

In less than one quarter, open OCR went from "good enough text extraction" to a crowded race where multiple compact vision-language models now compete on parsing quality, structure fidelity, grounding, and cost.

If you are choosing an OCR stack in February 2026, the bottleneck is no longer finding a capable model. The bottleneck is choosing the model that fails least on your actual documents.

TL;DR Top models are now close on headline benchmarks, but they are not interchangeable in production. GLM-OCR and PaddleOCR-VL-1.5 currently lead reported OmniDocBench scores, while LightOnOCR-2 has a strong speed/quality profile and DeepSeek-OCR-2 remains attractive for markdown-oriented workflows. Start with a use-case-first shortlist, then run a fixed 50-page bake-off before rollout.

Update (Feb 16, 2026): rednote-hilab/dots.ocr-1.5 was released with new reported comparisons against GLM-OCR and PaddleOCR-VL-1.5. We now treat it as a serious challenger, especially for teams that need OCR plus broader image parsing (SVG, web screens, and scene text). For a dedicated breakdown, see: <https://instavar.com/blog/Dots OCR 1 5 vs GLM OCR vs PaddleOCR VL 1 5>.

If you are short on time:

1. Read Section 1 for the first model to test.
2. Read Section 4 for use-case fit.
3. Use Section 5 as your production evaluation checklist.

1 Start here: which model should you test first?

Your priority	Recommended first model to test	Why	Second model to test
Highest headline benchmark performance	GLM-OCR	Top reported OmniDocBench score among current open releases	PaddleOCR-VL-1.5
OCR plus broader image parsing	dots.ocr-1.5	Extends beyond document parsing	GLM-OCR

(SVG, web, scene text)		into multi-task vision-language parsing	
Best speed/quality for heavy page volume	LightOnOCR-2	Strong reported OlmOCR-Bench + throughput profile	DeepSeek-OCR-2
Grounding reliability for extraction workflows	GutenOCR	Grounded OCR is core design target	HunyuanOCR
Broad multilingual compact deployment	HunyuanOCR	1B-class model with broad task framing	GLM-OCR
Safest portfolio strategy	GLM-OCR + LightOnOCR-2 + one grounding specialist	Covers accuracy, efficiency, and grounding diversity	Add DeepSeek-OCR-2 for structured-mode fallback

2 What changed in OCR by February 2026

Three shifts define the current state:

- OCR has converged with compact VLM design, and in some workflows these models reduce or replace parts of multi-stage OCR pipelines.
- Benchmarks increasingly reward document-level understanding, not just line-level text extraction.
- Open releases now include practical deployment paths (vLLM, SGLang, Hugging Face, and in some cases Ollama), reducing integration friction.

3 What the benchmark evidence says (reported)

Before comparing model cards, keep three filters in mind:

- Compare only like-for-like benchmarks.
- Treat low-sample live leaderboard results as directional, not final.
- Validate on your own corpus before production promotion.

3.1 OmniDocBench snapshot

The table below consolidates reported OmniDocBench scores from model papers/cards, using v1.5 where explicitly stated.

Model	Params	OmniDocBench (reported)	Notes	Source
GLM-OCR	0.9B	94.62	Strong all-round reported score; very recent release	GLM-OCR repo
PaddleOCR-VL-1.5	0.9B	94.50	Competitive accuracy with compact footprint	PaddleOCR-VL paper
HunyuanOCR	1B	94.10	High reported score and broad task framing	HunyuanOCR report
DeepSeek-OCR-2	3B MoE decoder (~500M active) + 80M image compressor	91.09	Notable jump over earlier DeepSeek OCR baseline in authors' report	DeepSeek-OCR-2 paper

Interpretation:

- The top reported scores are now close enough that cost, failure mode, and licensing often matter more than a small benchmark gap.

3.2 OlmOCR-Bench snapshot

Reported from LightOnOCR-2 benchmarking (headers/footers excluded setting):

Model	OlmOCR-Bench (reported)	Context	Source
LightOnOCR-2-1B	83.2 +/- 0.9	Reported as strongest in this comparison slice	LightOnOCR-2 paper
Chandra-9B	81.7 +/- 0.9	Large model baseline in same evaluation	LightOnOCR-2 paper

olmOCR-2-8B 80.4 +/- 1.1 Strong open baseline with robust ecosystem support [LightOnOCR-2 paper](#)

Interpretation:

- A tuned 1B OCR model can match or exceed larger document VLM baselines on task-specific OCR benchmarks.

3.3 Throughput snapshot

Throughput matters because OCR cost scales with page volume, not request count.

Model	Throughput (pages/s, reported)	Hardware context	Source
LightOnOCR-2-1B	5.71	Single H100 context in authors' report	LightOnOCR-2 paper
DeepSeek-OCR family	Varies by mode and output format	Public demos emphasize extraction-mode trade-offs	DeepSeek-OCR-2
GLM-OCR	Deployment-oriented serving options published; no single canonical throughput figure in repo	Depends on serving stack (vLLM, SGLang, Ollama)	GLM-OCR repo

3.4 dots.ocr-1.5 early evidence snapshot (author-reported)

The dots.ocr-1.5 release adds important new evidence, but most of it currently comes from the model card/repo evaluation stack.

Signal	Reported value	Source
Elo on olmOCR-Bench	1089.0	dots.ocr-1.5 model card
Elo on OmniDocBench (v1.5)	1025.8	dots.ocr-1.5 model card
Elo on XDocParse	1157.1	dots.ocr-1.5 model card

OmniDocBench (v1.5) TextEdit	0.031 (lower is better)	dots.ocr-1.5 model card
OmniDocBench (v1.5) ReadOrderEdit	0.029 (lower is better)	dots.ocr-1.5 model card

Interpretation:

- Useful new signal, but evaluate carefully.
- The Elo setup is judged by Gemini 3 Flash in the authors' pipeline and is not a drop-in replacement for independent leaderboard results.

4 Model fit by use case

4.1 Use-case fit matrix

Model	Choose first when	Why it wins there	Watch-outs
GLM-OCR	You need a strong default baseline across mixed documents	Top-tier reported OmniDocBench result in compact size; multiple serving paths	Very new release; long-tail behavior still needs broad replication
dots.ocr-1.5	You need one model for OCR plus web/screen/scene/SVG parsing	Broad task coverage in a single 3B model family and strong reported release benchmarks	Many benchmark claims are currently model-card/repo reported for this version
DeepSeek-OCR-2	You need markdown-oriented output and mode switching	Reading-order-focused design and dual extraction modes (Free OCR and structured conversion)	Validate complex tables and multilingual edge cases on your own corpus
LightOnOCR-2-1B	You process high page volume and care about cost per page	Strong reported OlmOCR-Bench + throughput profile at 1B scale	Check performance on your language/script distribution
GutenOCR	You need reliable text-to-location grounding for	Grounded OCR is core design	Weight license is CC-BY-NC;

	downstream extraction	objective and first-class output	commercial use may be constrained
HunyuanOCR	You want one compact model for broad document tasks	Strong reported compact-model results across parsing-oriented tasks	Custom community license requires legal/compliance review
PaddleOCR-VL-1.5	Your inputs are messy scans/photos and you already run Paddle tooling	Near-frontier reported OmniDocBench score with robustness framing	Confirm accuracy on your distortion mix and template families

4.2 Adoption and maturity signals (Feb 13, 2026 snapshot)

These are not quality scores. They are practical signals for implementation confidence and ecosystem support.

Model	Maturity signal	What it means for rollout
GLM-OCR	Rapid early GitHub/HF uptake after launch	Fast-moving ecosystem, but still early for stability assumptions
dots.ocr-1.5	Fresh Feb 16, 2026 release with expanded task scope	High upside for multi-task use cases, but treat current results as early-cycle evidence
DeepSeek-OCR-2	Strong HF traction soon after release	Good community momentum for tooling and examples
HunyuanOCR	High visibility and broad activity across channels	More examples in the wild for compact deployment patterns
GutenOCR	Growing technical interest from doc-AI builders	Strong relevance for grounding-heavy extraction workflows
LightOnOCR-2-1B	Attention driven by 1B speed/quality profile	Good candidate for throughput-first deployments
PaddleOCR-VL-1.5	Benchmark-competitive and aligned with Paddle stack users	Lower integration risk if your team already uses Paddle

5 A practical evaluation protocol (6 core models + challengers)

If you want one rigorous, reproducible process, run one fixed 50-page bake-off across six core models:

1. GutenOCR
2. HunyuanOCR
3. LightOnOCR-2-1B
4. DeepSeek-OCR-2
5. GLM-OCR
6. PaddleOCR-VL-1.5

Then add challenger tracks for newly released models. For this cycle:

7. dots.ocr-1.5 (especially if you need OCR plus web/screen/scene/SVG parsing)

5.1 Preflight gates (before benchmarking)

Filter models before inference:

- License/commercial gate
- Region/compliance gate
- Serving/runtime gate
- Output-format gate

Practical note:

- GutenOCR weights are CC-BY-NC, which often disqualifies direct commercial deployment.
- HunyuanOCR uses a custom community license with territory and usage constraints, so legal review should happen before production rollout.

5.2 50-page stratified set

Slice	Pages	Why this slice matters
Clean digital single-column PDFs	8	Baseline text fidelity
Multi-column + sidebars + footnotes	8	Reading-order stress
Table-heavy documents	8	Structure fidelity and cell ordering

Formula-heavy documents	6	Formula extraction and sequencing
Forms/invoices/receipts	6	Region association and key-value linking
Messy photos/scans	10	Skew, warping, glare, and capture artifacts
Multilingual mixed-script pages	4	Language/layout stability

Total: 50 pages.

5.3 Ground-truth package per page

Prepare three artifacts for each page:

- gt_text.txt
- gt_markdown.md
- gt_blocks.json with block_id, text, bbox, reading_index, and type

Quality control:

- Dual-annotate all messy-photo pages and all multi-column pages.
- Resolve disagreements before scoring.

5.4 Inference protocol (same policy for all models)

- Render all pages at one fixed resolution (for example, 200 DPI).
- Use deterministic decoding (temperature=0, no retries in the primary run).
- Freeze model versions/commits and prompts.
- Run one no-heuristic primary pass; report heuristic retries separately if used.

Mode recommendations:

- DeepSeek-OCR-2: run both Free OCR and markdown conversion mode.
- GLM-OCR: run markdown plus JSON layout output.
- PaddleOCR-VL-1.5: run full document parsing mode.
- dots.ocr-1.5: run document parsing mode first; if relevant, add web parsing and scene spotting prompts, and evaluate SVG output in a separate track.
- GutenOCR: run grounded mode (bbox outputs) and plain text mode.
- HunyuanOCR: run document parsing prompt and spotting-style prompt where applicable.
- LightOnOCR-2-1B: run standard OCR parsing mode.

5.5 Metrics

Reading-order metrics:

- RO-ED (normalized reading-order edit distance, lower is better)
- Kendall tau ON reading_index sequence (higher is better)
- Missing-block rate
- Duplicate-block rate

Content and structure metrics:

- CER and WER
- Table TEDS
- Formula metric (CDM or token-level F1, fixed across all runs)

Operational metrics:

- p50 and p95 latency per page
- Throughput (pages/sec)
- Parse failure rate (empty output, truncation, repetition loop)

5.6 Pass/fail thresholds

A model passes production gate only if all thresholds are met:

- Overall RO-ED ≤ 0.12
- Complex-layout RO-ED ≤ 0.16 (multi-column + table-heavy + messy slices)
- Overall Kendall tau ≥ 0.88
- Complex-layout Kendall tau ≥ 0.80
- Overall CER $\leq 6\%$
- Clean-slice CER $\leq 3\%$
- Messy-slice CER $\leq 10\%$
- Missing-block rate $\leq 5\%$
- Duplicate-block rate $\leq 2\%$
- Parse failure rate $\leq 2\%$
- Meets your explicit p95 latency SLA

5.7 Weighted winner rule

If multiple models pass, rank by weighted score:

- 45% reading-order composite
- 30% text fidelity

- 15% structure fidelity (tables and formulas)
- 10% operational efficiency and reliability

Declare a winner only when:

- Weighted score gap is at least 3 points, and
- The gap is stable under paired bootstrap confidence checks on primary metrics (RO-ED and/or CER).

5.8 Deployment outcome template

After scoring, select:

- One primary model (best weighted score with all gates passed)
- One fallback model (best performer on the slice where the primary is weakest)
- One router policy:

clean digital docs -> primary messy photos/scans -> robustness leader ordering-critical complex layouts -> reading-order leader

6 Risks and caveats before production rollout

- Benchmark overfitting risk: do not promote a model to primary production without document-type stratified tests.
- Layout drift risk: table structure quality can degrade faster than plain text quality across new templates.
- Grounding risk: extraction pipelines fail when text is correct but linked to the wrong box or wrong row.
- License risk: confirm commercial terms for each model/repo combination, not just the model card headline.
- Operations risk: define fallback modes (text-only, markdown, or dual-model checks) before first rollout.

7 Conclusion

The OCR frontier in February 2026 is no longer dominated by one giant model. It is a tight, fast-moving cluster of compact document VLMs that each optimize a different part of the real production equation.

If your team wants one practical rollout path:

1. Start with the use-case matrix in Section 1.
2. Shortlist three models with different strengths.
3. Run the fixed 50-page protocol.
4. Promote one primary model and one fallback model, then keep one challenger lane for fast releases like dots.ocr-1.5.

That gives you better reliability than chasing a single leaderboard winner in a market where rankings can shift quickly.

Sources

- [OCR Arena](#)
- [CodeOCR: On the Effectiveness of Vision Language Models in Code Understanding \(Feb 2026\)](#)
- [GLM-OCR website](#)
- [GLM-OCR GitHub](#)
- [GLM-OCR model card](#)
- [Chandra model card](#)
- [DeepSeek-OCR-2 paper](#)
- [DeepSeek-OCR-2 model card](#)
- [GutenOCR paper](#)
- [GutenOCR repository](#)
- [GutenOCR-3B](#)
- [GutenOCR-7B](#)
- [GutenOCR demo](#)
- [LightOnOCR-2 paper](#)
- [LightOnOCR / Chandra repository](#)
- [LightOnOCR-2 announcement](#)
- [HunyuanOCR repository](#)
- [HunyuanOCR technical report](#)
- [HunyuanOCR model card](#)
- [dots.ocr repository](#)
- [dots.ocr-1.5 model card](#)
- [dots.ocr paper \(base model\)](#)
- [PaddleOCR-VL-1.5 paper](#)
- [PaddleOCR repository](#)
- [Mistral OCR](#)
- [olmOCR 2 blog](#)

- [LlamaIndex analysis of olmOCR benchmark pitfalls](#)
- [Hugging Face open OCR models overview](#)