



Thai Document OCR v3.0: Accuracy Against Google Cloud Vision on Thai Documents

Service version v3.0.0-20261005 · Benchmarks run 4–5 October 2026 · iApp Technology Co., Ltd.,
Bangkok, Thailand

Executive summary. Since 5 October 2026 the iApp Thai Document OCR API reads pages with iApp’s own Thai recognition model. We compared it with Google Cloud Vision on the same page images and the same checked answers: 2,825 Thai pages from eleven sets, and 150 scanned Thai Administrative Court judgments. Measured as character error rate on the text in reading order, which is the text a customer receives, iApp scored 29.6% against 36.0% over all pages (difference –6.4 points, 95% CI [–6.9, –5.9]), and was better on 1,654 pages against 425. The gap is widest on government documents: 17.2% against 37.4% on parliamentary budget documents, and better on 217 of 221 Royal Gazette pages. On the 242-page court benchmark of WP-2026-01, character accuracy rose from 94.8% to 96.3% (Google Cloud Vision: 93.9%). Google Cloud Vision remains slightly ahead on short photographed text. The service runs on two servers and sustains about 27,000 pages of plain text an hour.

1 Introduction

The iApp Thai Document OCR API converts printed Thai and English documents — images, PDF files and Microsoft Office files — into plain text, layout-structured JSON, or generated Word, PowerPoint, Excel and searchable PDF files. Version 3.0.0, in production since 5 October 2026, replaces the recognition engine. The endpoints, the request format and the response format are unchanged; customers do not need to change any code.

Google Cloud Vision is the OCR service most Thai developers use for documents, and it was the cloud baseline of our previous benchmark, WP-2026-01 [1]. This paper asks one question: on Thai documents, which system returns the more accurate text? It reports every set we ran, including those where Google Cloud Vision is better, with paired confidence intervals, and states the limits of the evidence in Section 9.

2 Systems Compared

Google Cloud Vision

The `document_text_detection` feature with language hints `th` and `en`, given JPEG images at quality 95. This is the configuration iApp’s own service used when it called Google Cloud Vision. The model is Google’s as of 4 October 2026; Google Cloud Vision has no reading-order option.

iApp v3.0.0

The production service since 5 October 2026: iApp’s own Thai line-recognition model. When the model reports that it cannot read a line cleanly, the line is re-read by a secondary recognition engine running on the same servers. The share of pages that used it ranged from none on government documents to 3–6% on mixed documents and old books, and more on photographs and handwriting (Appendix A).

iApp v2.1.13

The previous production engine, in service until 4 October 2026, shown for reference: it is what WP-2026-01 measured.

Every engine received the same pixels for every page, with the longer side capped at 4,000 pixels.

3 Benchmark Data

Table 1: The benchmark sets. Pages are those with a usable checked answer; 420 pages without one were left out for every engine.

Set	Pages	Source and ground truth
Royal Gazette	221	Thai government gazette pages; ground truth from the publication’s PDF text layer
Royal Gazette, verified	54	A subset whose ground truth was checked by hand
Budget documents	539	Thai parliamentary budget documents, largely two-column tables; PDF text layer
Budget documents, verified	74	A hand-checked subset
SEA-DocBench (Thai)	1,225	The Thai pages of a public document-understanding benchmark
Old books (Wikisource)	608	Scanned old Thai books with proofread, validated transcriptions
OCR-104 (five sets)	104	Documents, Thai and English, handwriting, document photographs and scene text; 12–24 images each
Court judgments	150 docs	Scanned Thai Administrative Court judgments; ground truth from the court’s own DOC files

None of these pages or documents was used to train the iApp model. The model was, however, trained on *other* judgments of the same court, so the court sets are within its training domain; we report them separately for that reason.

4 Method

4.1 Metrics

- **Ordered character error rate (CER).** The edit distance between the engine’s text, in its reading order, and the ground truth, divided by the length of the ground truth. This is the measure of the text a customer receives. Lower is better.
- **Order-free character error rate.** A bag-of-characters error with spaces ignored. It measures character recognition alone, without reading order.

Both metrics use the same text normalisation for every engine. On the court benchmark of WP-2026-01 we also report word error rate and ROUGE-L, as that paper did.

4.2 Statistics

Engines are compared on the same pages. For each set we report the mean difference in CER (iApp v3.0.0 minus Google Cloud Vision) with a seeded bootstrap 95% confidence interval, and count the pages each engine reads better; a difference smaller than 0.5 points counts as a tie.

5 Results

5.1 Text in reading order

Table 2: Ordered CER (%), lower is better. The best of the three engines on each set is in bold. Δ is iApp v3.0.0 minus Google Cloud Vision, in points, with its 95% CI; the last column counts pages where iApp is better / Google is better / tie.

Set	Pages	Google	iApp v2.1.13	iApp v3.0.0	Δ [95% CI]	iApp / Google / tie
Royal Gazette, verified	54	4.1	1.9	2.2	-1.9 [-2.9, -1.1]	36 / 11 / 7
Budget documents, verified	74	24.9	0.1	0.5	-24.5 [-28.1, -20.8]	67 / 1 / 6
Royal Gazette	221	13.8	5.8	6.0	-7.8 [-9.1, -6.6]	217 / 0 / 4
Budget documents	539	37.4	16.1	17.2	-20.2 [-21.8, -18.7]	469 / 25 / 45
SEA-DocBench (Thai)	1,225	49.4	51.2	48.7	-0.7 [-1.3, -0.1]	443 / 243 / 539
Old books (Wikisource)	608	25.4	20.2	21.0	-4.4 [-5.1, -3.7]	413 / 124 / 71
OCR-104 documents	22	0.8	16.0	2.8	+1.9 [+0.0, +5.4]	0 / 2 / 20
OCR-104 Thai and English	12	5.1	17.7	13.9	+8.8 [-6.8, +29.4]	2 / 3 / 7
OCR-104 handwriting	22	7.6	7.1	7.4	-0.1 [-10.2, +8.7]	4 / 8 / 10
OCR-104 document photos	24	6.5	25.6	1.2	-5.4 [-16.2, +1.7]	2 / 4 / 18
OCR-104 scene text	24	2.4	13.9	3.9	+1.5 [-1.5, +4.4]	1 / 4 / 19
All pages	2,825	36.0	30.7	29.6	-6.4 [-6.9, -5.9]	1,654 / 425 / 746
Mean of the 11 sets		16.1	16.0	11.3	[-6.9, -2.5]	

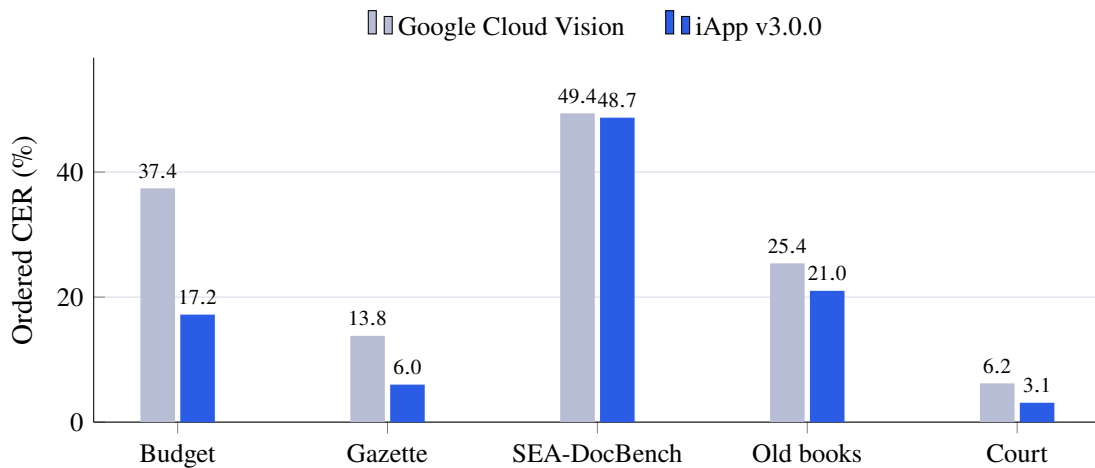


Figure 1: Ordered CER on the four largest page sets and the 150 court judgments.

5.2 Recognition without reading order

Table 3: Order-free CER (%), recognition only. Δ is iApp v3.0.0 minus Google Cloud Vision, with its 95% CI.

Set	Pages	Google	iApp v2.1.13	iApp v3.0.0	Δ [95% CI]
Royal Gazette, verified	54	0.18	0.03	0.37	+0.19 [+0.12, +0.26]
Budget documents, verified	74	0.37	0.09	0.38	+0.02 [−0.11, +0.13]
Royal Gazette	221	6.08	5.85	6.04	−0.04 [−0.10, +0.02]
Budget documents	539	6.23	5.83	6.14	−0.09 [−0.16, −0.02]
SEA-DocBench (Thai)	1,225	15.27	14.64	13.92	−1.34 [−1.80, −0.91]
Old books (Wikisource)	608	17.14	15.82	16.18	−0.96 [−1.44, −0.56]
OCR-104 documents	22	0.93	16.22	3.97	+3.04 [+0.00, +7.44]
OCR-104 Thai and English	12	10.31	24.76	17.12	+6.81 [−15.16, +30.77]
OCR-104 handwriting	22	9.99	8.18	9.70	−0.30 [−12.32, +9.40]
OCR-104 document photos	24	6.55	25.94	2.19	−4.36 [−15.48, +3.37]
OCR-104 scene text	24	3.66	15.21	5.75	+2.09 [−2.17, +7.22]
All pages	2,825	12.2	12.0	11.4	
Mean of the 11 sets		7.0	12.1	7.4	

5.3 Court judgments

Table 4: 150 scanned Administrative Court judgments, ordered CER (%) per document. This set was scored with the iApp model alone, without the secondary engine.

Engine	Mean	Median	Documents better than Google
Google Cloud Vision	6.17	5.50	
iApp v2.1.13	4.98	4.38	131 of 150 (19 ties, none worse)
iApp v3.0.0 model	3.12	2.53	150 of 150; Δ −3.05 [−3.19, −2.92]

The benchmark of WP-2026-01 — 20 judgments, 242 pages at 200 dpi, sampled with seed 1337 from the public *iApp Thai Administrative Court Judgment OCR* dataset, scored against the court’s DOC files with the same script — was run again on the production service on 5 October 2026. The Google Cloud Vision and iApp v2.1.13 figures are those of August 2026.

Table 5: The WP-2026-01 court benchmark (20 documents, 242 pages), measured through the production service.

Engine	CER mean (%)	CER median (%)	WER mean (%)	ROUGE-L
Google Cloud Vision (Aug 2026)	6.09	5.04	7.52	0.9576
iApp v2.1.13 (Aug 2026)	5.18	4.32	7.03	0.9609
iApp v3.0.0 (Oct 2026)	3.66	2.59	5.60	0.9693

Character accuracy rises from 94.8% to 96.3%, a 29% relative cut in CER, and word accuracy from 93.0% to 94.4%. iApp v3.0.0 has the lowest CER on 20 of the 20 documents against Google Cloud Vision and on 20 of 20 against the previous engine (Table 6).

Table 6: Per-document ordered CER (%) on the WP-2026-01 benchmark; the lowest in bold. Document names are in Appendix B.

Doc	Pages	Google	iApp v2.1.13	iApp v3.0.0
J1	9	5.11	4.90	3.11
J2	5	5.22	4.41	3.57
J3	21	3.96	3.31	2.29
J4	13	2.66	1.90	1.13
J5	5	23.83	22.77	22.09
J6	20	4.50	3.85	2.69
J7	11	4.55	4.34	2.09
J8	4	8.01	6.03	4.71
J9	8	4.96	4.22	1.73
J10	7	5.61	4.49	3.54
J11	27	4.00	3.68	2.40
J12	5	7.67	5.97	3.98
J13	7	6.06	4.05	1.77
J14	45	3.56	3.16	1.59
J15	8	5.43	4.40	3.45
J16	9	4.44	4.05	2.99
J17	5	6.35	5.24	2.40
J18	14	4.92	4.30	2.37
J19	9	4.93	4.43	2.48
J20	10	6.11	4.07	2.79

6 Discussion

Most of Google Cloud Vision’s gap on government documents is reading order, not reading. On Royal Gazette and budget pages the two systems recognise characters about equally well (Table 3). Google Cloud Vision, however, returns a two-column budget table as all the labels and then all the amounts, so each amount is separated from its line item; iApp keeps each row together. For a customer who needs the text, an amount detached from its line item is not usable, which is why the ordered CER on budget documents is 17.2% against 37.4%.

Against its own predecessor, v3.0 trades a little on government documents for gains elsewhere. On the Royal Gazette and budget sets iApp v2.1.13 remains marginally lower (Table 2); v3.0 is better on mixed documents, old books, document photographs and both court benchmarks, and lower over all pages (29.6% against 30.7%).

On mixed documents and old books, iApp also recognises better. The leads on SEA-DocBench and Wikisource hold in the order-free metric, and both confidence intervals exclude zero.

Google Cloud Vision is better on short photographed text. On the OCR-104 documents, Thai and English, and scene-text sets, Google Cloud Vision has the lowest error. These sets hold 12–24 images each and most intervals include zero, but the direction is consistent; it is where the iApp model will be improved next. Handwriting is not a target of this model.

The mean of the eleven sets points two ways. In reading order, iApp is ahead (11.3% against 16.1%). Order-free, Google Cloud Vision is ahead (7.0% against 7.4%), because each of the five small OCR-104 sets weighs as much as the 1,225-page SEA-DocBench. Over all pages, iApp is ahead on both metrics.

The claim the evidence supports: on Thai documents — government forms, gazettes, budget tables, court judgments, mixed web documents and old books — iApp Thai Document OCR v3.0 returns more accurate Thai text than Google Cloud Vision. On short photographed snippets, Google Cloud Vision is still slightly ahead.

7 Production Performance

Load was generated inside iApp’s data centre directly against the service, bypassing the API gateway so that nothing was billed, on 5 October 2026 with no other traffic. Twenty real document pages (Royal Gazette, budget, SEA-DocBench and Wikisource) were sent in a closed loop by *C* concurrent clients. The service runs as two replicas on separate servers.

Table 7: Throughput and latency, two replicas. No request failed at any load.

Output	Concurrency	Pages/s	p50 (s)	p95 (s)
Plain text	1	1.20	0.69	1.61
Plain text	4	4.35	0.80	1.75
Plain text	8	7.03	0.98	2.18
Plain text	16	7.30	1.91	4.34
Plain text	24	7.70	2.22	6.22
Word (layout path)	2	0.23	8.2	13.2
Word (layout path)	4	0.40	8.6	16.0
Word (layout path)	8	0.59	11.8	25.7

- **Capacity.** About 7.5 pages a second for plain text, roughly 27,000 pages an hour, twice one replica’s 3.7. Layout JSON, Word, PowerPoint and Excel share one layout path: at least 0.59 pages a second, roughly 2,100 pages an hour, not yet saturated at eight concurrent clients.
- **Time per page.** The service reports a median processing time of 0.60 s per page of plain text (90th percentile 0.97 s), against 0.53 s for v2.1.13 in August 2026: v3.0 is more accurate and has more capacity, but is not faster per page.
- **Availability.** With one replica deleted under load, 0 of 198 requests failed and a replacement was ready within 75 seconds. Across every deployment and test that night, 1 of 1,503 one-second health checks failed, during the first deployment, before a shutdown grace period was added.

8 Availability, Pricing and Data Handling

The service is available at <https://api.iapp.co.th/v3/store/ocr/document> under the same endpoints as before. From 5 October 2026 plain text costs 0.049 IC per page, and layout JSON, Word, PowerPoint, Excel and searchable PDF cost 0.15 IC per page; before, every output cost 1 IC per page. Documents are processed on iApp servers in Thailand and used only to produce the requested output; generated files are deleted within one hour. An on-premise deployment, in which no document leaves the customer’s network, is available.

9 Limitations

- **Ground truth.** The Royal Gazette and budget ground truth comes from the publications’ PDF text layers and follows their order, which on budget pages is row by row. Some pages carry garbled text layers, which costs every engine the same; the hand-checked subsets are reported separately.
- **Training domain.** The iApp model was trained on other judgments of the same court, so both court benchmarks are in its domain. The previous engine, with no court training, was already better than Google Cloud Vision on 131 of 150 judgments.

- **One configuration of Google Cloud Vision**, the one iApp’s service used, and one run: Google’s models change over time; these are its results as of 4 October 2026.
- **Small sets.** The OCR-104 sets hold 12–24 images each; their intervals are wide.

10 Reproducibility

The SEA-DocBench, Wikisource and court-judgment data are public, and the Royal Gazette and budget documents are public government publications. The metrics and the statistical method are defined in Section 4. The Google Cloud Vision run used 3,248 requests, US\$4.87 at the list price of US\$1.50 per 1,000. The page lists, the per-page results and the scoring code are available from iApp Technology on request: sale@iapp.co.th.

References

- [1] iApp Technology. *General Thai Document OCR: Accuracy Benchmark on Thai Government Documents*. Technical Whitepaper WP-2026-01, 22 August 2026. <https://iapp.co.th/whitepapers/thai-document-ocr-benchmark-2026.pdf>

A Pages Read by the Secondary Engine

Set	Share of pages
Royal Gazette, verified	0%
Budget documents, verified	0%
Royal Gazette	0%
Budget documents	0%
SEA-DocBench (Thai)	3%
Old books (Wikisource)	6%
OCR-104 documents	9%
OCR-104 Thai and English	33%
OCR-104 handwriting	82%
OCR-104 document photos	21%
OCR-104 scene text	29%

B Document File Names

The 20 judgments of the WP-2026-01 benchmark, by their file names in the public *iApp Thai Administrative Court Judgment OCR* dataset.

Doc	File
J1	01012-600387-2F-600824-0000604662
J2	01013-600011-2F-600823-0000604607
J3	01013-600125-2F-611122-0000631729
J4	01013-600163-2F-600810-0000603887
J5	01013-600216-2F-610131-0000614723
J6	01013-600245-2F-610809-0000626076
J7	01013-600276-2F-601208-0000611364

J8	01013-600315-2F-600921-0000606624
J9	01013-600427-2F-610201-0000614674
J10	02013-600008-2F-600803-0000603597
J11	03012-600477-1F-620502-0000639287
J12	03013-600092-2F-600816-0000604139
J13	03013-600204-2F-610213-0000615245
J14	08012-600443-1F-620221-0000636050
J15	08013-600022-2F-600816-0000604217
J16	08013-600140-2F-621018-0000648823
J17	09013-600008-2F-600627-0000606896
J18	09013-600096-2F-610427-0000620180
J19	09013-600227-2F-620110-0000634060
J20	09013-600261-2F-620513-0000639495
