Can we trust this model to edit our contracts?

5 cases × 3 models × 5 attempts = 75 candidate outputs.

Summary

Meets your essential requirements: Claude Opus 5.5 (25/25 runs successful), GPT-6 Astra (25/25 runs successful) and Gemini 4 Carbon (25/25 runs successful).

Best quality: a tie between Gemini 4 Carbon, Claude Opus 5.5 and GPT-6 Astra (each met 100% of checks).

Lowest cost per successful task: Gemini 4 Carbon, $0.011 (failed attempts and retries included).

Fastest end to end: GPT-6 Astra, median 5.3 s per answer (9 in 10 within 9.4 s).

Inspect the outputs below before you choose a baseline. A recommendation never changes your baseline by itself.

Who meets your essential requirements

Model	Version	Successful	Dealbreaker runs	Meets?
Gemini 4 Carbon	gemini-4-carbon	25/25	0/25	Yes
Claude Opus 5.5	claude-opus-5-5	25/25	0/25	Yes
GPT-6 Astra	gpt-6-astra	25/25	0/25	Yes

Important failures

No dealbreaker was missed and every call returned an answer.

Cost, speed and repeatability

Model	Quality	Cost / successful	Cost / attempt	Time med / p90	Same result all runs
Gemini 4 Carbon	100%	$0.011	$0.011	19 s / 32 s	5 of 5
Claude Opus 5.5	100%	$0.026	$0.026	9.8 s / 11 s	5 of 5
GPT-6 Astra	100%	$0.028	$0.028	5.3 s / 9.4 s	5 of 5

Case by case

1. Extend the initial term

Requirement	Claude Opus 5.5	GPT-6 Astra	Gemini 4 Carbon
The initial term is 24 months (dealbreaker)	5/5	5/5	5/5
The renewal term stays at 12 months (dealbreaker)	5/5	5/5	5/5
Pricing is unchanged (dealbreaker)	5/5	5/5	5/5
Successful runs	5/5	5/5	5/5

2. Fee increase from year two

Requirement	Claude Opus 5.5	GPT-6 Astra	Gemini 4 Carbon
The new fee is US$5,250 per month (dealbreaker)	5/5	5/5	5/5
The current fee still applies in the first year (dealbreaker)	5/5	5/5	5/5
The new amount is written in words and figures, like the existing one	5/5	5/5	5/5
The new fee starts only from the second year (month 13) (dealbreaker)	5/5	5/5	5/5

Successful runs	5/5	5/5	5/5

3. Delete a clause and renumber

Requirement	Claude Opus 5.5	GPT-6 Astra	Gemini 4 Carbon
Clause 6 is gone (dealbreaker)	5/5	5/5	5/5
Termination for Convenience is renumbered to clause 6 (dealbreaker)	5/5	5/5	5/5
Governing Law is renumbered to clause 7 (dealbreaker)	5/5	5/5	5/5
The cross-reference in clause 2 now points to clause 6 (dealbreaker)	5/5	5/5	5/5
Confidentiality is unchanged (dealbreaker)	5/5	5/5	5/5
Successful runs	5/5	5/5	5/5

4. Three negotiated changes and one conflict

Requirement	Claude Opus 5.5	GPT-6 Astra	Gemini 4 Carbon
Invoices payable within 45 days (dealbreaker)	5/5	5/5	5/5
Liability cap uses the previous 12 months of fees (dealbreaker)	5/5	5/5	5/5
Keeps fees fixed for the Initial Term (does not apply change d) (dealbreaker)	5/5	5/5	5/5
Adds termination for uncured material breach (15 days to cure)	5/5	5/5	5/5
Flags that change (d) conflicts with the fixed fees in clause 4 (dealbreaker)	5/5	5/5	5/5
Successful runs	5/5	5/5	5/5

5. Add a data protection clause

Requirement	Claude Opus 5.5	GPT-6 Astra	Gemini 4 Carbon
Adds clause 9 headed Data Protection (dealbreaker)	5/5	5/5	5/5
Names the Personal Data Protection Act 2012 (dealbreaker)	5/5	5/5	5/5
Governing law clause is unchanged (dealbreaker)	5/5	5/5	5/5
The obligation applies to both parties	5/5	5/5	5/5
Successful runs	5/5	5/5	5/5

Method

Same messages for every model through the Codegraff gateway; 5 independent attempts per case per model; failed calls retried up to 2 more times with every retry's cost counted. Deterministic requirements checked by code; the rest graded blind by GPT-6 Sol (gpt-6-sol, grading prompt tasks-judge-2026-11-16), never one of the models compared. Outputs and redlines are in the HTML report.