Sample report: which model extracts these invoice fields, and what

Decision

Claude Opus 5.5, GPT-6 Astra and Gemini 4 Carbon all captured every field you marked essential on these five documents, so they tie on your requirements.

Of those, Gemini 4 Carbon costs least, at $0.0081 per document against $0.0184 for Claude Opus 5.5 and $0.0239 for GPT-6 Astra.

This applies to these five documents only. It is not a general ranking of the models.

Essential fields

Essential field	Claude Opus 5.5	GPT-6 Astra	Gemini 4 Carbon
Invoice number	5/5	5/5	5/5
Invoice date	5/5	5/5	5/5
Supplier name	5/5	5/5	5/5
Purchase-order number	5/5	5/5	5/5
Total amount due	5/5	5/5	5/5
All essential values	25/25 met	25/25 met	25/25 met

Overall accuracy and cost

Model	All values right	Missed	Wrong	Unexpected	Ambiguous	Cost/doc	Time, 5 calls
Claude Opus 5.5 (claude-opus-5-5)	45/45	0	0	0	0	$0.0184	32.1 s
GPT-6 Astra (gpt-6-astra)	45/45	0	0	0	0	$0.0239	23.7 s
Gemini 4 Carbon (gemini-4-carbon)	45/45	0	0	0	0	$0.0081	45.7 s

Representative errors

No errors.

Per-field results

Field	Claude Opus 5.5	GPT-6 Astra	Gemini 4 Carbon
Invoice number (essential)	5/5	5/5	5/5
Invoice date (essential)	5/5	5/5	5/5
Due date	5/5	5/5	5/5
Supplier name (essential)	5/5	5/5	5/5
Customer name	5/5	5/5	5/5
Purchase-order number (essential)	5/5	5/5	5/5
Subtotal	5/5	5/5	5/5
Tax amount	5/5	5/5	5/5
Total amount due (essential)	5/5	5/5	5/5

Method

Same five documents, same instructions (cmp-extract-v1), same output format for all three models; one document per call through the Codegraff gateway (The Standard Harness Company of Singapore). Models:

claude-opus-5-5, gpt-6-astra, gemini-4-carbon. No sampling settings sent; no model substituted. Your expected values were never sent to any model. Scoring is by rule per field type: identifiers exact (leading zeros, prefixes and hyphens count), amounts by value and currency, dates as calendar dates, names ignoring case and punctuation.