anthropic/claude-opus-4.8

answered like

deepseek/deepseek-v4.1-flash

<redacted-1> > <redacted-1> · the model's default reasoning, max_tokens 400

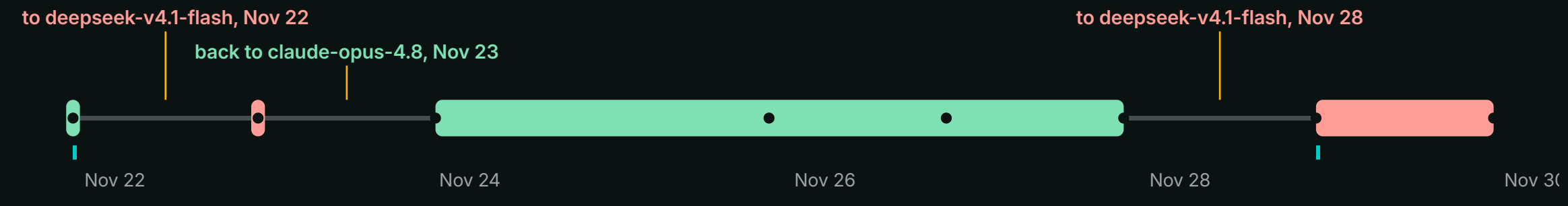
Compared with deepseek-v4.1-flash: the route's best fit now.

Prepared on November 29, 2026 at 20:35:54 UTC

8	5	3	3
cards, Nov 21 to Nov 29, 2026	answer like claude-opus-4.8	answer like deepseek-v4.1-flash	changes

Observation

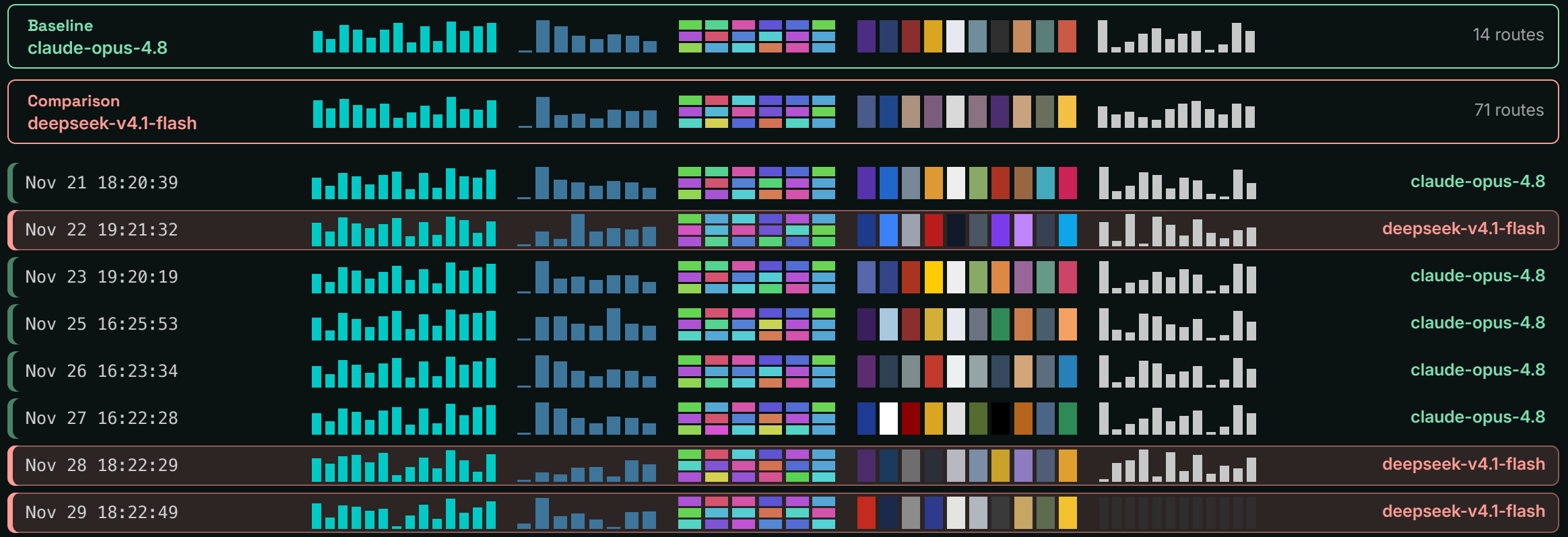
Timeline



Change	Last card before (UTC)	First card after (UTC)
to deepseek-v4.1-flash	2026-11-21 18:20:39.254	2026-11-22 19:21:32.385
back to claude-opus-4.8	2026-11-22 19:21:32.385	2026-11-23 19:20:19.543
to deepseek-v4.1-flash	2026-11-27 16:22:28.289	2026-11-28 18:22:29.783

Observation

Behavior history for claude-opus-4.8

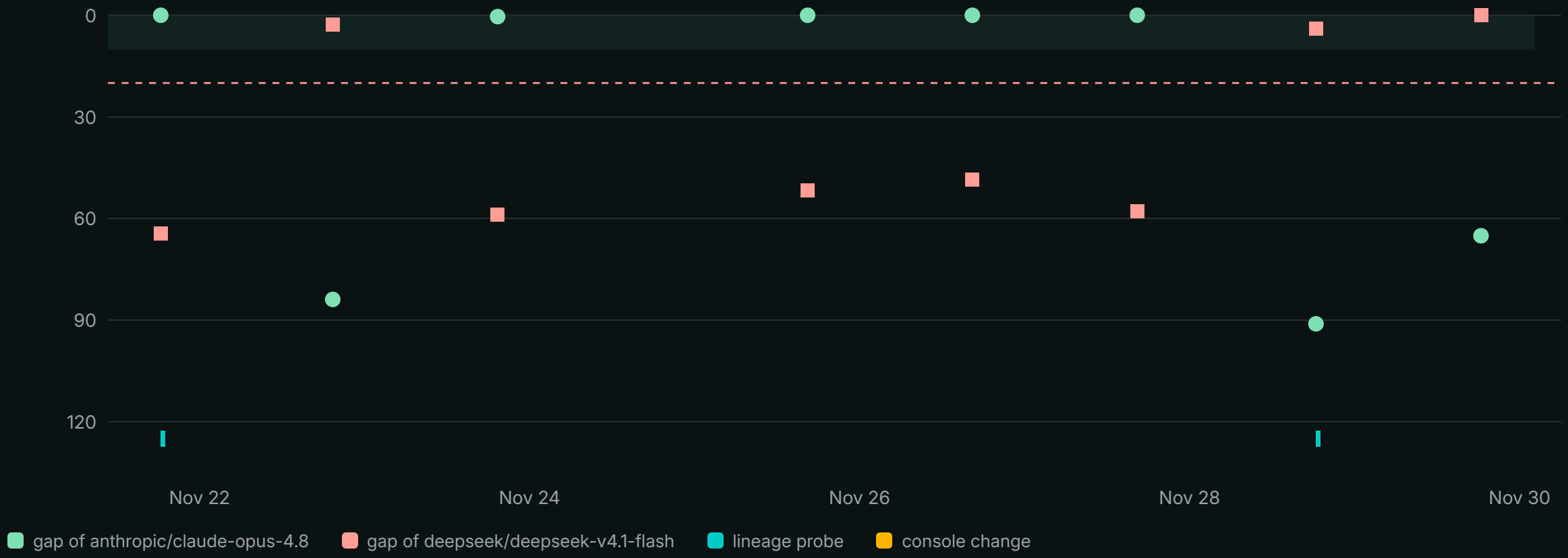


One row per card: ratings, guesses, words, colors, ranks. tinted: answers like deepseek-v4.1-flash

Observation

Card scores

Gap under each card's best-fitting model. 10.0 or less: answers like that model.



Observation

Each card

#	Received (UTC)	Fits	Best model for this card	Gap, claude-opus-4.8	Gap, deepseek-v4.1-flash	Card
1	2026-11-21 18:20:39.254	claude-opus-4.8	claude-opus-4.8	0.0	64.5	
2	2026-11-22 19:21:32.385	deepseek-v4.1-flash	deepseek-v4.1-flash:thinking	83.9	2.8	
3	2026-11-23 19:20:19.543	claude-opus-4.8	claude-opus-4.8	0.4	58.8	
4	2026-11-25 16:25:53.629	claude-opus-4.8	claude-opus-4.8	0.0	51.7	
5	2026-11-26 16:23:34.796	claude-opus-4.8	claude-opus-4.8	0.0	48.5	
6	2026-11-27 16:22:28.289	claude-opus-4.8	claude-opus-4.8	0.0	57.8	
7	2026-11-28 18:22:29.783	deepseek-v4.1-flash	deepseek-v4.1-flash-1017	91.1	4.0	
8	2026-11-29 18:22:49.865	deepseek-v4.1-flash	deepseek-v4.1-flash	65.1	0.0	

Gap under the card's best model. 10.0 or less: answers like it.

Comparison

Nearest routes

Of 4,479 routes, any claimed model.

The 3 cards that fit deepseek/deepseek-v4.1-flash

anthropic/claude-haiku-4.5-20251001-aws <redacted...	0.0
deepseek/deepseek-v4.1-flash <redacted-2> > <redact...	0.0
deepseek/deepseek-v4.1-flash-1017 <redacted-2> > <r...	0.0
deepseek/deepseek-v4.1-flash <redacted-2> > <redact...	0.0
deepseek-v4p1-flash <redacted-6> > <redacted-6>	0.0
deepseek/deepseek-v4.1-flash-1017 <redacted-7> > <r...	0.0
deepseek/deepseek-v4.1-flash-1017 <redacted-7> > <r...	0.0
deepseek/deepseek-v4-1-flash-1017 <redacted-10> > ...	0.0
deepseek/deepseek-v4.1-flash <redacted-7> > <redact...	0.1
deepseek/deepseek-v4.1-flash-1017 <redacted-7> > <r...	0.5
deepseek/deepseek-v4.1-flash <redacted-7> > <redact...	0.6
deepseek/deepseek-v4.1-flash <redacted-7> > <redact...	1.0

The 5 cards that fit anthropic/claude-opus-4.8

claude-opus-4-8 <redacted-17> > <redacted-17>	0.0
claude-opus-4.8 <redacted-18> > <redacted-19>	0.0
claude-opus-4.8 <redacted-18> > <redacted-11> > <reda...	0.0
anthropic/claude-opus-4.8 <redacted-7> > <redacted-1...	0.0
anthropic/claude-opus-4.8 <redacted-2> > <redacted-1...	0.0
anthropic/claude-opus-4.8 <redacted-7> > <redacted-2...	0.0
anthropic/claude-opus-4.8 <redacted-7> > <redacted-1...	0.0
anthropic/claude-opus-4.8 <redacted-2> > <redacted-2...	0.0
anthropic/claude-opus-4.8 <redacted-10> > <redacted-...	0.0
anthropic/claude-opus-4.8 <redacted-2> > <redacted-1...	0.4
claude-opus-4.8 <redacted-18> > <redacted-17>	0.8
claude-opus-4.8 <redacted-18> > <redacted-23>	0.8

route claims deepseek-v4.1-flash route claims claude-opus-4.8 another model Bar: sameness gap. 10.0 or less: one model.

Comparison

claude-opus-4.8 elsewhere

21 routes, 12 providers. Spread 1.9. claude-opus-4.8 cards: 0.8. deepseek-v4.1-flash cards: 41.5.



Comparison

deepseek-v4.1-flash elsewhere

394 routes, 98 providers. Spread 5.9. claude-opus-4.8 cards: 56.2. deepseek-v4.1-flash cards: 7.1.



Comparison

Telling answers

Card item	claude-opus-4.8, other routes	This route, claude-opus-4.8 cards	This route, deepseek-v4.1-flash cards	deepseek-v4.1-flash, other routes
prime	65	**65** C 95% · O 1%	72 C 39% · O 19%	85
rain	30	**30** C 80% · O 7%	40 C 29% · O 44%	47
sand	5,000	**5,000** C 61% · O 7%	**8,000** C 0% · O 48%	8,000
Paris steps	500,000	**500,000** C 88% · O 4%	12,000 C 0% · O 3%	120,000
novel pages	20,000	**15,000** C 34% · O 3%	**1,400** C 0% · O 46%	3,000
bricks	5,000	**5,000** C 66% · O 1%	**12,000** C 2% · O 91%	12,000
luck 2	chance	**chance** C 80% · O 9%	**coin** C 0% · O 49%	coin
luck 3	charm	**charm** C 51% · O 3%	clover C 2% · O 5%	fortune
king 2	queen	**queen** C 76% · O 1%	**crown** C 7% · O 60%	crown
king 3	throne	**throne** C 83% · O 6%	lion C 0% · O 4%	ruler

C, O: share of claude-opus-4.8 and deepseek-v4.1-flash cards near this value. 10 of 15 cells shown.

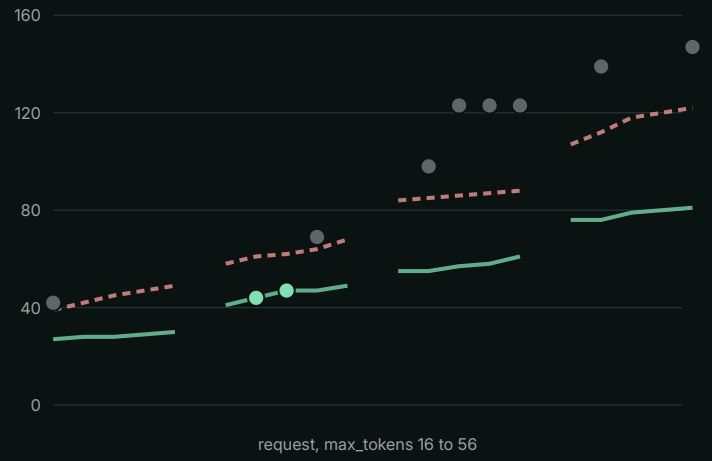
Observation

Tokenizer probes

20 requests repeat one text, max_tokens 16 to 56. Each reply stops at a token edge.

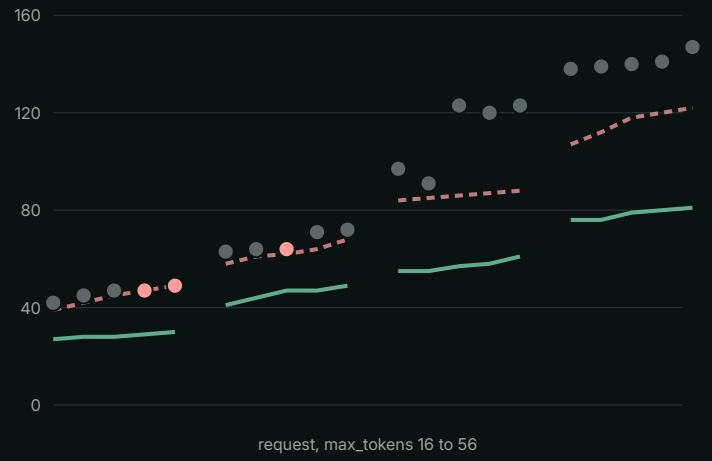
2026-11-21 18:21:22

2 of 10 edges like claude-opus-4.8. Nearest card 1, 43s before, fits claimed.



2026-11-28 18:23:14

0 of 20 edges like claude-opus-4.8. Nearest card 7, 44s before, fits other.



claude-opus-4.8 tokenizer: claude #29 (337 routes) deepseek-v4.1-flash tokenizer: deepseek #11 (961 routes) matches neither

Observation

Same provider

Other <redacted-1> routes with deepseek-v4.1-flash cards.

anthropic/claude-opus-4.8
<redacted-1> > <redacted-1> (this report)

anthropic/claude-haiku-4.5-20251001-aws
<redacted-1> > <redacted-1>

anthropic/claude-opus-4.6
<redacted-1> > <redacted-1>

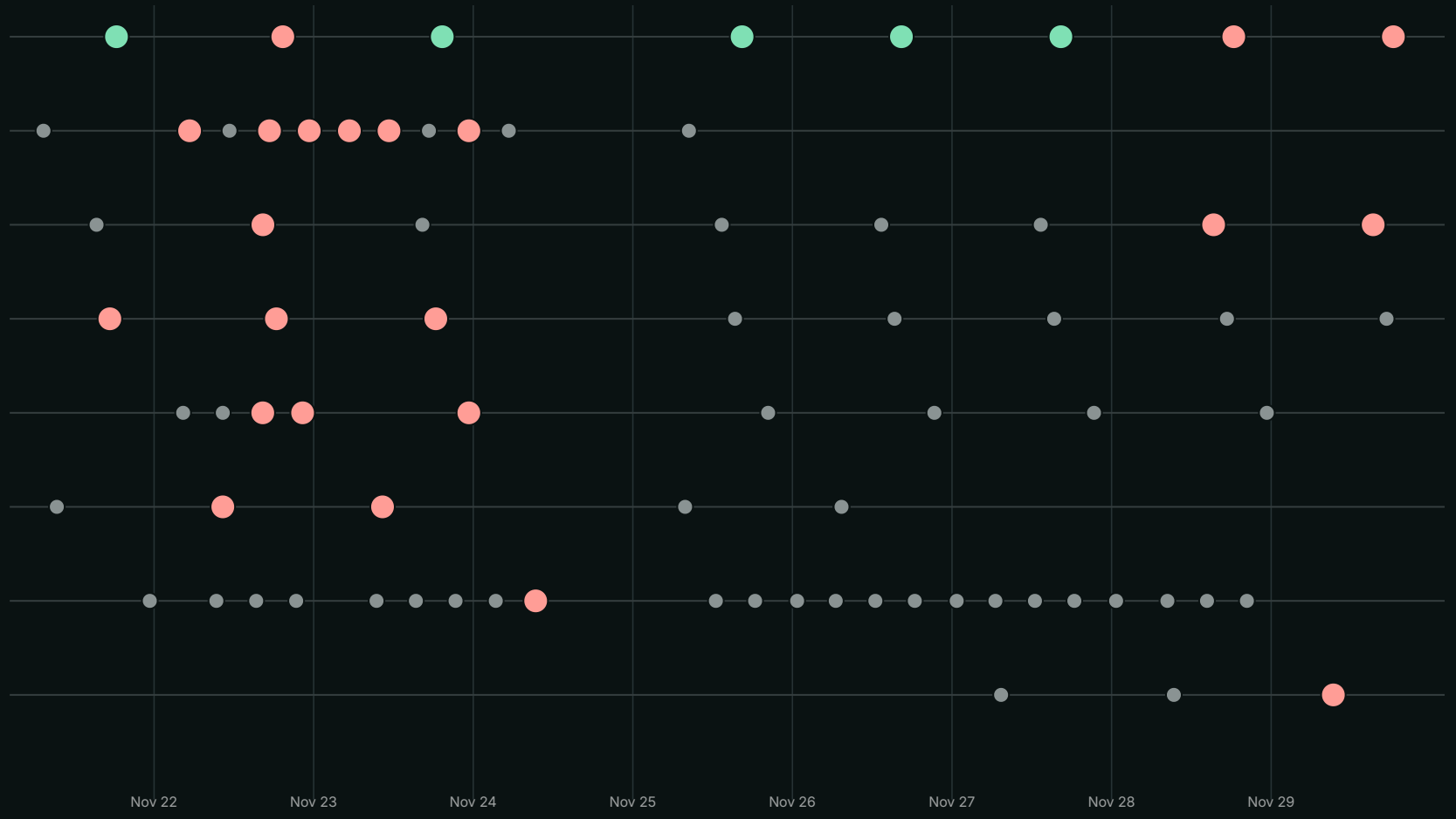
anthropic/claude-opus-4.7
<redacted-1> > <redacted-1>

xiaomi/mimo-v2.5-pro
<redacted-1> > <redacted-1>

anthropic/claude-sonnet-4.6
<redacted-1> > <redacted-1>

qwen/qwen3.5-122b-a10b
<redacted-1> > <redacted-1>

anthropic/claude-sonnet-4.6
<redacted-1> > <redacted-1> > <redacted-17>



fits deepseek-v4.1-flash this route, fits claude-opus-4.8 other card

Summary

Key measurements

10 of 12

nearest routes claim deepseek-v4.1-flash

0.8 / 41.5

gap to claude-opus-4.8 routes: claude-opus-4.8 / deepseek-v4.1-flash cards (spread 1.9)

7.1

gap to deepseek-v4.1-flash routes (spread 5.9)

0 of 20

token edges like claude-opus-4.8, Nov 28 18:23

7

other <redacted-1> routes with deepseek-v4.1-flash cards

Summary

Reasoning

Cards	**5 of 8 fit claude-opus-4.8, 3 fit deepseek-v4.1-flash**	Nov 21 to Nov 29, 2026
Changes	**3 changes**	deepseek-v4.1-flash by Nov 22 19:21, claude-opus-4.8 by Nov 23 19:20, deepseek-v4.1-flash by Nov 28 18:22
Rank of claude-opus-4.8	**298 of 561 models**	deepseek-v4.1-flash cards, 77.8 under the best
Nearest routes	**10 of 12 claim deepseek-v4.1-flash**	of 4,479 routes, 1 at <redacted-1>
Gap to claude-opus-4.8 routes	**0.8 for claude-opus-4.8 cards, 41.5 for deepseek-v4.1-flash cards**	21 routes, 12 providers, spread 1.9
Gap to deepseek-v4.1-flash routes	**7.1 for deepseek-v4.1-flash cards**	394 routes, 98 providers, spread 5.9
Token edges like claude-opus-4.8	**2/10 Nov 21, 0/20 Nov 28**	like deepseek-v4.1-flash: 0/10, 3/20
Host token counts	**as claude-opus-4.8 in 2 of 2 probes**	
Other <redacted-1> routes	**7 routes with deepseek-v4.1-flash cards**	claude-haiku-4.5-20251001-aws 6/11, claude-opus-4.6 5/10, claude-opus-4.7 3/8, mimo-v2.5-pro 3/9, claude-sonnet-4.6 2/7, qwen3.5-122b-a10b 1/23, claude-sonnet-4.6 1/3

Appendix: reading the measurements.

Summary

Limits

Not shown

- Why the answers changed, or who in the path changed them.
- Intent: only which model the answers fit, and when.
- Proof from one card: this rests on 8 cards here and 4,479 other routes.

- **Card:** one request, a fixed 62-item questionnaire, the model's default reasoning, max_tokens 400.
- **Gap:** how far a model is under the best, per card.
- **Fit:** gap 10.0 or less. The other model counts only with the claimed model 20.0 or more under.
- **Route gap:** the console's sameness gap, both directions.
- **Times:** when the seeder received each reply, UTC.

Appendix

Method: cards

What a card is

One chat request: a fixed questionnaire of 62 items. The model rates 14 statements, guesses 8 quantities, gives 3 words for 6 cues, picks 10 colors and ranks 12 options.

No item has a right answer. Each model answers in its own steady way, and 2 models differ more than one model at 2 providers.

Every route gets the same text, default temperature, max_tokens 400, about every 6 hours.

How a card scores

A model's reference is the newest cards, up to 160, of every route that claims it. A route never scores against its own cards.

A card's score is its log-likelihood under a reference. The gap is how far a model is under the best, per card. 10 or less: answers like that model. 20 or more: does not.

Why it works

- One request of about 400 tokens, through any chat API.
- It reads the answers, not the name, headers or price.
- A host cannot fake a model's answers without running that model or one like it.

Appendix

Method: lineage

The lineage probe

20 requests, about every 7 days. Each asks the route to repeat one text of mixed scripts, digits and code, with max_tokens from 16 to 56.

Cut mark. Each reply stops at a token edge. The characters in the 20 replies give the edges of the tokenizer that writes the text.

Count mark. Each request carries a note in one of 8 scripts. The prompt token counts the host reports read the tokenizer the host counts with. That can be a gateway's, not the model's.

Lineage groups

Routes with one cut mark share a group: one tokenizer. A route with no usable cut mark joins the nearest count mark. A shift of a few tokens is allowed for text the host adds first.

What it shows

- A cut mark unlike every route of the claimed model: another tokenizer wrote the reply.
- The claimed model's count mark with another cut mark: the host counts one way, another model writes.
- Reasoning tokens move the edges. The report compares edges request by request.

Method: limits

Where cards break down

- **Changed system prompt.** A prompt the host adds shifts the answers. The cards then fit neither model.
- **New model version.** Every route changes at once. The report shows other providers on the same days.
- **Reasoning settings.** They change answers and edges. Each route's setting is shown.
- **Serving differences.** Quantization moves answers a little. Gaps under 10 cover it.
- **Distilled models.** A model trained on another's output can answer like it.

Where the report needs care

- **Few cards.** One card can fit by chance. Every card is shown with its time.
- **Thin references.** A model with few routes has a weak reference. A model with none cannot be named.
- **Hosts that know the card.** They can answer it honestly and route the rest elsewhere. Cards then understate a swap. They cannot create one.
- **Caches and splits.** A cache repeats a reply. A router can split a route. Both show as a mix.
- **Intent.** Cards show which model answered and when, not why or who.

Appendix

Method: reading the measurements

- **Cards.** Each card scores alone against every model with 12 or more cards elsewhere. It fits at a gap of 10 or less. It fits the other model only with the claimed model 20 or more under.
- **Changes.** A change lies between the last card of one fit and the first of the next.
- **Rank.** The cards of one fit score together. Rank is the claimed model's place in that list.
- **Nearest routes.** The cards against every route with 3 or more cards, by sameness gap. 10 or less: one model.
- **Gap to routes.** The median gap to a model's routes, beside the spread between those routes. Near the spread: that model. Far above: another model.
- **Token edges.** Probe replies that stop where the lineage's usual cut mark stops, out of those with an edge.
- **Host token counts.** Probes whose count mark matches the claimed model's lineage.
- **Other routes of the provider.** Their cards that fit the other model by the rule for cards.

Appendix

Same provider: card times (1 of 2)

Route	Cards that fit deepseek-v4.1-flash	Times of those cards (UTC)
anthropic/claude-haiku-4.5-20251001-aws <redacted-1> > <redacted-1>	6 of 11	2026-11-22 05:20:27.345, 2026-11-22 17:21:28.938, 2026-11-22 23:20:06.240, 2026-11-23 05:22:15.612, 2026-11-23 11:19:45.968, 2026-11-23 23:20:23.920
anthropic/claude-opus-4.6 <redacted-1> > <redacted-1>	5 of 10	2026-11-22 16:21:55.241, 2026-11-28 15:20:41.566, 2026-11-29 15:20:27.807, 2026-11-29 15:20:34.761, 2026-11-29 15:20:34.762
anthropic/claude-opus-4.7 <redacted-1> > <redacted-1>	3 of 8	2026-11-21 17:20:49.009, 2026-11-22 18:23:35.652, 2026-11-23 18:22:04.663
xiaomi/mimo-v2.5-pro <redacted-1> > <redacted-1>	3 of 9	2026-11-22 16:20:35.888, 2026-11-22 22:20:17.814, 2026-11-23 23:20:14.550
anthropic/claude-sonnet-4.6 <redacted-1> > <redacted-1>	2 of 7	2026-11-22 10:20:08.406, 2026-11-23 10:21:18.180
qwen/qwen3.5-122b-a10b <redacted-1> > <redacted-1>	1 of 23	2026-11-24 09:23:17.115

Appendix

Same provider: card times (2 of 2)

Route	Cards that fit deepseek-v4.1-flash	Times of those cards (UTC)
anthropic/claude-sonnet-4.6 <redacted-1> > <redacted-1> > <redacted-17>	1 of 3	2026-11-29 09:21:05.561

Appendix

Lineage probe marks

Probe (UTC)	Lineage group	Cut mark (characters per reply)	Count mark
2026-11-21 18:21:22.396	claude #29 (337 routes) by count	42 - - - - - 44 47 69 - - 98 123 123 123 - 139 - - 147	-46 -39 -26 -39 -56 -60 -12
2026-11-28 18:23:14.119	claude #29 (337 routes) by count	42 45 47 47 49 63 64 64 71 72 97 91 123 120 123 138 139 140 141 147	-46 -39 -26 -39 -56 -60 -12

claude-opus-4.8 lineage: claude #29 (337 routes) deepseek-v4.1-flash lineage: deepseek #11 (961 routes) "-": no token edge.

All cards

Card 1 2026-11-21 18:20:39.254 — claude-opus-4.8 · best: anthropic/claude-opus-4.8

A 65 40 80 70 45 75 85 30 80 35 95 70 65 90
B 2000 500000 8000 3 100 500000 200000 5000
C wave blue deep clover chance gold crown queen throne dark ghost figure time tick round river cross span
D 5533aa 2266cc 778899 dd9933 eeeeee 88aa66 aa3322 996644 44aabb cc2255
E AKDEGHLCFBIJ

Card 2 2026-11-22 19:21:32.385 — deepseek-v4.1-flash · best: deepseek/deepseek-v4.1-flash:thinking

A 72 45 88 70 55 68 86 30 74 35 91 80 40 77
B 8000 12000 1400 2500 125 500000 200000 15000
C wave tide blue charm coin clover throne crown queen dark shade ghost tick hands time span river crossing
D 1e3a8a 3b82f6 9ca3af b91c1c 111827 4b5563 7c3aed c084fc 374151 0ea5e9
E CEHAFLKIGJBD

Card 3 2026-11-23 19:20:19.543 — claude-opus-4.8 · best: claude-opus-4.8

A 60 35 80 70 45 65 85 40 75 30 95 70 80 90
B 5000 500000 4000 3 100 500000 250000 5000
C wave blue deep clover chance charm crown queen throne dark light ghost time tick round river cross span
D 5566aa 334488 aa3322 ffcc00 eeeeee 88aa66 dd8844 996699 669988 cc4466
E AKLGEHFDCJBI

Card 4 2026-11-25 16:25:53.629 — claude-opus-4.8 · best: anthropic/claude-opus-4.8

A 70 30 85 65 40 75 90 35 80 45 95 80 85 90
B 2000 45000 15000 3 125 1.6e+07 200000 8000
C wave blue salt chance clover charm crown queen throne dark night figure time tick round river cross span
D 3a1f5c a8c8e0 8b2e2e d4af37 e8e8f0 6b7280 2e8b57 c97b4a 4a5d6b f4a261
E AKDEHLFGBCJI

All cards

Card 5 2026-11-26 16:23:34.796 — claude-opus-4.8 · best: anthropic/claude-opus-4.8

A 65 40 85 70 45 75 90 30 80 35 95 70 80 90
B 5000 500000 20000 3 80 500000 200000 5000
C wave blue deep chance charm clover crown queen throne dark light figure time tick round river cross span
D 5b2c6f 2e4053 7f8c8d c0392b ecf0f1 95a5a6 34495e d2a679 5d6d7e 2980b9
E AKDEHLGFCJBI

Card 6 2026-11-27 16:22:28.289 — claude-opus-4.8 · best: anthropic/claude-opus-4.8

A 65 40 80 70 45 75 85 30 80 35 95 60 85 90
B 5000 500000 20000 3 80 500000 200000 5000
C wave blue deep charm chance four crown queen throne dark figure night time tick hands river cross span
D 1f3a93 ffffff 8b0000 daa520 e0e0e0 556b2f 000000 b5651d 4b6587 2e8b57
E AKEHLGDFCJBI

Card 7 2026-11-28 18:22:29.783 — deepseek-v4.1-flash · best: deepseek/deepseek-v4.1-flash-1017

A 72 55 78 82 60 88 20 45 74 40 96 65 30 85
B 8000 6000 1400 2 120 80000 400000 12000
C wave salt blue chance dice fortune throne crown reign dark cast echo tick time hand span cross arch
D 4a2c6b 1b3a5c 6e6e6e 2e2e3a b8b8c0 7a8fa6 c9a227 8e7cc3 4f5d75 e0a030
E DFHLCBJKGIEA

Card 8 2026-11-29 18:22:49.865 — deepseek-v4.1-flash · best: deepseek/deepseek-v4.1-flash

A 78 35 72 68 55 62 8 40 75 30 92 50 65 84
B 20000 120000 3000 2.5 60 40000 200000 12000
C blue wave salt chance coin fourleaf crown throne lion dark cast follow tick hands time span gap arch
D c42a1e 1b2a4a 8c8c8c 2e3b8c e4e4e0 b0b7be 3a3a3a c6a664 5e6b4f f2c230
E –

Appendix

Card request

The only user message, with the model's default reasoning, max_tokens 400.

```
Fill in this card. Reply with exactly 5 lines. Each line starts with its letter and a colon. Write nothing else.
A: rate each A item from 0 to 100. 14 integers separated by commas.
B: your best single-number guess for each B item, digits only, no units, no working. 8 numbers separated by semicolons.
C: for each C cue, the three single words that first come to mind, most likely first. Cues separated by semicolons, words by commas.
D: one hex color for each D concept. 10 codes separated by commas.
E: the E options in order from your most favorite to your least favorite. 12 letters with no spaces.

A items: 1 how beautiful a prime number is; 2 how hard chess is to learn; 3 how round an egg is; 4 how loud a vacuum cleaner is; 5 how funny a pun is; 6 how formal the word "hence" is; 7 how boring a tax form is; 8 how likely rain is tomorrow; 9 how dangerous skydiving is; 10 how trustworthy a stranger's promise is; 11 how strange the platypus is; 12 how lucky the number 7 is; 13 how loud the color red is; 14 how honest a mirror is
B items: 1 grains of sand in a teaspoon; 2 steps to walk across Paris; 3 pages in the longest novel; 4 heartbeats in a human lifetime, in billions; 5 piano tuners in Chicago; 6 golf balls that fit in a school bus; 7 leaves on a mature oak tree; 8 bricks in a typical house
C cues: ocean, luck, king, shadow, clock, bridge
D concepts: jazz, north, Monday, seven, silence, Tuesday, the letter K, nostalgia, Thursday, the number 3
E options: A violin, B granite, C lemon, D fox, E harbor, F copper, G thunder, H velvet, I ladder, J pepper, K comet, L orchid
```