60-second takeaway

DeepSeek OCR-2 is a solid markdown-oriented OCR with the best blank-page detection in our benchmark. But it is the slowest model (14.8 s/page) and its aggregate CER (39.34%) puts it 4th of 5 models. Use it when blank detection or grounded output matters; avoid it as a general-purpose default.

Where this fits

- **For founders:** DeepSeek OCR-2 has a niche role in pipelines that need reliable blank detection and grounded coordinates. It is not your first-choice general OCR. If you are building a multi-model pipeline, it earns a lane - but a narrow one.
- **For engineers:** Use this page to understand where DeepSeek fits in a multi-model routing pipeline, what its failure modes look like on degraded scans, and when to route pages away from it.

For the full workflow routing guide across all promoted models, see: [Which OCR Model Fits Which Workflow in 2026](#).

For the wider leaderboard and shortlist builder, see: [OCR Model Leaderboard 2026](#).

1 What DeepSeek OCR-2 is

DeepSeek OCR-2 is a vision-language model (VLM) from DeepSeek designed for document OCR. Like several other models in the current open-source OCR wave, it combines a vision encoder with a language decoder to produce structured text output from page images.

Two properties set it apart from the other models in our benchmark:

- **Markdown-oriented output.** DeepSeek produces clean, structured markdown by default. Headings, lists, and tables come through with reasonable fidelity, which reduces downstream cleanup cost when the output feeds into a markdown-native pipeline.
- **Grounded output with bounding boxes.** DeepSeek supports coordinate grounding - it can return bounding box positions alongside extracted text.

In our benchmark, it was the second-best grounded workflow after Hunyuan.

2 Benchmark scores vs reality

The aggregate numbers for DeepSeek look mediocre at first glance:

Metric	DeepSeek OCR-2
Average CER	39.34%
Average WER	33.39%
Pages processed	49
Speed	14.8 s/page

A 39.34% CER puts DeepSeek 4th out of 5 models in our full-50 benchmark. That is not a competitive aggregate number.

But the aggregate hides the interesting finding. DeepSeek's errors are not evenly distributed across page types - they are concentrated in specific archetypes where it is a poor fit, while it performs competitively on others.

Per-archetype CER breakdown

Archetype	Pages	DeepSeek CER	Best model	Best CER
text_first_notes	10	8.3%	Qianfan	5.9%
diagram_question	10	30.2%	GLM	6.1%
formula_heavy	8	76.6%	Qianfan	20.7%
table_heavy	8	43.8%	Qianfan	15.7%
worksheet_options	8	46.5%	Qianfan	7.1%
low_contrast_or_faint_scan	3	69.1%	Qianfan	0.0%
blank_or_near_blank	2	0.0%	DeepSeek/Hunyuan/Qianfan	0.0%

The 8.3% CER on text-first notes is within striking distance of the best model. The 0.0% on blank pages is perfect. Everything else is a weakness.

For the benchmarking method behind these numbers, see: [How We Benchmark OCR Models on Scan-Heavy PDFs](#).

3 Where it excels

3.1 Blank page detection

DeepSeek detected all 3 blank pages in our 50-page corpus with 0.0% CER. This sounds trivial, but blank detection is a real pipeline hygiene problem. Models that hallucinate text on blank pages inject noise into downstream processing, create phantom entries in document indices, and waste compute on non-content.

GLM, for comparison, failed all 3 blank pages - it hallucinated content on every one.

Only three models in our benchmark achieved 3/3 blank detection: DeepSeek, Hunyuan, and Qianfan.

3.2 Text-first notes

On the text_first_notes archetype (10 pages of handwritten and printed notes, bullets, and worked answers), DeepSeek scored 8.3% CER. Qianfan led at 5.9%, but DeepSeek's result is competitive and usable in production without heavy post-processing.

3.3 Grounded output

DeepSeek is the second-best grounded workflow in our benchmark, after Hunyuan. If your pipeline needs bounding box coordinates alongside extracted text - for anchor overlays, region-level confidence scoring, or spatial search - DeepSeek is one of only two models that deliver this reliably.

3.4 Markdown output quality

The markdown DeepSeek produces is structurally clean. Tables render correctly, heading hierarchy is preserved, and list formatting is consistent. For pipelines that consume markdown directly (rendering, indexing, downstream LLM input), this reduces the cleanup step.

4 Where it fails

4.1 Low-contrast scans

CER of 69.1% on low-contrast and faint scans. Qianfan gets 0.0% on the same pages. Hunyuan gets 6.6%. DeepSeek's performance on degraded input is not competitive.

If your corpus includes photocopied worksheets, faded thermal prints, or low-DPI scans, route these pages away from DeepSeek.

4.2 Worksheets

46.5% CER on worksheet_options pages (multiple-choice layouts, grid-style answers). Qianfan gets 7.1% on the same pages. The gap is large enough that DeepSeek should not be used as the primary model for worksheet-heavy documents.

4.3 Formulas

76.6% CER on formula_heavy pages. This is the worst result among DeepSeek's archetypes. Qianfan leads at 20.7%, and even that is not great - formula OCR remains hard across all models. But DeepSeek's result here is unusable without significant post-correction.

4.4 Diagrams

30.2% CER on diagram_question pages, where GLM leads at 6.1%. DeepSeek does not preserve question-local visuals or diagram-linked text as reliably as GLM.

5 Speed and cost

At 14.8 seconds per page, DeepSeek is the slowest model in our benchmark by a wide margin.

Model	Speed
GLM	0.9 s/page
DeepSeek	14.8 s/page

GLM is roughly 16x faster. A 1,000-page document takes DeepSeek over 4 hours. The same document takes GLM about 15 minutes.

This makes DeepSeek impractical for high-volume processing. It is only viable for small-batch workflows or quality-critical lanes where its specific strengths (blank detection, grounding) justify the latency cost.

6 Comparison: DeepSeek vs GLM vs Qianfan

These are the three models most commonly compared for markdown-oriented OCR in our pipeline. Here is the head-to-head:

Metric	DeepSeek	GLM	Qianfan
Avg CER	39.34%	33.84%	12.80%
Avg WER	33.39%	27.59%	13.18%
Speed	14.8 s/page	0.9 s/page	N/A
Blank detection	3/3 (100%)	0/3 (failed)	3/3 (100%)
Best archetype	blank_or_near_blank, text notes	diagram_question	5 of 7 archetypes

Qianfan wins 5 of 7 archetypes outright. GLM wins diagrams and is dramatically faster. DeepSeek's wins are narrower: blank detection (shared with Qianfan) and grounded output (second to Hunyuan).

The honest read: if you do not need grounded output or bounding box coordinates, Qianfan is the stronger default and GLM is the faster one. DeepSeek earns its lane only when its specific strengths matter to your pipeline.

7 When to use DeepSeek OCR-2

Route to DeepSeek when:

- **Blank detection matters.** If your pipeline processes mixed documents with intermittent blank or near-blank pages, DeepSeek's 3/3 detection prevents phantom entries.
- **You need grounded coordinates.** For anchor overlays, spatial search, or region-level extraction, DeepSeek is the second-best grounded workflow.
- **The pages are text-first.** On notes, bullets, and worked answers, DeepSeek's 8.3% CER is competitive.

Route away from DeepSeek when:

- **Formulas are present.** 76.6% CER is not salvageable.
- **Worksheets dominate.** 46.5% CER - use Qianfan instead.

- **Scans are degraded.** Low-contrast, faded, or faint pages break DeepSeek badly.
- **Volume is high.** 14.8 s/page makes it impractical for batch processing above a few hundred pages.

Default alternative

Qianfan wins 5 of 7 archetypes in our benchmark and has competitive blank detection. If you need a single markdown-oriented OCR model and do not require grounded output, start with Qianfan.

For the full multi-model routing recommendation, see: [Which OCR Model Fits Which Workflow in 2026](#).

8 FAQ

Is DeepSeek OCR-2 the best open-source OCR model?

No. In our 50-page benchmark, it places 4th of 5 models by aggregate CER. It has specific strengths (blank detection, grounded output, text notes) but is not the strongest general-purpose choice.

Can I use DeepSeek OCR-2 as my only OCR model?

You can, but you will get poor results on formulas, worksheets, and low-contrast scans. A multi-model routing pipeline that sends different page types to different models will outperform any single-model deployment.

How does DeepSeek compare to Hunyuan for grounded output?

Hunyuan is the strongest grounded workflow in our benchmark. DeepSeek is second. If grounded output is your primary requirement, test Hunyuan first. If you also need blank detection, DeepSeek adds value as a complementary lane.

Why is DeepSeek so slow?

At 14.8 s/page, DeepSeek is the slowest model we benchmarked. The exact cause is architectural - larger model size and inference cost per page. For comparison, GLM processes pages at 0.9 s/page (16x faster).

Should I use DeepSeek or Qianfan for markdown OCR?

Qianfan wins 5 of 7 archetypes and has lower aggregate CER (12.80% vs 39.34%). Unless you specifically need grounded output or bounding box coordinates, Qianfan is the better markdown-oriented default.

9 Sources and further reading

- **OCR workflow routing guide:** [Which OCR Model Fits Which Workflow in 2026](#)
- **OCR model leaderboard:** [OCR Model Leaderboard 2026](#)
- **Scan-heavy benchmark method:** [How We Benchmark OCR Models on Scan-Heavy PDFs](#)
- **Hunyuan vs FireRed comparison:** [Hunyuan OCR vs FireRed OCR 2026](#)
- **Scanned PDFs practical guide:** [OCR for Scanned PDFs - Practical Guide 2026](#)