark":true, "task_id":"en", "value":86.2, "source":{"url":"https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name":"MDPBench leaderboard", "isExternal":true, "author":

```

```

{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":4,"label":"En"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"es"},"value":72.7,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Es"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"fr"},"value":71,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Fr"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"id"},"value":80.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":7,"label":"Id"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"it"},"value":86.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"It"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"nl"},"value":77.4,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":9,"label":"Nl"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"pt"},"value":87.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":3,"label":"Pt"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"vi"},"value":86.2,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":1,"label":"Vi"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"non_latin"},"value":72.9,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":8,"label":"Non Latin"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"ar"},"value":75.8,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-

```

```

Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 6, "label": "Ar"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "hi", "value": 74.5, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Hi"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "jp", "value": 72.5, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 3, "label": "Jp"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "ko", "value": 70.9, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 8, "label": "Ko"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "ru", "value": 61.8, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 9, "label": "Ru"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "th", "value": 67, "source": {"url": "https://huggingface.co/datasets/Delores-
Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Th"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "zh", "value": 81.7, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 4, "label": "Zh"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "zh_t", "value": 78.6, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 2, "label": "Zh T"}, {"dataset": {"id": "Delores-Lin/
MDPBench", "isBenchmark": true, "task_id": "private", "value": 81.2, "source": {"url": "https://huggingface.co/datasets/
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-
auth/LgZxNkDzfd9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 4, "label": "Private"}, {"dataset": {"id": "cais/
hle", "isBenchmark": false, "task_id": "default", "value": 24.37, "source": {"url": "https://huggingface.co/datasets/
evaleval/EEE_datastore/blob/192329fb7d6b15b7b0936a1a58ae862aa7e8ba24/flat/objects/aa/05/
aa055fe5-5174-4517-8b10-19d46da94f3e.json", "name": "EvalEval", "isExternal": true, "filename": ".eval_results/
hle.yaml", "verified": false, "pullRequest": 126, "rank": 1, "label": "Default"}, {"dataset": {"id": "internlm/
WildClawBench", "isBenchmark": false, "task_id": "overall", "value": 30.8, "source": {"url": "https://internlm.github.io/
WildClawBench", "name": "WildClawBench", "isExternal": true, "author":

```

```

{"_id":"64a2d5fa81252883206f24c9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png","fullname":"Intern Large Models","name":"internlm","type":"org","isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":1347,"isUserFollow wildclawbench.yaml","verified":false,"pullRequest":121,"rank":20,"label":"Overall"},{"dataset":{"id":"internlm/WildClawBench"},"isBenchmark":false,"task_id":"avg_time"},"value":406,"source":{"url":"https://internlm.github.io/WildClawBench"},"name":"WildClawBench","isExternal":true,"author":
{"_id":"64a2d5fa81252883206f24c9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png","fullname":"Intern Large Models","name":"internlm","type":"org","isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":1347,"isUserFollow wildclawbench.yaml","verified":false,"pullRequest":121,"rank":15,"label":"Avg Time"},{"dataset":{"id":"internlm/WildClawBench"},"isBenchmark":false,"task_id":"avg_cost"},"value":6.6,"source":{"url":"https://internlm.github.io/WildClawBench"},"name":"WildClawBench","isExternal":true,"author":
{"_id":"64a2d5fa81252883206f24c9","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png","fullname":"Intern Large Models","name":"internlm","type":"org","isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":1347,"isUserFollow wildclawbench.yaml","verified":false,"pullRequest":121,"rank":11,"label":"Avg Cost"},{"dataset":{"id":"InternScience/ResearchClawBench"},"isBenchmark":false,"task_id":"overall"},"value":13.96,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.5"},"name":"Model Card","isExternal":false},"filename":".eval_results/researchclawbench.yaml","verified":false,"pullRequest":120,"rank":7,"label":"Overall","notes":"ResearchHarness evaluation with tools enabled, code execution, and a file-system workspace; completed 39/40 ResearchClawBench tasks."},"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"overall"},"value":0.3026,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Overall","notes":"agent: Claude Code"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"appworld"},"value":0.08,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Appworld","notes":"agent: Claude Code"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"browsecomp_plus"},"value":0.56,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"rank":3,"label":"Browsecomp Plus","notes":"agent: Claude Code"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"swebench"},"value":0.5204,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Swebench","notes":"agent: Claude Code"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"taubench_airline"},"value":0.12,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Taubench Airline","notes":"agent: Claude Code"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"taubench_retail"},"value":0.03,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Taubench Retail","notes":"agent: Claude Code"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"taubench_telecom"},"value":0,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_claude_code.yaml","verified":false,"pullRequest":118,"label":"Taubench Telecom","notes":"agent: Claude Code"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"overall"},"value":0.2501,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/open_agent_leaderboard_openai_solo.yaml","verified":false,"pullRequest":118,"label":"Overall","notes":"agent: OpenAI Solo"},{"dataset":{"id":"open-agent-leaderboard/results"},"isBenchmark":false,"task_id":"appworld"},"value":0.08,"source":{"url":"https://www.exgentic.ai"},"name":"Open Agent Leaderboard","isExternal":true},"filename":".eval_results/

```

```

open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.35, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.567, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Airline", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.01, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Telecom", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall"}, "value": 0.42, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent:
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.11, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Appworld", "note
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.33, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.5761, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Swebench", "noi
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0.56, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Airline", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.7245, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taubench
Retail", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0.7071, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Telecom", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall"}, "value": 0.4251, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent:
React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.09, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/

```

```

open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent:
React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.34, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.5714, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent:
React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0.62, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taubench
Airline", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.6465, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0.83, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "rank": 4, "label": "Taubench
Telecom", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall"}, "value": 0.4276, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Over
React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.1, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "n
React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.34, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.5714, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "i
React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0.62, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taut
Airline", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.6465, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0.83, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true}, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 4, "label": "Taut
Telecom", "notes": "agent: React + Shortlisting"}, {"dataset": {"id": "claw-eval/Claw-
Eval", "isBenchmark": true, "task_id": "general"}, "value": 52.8, "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/ryhvAX5djQpD5OrIIQZQ1f.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": 117, "rank": 20, "label": "General", "notes": "Pass3% | N=3 | 161 tasks"},

```

```

{"dataset":{"id":"claw-eval/Claw-Eval","isBenchmark":true,"task_id":"multimodal"},"value":14.8,"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":117,"rank":6,"label":"Multimodal","notes":"Pass3% | N=3 | 101
tasks"},"dataset":{"id":"claw-eval/Claw-Eval","isBenchmark":true,"task_id":"multi_turn"},"value":50,"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":117,"rank":10,"label":"Multi Turn","notes":"Pass3% | N=3 | 38 tasks"},
{"dataset":{"id":"MME-Benchmarks/Video-MME-v2","isBenchmark":true,"task_id":"video-mme-
v2"},"value":61.1,"source":{"url":"https://video-mme-v2.netlify.app/#leaderboard","name":"VideoMMEv2
leaderboard","isExternal":true,"author":{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-
avatars.huggingface.co/v1/production/
uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullname":"Niels
Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isI
videomme_v2.yaml","verified":false,"pullRequest":116,"rank":3,"label":"Video Mme V2","notes":"512 frames"},
{"dataset":{"id":"MME-Benchmarks/Video-MME-v2","isBenchmark":true,"task_id":"video-mme-
v2"},"value":54.4,"source":{"url":"https://video-mme-v2.netlify.app/#leaderboard","name":"VideoMMEv2
leaderboard","isExternal":true,"author":{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-
avatars.huggingface.co/v1/production/
uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullname":"Niels
Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isI
videomme_v2.yaml","verified":false,"pullRequest":116,"label":"Video Mme V2","notes":"64 frames"},"dataset":
{"id":"mercor/APEX-SWE","isBenchmark":false,"task_id":"apex-swe"},"value":30.3,"source":{"url":"https://
www.mercor.com/apex/apex-swe-leaderboard/","name":"APEX-SWE Leaderboard","isExternal":true,"author":
{"_id":"6531a6d5c93c85e11c82d495","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/
uploads/63b4a97834df1f9a90a0da76/
HHWjFJmEitVTMv3Fk85wu.jpeg","fullname":"Mercor","name":"mercor","type":"org","isHf":false,"isHfAdmin":false,"isI
apex-swe.yaml","verified":false,"pullRequest":114,"rank":1,"label":"Apex Swe","notes":"harness: claude_code;
reasoning_effort: thinking"},"dataset":{"id":"collinear-ai/yc-
bench","isBenchmark":true,"task_id":"medium"},"value":408822,"source":{"url":"https://github.com/collinear-ai/yc-
bench","name":"YC-Bench eval","isExternal":true,"filename":"eval_results/yc-
bench.yaml","verified":false,"pullRequest":104,"rank":6,"label":"Medium","notes":"avg final funds (USD) across
seeds 1,2,3. Kimi K2.5 (via OpenRouter moonshotai/kimi-k2.5)"},"dataset":{"id":"MathArena/
hmmmt_feb_2026","isBenchmark":true,"task_id":"MathArena/hmmmt_feb_2026"},"value":87.12,"source":{"url":"https://
matharena.ai/?comp=hmmmt-hmmmt_feb_2026","name":"Official MathArena
Evaluation","isExternal":true,"filename":"eval_results/MathArena-
hmmmt_feb_2026.yaml","verified":false,"pullRequest":95,"rank":6,"label":"MathArena Hmmt Feb
2026"},"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","deploy:sagemaker","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"},
{"id":"deploy:sagemaker","label":"sagemaker","type":"deploy"},"{"type":"region","label":"🌐 Region:
US","id":"region:us"}],"transformersInfo":

```

```

{"auto_model":"AutoModel","custom_class":"modeling_kimi_k25.KimiK25ForConditionalGeneration","pipeline_tag":"text-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":16,"isReleased":true,"accuratePricing":true}, {"isOriginalProvider":false,"name":"featherless-ai","enabled":true,"position":7,"isReleased":true,"accuratePricing":true}],"canWrite":false}> 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\n# Find last non-tool-call assistant message #\n{%- set ns = namespace(last_non_tool_call_assistant_msg=-1) -%>\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\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{%- 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 -%>\n <think></think>{{render_content(message)}}\n {%- else -%>\n {%- set rc = 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\n%}"},"created_at":"2026-01-01T06:06:03.000Z","discussionsDisabled":false,"discussionsSorting":"recently-created","downloads":327584,"downloadsAllTime":16128649,"id":"moonshotai/Kimi-K2.5","isLikedByUser":false,"availableInferenceProviders":[{"provider":"novita","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/kimi-k2.5","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","features":{"structuredOutput":false,"toolCalling":true},"isCheapestPricingOutput":false,"isFastestThroughput":true,"isModelAuth":{"provider":"featherless-ai","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/Kimi-K2.5","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","isCheapestPricingOutput":false,"isFastestThroughput":false,"isModelAuthor":false},

```

```

{"provider":"deepinfra","modelStatus":"live","providerStatus":"live","providerId":"moonshotai/Kimi-K2.5","task":"conversational","adapterWeightsPath":"model-00001-of-000064.safetensors","features":{"structuredOutput":true,"toolCalling":true},"isCheapestPricingOutput":true,"isFastestThroughput":false,"isModelAuthentic":true,"library_name":"transformers","librariesOther":[],"trackDownloads":true,"model-index":null,"evalResults":[{"dataset":{"id":"TIGER-Lab/MMLU-Pro"},"isBenchmark":true,"task_id":"mmlu_pro"},"value":87.1,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.5"},"name":"Model Card","isExternal":false,"author":{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullName":"Niels Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isMmlu_pro.yaml","verified":false,"pullRequest":51,"rank":6,"label":"Mmlu Pro"},"dataset":{"id":"SWE-bench/SWE-bench_Verified"},"isBenchmark":true,"task_id":"swe_bench_%_resolved"},"value":70.8,"source":{"url":"https://www.swebench.com/"},"name":"SWE-Bench official evaluation","isExternal":true,"author":{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullName":"Niels Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isSWE_bench_verified.yaml","verified":false,"pullRequest":97,"rank":38,"label":"Swe Bench Resolved","notes":"high reasoning"},"dataset":{"id":"Idavidrein/gpqa"},"isBenchmark":true,"task_id":"diamond"},"value":87.3737373737,"source":{"url":"https://huggingface.co/datasets/eval/EEE_datastore/blob/192329fb7d6b15b7b0936a1a58ae862aa7e8ba24/flat/objects/5c/4b/5c4b9bd9-14db-4c11-bf99-9661461459f6.json"},"name":"EvalEval","isExternal":true},"filename":"eval_results/gpqa-diamond.yaml","verified":false,"pullRequest":126,"label":"Diamond","notes":"GPQA Diamond"},"dataset":{"id":"ScaleAI/SWE-bench_Pro"},"isBenchmark":true,"task_id":"SWE_Bench_Pro"},"value":50.7,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.5"},"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,"isSWE_bench_pro.yaml","verified":false,"pullRequest":101,"rank":31,"label":"SWE Bench Pro"},"dataset":{"id":"Idavidrein/gpqa"},"isBenchmark":true,"task_id":"diamond"},"value":87.6,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.5"},"name":"Model Card","isExternal":false,"author":{"_id":"5f1158120c833276f61f1a84","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg","fullName":"Niels Rogge","name":"nielsr","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":1336,"isGPQA_diamond.yaml","verified":false,"pullRequest":50,"rank":28,"label":"Diamond"},"dataset":{"id":"mercor/apex-agents"},"isBenchmark":true,"task_id":"apex-agents"},"value":14.4,"source":{"url":"https://www.mercor.com/apex/apex-agents-leaderboard/"},"name":"APEX-Agents Leaderboard","isExternal":true,"author":{"_id":"6531a6d5c93c85e11c82d495","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/63b4a97834df1f9a90a0da76/HHWjFJmEitVTMv3Fk85wu.jpeg","fullName":"Mercor","name":"mercor","type":"org","isHf":false,"isHfAdmin":false,"isApex-agents.yaml","verified":false,"pullRequest":113,"rank":8,"label":"Apex Agents","notes":"harness: react_toolbelt_agent; reasoning_effort: thinking"},"dataset":{"id":"MMMU/MMMU_Pro"},"isBenchmark":true,"task_id":"mmmu_pro_vision"},"value":78.5,"source":{"url":"https://huggingface.co/moonshotai/Kimi-K2.5"},"name":"Model Card","isExternal":false},"filename":"eval_results/mmму_pro.yaml","verified":false,"pullRequest":119,"rank":3,"label":"Mmmu Pro Vision"},"dataset":{"id":"harborframework/terminal-bench-2.0"},"isBenchmark":true,"task_id":"terminalbench_2"},"value":43.2,"source":{"url":"https://www.tbench.ai/leaderboard/terminal-bench/2.0"},"name":"Terminal-Bench Leaderboard","isExternal":true,"author":{"_id":"62d648291fa3e4e7ae3fa6e8","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/62d648291fa3e4e7ae3fa6e8/oatOwF8Xqe5eDbCSuYqCd.png","fullName":"ben burtenshaw","name":"burtenshaw","type":"user","isPro":false,"isHf":true,"isHfAdmin":false,"isMod":false,"followerCount":190,"isTerminal_bench.yaml","verified":false,"pullRequest":90,"rank":19,"label":"Terminalbench 2"},"notes":"agent: Terminus 2"},"dataset":{"id":"MathArena/aima_2026"},"isBenchmark":true,"task_id":"MathArena/aima_2026"},"value":95.83,"source":{"url":"https://matharena.ai/?comp=aima--aima_2026"},"name":"Official MathArena Evaluation","isExternal":true},"filename":"eval_results/MathArena--aima_2026.yaml","verified":false,"pullRequest":88,"rank":7,"label":"MathArena Aima 2026"},"dataset":{"id":"cais/

```

hle", "isBenchmark": false, "task_id": "hle", "value": 50.2, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "is_hle_with_tools.yaml", "verified": false, "pullRequest": 49, "rank": 17, "label": "Hle", "notes": "With tools"}, {"dataset": {"id": "SWE-bench/SWE-bench_Multilingual", "isBenchmark": true, "task_id": "swe_bench_multilingual_%_resolved", "value": 73, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "filename": ".eval_results/swe-bench_multilingual.yaml", "verified": false, "pullRequest": 131, "rank": 14, "label": "Swe Bench Multilingual Resolved", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "overall", "value": 89.09, "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": 130, "rank": 7, "label": "Overall", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "scanning", "value": 89.67, "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": 130, "rank": 9, "label": "Scanning", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "warping", "value": 88.86, "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": 130, "rank": 8, "label": "Warping", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "screen_photography", "value": 88.39, "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": 130, "rank": 10, "label": "Screen Photography", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "illumination", "value": 89.66, "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": 130, "rank": 8, "label": "Illumination", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "skew", "value": 88.86, "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": 130, "rank": 6, "label": "Skew", "dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "overall", "value": 77.5, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}}, "mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 7, "label": "Overall", "dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "digital", "value": 85, "source": {"url": "https://huggingface.co/datasets/

```

Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 9, "label": "Digital", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "photographed"}, "value": 75, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 7, "label": "Photographed", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "latin", "value": 81.6, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Latin", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "de", "value": 85.9, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 3, "label": "De", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "en", "value": 86.2, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 4, "label": "En", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "es", "value": 72.7, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Es", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "fr", "value": 71, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Fr", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "id", "value": 80.6, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 7, "label": "Id", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "it", "value": 86.6, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false},
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "It", {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "nl", "value": 77.4, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-

```

```
auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":9,"label":"Nl"),{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"pt"},"value":87.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":3,"label":"Pt"),{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"vi"},"value":86.2,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":1,"label":"Vi"),{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"non_latin"},"value":72.9,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":8,"label":"Non Latin"},{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"ar"},"value":75.8,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":6,"label":"Ar"},{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"hi"},"value":74.5,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Hi"},{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"jp"},"value":72.5,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":3,"label":"Jp"},{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"ko"},"value":70.9,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":8,"label":"Ko"},{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"ru"},"value":61.8,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}}, mdpbench.yaml","verified":false,"pullRequest":129,"rank":9,"label":"Ru"},{"dataset":{"id":"Delores-Lin/MDPBench","isBenchmark":true,"task_id":"th"},"value":67,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench","name":"MDPBench leaderboard","isExternal":true,"author":{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullname":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
```

```

mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 5, "label": "Th"}, {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "zh"}, "value": 81.7, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png", "fullname": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}}, mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 4, "label": "Zh"}, {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "zh_t"}, "value": 78.6, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png", "fullname": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}}, mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 2, "label": "Zh T"}, {"dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "private"}, "value": 81.2, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png", "fullname": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}}, mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 4, "label": "Private"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "default"}, "value": 24.37, "source": {"url": "https://huggingface.co/datasets/evaleval/EEE_datastore/blob/192329fb7d6b15b7b0936a1a58ae862aa7e8ba24/flat/objects/aa/05/aa055fe5-5174-4517-8b10-19d46da94f3e.json", "name": "EvalEval", "isExternal": true, "filename": ".eval_results/hle.yaml", "verified": false, "pullRequest": 126, "rank": 1, "label": "Default"}, {"dataset": {"id": "internlm/WildClawBench", "isBenchmark": false, "task_id": "overall"}, "value": 30.8, "source": {"url": "https://internlm.github.io/WildClawBench", "name": "WildClawBench", "isExternal": true, "author": {"_id": "64a2d5fa81252883206f24c9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png", "fullname": "Intern Large Models", "name": "internlm", "type": "org", "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 1347, "isUserFollow": true}, wildclawbench.yaml", "verified": false, "pullRequest": 121, "rank": 20, "label": "Overall"}, {"dataset": {"id": "internlm/WildClawBench", "isBenchmark": false, "task_id": "avg_time"}, "value": 406, "source": {"url": "https://internlm.github.io/WildClawBench", "name": "WildClawBench", "isExternal": true, "author": {"_id": "64a2d5fa81252883206f24c9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png", "fullname": "Intern Large Models", "name": "internlm", "type": "org", "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 1347, "isUserFollow": true}, wildclawbench.yaml", "verified": false, "pullRequest": 121, "rank": 15, "label": "Avg Time"}, {"dataset": {"id": "internlm/WildClawBench", "isBenchmark": false, "task_id": "avg_cost"}, "value": 6.6, "source": {"url": "https://internlm.github.io/WildClawBench", "name": "WildClawBench", "isExternal": true, "author": {"_id": "64a2d5fa81252883206f24c9", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/6432683407bad11484a68457/Q3Y0dL79GcsnaBCGRMooZ.png", "fullname": "Intern Large Models", "name": "internlm", "type": "org", "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 1347, "isUserFollow": true}, wildclawbench.yaml", "verified": false, "pullRequest": 121, "rank": 11, "label": "Avg Cost"}, {"dataset": {"id": "InternScience/ResearchClawBench", "isBenchmark": false, "task_id": "overall"}, "value": 13.96, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "filename": ".eval_results/researchclawbench.yaml", "verified": false, "pullRequest": 120, "rank": 7, "label": "Overall", "notes": "ResearchHarness evaluation with tools enabled, code execution, and a file-system workspace; completed 39/40 ResearchClawBench tasks."}, {"dataset": {"id": "open-agent-leaderboard/results", "isBenchmark": false, "task_id": "overall"}, "value": 0.3026, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.08, "source": {"url": "https://www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.56, "source": {"url": "https://

```

```

www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "rank": 3, "label": "Browsecomp
Plus", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.5204, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent:
Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0.12, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Airline", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.03, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_claude_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Telecom", "notes": "agent: Claude Code"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall"}, "value": 0.2501, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.08, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.35, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.567, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent:
OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Airline", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.01, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_openai_solo.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Telecom", "notes": "agent: OpenAI Solo"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall"}, "value": 0.42, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent:
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.11, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Appworld", "note
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.33, "source": {"url": "https://

```

```

www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.5761, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Swebench", "noi
Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0.56, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taubench
Airline", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.7245, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taubench
Retail", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0.7071, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_smolagents_code.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Telecom", "notes": "agent: Smolagent"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall"}, "value": 0.4251, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Overall", "notes": "agent:
React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.09, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "notes": "agent:
React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.34, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
Plus", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "swebench"}, "value": 0.5714, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "notes": "agent:
React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_airline"}, "value": 0.62, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taubench
Airline", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_retail"}, "value": 0.6465, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
Retail", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "taubench_telecom"}, "value": 0.83, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling.yaml", "verified": false, "pullRequest": 118, "rank": 4, "label": "Taubench
Telecom", "notes": "agent: React"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "overall"}, "value": 0.4276, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Over
React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "appworld"}, "value": 0.1, "source": {"url": "https://
www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Appworld", "n
React + Shortlisting"}, {"dataset": {"id": "open-agent-leaderboard/
results", "isBenchmark": false, "task_id": "browsecomp_plus"}, "value": 0.34, "source": {"url": "https://

```

www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
 open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Browsecomp
 Plus", "notes": "agent: React + Shortlisting", {"dataset": {"id": "open-agent-leaderboard/
 results", "isBenchmark": false, "task_id": "swebench", "value": 0.5714, "source": {"url": "https://
 www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
 open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Swebench", "i
 React + Shortlisting", {"dataset": {"id": "open-agent-leaderboard/
 results", "isBenchmark": false, "task_id": "taubench_airline", "value": 0.62, "source": {"url": "https://
 www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
 open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 6, "label": "Taut
 Airline", "notes": "agent: React + Shortlisting", {"dataset": {"id": "open-agent-leaderboard/
 results", "isBenchmark": false, "task_id": "taubench_retail", "value": 0.6465, "source": {"url": "https://
 www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
 open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "label": "Taubench
 Retail", "notes": "agent: React + Shortlisting", {"dataset": {"id": "open-agent-leaderboard/
 results", "isBenchmark": false, "task_id": "taubench_telecom", "value": 0.83, "source": {"url": "https://
 www.exgentic.ai", "name": "Open Agent Leaderboard", "isExternal": true, "filename": ".eval_results/
 open_agent_leaderboard_tool_calling_with_shortlisting.yaml", "verified": false, "pullRequest": 118, "rank": 4, "label": "Taut
 Telecom", "notes": "agent: React + Shortlisting", {"dataset": {"id": "claw-eval/Claw-
 Eval", "isBenchmark": true, "task_id": "general", "value": 52.8, "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/ryhvAX5djQpD5OrlZQ1f.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": 117, "rank": 20, "label": "General", "notes": "Pass% | N=3 | 161 tasks"},
 {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multimodal", "value": 14.8, "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/ryhvAX5djQpD5OrlZQ1f.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": 117, "rank": 6, "label": "Multimodal", "notes": "Pass% | N=3 | 101
 tasks"}, {"dataset": {"id": "claw-eval/Claw-Eval", "isBenchmark": true, "task_id": "multi_turn", "value": 50, "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/ryhvAX5djQpD5OrlZQ1f.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": 117, "rank": 10, "label": "Multi Turn", "notes": "Pass% | N=3 | 38 tasks"},
 {"dataset": {"id": "MME-Benchmarks/Video-MME-v2", "isBenchmark": true, "task_id": "video-mme-
 v2", "value": 61.1, "source": {"url": "https://video-mme-v2.netlify.app/#leaderboard", "name": "VideoMMEv2
 leaderboard", "isExternal": true, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels
 Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isl
 videomme_v2.yaml", "verified": false, "pullRequest": 116, "rank": 3, "label": "Video Mme V2", "notes": "512 frames"},
 {"dataset": {"id": "MME-Benchmarks/Video-MME-v2", "isBenchmark": true, "task_id": "video-mme-
 v2", "value": 54.4, "source": {"url": "https://video-mme-v2.netlify.app/#leaderboard", "name": "VideoMMEv2
 leaderboard", "isExternal": true, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-
 avatars.huggingface.co/v1/production/
 uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels
 Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isl
 videomme_v2.yaml", "verified": false, "pullRequest": 116, "label": "Video Mme V2", "notes": "64 frames"}, {"dataset":
 {"id": "mercor/APEX-SWE", "isBenchmark": false, "task_id": "apex-swe", "value": 30.3, "source": {"url": "https://
 www.mercor.com/apex/apex-swe-leaderboard/", "name": "APEX-SWE Leaderboard", "isExternal": true, "author":
 {"_id": "6531a6d5c93c85e11c82d495", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/
 uploads/63b4a97834df1f9a90a0da76/
 HHWjFJmEitVTMv3Fk85wu.jpeg", "fullName": "Mercor", "name": "mercor", "type": "org", "isHf": false, "isHfAdmin": false, "isA

```
apex-swe.yaml", "verified": false, "pullRequest": 114, "rank": 1, "label": "Apex Swe", "notes": "harness: claude_code; reasoning_effort: thinking"}, {"dataset": {"id": "collinear-ai/yc-bench", "isBenchmark": true, "task_id": "medium"}, "value": 408822, "source": {"url": "https://github.com/collinear-ai/yc-bench", "name": "YC-Bench eval", "isExternal": true, "filename": ".eval_results/yc-bench.yaml"}, "verified": false, "pullRequest": 104, "rank": 6, "label": "Medium", "notes": "avg final funds (USD) across seeds 1,2,3. Kimi K2.5 (via OpenRouter moonshotai/kimi-k2.5)"}, {"dataset": {"id": "MathArena/hmmt_feb_2026", "isBenchmark": true, "task_id": "MathArena/hmmt_feb_2026"}, "value": 87.12, "source": {"url": "https://matharena.ai/?comp=hmmt-hmmt_feb_2026", "name": "Official MathArena Evaluation", "isExternal": true, "filename": ".eval_results/MathArena-hmmt_feb_2026.yaml"}, "verified": false, "pullRequest": 95, "rank": 6, "label": "MathArena Hmmt Feb 2026"}], "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", "deploy:sagemaker", "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"}, {"id": "deploy:sagemaker", "label": "sagemaker", "type": "deploy"}, {"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": 16, "isReleased": true, "accuratePricing": true}, {"isOriginalProvider": false, "name": "featherless-ai", "enabled": true, "position": 7, "isReleased": true, "accuratePricing": true}], "linkToInferenceDiscussion": {"url": "/spaces/huggingface/InferenceSupport/discussions/7667", "numReactions": 14}, "isLoggedIn": false, "canWrite": false, "examples": [{"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."}], "initialSelectedProvider": "novita", "widgetType": "conversational", "isMaximized": false, "isMac": true}}>
```

Image-Text-to-Text

Examples

Input a message to start chatting with moonshotai/Kimi-K2.5 .

Send

View Code Snippets

Compare providers

Model tree for moonshotai/Kimi-K2.5

Adapters

29 models

Finetunes

45 models

Merges

1 model

Quantizations

40 models

Spaces using moonshotai/Kimi-K2.5 100

Collection including moonshotai/Kimi-K2.5 Moonshot's large visual-language model • 4 items • Updated Jul 27 • 84

Paper for moonshotai/Kimi-K2.5 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\n# Find last non-tool-call assistant message #\n{%- set ns =
namespace(last_non_tool_call_assistant_msg=-1) -%}\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\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\n{%- for message in suffix_msgs -
%}\n {{set_roles(message)}}\n {%- if message['role'] == 'assistant' -%}\n {%- if thinking is defined and thinking is
false -%}\n <think></think>{{render_content(message)}}\n {%- else -%}\n {%- set rc =
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\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": "2026-01-01T06:06:03.000Z", "discussionsDisabled": false, "discussionsSorting": "recently-
created", "downloads": 327584, "downloadsAllTime": 16128649, "id": "moonshotai/Kimi-
K2.5", "isLikedByUser": false, "availableInferenceProviders":
[{"provider": "novita", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/kimi-
k2.5", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features":
{"structuredOutput": false, "toolCalling": true}, "isCheapestPricingOutput": false, "isFastestThroughput": true, "isModelAuth
{"provider": "featherless-ai", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
K2.5", "task": "conversational", "adapterWeightsPath": "model-00001-
of-000064.safetensors", "isCheapestPricingOutput": false, "isFastestThroughput": false, "isModelAuthor": false},
{"provider": "deepinfra", "modelStatus": "live", "providerStatus": "live", "providerId": "moonshotai/Kimi-
```

K2.5", "task": "conversational", "adapterWeightsPath": "model-00001-of-000064.safetensors", "features": {"structuredOutput": true, "toolCalling": true}, "isCheapestPricingOutput": true, "isFastestThroughput": false, "isModelAuthentic": true, "library_name": "transformers", "librariesOther": [], "trackDownloads": true, "model-index": null, "evalResults": [{"dataset": {"id": "TIGER-Lab/MMLU-Pro", "isBenchmark": true, "task_id": "mmlu_pro"}, "value": 87.1, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isMmlu_pro_verified": false, "pullRequest": 51, "rank": 6, "label": "Mmlu Pro"}, {"dataset": {"id": "SWE-bench/SWE-bench_Verified", "isBenchmark": true, "task_id": "swe_bench_%_resolved", "value": 70.8, "source": {"url": "https://www.swebench.com/", "name": "SWE-Bench official evaluation", "isExternal": true, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isSwe_bench_verified": false, "pullRequest": 97, "rank": 38, "label": "Swe Bench Resolved", "notes": "high reasoning"}, {"dataset": {"id": "Idavidrein/gpqa", "isBenchmark": true, "task_id": "diamond", "value": 87.3737373737, "source": {"url": "https://huggingface.co/datasets/eval/EEE_datastore/blob/192329fb7d6b15b7b0936a1a58ae862aa7e8ba24/flat/objects/5c/4b/5c4b9bd9-14db-4c11-bf99-9661461459f6.json", "name": "EvalEval", "isExternal": true}, "filename": ".eval_results/gpqa-diamond.yaml", "verified": false, "pullRequest": 126, "label": "Diamond", "notes": "GPQA Diamond"}, {"dataset": {"id": "ScaleAI/SWE-bench_Pro", "isBenchmark": true, "task_id": "SWE_Bench_Pro", "value": 50.7, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "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, "isSwe_bench_pro_verified": false, "pullRequest": 101, "rank": 31, "label": "SWE Bench Pro"}, {"dataset": {"id": "Idavidrein/gpqa", "isBenchmark": true, "task_id": "diamond", "value": 87.6, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isGpqa_diamond_verified": false, "pullRequest": 50, "rank": 28, "label": "Diamond"}, {"dataset": {"id": "mercor/apex-agents", "isBenchmark": true, "task_id": "apex-agents", "value": 14.4, "source": {"url": "https://www.mercor.com/apex/apex-agents-leaderboard/", "name": "APEX-Agents Leaderboard", "isExternal": true, "author": {"_id": "6531a6d5c93c85e11c82d495", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/63b4a97834df1f9a90a0da76/HHWjFJmEitVTMv3Fk85wu.jpeg", "fullName": "Mercor", "name": "mercor", "type": "org", "isHf": false, "isHfAdmin": false, "isApex_agents_verified": false, "pullRequest": 113, "rank": 8, "label": "Apex Agents", "notes": "harness: react_toolbelt_agent; reasoning_effort: thinking"}, {"dataset": {"id": "MMMU/MMMU_Pro", "isBenchmark": true, "task_id": "mmmu_pro_vision", "value": 78.5, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false}, "filename": ".eval_results/mmmu_pro.yaml", "verified": false, "pullRequest": 119, "rank": 3, "label": "Mmmu Pro Vision"}, {"dataset": {"id": "harborframework/terminal-bench-2.0", "isBenchmark": true, "task_id": "terminalbench_2", "value": 43.2, "source": {"url": "https://www.tbench.ai/leaderboard/terminal-bench/2.0", "name": "Terminal-Bench Leaderboard", "isExternal": true, "author": {"_id": "62d648291fa3e4e7ae3fa6e8", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/62d648291fa3e4e7ae3fa6e8/oatOwF8Xqe5eDbCSuYqCd.png", "fullName": "ben burtenshaw", "name": "burtenshaw", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 90, "rank": 19, "label": "Terminalbench 2", "notes": "agent: Terminus 2"}, {"dataset": {"id": "MathArena/aima_2026", "isBenchmark": true, "task_id": "MathArena/aima_2026", "value": 95.83, "source": {"url": "https://matharena.ai/?comp=aima--aima_2026", "name": "Official MathArena Evaluation", "isExternal": true}, "filename": ".eval_results/MathArena--aima_2026.yaml", "verified": false, "pullRequest": 88, "rank": 7, "label": "MathArena Aima 2026"}, {"dataset": {"id": "cais/hle", "isBenchmark": false, "task_id": "hle", "value": 50.2, "source": {"url": "https://huggingface.co/moonshotai/Kimi-

K2.5", "name": "Model Card", "isExternal": false, "author": {"_id": "5f1158120c833276f61f1a84", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/1608042047613-5f1158120c833276f61f1a84.jpeg", "fullName": "Niels Rogge", "name": "nielsr", "type": "user", "isPro": false, "isHf": true, "isHfAdmin": false, "isMod": false, "followerCount": 1336, "isHle_with_tools.yaml", "verified": false, "pullRequest": 49, "rank": 17, "label": "Hle", "notes": "With tools"}, {"dataset": {"id": "SWE-bench/SWE-bench_Multilingual", "isBenchmark": true, "task_id": "swe_bench_multilingual_%_resolved", "value": 73, "source": {"url": "https://huggingface.co/moonshotai/Kimi-K2.5", "name": "Model Card", "isExternal": false, "filename": ".eval_results/swe-bench_multilingual.yaml", "verified": false, "pullRequest": 131, "rank": 14, "label": "Swe Bench Multilingual Resolved", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "overall", "value": 89.09, "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": 130, "rank": 7, "label": "Overall", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "scanning", "value": 89.67, "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": 130, "rank": 9, "label": "Scanning", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "warping", "value": 88.86, "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": 130, "rank": 8, "label": "Warping", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "screen_photography", "value": 88.39, "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": 130, "rank": 10, "label": "Screen Photography", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "illumination", "value": 89.66, "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": 130, "rank": 8, "label": "Illumination", "dataset": {"id": "PaddlePaddle/Real5-OmniDocBench", "isBenchmark": true, "task_id": "skew", "value": 88.86, "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": 130, "rank": 6, "label": "Skew", "dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "overall", "value": 77.5, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfr9DP6tqb8bh.png", "fullName": "Delores Lin", "name": "Delores-Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}, "mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 7, "label": "Overall", "dataset": {"id": "Delores-Lin/MDPBench", "isBenchmark": true, "task_id": "digital", "value": 85, "source": {"url": "https://huggingface.co/datasets/Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":

```

{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":9,"label":"Digital"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"photographed"},"value":75,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":7,"label":"Photographed"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"latin"},"value":81.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Latin"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"de"},"value":85.9,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":3,"label":"De"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"en"},"value":86.2,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":4,"label":"En"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"es"},"value":72.7,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Es"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"fr"},"value":71,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"Fr"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"id"},"value":80.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":7,"label":"Id"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"it"},"value":86.6,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-Lin","type":"user","isPro":false,"isHf":false,"isHfAdmin":false,"isMod":false,"followerCount":2,"isUserFollowing":false}},
mdpbench.yaml","verified":false,"pullRequest":129,"rank":5,"label":"It"},{"dataset":{"id":"Delores-Lin/MDPBench"},"isBenchmark":true,"task_id":"nl"},"value":77.4,"source":{"url":"https://huggingface.co/datasets/Delores-Lin/MDPBench"},"name":"MDPBench leaderboard","isExternal":true,"author":
{"_id":"687a14565d8f9617af6ab38b","avatarUrl":"https://cdn-avatars.huggingface.co/v1/production/uploads/no-auth/LgZxNkDzfd9DP6tqb8bh.png","fullName":"Delores Lin","name":"Delores-

```

```
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},  
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 9, "label": "NI"}, {"dataset": {"id": "Delores-Lin/  
MDPBench", "isBenchmark": true, "task_id": "pt"}, "value": 87.6, "source": {"url": "https://huggingface.co/datasets/  
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author":  
{"_id": "687a14565d8f9617af6ab38b", "avatarUrl": "https://cdn-avatars.huggingface.co/v1/production/uploads/no-  
auth/LgZxNkDzldr9DP6tqb8bh.png", "fullname": "Delores Lin", "name": "Delores-  
Lin", "type": "user", "isPro": false, "isHf": false, "isHfAdmin": false, "isMod": false, "followerCount": 2, "isUserFollowing": false}},  
mdpbench.yaml", "verified": false, "pullRequest": 129, "rank": 3, "label": "Pt"}, {"dataset": {"id": "Delores-Lin/  
MDPBench", "isBenchmark": true, "task_id": "vi"}, "value": 86.2, "source": {"url": "https://huggingface.co/datasets/  
Delores-Lin/MDPBench", "name": "MDPBench leaderboard", "isExternal": true, "author": {"_id": "687a1
```