Benchmark results

Scores for Patram 1.5 and the 27 models it was evaluated alongside, on PatramBench's 15 tasks and 34 public benchmarks.

83

Patram 1.5's overall score. 21 models that read images.

1	**Patram 1.5**	**83**
1	DeepSeek-V4.1-Flash	83
1	Qwen3.5-397B-A17B	83
4	Gemma 4 31B	82
4	Muse Glimmer 30B	82
6	InternVL3.5-241B-A28B	80
6	MiniMax-M3.1-Flash	80
6	Nemotron-3.5-Super-VL	80
9	Qwen3.5-9B	79
10	GLM-5.3-Flash	77
11	ERNIE-4.5-VL-424B-A47B	76
12	InternVL3.5-8B	75
13	Gemma 4 12B	74
14	GLM-4.6V-Flash	71
14	Mistral-Small-4-119B	71
16	Kimi-VL-A3B-Thinking-2506	70
17	Ministral-3-8B	68
18	Phi-4-reasoning-vision-15B	63
19	LensVLM-9B	55
20	DeepSeek-VL2-small	54
21	Molmo2-8B	51

85

Patram 1.5's PatramBench score. 15 tasks.

1	**Patram 1.5**	**85**
1	GLM-5.3-Flash	85
1	Gemma 4 31B	85
1	Muse Glimmer 30B	85
5	DeepSeek-V4.1-Flash	84
5	MiniMax-M3.1-Flash	84
5	Qwen3.5-397B-A17B	84
8	InternVL3.5-241B-A28B	80
9	Gemma 4 12B	78
9	Nemotron-3.5-Super-VL	78
9	Qwen3.5-9B	78
12	ERNIE-4.5-VL-424B-A47B	77
13	InternVL3.5-8B	74
13	Mistral-Small-4-119B	74
15	Ministral-3-8B	72
16	Kimi-VL-A3B-Thinking-2506	70
17	GLM-4.6V-Flash	69
18	Phi-4-reasoning-vision-15B	65
19	Molmo2-8B	54
20	LensVLM-9B	52
21	DeepSeek-VL2-small	47

81

Patram 1.5's public benchmark score. 34 benchmarks.

1	DeepSeek-V4.1-Flash	82
1	Qwen3.5-397B-A17B	82
3	**Patram 1.5**	**81**
3	Nemotron-3.5-Super-VL	81
3	Qwen3.5-9B	81
6	Gemma 4 31B	80
6	InternVL3.5-241B-A28B	80
8	Muse Glimmer 30B	79
9	MiniMax-M3.1-Flash	77
10	ERNIE-4.5-VL-424B-A47B	76
10	InternVL3.5-8B	76
12	GLM-4.6V-Flash	72
13	GLM-5.3-Flash	69
13	Gemma 4 12B	69
13	Kimi-VL-A3B-Thinking-2506	69
13	Mistral-Small-4-119B	69
17	Ministral-3-8B	64
18	Phi-4-reasoning-vision-15B	61
19	DeepSeek-VL2-small	60
20	LensVLM-9B	57
21	Molmo2-8B	47

Patram 1.5 other models Patram 1.5's score, as a guide line. Whole numbers, rounded half up.

How the numbers were measured

- Our own runs are scored on the Patram Eval board, official-metric mode, thinking off, with the same item sets and the same scorers.
- English items only. Serving: vLLM, Hugging Face, or the provider's API.
- Every score is our own run, for every model.
- Whole numbers, rounded half up. Ranks are taken on the shown numbers, so a tie shares the place.
- A capability score is the mean of its tests. A model needs all of them when there are 4 or fewer, at least 80% when there are more.
- PatramBench overall is the mean of the 15 tasks. Public overall is the mean of the 5 capability scores (a model needs at least 3). Overall score = (PatramBench + public) / 2. PatramBench is BharatGen's own benchmark: 11,828 test items from 8,773 pages of Indian documents, across 20 domains and 151 document types. Dataset: to be released on Hugging Face in the coming days.
- Field: Patram 1.5 plus 27 models (20 models that read images, both specialist and general-purpose, and 7 OCR and parsing systems), from 0.63B to 763B parameters. OCR and parsing systems are scored on the parsing tests only, so they have no overall score. Sizes are measured parameter counts. Patram 1.5 is released as a 7B model; its measured size, 7.8B, is the figure shown in the charts.

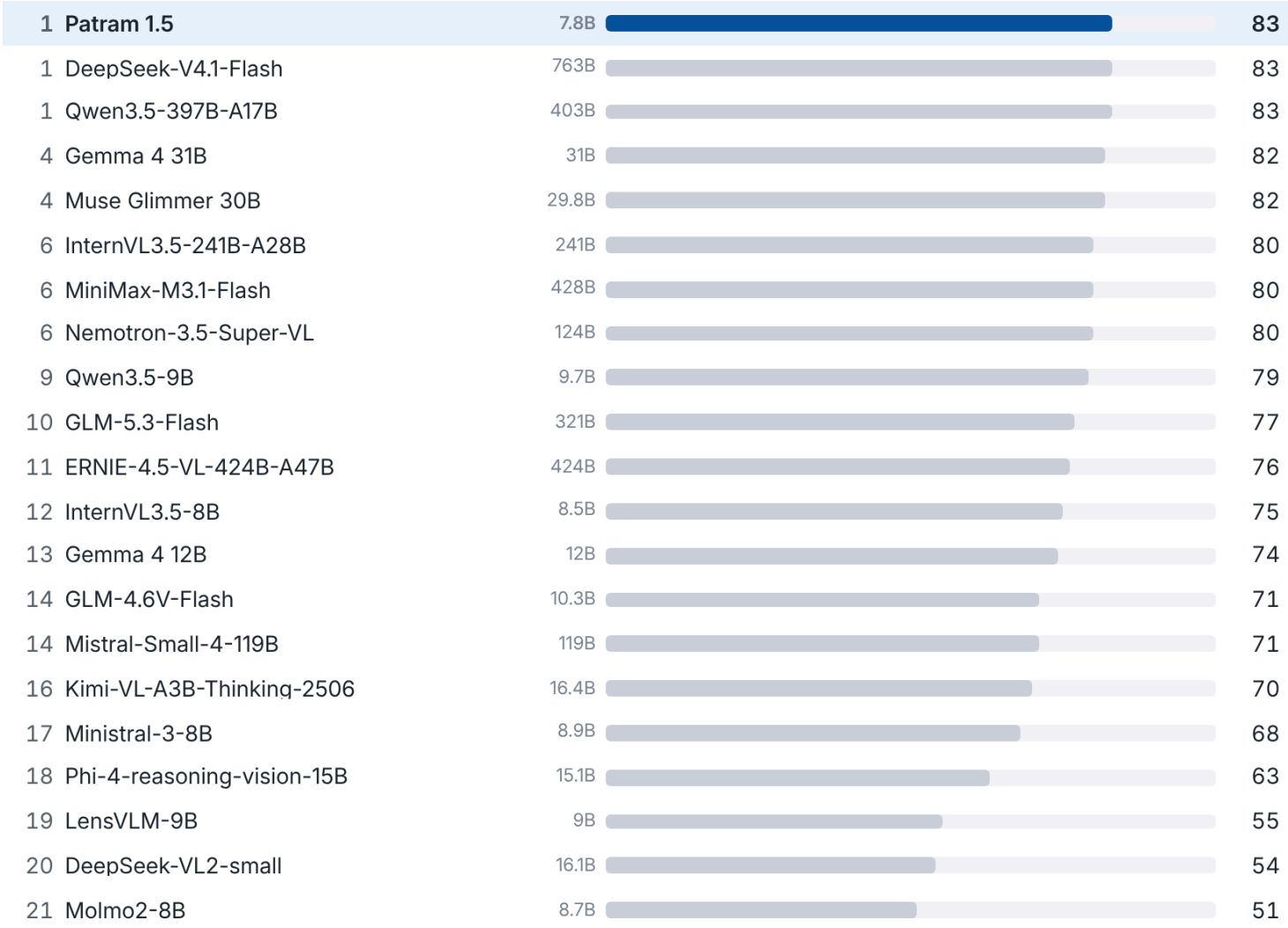
Contents. 1 What we measure · 2 Overall score · 3 Models up to 10B · 4 Models from 12B to 763B · 5 OCR and parsing systems · 6 Every result · 7 Every PatramBench task · 8 Every public benchmark, with an index of the benchmarks

1 What we measure

Capability	PatramBench tasks (15)	Public benchmarks (34)
Doc parsing	Full-page OCR; Page to Markdown; Page to HTML	OCRBench, TextOCR, HierText, Union14M-Benchmark, WordArt-Bench, CC-OCR multi-scene, IAM (handwriting lines), OmniDocBench, olmOCR-Bench, olmOCR-Bench v2, OmniParseBench, DP-Bench, OCRFlux-bench (single page), Real5-OmniDocBench, PureDocBench, CC-OCR Doc Parsing, MDPBench, RealDoc-Bench, AmbiHTMLBench
Element parsing	Table to HTML; Table to Markdown; Filling a form; Chart to data; Chart to code; Flowchart to Mermaid; Formula to LaTeX	PubTabNet, FinTabNet, PulseBench-Tab
Ask anything	Questions on a page	DocVQA, InfographicVQA, ChartQA (test), FinChart-Bench, TabFact, TableVQA, TabMWP
Key-value extraction	Page to JSON; Form fields to key-values	CC-OCR KIE, SROIE, DeepForm, UNIKIE-Bench
Abstention	Knowing when not to answer	HallusionBench
OCR 3.0	OCR on request	None exists yet

2 Overall score

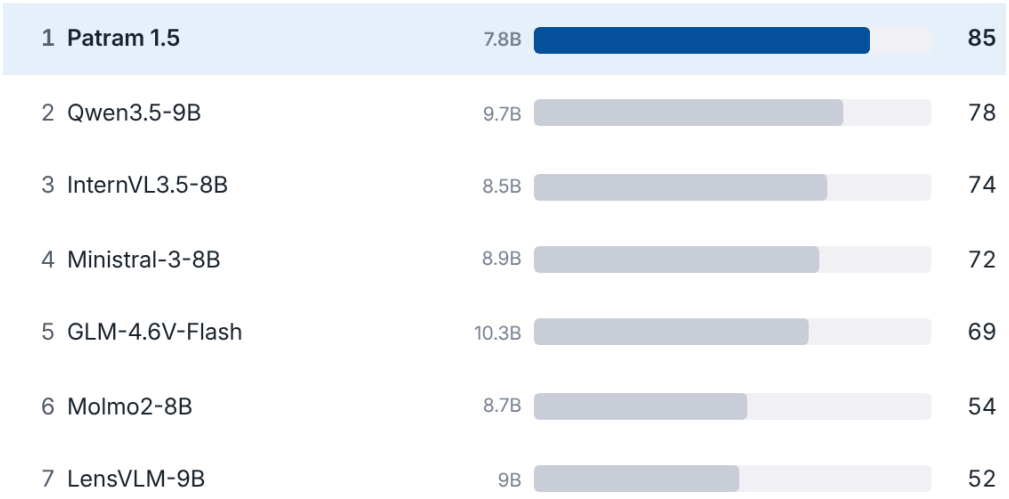
Overall score 21 models that read images · (PatramBench + public) / 2



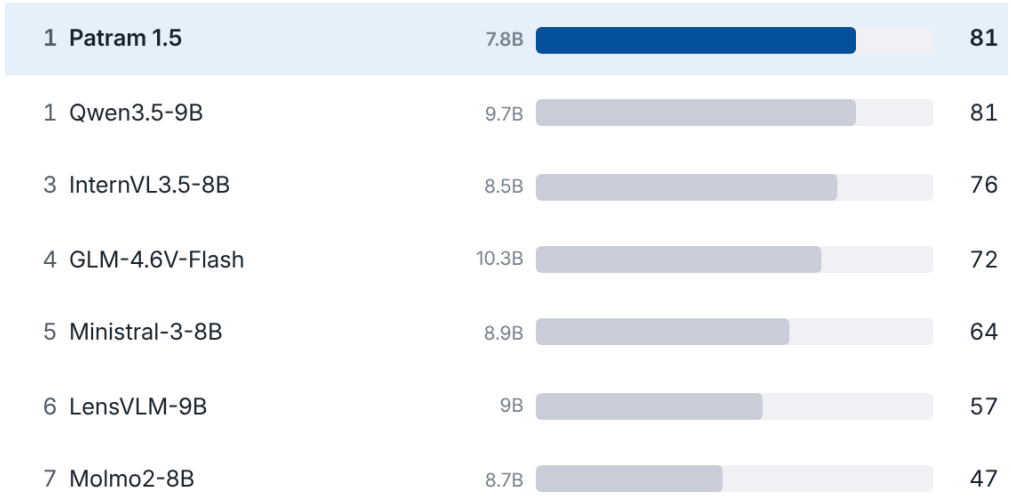
1	**Patram 1.5**	7.8B	**83**
1	DeepSeek-V4.1-Flash	763B	83
1	Qwen3.5-397B-A17B	403B	83
4	Gemma 4 31B	31B	82
4	Muse Glimmer 30B	29.8B	82
6	InternVL3.5-241B-A28B	241B	80
6	MiniMax-M3.1-Flash	428B	80
6	Nemotron-3.5-Super-VL	124B	80
9	Qwen3.5-9B	9.7B	79
10	GLM-5.3-Flash	321B	77
11	ERNIE-4.5-VL-424B-A47B	424B	76
12	InternVL3.5-8B	8.5B	75
13	Gemma 4 12B	12B	74
14	GLM-4.6V-Flash	10.3B	71
14	Mistral-Small-4-119B	119B	71
16	Kimi-VL-A3B-Thinking-2506	16.4B	70
17	Ministral-3-8B	8.9B	68
18	Phi-4-reasoning-vision-15B	15.1B	63
19	LensVLM-9B	9B	55
20	DeepSeek-VL2-small	16.1B	54
21	Molmo2-8B	8.7B	51

3 Models up to 10B

PatramBench models up to 10B



Public benchmarks models up to 10B



4 Models from 12B to 763B

PatramBench Patram 1.5 and the models from 12B to 763B

1	**Patram 1.5**	7.8B	**85**
1	GLM-5.3-Flash	321B	85
1	Gemma 4 31B	31B	85
1	Muse Glimmer 30B	29.8B	85
5	DeepSeek-V4.1-Flash	763B	84
5	MiniMax-M3.1-Flash	428B	84
5	Qwen3.5-397B-A17B	403B	84
8	InternVL3.5-241B-A28B	241B	80
9	Gemma 4 12B	12B	78
9	Nemotron-3.5-Super-VL	124B	78
11	ERNIE-4.5-VL-424B-A47B	424B	77
12	Mistral-Small-4-119B	119B	74
13	Kimi-VL-A3B-Thinking-2506	16.4B	70
14	Phi-4-reasoning-vision-15B	15.1B	65
15	DeepSeek-VL2-small	16.1B	47

Public benchmarks Patram 1.5 and the models from 12B to 763B

1	DeepSeek-V4.1-Flash	763B	82
1	Qwen3.5-397B-A17B	403B	82
3	**Patram 1.5**	7.8B	**81**
3	Nemotron-3.5-Super-VL	124B	81
5	Gemma 4 31B	31B	80
5	InternVL3.5-241B-A28B	241B	80
7	Muse Glimmer 30B	29.8B	79
8	MiniMax-M3.1-Flash	428B	77
9	ERNIE-4.5-VL-424B-A47B	424B	76
10	GLM-5.3-Flash	321B	69
10	Gemma 4 12B	12B	69
10	Kimi-VL-A3B-Thinking-2506	16.4B	69
10	Mistral-Small-4-119B	119B	69
14	Phi-4-reasoning-vision-15B	15.1B	61
15	DeepSeek-VL2-small	16.1B	60

5 OCR and parsing systems: parsing as a whole

PatramBench parsing tasks

1	**Patram 1.5**	7.8B	**88**
2	Chandra OCR 2	5.3B	77
3	Nanonets-OCR2-3B	3.8B	76
4	Nemotron-3.5-Parse	0.63B	74
5	DeepSeek-OCR-2	3.4B	73
5	dots.mocr	3B	73
7	Bodhan Indic OCR	undisclosed	63
8	OXR-1.0	1.5B	61

Public benchmarks parsing benchmarks

1	**Patram 1.5**	7.8B	**83**
2	Chandra OCR 2	5.3B	76
3	Nanonets-OCR2-3B	3.8B	75
4	Nemotron-3.5-Parse	0.63B	72
5	OXR-1.0	1.5B	69
5	dots.mocr	3B	69
7	DeepSeek-OCR-2	3.4B	68
8	Bodhan Indic OCR	undisclosed	59

OCR and parsing systems: by capability

Doc parsing · PatramBench

1	**Patram 1.5**	7.8B	**82**
2	Chandra OCR 2	5.3B	75
3	dots.mocr	3B	71
4	Nanonets-OCR2-3B	3.8B	68
5	DeepSeek-OCR-2	3.4B	67
5	Bodhan Indic OCR	undisclosed	67
7	OXR-1.0	1.5B	66
8	Nemotron-3.5-Parse	0.63B	63

Doc parsing · public

1	**Patram 1.5**	7.8B	**80**
2	dots.mocr	3B	78
3	OXR-1.0	1.5B	70
4	Chandra OCR 2	5.3B	69
4	Nanonets-OCR2-3B	3.8B	69
6	Nemotron-3.5-Parse	0.63B	65
7	DeepSeek-OCR-2	3.4B	61
8	Bodhan Indic OCR	undisclosed	58

Element parsing · PatramBench

1	**Patram 1.5**	7.8B	**94**
2	Nemotron-3.5-Parse	0.63B	85
3	Nanonets-OCR2-3B	3.8B	83
4	Chandra OCR 2	5.3B	79
4	DeepSeek-OCR-2	3.4B	79
6	dots.mocr	3B	74
7	Bodhan Indic OCR	undisclosed	59
8	OXR-1.0	1.5B	57

Element parsing · public

1	**Patram 1.5**	7.8B	**87**
2	Chandra OCR 2	5.3B	82
3	Nanonets-OCR2-3B	3.8B	80
4	Nemotron-3.5-Parse	0.63B	79
5	DeepSeek-OCR-2	3.4B	76
6	OXR-1.0	1.5B	67
7	dots.mocr	3B	61
7	Bodhan Indic OCR	undisclosed	61

OCR and parsing systems: by subtask

Text reading · PatramBench

1	Chandra OCR 2	5.3B		99
2	**Patram 1.5**	7.8B		**95**
3	dots.mocr	3B		93
4	Nanonets-OCR2-3B	3.8B		91
5	OXR-1.0	1.5B		88
6	Bodhan Indic OCR	undisclosed		87
7	DeepSeek-OCR-2	3.4B		86
8	Nemotron-3.5-Parse	0.63B		72

Text reading · public

1	**Patram 1.5**	7.8B		**78**
2	dots.mocr	3B		71
3	OXR-1.0	1.5B		70
4	Nanonets-OCR2-3B	3.8B		61
5	Nemotron-3.5-Parse	0.63B		53
6	Chandra OCR 2	5.3B		49
7	DeepSeek-OCR-2	3.4B		39
8	Bodhan Indic OCR	undisclosed		29

Page to Markdown and HTML · PatramBench

1	**Patram 1.5**	7.8B		**75**
2	Chandra OCR 2	5.3B		63
3	dots.mocr	3B		60
4	Nemotron-3.5-Parse	0.63B		59
5	DeepSeek-OCR-2	3.4B		57
5	Nanonets-OCR2-3B	3.8B		57
5	Bodhan Indic OCR	undisclosed		57
8	OXR-1.0	1.5B		54

Page to Markdown and HTML · public

1	dots.mocr	3B		82
2	**Patram 1.5**	7.8B		**81**
2	Chandra OCR 2	5.3B		81
4	Bodhan Indic OCR	undisclosed		75
5	DeepSeek-OCR-2	3.4B		74
5	Nanonets-OCR2-3B	3.8B		74
7	Nemotron-3.5-Parse	0.63B		72
8	OXR-1.0	1.5B		70

OCR and parsing systems: by subtask

Tables · PatramBench

1	**Patram 1.5**	7.8B	**93**
2	Nemotron-3.5-Parse	0.63B	90
3	Nanonets-OCR2-3B	3.8B	88
4	DeepSeek-OCR-2	3.4B	82
5	Chandra OCR 2	5.3B	81
6	Bodhan Indic OCR	undisclosed	79
7	dots.mocr	3B	65
8	OXR-1.0	1.5B	53

Tables · public

1	**Patram 1.5**	7.8B	**87**
2	Chandra OCR 2	5.3B	82
3	Nanonets-OCR2-3B	3.8B	80
4	Nemotron-3.5-Parse	0.63B	79
5	DeepSeek-OCR-2	3.4B	76
6	OXR-1.0	1.5B	67
7	dots.mocr	3B	61
7	Bodhan Indic OCR	undisclosed	61

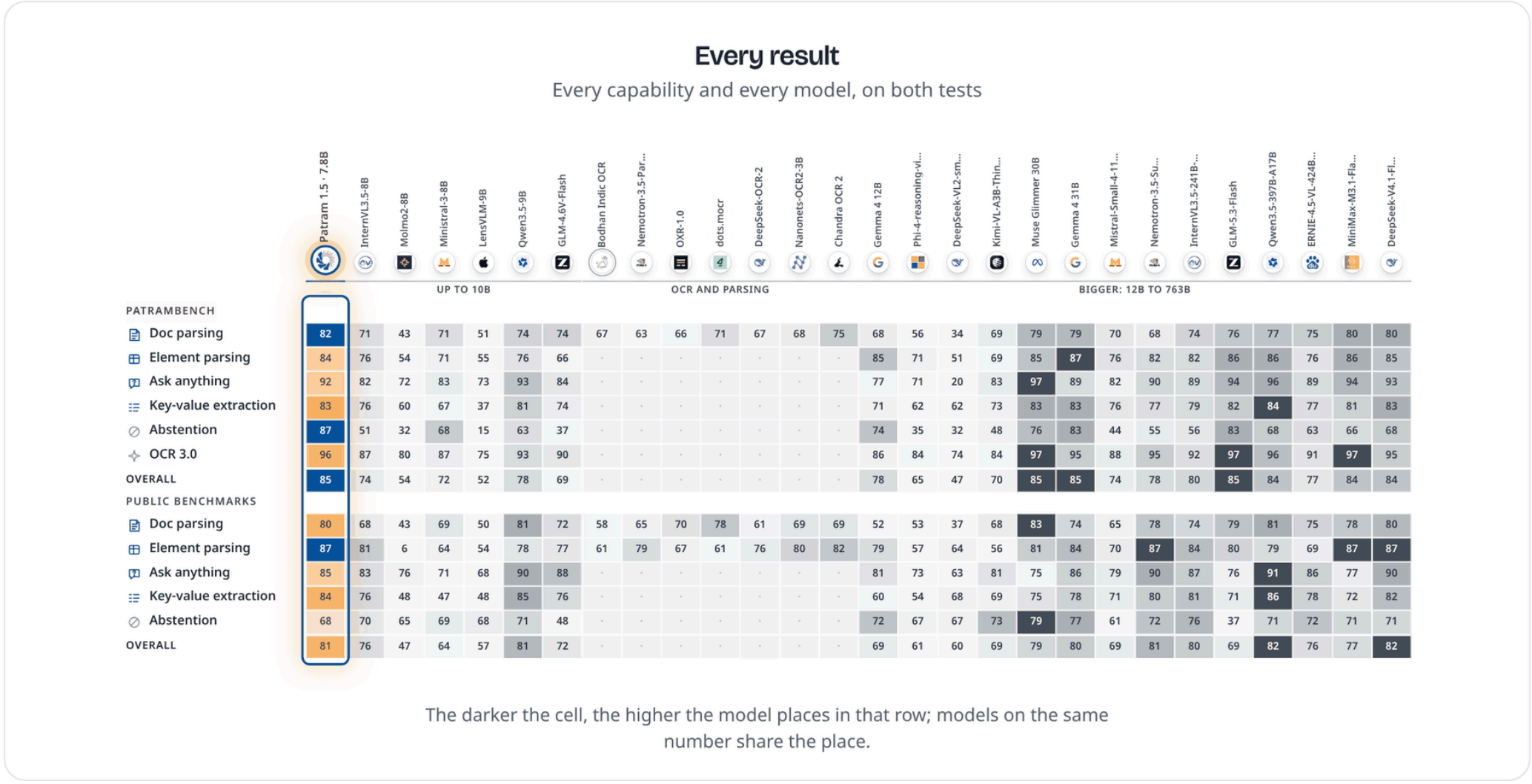
Formulas · PatramBench

no public benchmark yet

1	**Patram 1.5**	7.8B	**94**
2	dots.mocr	3B	92
3	Chandra OCR 2	5.3B	75
3	Nemotron-3.5-Parse	0.63B	75
5	Nanonets-OCR2-3B	3.8B	74
6	DeepSeek-OCR-2	3.4B	73
7	OXR-1.0	1.5B	65
8	Bodhan Indic OCR	undisclosed	21

6 Every result

Every model's overall score, PatramBench score and public score.



7 Every PatramBench task

The PatramBench tasks used in the evaluation, every model. The darker the cell, the higher the model places on that task.

		UP TO 10B						OCR AND PARSING							BIGGER: 12B TO 763B												
	Patram 1.5 7.8B	InternVL3.5-8B 8.5B	Molmo2-8B 8.7B	Ministral-3-8B 8.9B	LensVLM-9B 9B	Qwen3.5-9B 9.7B	GLM-4.6V-Flash 10.3B	Bodhan Indic OCR	Nemotron-3.5-Parse 0.63B	OXR-1.0 1.5B	dots.mocr 3B	DeepSeek-OCR-2 3.4B	Nanonets-OCR2-3B 3.8B	Chandra OCR 2 5.3B	Gemma 4 12B 12B	Phi-4-reasoning-vision-15B 15.1B	DeepSeek-VL2-small 16.1B	Kimi-VL-A3B-Thinking-2506 16.4B	Muse Glimmer 30B 29.8B	Gemma 4 31B 31B	Mistral-Small-4-119B 119B	Nemotron-3.5-Super-VL 124B	InternVL3.5-241B-A28B 241B	GLM-5.3-Flash 321B	Qwen3.5-397B-A17B 403B	ERNIE-4.5-VL-424B-A47B 424B	MiniMax-M3.1-Flash 428B	DeepSeek-V4.1-Flash 763B
DOC PARSING																													
Page OCR	95	89	75	89	64	92	91	87	72	88	93	86	91	99	82	82	82	89	96	93	88	87	92	95	96	93	95	95
Page to Markdown	91	80	54	82	59	86	83	77	83	74	83	77	80	80	77	71	1	76	88	88	83	84	84	88	89	87	89	90
Page to HTML	60	44	0	44	30	44	48	37	35	35	38	37	33	45	45	14	19	41	54	56	38	33	46	46	47	47	55	53
ELEMENT PARSING																													
Table to HTML	94	84	0	72	77	88	93	78	90	53	64	83	88	82	88	79	73	80	96	94	78	94	89	94	93	88	94	93
Table to Markdown	93	86	60	75	73	87	89	80	90	53	67	81	88	81	89	82	73	81	93	91	88	92	88	92	93	88	91	90
Chart to JSON	85	59	63	64	22	52	57	·	·	·	·	·	·	·	85	65	36	65	84	87	62	61	81	87	82	76	86	86
Chart to code	68	73	58	75	54	54	79	·	·	·	·	·	·	·	85	62	0	59	79	88	78	86	79	76	81	73	74	77
Flowchart to Mermaid	76	65	59	67	57	78	68	·	·	·	·	·	·	·	80	49	33	62	84	84	69	77	69	81	80	70	80	80
Form to JSON	78	75	62	59	16	76	66	·	·	·	·	·	·	·	73	67	58	64	77	85	73	76	76	83	78	71	81	75
Formula to LaTeX	94	90	77	88	83	94	8	21	75	65	92	73	74	75	95	95	87	69	83	81	81	90	94	93	92	62	92	95
ASK ANYTHING																													
Questions on documents	92	82	72	83	73	93	84	·	·	·	·	·	·	·	77	71	20	83	97	89	82	90	89	94	96	89	94	93
KEY-VALUE EXTRACTION																													
Page to JSON	79	75	56	67	33	80	69	·	·	·	·	·	·	·	74	65	59	72	82	82	75	73	79	82	83	77	81	81
Key-value pairs	87	77	64	68	41	83	80	·	·	·	·	·	·	·	67	59	65	73	85	83	77	81	80	83	85	77	81	84
ABSTENTION																													
When not to answer	87	51	32	68	15	63	37	·	·	·	·	·	·	·	74	35	32	48	76	83	44	55	56	83	68	63	66	68
OCR 3.0																													
Promptable OCR	96	87	80	87	75	93	90	·	·	·	·	·	·	·	86	84	74	84	97	95	88	95	92	97	96	91	97	95
OVERALL																													
Mean of the 15 tasks	85	74	54	72	52	78	69	·	·	·	·	·	·	·	78	65	47	70	85	85	74	78	80	85	84	77	84	84

The 34 public benchmarks

Each name links to the benchmark's official page, with the metric used. olmOCR-Bench v2 has no public page.

Doc parsing 19

OCRBench OCRBench score
TextOCR word F1
HierText word F1
Union14M word accuracy
WordArt word accuracy
CC-OCR Scene macro F1
IAM Handwriting character error, as a score
OmniDocBench overall
olmOCR-Bench overall
olmOCR-Bench v2 overall
OmniParseBench text and tables
DP-Bench NID
OCRFlux edit-distance similarity
CC-OCR Parsing score
MDPBench overall
PureDocBench overall
Real5-OmniDocBench overall
RealDoc-Bench field accuracy
AmbiHTMLBench objective score

Element parsing 3

PubTabNet TEDS
FinTabNet TEDS
PulseBench-Tab TLAG

Key-value extraction 4

CC-OCR KIE F1
SROIE F1 (task 3)
DeepForm ANLS
UNIKIE F1, mean over categories

Ask anything 7

DocVQA ANLS
InfographicVQA ANLS
ChartQA relaxed accuracy
FinChart-Bench accuracy
TabFact accuracy
TableVQA exact match
TabMWP accuracy

Abstention 1

HallusionBench accuracy

8 Every public benchmark: doc parsing

The darker the cell, the higher the model places on that benchmark. Every cell is our own run. Benchmarks marked † are shown without places.

		UP TO 10B						OCR AND PARSING							BIGGER: 12B TO 763B												
	Patram 1.5 7.8B	InternVL3.5-8B 8.5B	Molmo2-8B 8.7B	Ministral-3-8B 8.9B	LensVLM-9B 9B	Qwen3.5-9B 9.7B	GLM-4.6V-Flash 10.3B	*Bodhan Indic OCR*	*Nemotron-3.5-Parse* 0.63B	*OXR-1.0* 1.5B	*dots.mocr* 3B	*DeepSeek-OCR-2* 3.4B	*Nanonets-OCR2-3B* 3.8B	*Chandra OCR 2* 5.3B	Gemma 4 12B 12B	Phi-4-reasoning-vision-15B 15.1B	DeepSeek-VL2-small 16.1B	Kimi-VL-A3B-Thinking-2506 16.4B	Muse Glimmer 30B 29.8B	Gemma 4 31B 31B	Mistral-Small-4-119B 119B	Nemotron-3.5-Super-VL 124B	InternVL3.5-241B-A28B 241B	GLM-5.3-Flash 321B	Qwen3.5-397B-A17B 403B	ERNIE-4.5-VL-424B-A47B 424B	MiniMax-M3.1-Flash 428B	DeepSeek-V4.1-Flash 763B
DOC PARSING																												
OCRBench OCRBench score	87	84	81	78	72	87	87	41	68	77	73	49	87	70	81	76	82	86	88	89	72	86	89	90	88	85	81	80
TextOCR word F1	68	63	58	58	54	73	64	24	43	48	59	20	42	42	49	54	56	69	75	71	52	73	71	76	74	68	72	73
HierText word F1	63	62	47	65	45	71	66	35	55	49	68	32	49	60	41	50	48	69	77	71	60	68	70	79	73	70	77	76
Union14M word accuracy	71	54	59	58	57	83	83	1	28	72	58	18	48	8	71	64	69	66	73	80	37	74	73	86	84	79	65	76
WordArt word accuracy	80	64	75	71	76	83	83	2	32	76	74	22	52	21	78	76	82	78	81	82	44	81	72	84	82	79	69	75
CC-OCR Scene macro F1	80	62	38	54	67	82	77	38	55	72	67	46	50	66	47	69	72	66	75	74	62	76	69	47	83	48	46	52
IAM Handwriting character error, as a score	95	85	70	93	95	97	97	60	93	96	96	86	100	77	96	91	92	96	97	97	90	95	95	98	97	96	92	96
OmniDocBench overall	92	83	39	81	46	91	71	90	83	84	94	93	85	93	42	43	8	71	92	82	84	87	85	89	91	88	90	92
olmOCR-Bench overall	79	62	32	72	43	79	68	73	56	75	81	77	67	85	38	34	41	65	79	61	67	72	66	78	76	68	78	80
olmOCR-Bench v2 overall	80	61	30	71	42	77	66	73	56	75	80	78	68	87	38	33	42	64	78	62	66	70	66	78	75	67	76	80
OmniParseBench text and tables	78	47	3	48	26	58	73	73	73	64	84	67	65	87	14	16	17	35	81	53	51	78	56	78	64	59	77	84
DP-Bench NID	97	93	73	91	53	94	86	94	94	90	94	92	86	90	79	86	6	82	97	95	94	91	91	96	93	91	96	95
OCRFlux edit-distance similarity	90	80	51	77	43	85	84	84	78	83	91	91	78	86	63	71	11	76	90	87	82	85	82	90	86	82	90	87
CC-OCR Parsing score	72	73	24	67	44	75	66	69	69	53	73	70	71	38	63	56	41	63	83	76	57	76	73	71	76	72	73	78
MDPBench overall	76	64	32	66	32	82	35	74	81	67	84	65	76	85	29	41	3	48	85	61	68	78	·	70	82	72	82	84
PureDocBench overall	56	47	16	50	22	77	44	51	65	53	74	53	69	76	16	24	0	53	67	49	46	62	50	72	77	65	69	69
Real5-OmniDocBench overall	86	70	33	60	40	84	61	85	77	78	88	74	76	89	41	30	2	70	88	72	62	80	72	80	84	75	81	87
RealDoc-Bench field accuracy	86	73	44	79	42	87	80	75	73	61	79	61	84	80	38	53	5	71	88	75	79	85	79	86	88	80	88	89
AmbiHTMLBench objective score	75	70	9	67	52	67	71	59	64	61	63	64	60	72	62	37	36	63	78	75	65	61	71	59	69	74	76	76
Doc parsing, public score	80	68	43	69	50	81	72	58	65	70	78	61	69	69	52	53	37	68	83	74	65	78	74	79	81	75	78	80

Benchmarks on this page

OCRBench OCRBench score
TextOCR word F1
HierText word F1
Union14M word accuracy
WordArt word accuracy
CC-OCR Scene macro F1
IAM Handwriting character error, as a score
OmniDocBench overall
olmOCR-Bench overall
olmOCR-Bench v2 overall
OmniParseBench text and tables
DP-Bench NID
OCRFlux edit-distance similarity
CC-OCR Parsing score
MDPBench overall
PureDocBench overall
Real5-OmniDocBench overall
RealDoc-Bench field accuracy
AmbiHTMLBench objective score

Every public benchmark: element parsing and ask anything

	Patram 1.5 7.8B	UP TO 10B						OCR AND PARSING							BIGGER: 12B TO 763B												
		InternVL3.5-8B 8.5B	Molmo2-8B 8.7B	Ministral-3-8B 8.9B	LensVLM-9B 9B	Qwen3.5-9B 9.7B	GLM-4.6V-Flash 10.3B	*Bodhan Indic OCR*	*Nemotron-3.5-Parse* 0.63B	*OXR-1.0* 1.5B	*dots.mocr* 3B	*DeepSeek-OCR-2* 3.4B	*Nanonets-OCR2-3B* 3.8B	*Chandra OCR 2* 5.3B	Gemma 4 12B 12B	Phi-4-reasoning-vision-15B 15.1B	DeepSeek-VL2-small 16.1B	Kimi-VL-A3B-Thinking-2506 16.4B	Muse Glimmer 30B 29.8B	Gemma 4 31B 31B	Mistral-Small-4-119B 119B	Nemotron-3.5-Super-VL 124B	InternVL3.5-241B-A28B 241B	GLM-5.3-Flash 321B	Qwen3.5-397B-A17B 403B	ERNIE-4.5-VL-424B-A47B 424B	MiniMax-M3.1-Flash 428B	DeepSeek-V4.1-Flash 763B
ELEMENT PARSING																												
PubTabNet TEDS	89	84	7	72	50	80	82	68	83	91	65	88	86	91	87	64	73	61	84	90	78	90	87	76	81	78	87	93
FinTabNet TEDS	94	87	0	60	50	81	77	49	82	70	55	83	83	88	79	56	64	50	82	81	65	91	94	81	81	67	91	87
PulseBench-Tab TLAG	79	72	12	61	63	74	72	65	71	40	63	56	72	69	71	49	56	58	78	80	69	79	72	83	75	63	83	81
Element parsing, public score	87	81	6	64	54	78	77	61	79	67	61	76	80	82	79	57	64	56	81	84	70	87	84	80	79	69	87	87
ASK ANYTHING																												
DocVQA ANLS	95	92	85	61	79	96	96	·	·	·	·	·	·	·	80	81	88	89	74	88	86	95	94	66	96	94	72	95
InfographicVQA ANLS	84	73	65	37	50	84	84	·	·	·	·	·	·	·	49	57	65	75	45	66	53	86	77	26	86	78	41	83
ChartQA † relaxed accuracy	86	80	85	61	76	85	86	·	·	·	·	·	·	·	79	73	·	84	48	79	77	89	83	72	86	85	69	87
FinChart-Bench accuracy	83	84	79	83	76	88	84	·	·	·	·	·	·	·	86	76	75	81	84	88	80	89	86	87	88	80	84	88
TabFact accuracy	91	93	87	94	69	96	94	·	·	·	·	·	·	·	92	83	64	94	94	98	92	94	95	98	94	97	96	96
TableVQA † exact match	64	71	62	69	40	83	75	·	·	·	·	·	·	·	80	49	24	69	84	84	74	84	79	86	84	82	83	82
TabMWP † accuracy	95	89	66	94	82	100	97	·	·	·	·	·	·	·	99	92	64	76	99	100	91	95	92	100	100	85	95	98
Ask anything, public score	85	83	76	71	68	90	88	·	·	·	·	·	·	·	81	73	63	81	75	86	79	90	87	76	91	86	77	90

Benchmarks on this page

PubTabNet TEDS
FinTabNet TEDS
PulseBench-Tab TLAG
DocVQA ANLS
InfographicVQA ANLS
ChartQA relaxed accuracy
FinChart-Bench accuracy
TabFact accuracy
TableVQA exact match
TabMWP accuracy

Every public benchmark: key-value extraction, abstention and overall

		UP TO 10B						OCR AND PARSING							BIGGER: 12B TO 763B												
	Patram 1.5 7.8B	InternVL3.5-8B 8.5B	Molmo2-8B 8.7B	Ministral-3-8B 8.9B	LensVLM-9B 9B	Qwen3.5-9B 9.7B	GLM-4.6V-Flash 10.3B	*Bodhan Indic OCR*	*Nemotron-3.5-Parse* 0.63B	*OXR-1.0* 1.5B	*dots.mocr* 3B	*DeepSeek-OCR-2* 3.4B	*Nanonets-OCR2-3B* 3.8B	*Chandra OCR 2* 5.3B	Gemma 4 12B 12B	Phi-4-reasoning-vision-15B 15.1B	DeepSeek-VL2-small 16.1B	Kimi-VL-A3B-Thinking-2506 16.4B	Muse Glimmer 30B 29.8B	Gemma 4 31B 31B	Mistral-Small-4-119B 119B	Nemotron-3.5-Super-VL 124B	InternVL3.5-241B-A28B 241B	GLM-5.3-Flash 321B	Qwen3.5-397B-A17B 403B	ERNIE-4.5-VL-424B-A47B 424B	MiniMax-M3.1-Flash 428B	DeepSeek-V4.1-Flash 763B
KEY-VALUE EXTRACTION																												
CC-OCR KIE F1	80	74	39	44	40	86	72	·	·	·	·	·	·	·	64	49	58	61	77	75	61	76	78	78	87	71	75	79
SROIE F1 (task 3)	85	83	48	70	43	88	74	·	·	·	·	·	·	·	66	57	72	80	83	81	75	81	85	77	89	78	71	82
DeepForm † ANLS	93	81	67	29	63	88	94	·	·	·	·	·	·	·	54	70	86	75	61	83	88	89	91	53	89	94	71	93
UNIKIE F1, mean over categories	75	65	38	45	44	77	66	·	·	·	·	·	·	·	55	40	54	61	77	72	59	73	72	76	77	68	70	75
Key-value extraction, public score	84	76	48	47	48	85	76	·	·	·	·	·	·	·	60	54	68	69	75	78	71	80	81	71	86	78	72	82
ABSTENTION																												
HallusionBench accuracy	68	70	65	69	68	71	48	·	·	·	·	·	·	·	72	67	67	73	79	77	61	72	76	37	71	72	71	71
Abstention, public score	68	70	65	69	68	71	48	·	·	·	·	·	·	·	72	67	67	73	79	77	61	72	76	37	71	72	71	71
OVERALL																												
Public overall, mean of the 5 capabilities	81	76	47	64	57	81	72	·	·	·	·	·	·	·	69	61	60	69	79	80	69	81	80	69	82	76	77	82

Benchmarks on this page

CC-OCR KIE F1
SROIE F1 (task 3)
DeepForm ANLS
UNIKIE F1, mean over categories
HallusionBench accuracy