

PRISM review: what AI launches publicly evidence

Abstract

Background. AI launch pages say a lot about how a product is built and checked. This review asks which checks launches publicly evidence on their own site or repository with an artefact linked, and which they only claim.

Methods. The sampling frame is Show HN posts in the sampling window, 2025-09-01 to 2026-08-31, drawn in a seeded order. 300 launches were included. Each of 24 checks was coded by agent roles, every positive was confirmed by a separate role, and a deterministic gate decided each decision. Every figure is a count of launches under disclosure control: a count, its complement or its group size under 10 is suppressed, and an exact audit requires at least 5 possible values for every suppressed count.

Results. H1 (a majority of launches claim at least one check without linking an artefact, so they do not publicly evidence it) is not supported: 96 of 300 launches (32%), one-sided exact binomial, Holm-adjusted $p = 1.000$; with the recall bound, 85 of 300, Holm-adjusted $p = 1.000$. H3 (fewer than half of launches publicly evidence at least one rely-decision check) is not supported: at least 212 of 300 launches (71%) publicly evidence one, one-sided exact binomial, Holm-adjusted $p = 1.000$; with the recall bound, 238 of 300, Holm-adjusted $p = 1.000$.

Limitations. Screening missed its pre-registered target: 12.2% of audited exclusions were in error (5 of 41, 95% CI 5.3–25.5%) against a 10% bar. This is a selection limitation. Not publicly evidenced does not mean not done.

Introduction

Launch pages for AI products make many statements about how a product is tested, limited, operated and measured. A reader deciding whether to explore, rely on or pay for a product needs to know which of those statements the launch backs with something they can inspect.

PRISM reads each launch's own site and own repository and records, for each of 24 checks grouped under 5 decisions (explore, rely, operate, measure–outcome, scale–renew–stop), whether the launch publicly evidences the check with an artefact linked, only claims it, or shows nothing. The questions, tests and thresholds were registered in the protocol before any launch was coded. The review reports aggregate figures only: no launch is named, counted alone or linked.

Methods

Eligibility

A launch is included when it is an AI product with its own site, screened in two stages: first its title and URL, then the text of its own page.

Sampling frame and window

The sampling frame is Show HN posts in the sampling window, 2025-09-01 to 2026-08-31, drawn in a seeded order.

Collection

Evidence is collected from each launch's own site and own repository only. The paid search collector was replaced by a free own-site collector for all bulk coding (addendum A8), so every launch was collected the same way.

Coding states and the self-claim rule

A check counts as publicly evidenced only when the launch's own page quotes it and links an artefact; a quote with no linked artefact is coded as claimed. Every check on every launch is in one of four states:

- `pass`: publicly evidenced (artefact linked)
- `claimed`: claimed, no artefact linked
- `unknown`: no public evidence found
- `fail`: own page shows the check is not met

AI use

Every role below is an agent, named by its model id and host. Agreement figures are agreement with an independent labeller.

AI use: every role, its model, host and run settings

ROLE	MODEL ID	HOST	RUN SETTINGS
screeener	nvidia/Nemotron-3_5-Lightning	Nebius Token Factory	prompt inclusion rules (protocol section 2) · codebook v1.0 · seed ae0b3a26c3b17672 · token cap 20,000
coder	deepseek-ai/DeepSeek-V4-Flash-0731	Nebius Token Factory	prompt codebook prompt, guidance c1 on one check · codebook v1.0 · seed none (temperature 0) · token cap 20,000
second coder	nemotron-3-super	Ollama Cloud	prompt codebook prompt · codebook v1.0 · seed 1a8a5e2c0f6afdcdb · token cap 20,000
independent labeller	glm-5.3-flash	Ollama Cloud	prompt codebook prompt · codebook v1.0 · seed d1481ad50ed309ea · token cap 20,000
confirmer	nemotron-3-ultra	Ollama Cloud	prompt confirmation prompt · codebook v1.0 · seed none (temperature 0) · token cap 20,000
screening auditor	kimi-k3	Ollama Cloud	prompt screening audit prompt · codebook v1.0 · seed fa23e4fccbc39ce5 · token cap 20,000
recall gatherer	kimi-k3	Ollama Cloud	prompt recall gathering prompt · codebook v1.0 · seed cc45c03dd0094b92 · token cap 20,000

Confirmation

Every positive is confirmed by a confirmer that did not code it before it enters any figure. A rejected positive counts as not evidenced and stays in the denominator; a downgraded pass counts as claimed.

The deterministic gate

The deterministic gate, not a model, decides each decision from the confirmed check states.

Disclosure control

k = 10. Every figure is a count of launches under disclosure control: a count, its complement or its group size under 10 is suppressed, and an exact audit requires at least 5 possible values for every suppressed count. The complement clause applies: a count whose complement is under 10 is suppressed even when the count itself is not.

Exception (addendum A14): 4 process figures describe the instrument, not any launch, and are published outside the k = 10 rule: screening error of the audited exclusions; supported and raw collection recall; the count of labels edited by a person; the total share of proposed positives changed at confirmation. Per-check precision and per-check changes at confirmation stay under the rule.

Hypotheses and tests

H1 and H3 use one-sided exact binomial tests and H2 a permutation test with Cramér's V. Holm's correction holds the family-wise level at 0.05, and a hypothesis is supported only when it also holds with the measured collection recall as a bound. Each headline figure is repeated excluding the 0 launches whose site could not be read (protocol section 3).

Release ledger

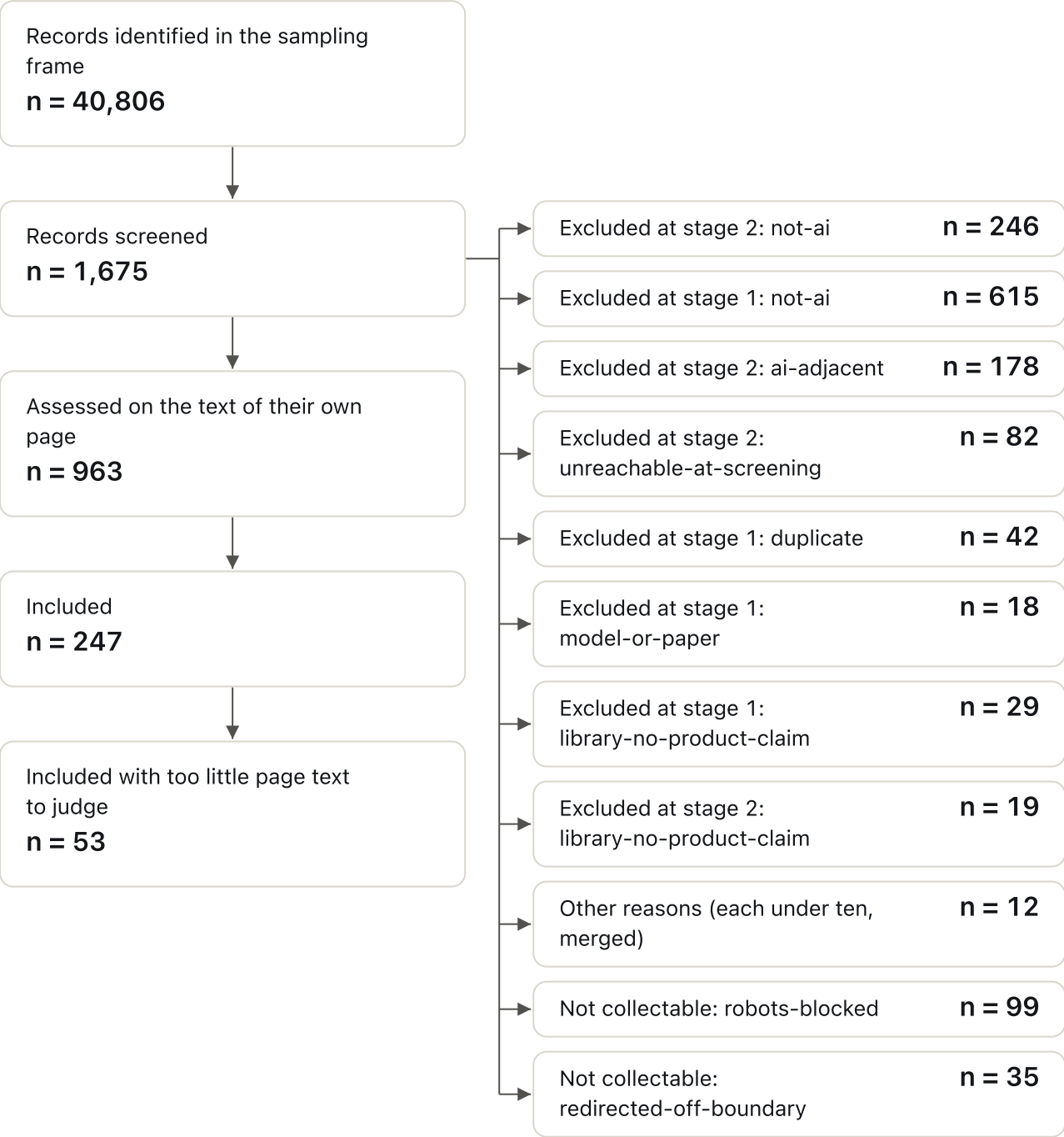
Release v1 is recorded in disclosure ledger v1: later releases are audited against every figure already published, so no hidden count can be worked out by comparing releases.

Results

Selection

Figure 1. How launches were selected

Records at each stage of selection and the reasons for exclusion; reasons with fewer than 10 records are merged.



How we know

Launches counted	N = 300 included launches.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Selection	Screening error 12.2% (5 of 41, 95% CI 5.3–25.5%), above the 10% target; see Limitations.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Hypotheses

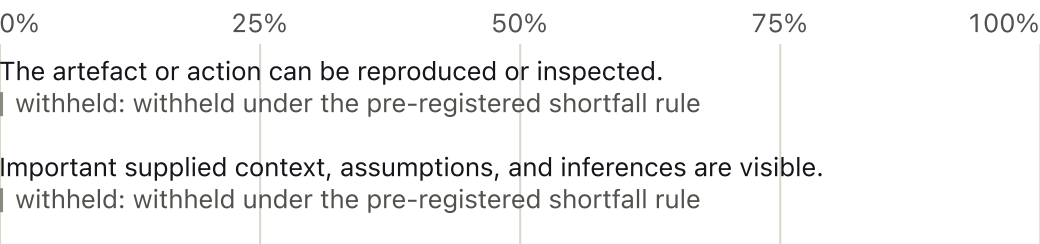
- H1 (a majority of launches claim at least one check without linking an artefact, so they do not publicly evidence it) is not supported: 96 of 300 launches (32%), one-sided exact binomial, Holm-adjusted $p = 1.000$; with the recall bound, 85 of 300, Holm-adjusted $p = 1.000$.
- Reported descriptively: at least one source has fewer than 30 included launches, so the pre-registered test was not run. The Holm correction covers the remaining hypotheses.
- H3 (fewer than half of launches publicly evidence at least one rely-decision check) is not supported: at least 212 of 300 launches (71%) publicly evidence one, one-sided exact binomial, Holm-adjusted $p = 1.000$; with the recall bound, 238 of 300, Holm-adjusted $p = 1.000$.

Evidence by check

Each bar is the share of all 300 included launches that publicly evidence the check with an artefact linked.

Figure 2. Checks for the decision: Keep exploring

share of included launches that publicly evidence the check (confirmed pass)

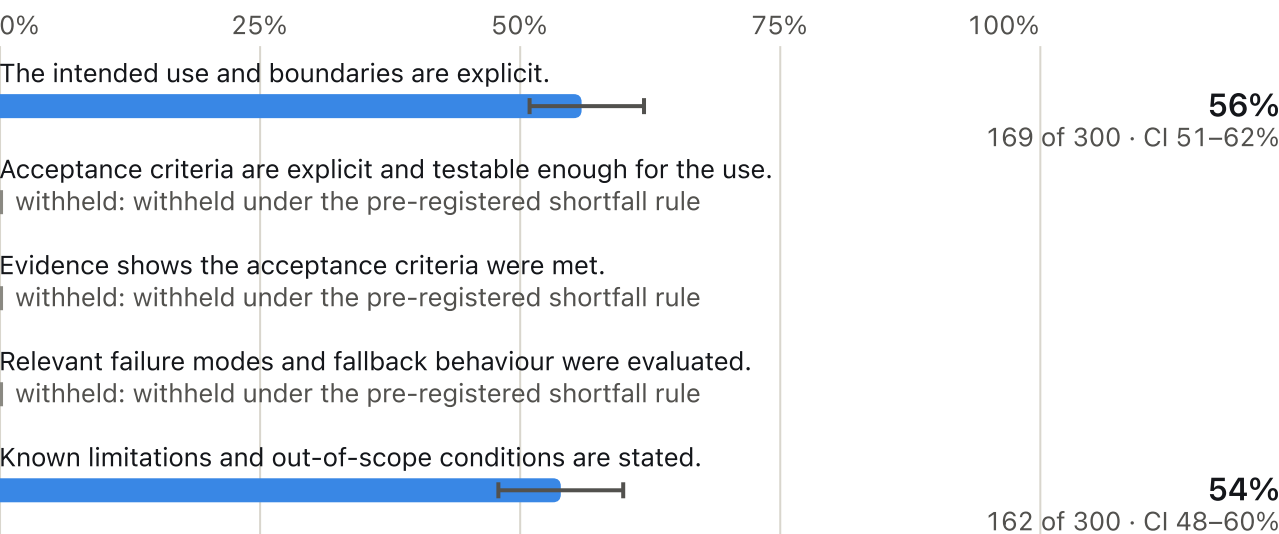


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 3. Checks for the decision: Permit reliance for the named use

share of included launches that publicly evidence the check (confirmed pass)



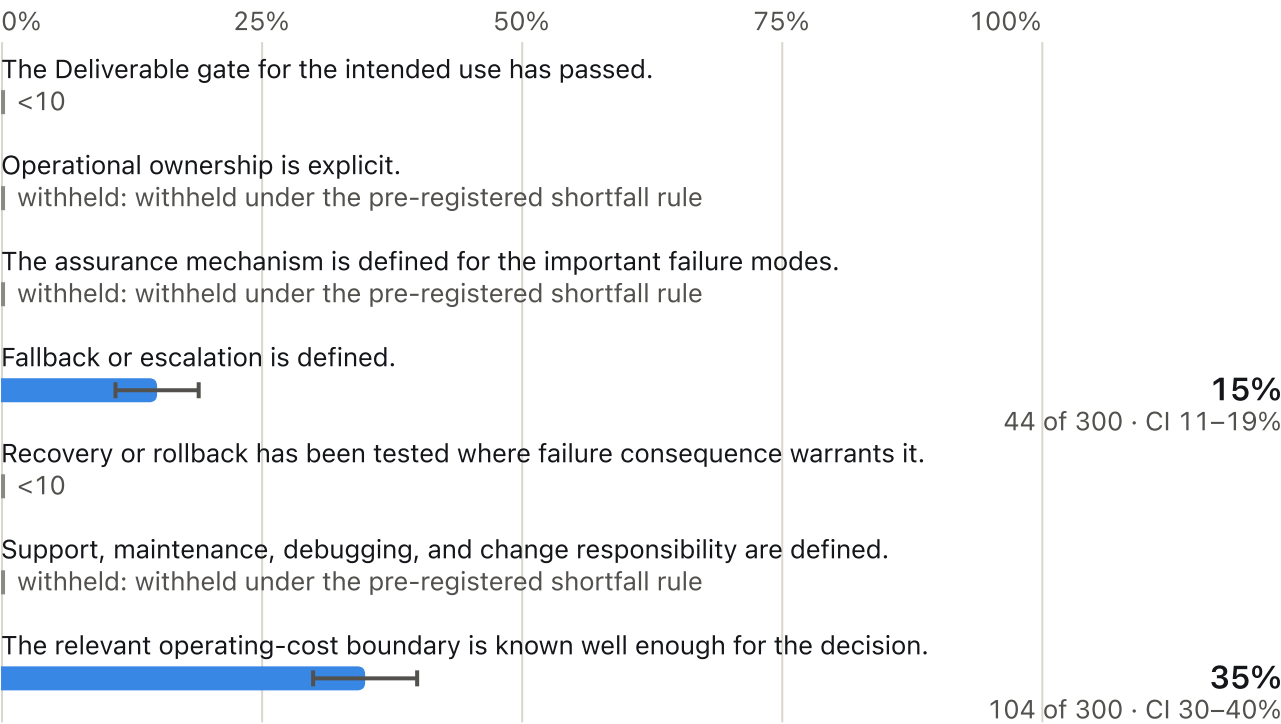
How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.86–0.89 across the shown checks (not estimable for 3); per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

- At least 169 of 300 launches publicly evidence The intended use and boundaries are explicit (56%, 95% CI 51–62%).
- At least 162 of 300 launches publicly evidence Known limitations and out-of-scope conditions are stated (54%, 95% CI 48–60%).

Figure 4. Checks for the decision: Sell, operate, support, or staff repeatedly

share of included launches that publicly evidence the check (confirmed pass)



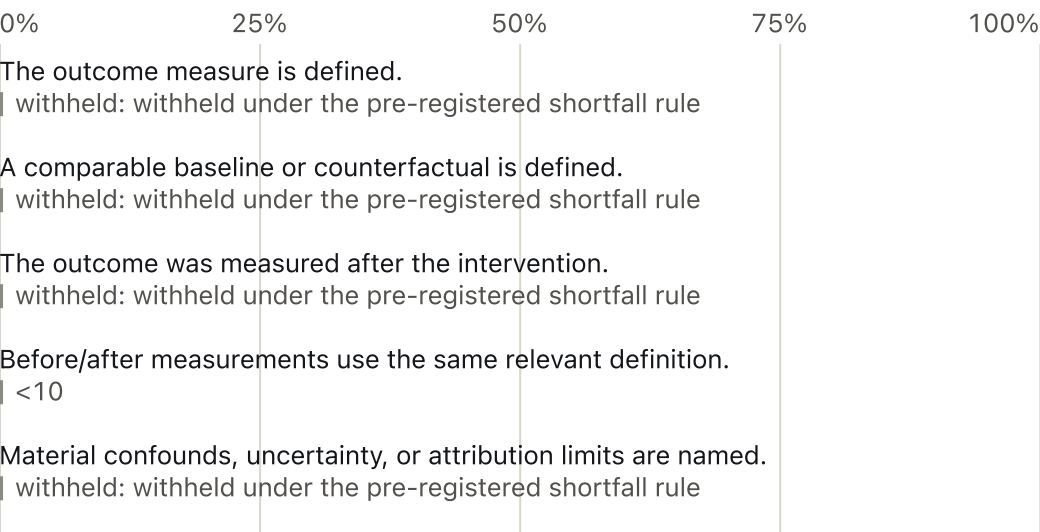
How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.72–0.82 across the shown checks (not estimable for 5); per check on the Validation page. Agreement not estimable for 2 shown checks (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

- At least 44 of 300 launches publicly evidence Fallback or escalation is defined (15%, 95% CI 11–19%).
- At least 104 of 300 launches publicly evidence The relevant operating-cost boundary is known well enough for the decision (35%, 95% CI 30–40%).

Figure 5. Checks for the decision: Determine whether the initiative changed the result

share of included launches that publicly evidence the check (confirmed pass)

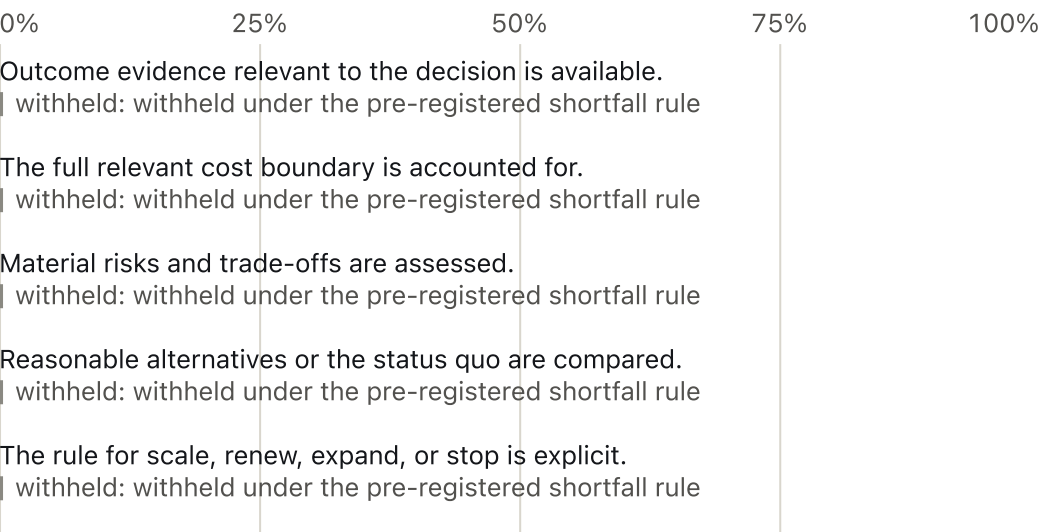


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Agreement not estimable for 1 shown check (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 6. Checks for the decision: Scale, renew, expand, or stop

share of included launches that publicly evidence the check (confirmed pass)

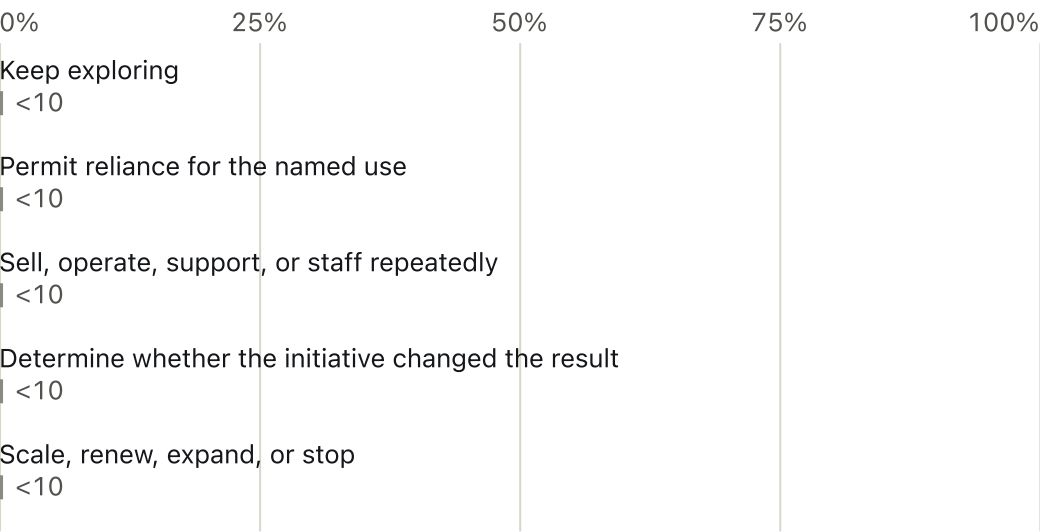


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 7. Evidence by decision

share of included launches whose confirmed check states let the deterministic gate support the decision

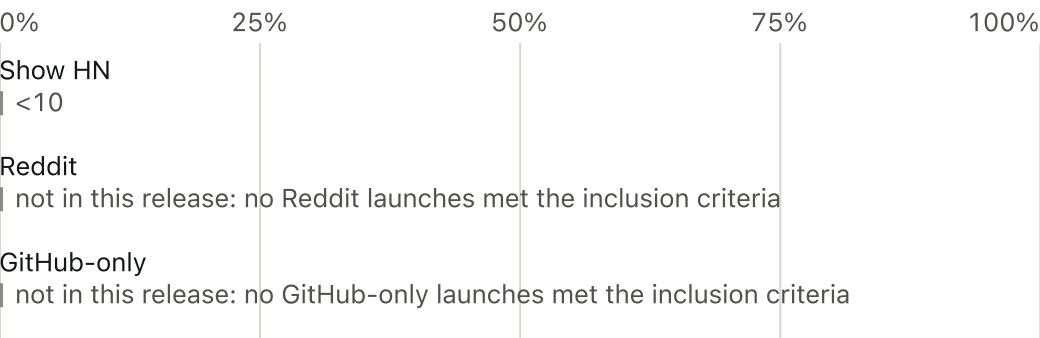


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.72–0.89 across the shown checks (not estimable for 20); per check on the Validation page. Agreement not estimable for 3 shown checks (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Sensitivity	Excluding the 0 launches whose site could not be read: Keep exploring: <10; Permit reliance for the named use: <10; Sell, operate, support, or staff repeatedly: <10; Determine whether the initiative changed the result: <10; Scale, renew, expand, or stop: <10.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 8. Discovery sources: Keep exploring

share of the source's included launches

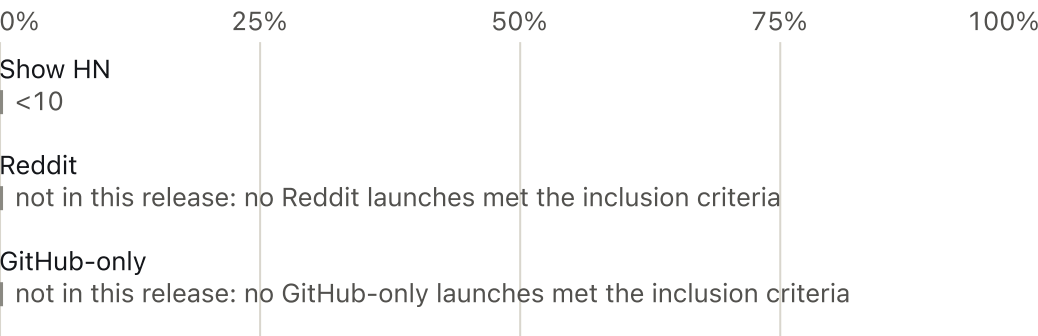


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 9. Discovery sources: Permit reliance for the named use

share of the source's included launches

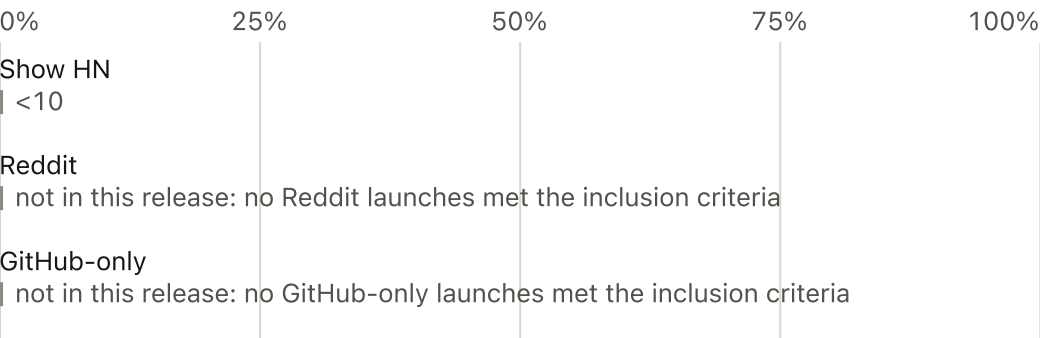


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.86–0.89 across the shown checks (not estimable for 3); per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 10. Discovery sources: Sell, operate, support, or staff repeatedly

share of the source's included launches

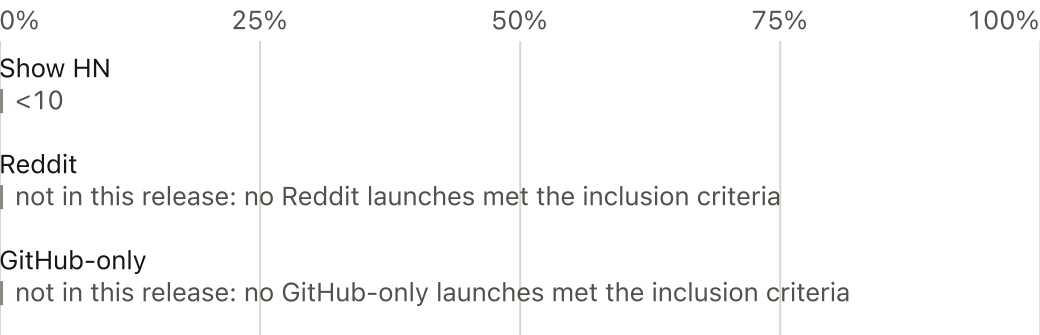


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.72–0.82 across the shown checks (not estimable for 5); per check on the Validation page . Agreement not estimable for 2 shown checks (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 11. Discovery sources: Determine whether the initiative changed the result

share of the source's included launches

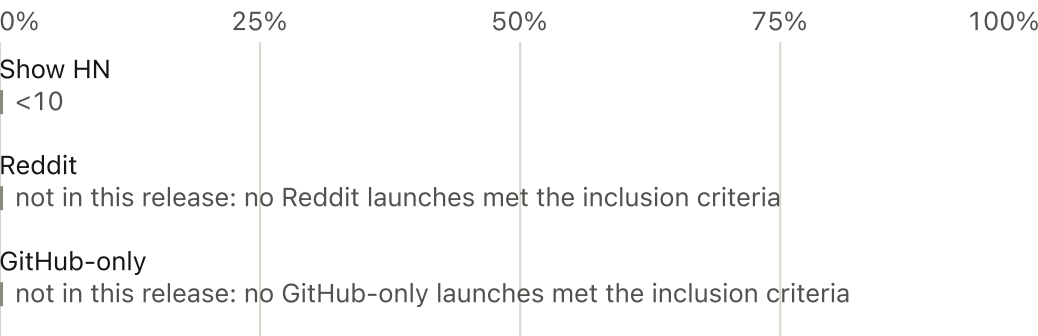


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Agreement not estimable for 1 shown check (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 12. Discovery sources: Scale, renew, expand, or stop

share of the source's included launches



How we know

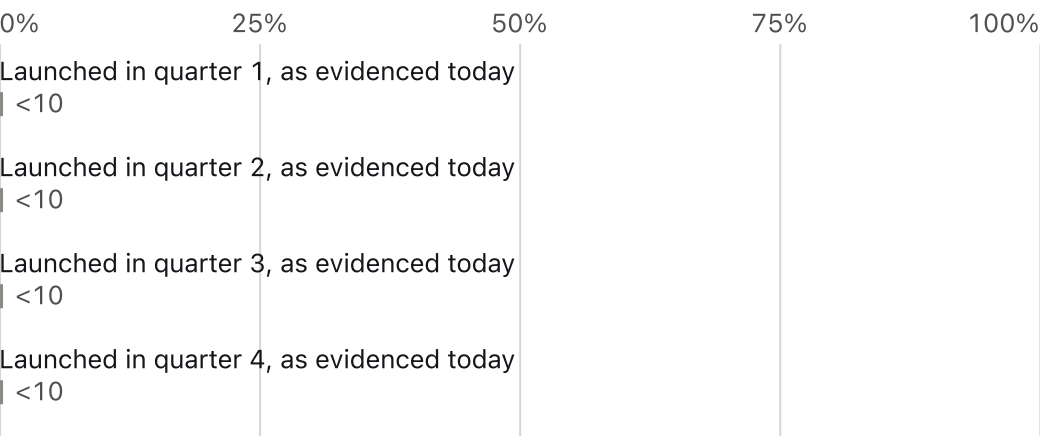
Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Evidence by launch cohort

Each cohort is the launches of one quarter, described by what they publicly evidence today. Older launches have had longer to publish, and launches that disappeared are not in the sample, so these figures are not a trend and are not compared as one.

Figure 13. Launch cohorts: Keep exploring

share of the quarter's included launches, as evidenced today

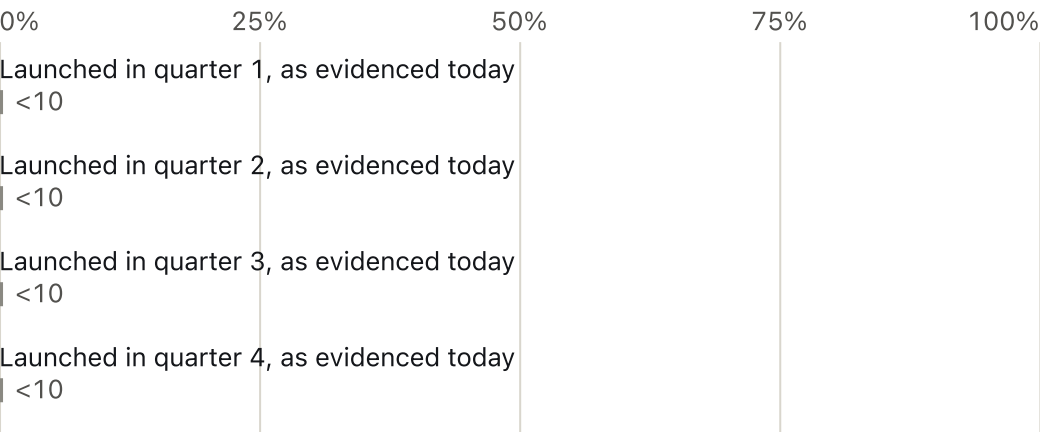


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 14. Launch cohorts: Permit reliance for the named use

share of the quarter's included launches, as evidenced today

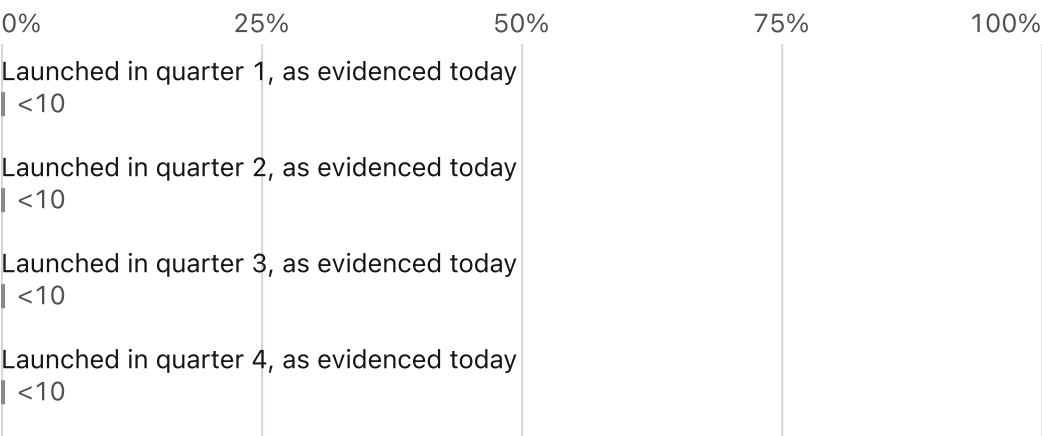


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.86–0.89 across the shown checks (not estimable for 3); per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 15. Launch cohorts: Sell, operate, support, or staff repeatedly

share of the quarter's included launches, as evidenced today

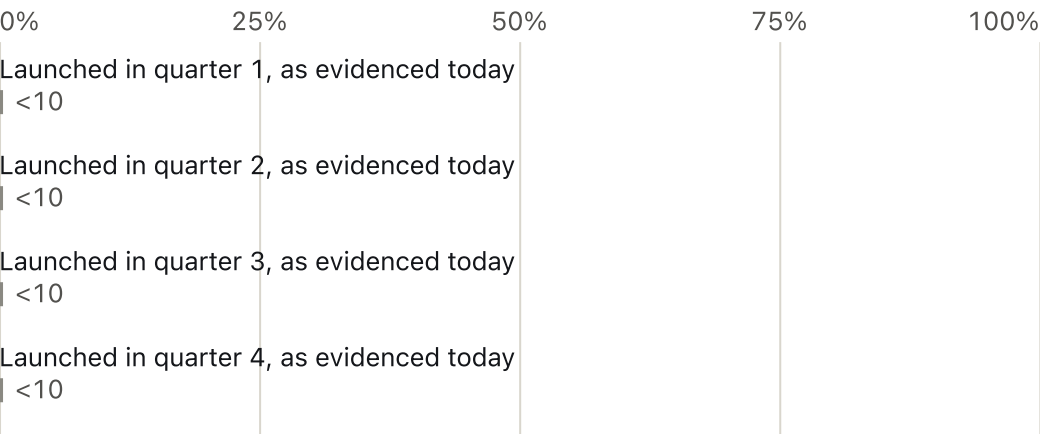


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.72–0.82 across the shown checks (not estimable for 5); per check on the Validation page . Agreement not estimable for 2 shown checks (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 16. Launch cohorts: Determine whether the initiative changed the result

share of the quarter's included launches, as evidenced today

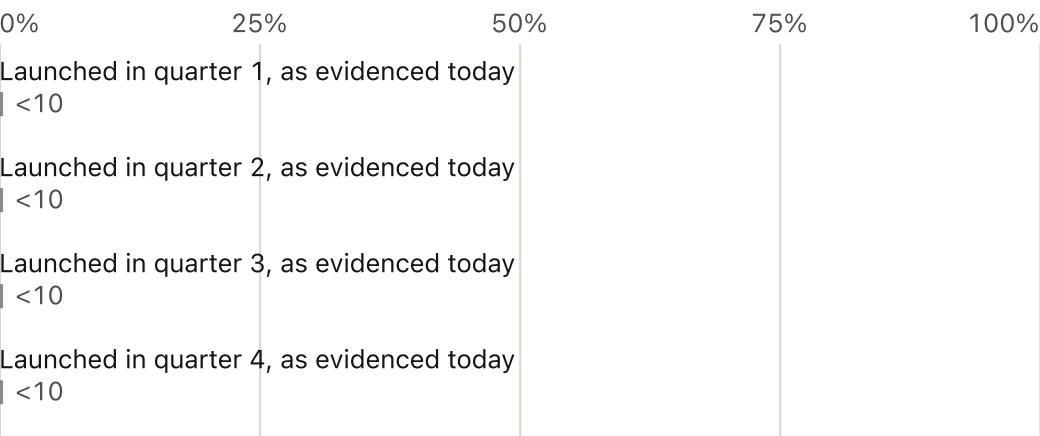


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Agreement not estimable for 1 shown check (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 17. Launch cohorts: Scale, renew, expand, or stop

share of the quarter's included launches, as evidenced today

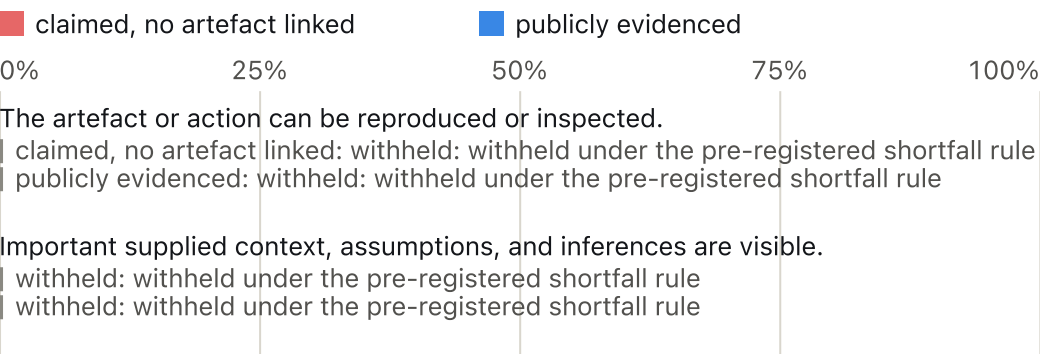


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 18. Claims and evidence (part 1 of 5: explore)

per check: share that claim it without linking an artefact, then share that publicly evidence it

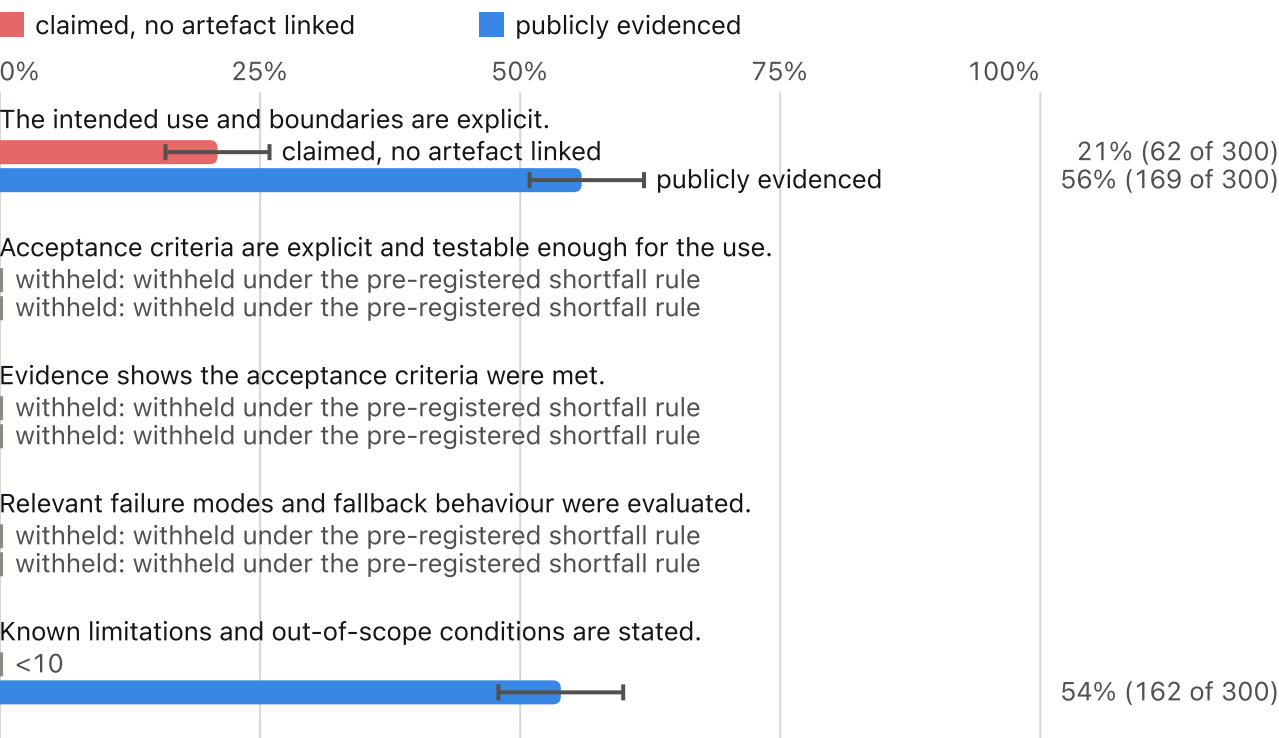


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 18. Claims and evidence (part 2 of 5: rely)

per check: share that claim it without linking an artefact, then share that publicly evidence it

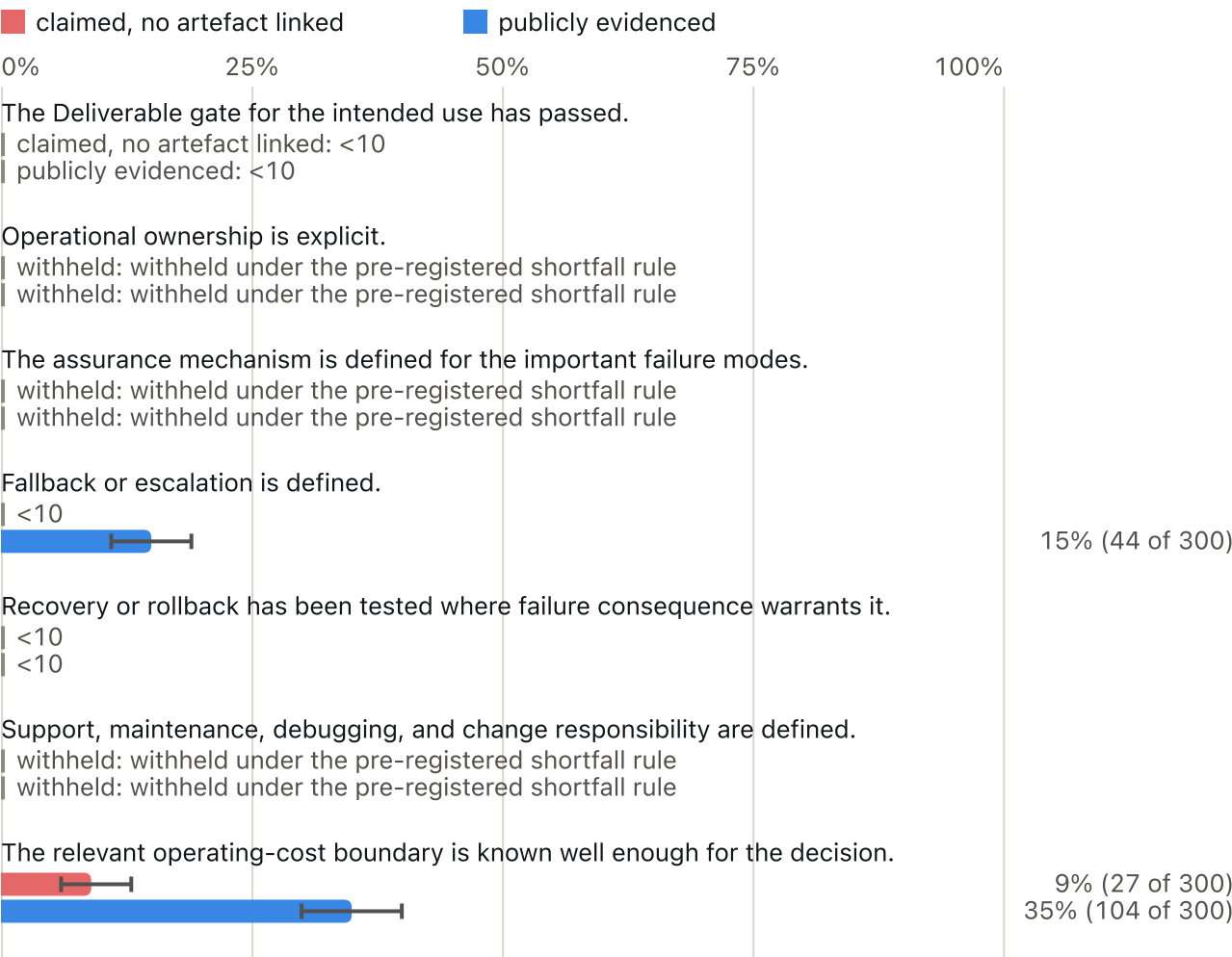


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.86–0.89 across the shown checks (not estimable for 3); per check on the Validation page. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 18. Claims and evidence (part 3 of 5: operate)

per check: share that claim it without linking an artefact, then share that publicly evidence it

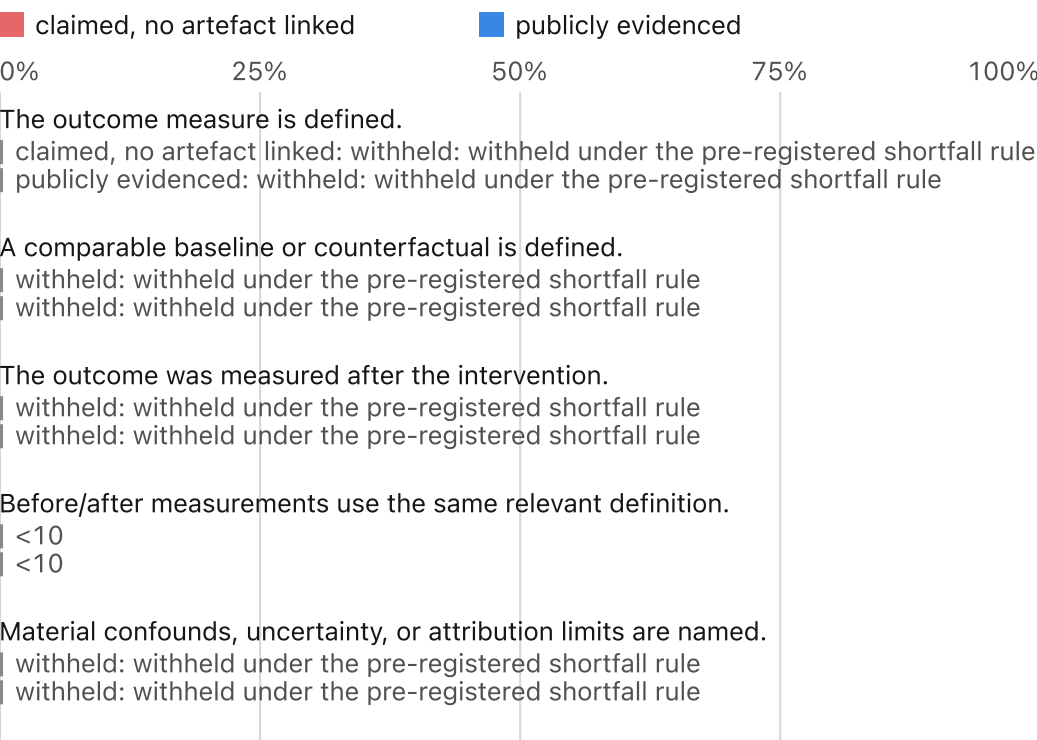


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.72–0.82 across the shown checks (not estimable for 5); per check on the Validation page. Agreement not estimable for 2 shown checks (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 18. Claims and evidence (part 4 of 5: measure-outcome)

per check: share that claim it without linking an artefact, then share that publicly evidence it

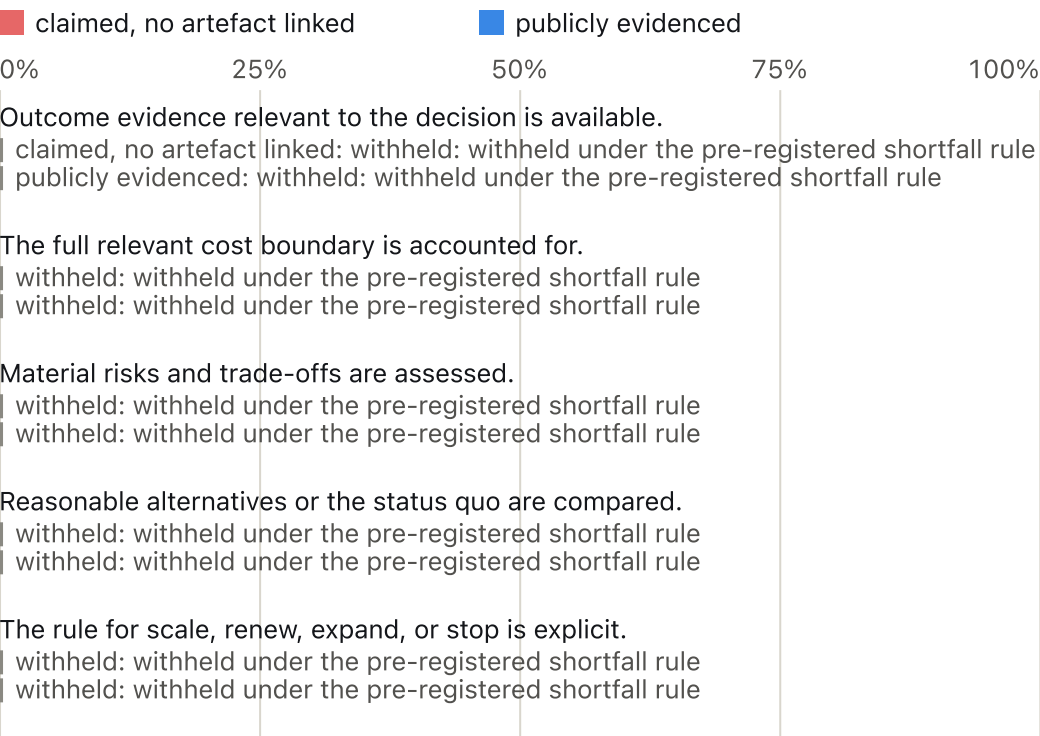


How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page. Agreement not estimable for 1 shown check (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Figure 18. Claims and evidence (part 5 of 5: scale-renew-stop)

per check: share that claim it without linking an artefact, then share that publicly evidence it



How we know

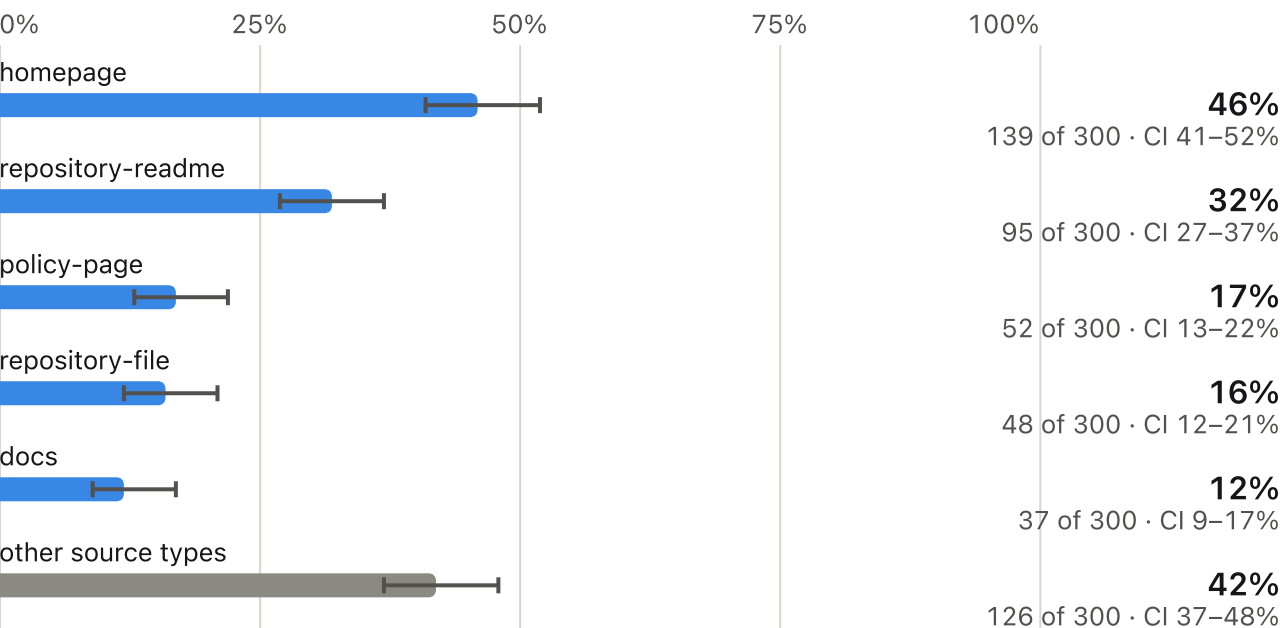
Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 not estimable for the shown checks; per check on the Validation page . Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Where launches put their evidence

One launch can count under several source types, so the shares do not sum to 100%.

Figure 19. Where launches put their evidence

share of included launches with at least one confirmed positive quoted from the source type; one launch can count under several types



How we know

Launches counted	N = 300 included launches. Launches with no evidence found stay in the denominator.
Agreement	Agreement with an independent labeller: AC1 0.72–0.89 across the shown checks (not estimable for 20); per check on the Validation page. Agreement not estimable for 3 shown checks (fewer than ten gold passes); those figures are descriptive. Labeller: glm-5.3-flash on Ollama Cloud.
Changed at confirmation	39.4% of proposed positives changed at confirmation (26% rejected, 14% downgraded to claimed). Confirmer: nemotron-3-ultra on Ollama Cloud.
Who checked	Agent roles, no human edits. Share human-edited: 0%.
Collection	Supported recall 0.894 (59 of 66); raw recall 0.887 (141 of 159).
Release	Release v1, 2026-10-25. Data: public-data/release.json.

Validation

Agreement with an independent labeller

Labeller: glm-5.3-flash on Ollama Cloud. Confirmer: nemotron-3-ultra on Ollama Cloud. Pass precision and the changes at confirmation per check are counts of launches and stay under disclosure control, so a small or suppressed value reads as words.

Agreement, pass precision and changes at confirmation: decision explore

CHECK	AC1 (95% CI)	PASS PRECISION	CHANGED AT CONFIRMATION
The artefact or action can be reproduced or inspected.	not estimable	suppressed (disclosure control)	30% (59 of 195)
Important supplied context, assumptions, and inferences are visible.	not estimable	87% (83 of 95)	35% (33 of 95)

Agreement, pass precision and changes at confirmation: decision rely

CHECK	AC1 (95% CI)	PASS PRECISION	CHANGED AT CONFIRMATION
The intended use and boundaries are explicit.	0.89 (0.84–0.94)	93% (229 of 246)	31% (78 of 252)
Acceptance criteria are explicit and testable enough for the use.	not estimable	75% (60 of 80)	55% (47 of 85)
Evidence shows the acceptance criteria were met.	not estimable	suppressed (disclosure control)	49% (42 of 86)
Relevant failure modes and fallback behaviour were evaluated.	not estimable	suppressed (disclosure control)	<10
Known limitations and out-of-scope conditions are stated.	0.86 (0.80–0.92)	95% (172 of 182)	11% (21 of 183)

Agreement, pass precision and changes at confirmation: decision operate

CHECK	AC1 (95% CI)	PASS PRECISION	CHANGED AT CONFIRMATION
The Deliverable gate for the intended use has passed.	agreement not estimable: fewer than ten gold passes	not estimable: agreement not estimable: fewer than ten gold passes	<10
Operational ownership is explicit.	not estimable	suppressed (disclosure control)	suppressed (disclosure control)
The assurance mechanism is defined for the important failure modes.	not estimable	suppressed (disclosure control)	suppressed (disclosure control)
Fallback or escalation is defined.	0.72 (0.64–0.80)	suppressed (disclosure control)	33% (22 of 66)
Recovery or rollback has been tested where failure consequence warrants it.	agreement not estimable: fewer than ten gold passes	<10	<10
Support, maintenance, debugging, and change responsibility are defined.	not estimable	83% (160 of 192)	46% (88 of 193)
The relevant operating-cost boundary is known well enough for the decision.	0.82 (0.75–0.88)	88% (130 of 148)	30% (45 of 150)

Agreement, pass precision and changes at confirmation: decision measure–outcome

CHECK	AC1 (95% CI)	PASS PRECISION	CHANGED AT CONFIRMATION
The outcome measure is defined.	not estimable	suppressed (disclosure control)	suppressed (disclosure control)
A comparable baseline or counterfactual is defined.	not estimable	<10	<10
The outcome was measured after the intervention.	not estimable	<10	suppressed (disclosure control)
Before/after measurements use the same relevant definition.	agreement not estimable: fewer than ten gold passes	<10	<10
Material confounds, uncertainty, or attribution limits are named.	not estimable	<10	<10

CHECK	AC1 (95% CI)	PASS PRECISION	CHANGED AT CONFIRMATION
Outcome evidence relevant to the decision is available.	not estimable	<10	suppressed (disclosure control)
The full relevant cost boundary is accounted for.	not estimable	<10	suppressed (disclosure control)
Material risks and trade-offs are assessed.	not estimable	suppressed (disclosure control)	38% (14 of 37)
Reasonable alternatives or the status quo are compared.	not estimable	60% (31 of 52)	75% (39 of 52)
The rule for scale, renew, expand, or stop is explicit.	not estimable	withheld: too few coder passes to publish under the disclosure rule	<10

Process figures (addendum A14)

Screening error of the audited exclusions

Screening error 12.2% (5 of 41, 95% CI 5.3–25.5%), above the 10% target; see Limitations. Published under addendum A14, outside the k = 10 rule: a property of the screener, not of any launch.

Supported and raw collection recall

Supported recall 0.894 (59 of 66); raw recall 0.887 (141 of 159). Published under addendum A14, outside the k = 10 rule: a property of the collector, not of any launch.

The count of labels edited by a person

0 of 7200 labels edited by a person. Agent roles, no human edits. Published under addendum A14, outside the k = 10 rule: a property of the process, not of any launch.

The total share of proposed positives changed at confirmation

634 of 1610 proposed positives (39.4%) changed at confirmation: Proposed positives rejected at confirmation: 411 of 1610 (26%, 95% CI 23–28%); Proposed passes downgraded to claimed at confirmation: 223 of 1610 (14%, 95% CI 12–16%). Published under addendum A14, outside the k = 10 rule: a property of the coder and confirmer, not of any launch.

Limitations

1. Screening missed its pre-registered target: 12.2% of audited exclusions were in error (5 of 41, 95% CI 5.3–25.5%) against a 10% bar. This is a selection limitation. A re-measure at 2.4% is exploratory: the rule was tightened on the same 41 cases and was not re-applied to the 300.
2. Supported recall 0.894 (59 of 66; 10 of 10 projects contributed); raw recall 0.887 (141 of 159).
3. Of 159 gathered items, 66 (42%) were supported and 93 (58%) unsupported, 27 (17%) of them boilerplate such as calls to action.
4. The paid search collector was replaced by a free own-site collector for all bulk coding (addendum A8), so every launch was collected the same way. Exploratory: on 142 projects the free and the paid

collectors agree at a positive-level agreement of about 0.78. The all-states agreement of 0.915 is raised by states both found nothing for and is not the headline.

5. A collector defect produced whole-page evidence spans. Every project was re-segmented into spans of at most 1,500 characters, and the decisions made on the defective spans are archived as exploratory and feed no published figure (addendum A9).
6. Every role is an agent: screener (nvidia/Nemotron-3_5-Lightning on Nebius Token Factory); coder (deepseek-ai/DeepSeek-V4-Flash-0731 on Nebius Token Factory); second coder (nemotron-3-super on Ollama Cloud); independent labeller (glm-5.3-flash on Ollama Cloud); confirmer (nemotron-3-ultra on Ollama Cloud); screening auditor (kimi-k3 on Ollama Cloud); recall gatherer (kimi-k3 on Ollama Cloud). No human edited a label. Agreement figures are agreement with an independent labeller.
7. Evidence is read from text extracted from each launch's own pages. What the extractor could not see, such as pages behind a login or text inside images, does not count as public evidence. Launches whose evidence came only from provider-extracted text: 218 of 300 (73%, 95% CI 67–77%).
8. Cohorts are described as they are evidenced today. Older launches have had longer to publish, and launches whose sites disappeared are not in the sample.
9. Not publicly evidenced does not mean not done.
0. The sampling frame is led by Show HN launches, so the figures describe launches posted there and not AI launches in general.
1. Every decision requires at least one check that was withheld at validation. A withheld check enters the deterministic gate as unknown, so no launch can reach PASS for any decision on published states alone. The decision figures show what the released checks can support, not how far launches meet each decision.

Exploratory analyses

Exploratory: collector cross-check (addendum A8)

Exploratory: on 142 projects the free and the paid collectors agree at a positive-level agreement of about 0.78. The all-states agreement of 0.915 is raised by states both found nothing for and is not the headline.

Exploratory: whole-page-span archive (addendum A9)

A collector defect produced whole-page evidence spans. Every project was re-segmented into spans of at most 1,500 characters, and the decisions made on the defective spans are archived as exploratory and feed no published figure (addendum A9).

Exploratory: screening re-measure

A re-measure at 2.4% is exploratory: the rule was tightened on the same 41 cases and was not re-applied to the 300.

Exploratory: community layer (addendum A4)

The community layer (addendum A4) is not part of this release.

Appendix A: codebook and crosswalk

The 24 checks by decision. Each crosswalk entry names a standard, a clause id and how the check relates to it; no standard text is reproduced.

Decision explore

CHECK	CROSSWALK (STANDARD, CLAUSE ID: RELATION)
The artefact or action can be reproduced or inspected. output-reproducible	iso42001 Annex A.6.2.7: supports; gaicc AIMS artefact: Operational documentation: supports; nistAiRmf GOVERN 1: supports; euAiAct Article 11: supports; oecd Principle 1.3: supports
Important supplied context, assumptions, and inferences are visible. assumptions-catalogued	iso42001 Annex A.6.2.3: addresses; gaicc AIMS artefact: Scope: supports; nistAiRmf MAP 1: supports; euAiAct Article 13: supports; oecd Principle 1.3: supports

Decision rely

CHECK	CROSSWALK (STANDARD, CLAUSE ID: RELATION)
The intended use and boundaries are explicit. intended-use-defined	iso42001 Annex A.9.4: addresses; gaicc AIMS artefact: Scope: addresses; nistAiRmf MAP 3: supports; euAiAct Article 13: supports; oecd Principle 1.5: supports
Acceptance criteria are explicit and testable enough for the use. acceptance-criteria-defined	iso42001 Annex A.6.2.2: addresses; gaicc AIMS artefact: Objectives: addresses; nistAiRmf MEASURE 1: supports; euAiAct Article 9: supports; oecd Principle 1.4: supports
Evidence shows the acceptance criteria were met. acceptance-criteria-met	iso42001 Annex A.6.2.4: addresses; gaicc AIMS artefact: Evidence of implementation: addresses; nistAiRmf MEASURE 2: supports; euAiAct Article 15: supports; oecd Principle 1.3: partial
Relevant failure modes and fallback behaviour were evaluated. failure-modes-tested	iso42001 Annex A.6.2.4: supports; gaicc AIMS artefact: AI risk assessment: addresses; nistAiRmf MEASURE 2: addresses; euAiAct Article 9: supports; oecd Principle 1.4: supports
Known limitations and out-of-scope conditions are stated. limitations-stated	iso42001 Annex A.8.2: addresses; gaicc AIMS artefact: Statement of Applicability: supports; nistAiRmf MEASURE 4: supports; euAiAct Article 13: supports; oecd Principle 1.3: addresses

CHECK	CROSSWALK (STANDARD, CLAUSE ID: RELATION)
The Deliverable gate for the intended use has passed. deliverable-gate-passed	iso42001 Clause 9: supports; gaicc AIMS artefact: Management review: supports; nistAiRmf GOVERN 1: supports; euAiAct Article 17: supports; oecd Principle 1.5: supports
Operational ownership is explicit. operational-owner-assigned	iso42001 Annex A.3.2: addresses; gaicc AIMS artefact: Roles: addresses; nistAiRmf GOVERN 2: addresses; euAiAct Article 14: supports; oecd Principle 1.5: addresses
The assurance mechanism is defined for the important failure modes. assurance-defined	iso42001 Annex A.6.2.6: addresses; gaicc AIMS artefact: Evidence of implementation: addresses; nistAiRmf GOVERN 1: addresses; euAiAct Article 9: supports; oecd Principle 1.4: supports
Fallback or escalation is defined. fallback-defined	iso42001 Annex A.6.2.6: supports; gaicc AIMS artefact: Risk treatment: addresses; nistAiRmf MANAGE 2: addresses; euAiAct Article 14: supports; oecd Principle 1.4: supports
Recovery or rollback has been tested where failure consequence warrants it. recovery-tested	iso42001 Clause 8: supports; gaicc AIMS artefact: Internal audit: supports; nistAiRmf MANAGE 4: addresses; euAiAct Article 72: supports; oecd Principle 1.4: supports
Support, maintenance, debugging, and change responsibility are defined. support-maintenance-defined	iso42001 Annex A.4.6: addresses; gaicc AIMS artefact: Operational documentation: addresses; nistAiRmf GOVERN 6: supports; euAiAct Article 17: supports; oecd Principle 1.5: supports
The relevant operating-cost boundary is known well enough for the decision. operating-cost-known	iso42001 Annex A.4.5: supports; gaicc AIMS artefact: AI risk assessment: partial; nistAiRmf MAP 3: supports; euAiAct Article 17: partial; oecd Principle 1.1: partial

Decision measure–outcome

CHECK	CROSSWALK (STANDARD, CLAUSE ID: RELATION)
The outcome measure is defined. outcome–measure–defined	iso42001 Clause 9: addresses; gaicc AIMS artefact: Objectives: supports; nistAiRmf MEASURE 1: supports; euAiAct Article 72: supports; oecd Principle 1.1: supports
A comparable baseline or counterfactual is defined. baseline–defined	iso42001 Clause 6: supports; gaicc AIMS artefact: Scope: supports; nistAiRmf MAP 1: supports; euAiAct Article 9: supports; oecd Principle 1.5: supports
The outcome was measured after the intervention. outcome–measured	iso42001 Annex A.6.2.8: addresses; gaicc AIMS artefact: Evidence of implementation: supports; nistAiRmf MEASURE 3: addresses; euAiAct Article 72: addresses; oecd Principle 1.1: supports
Before/after measurements use the same relevant definition. measurement–definition–consistent	iso42001 Annex A.6.2.8: supports; gaicc AIMS artefact: Operational documentation: supports; nistAiRmf MEASURE 1: supports; euAiAct Article 12: supports; oecd Principle 1.3: supports
Material confounds, uncertainty, or attribution limits are named. important–confounds–named	iso42001 Annex A.5.3: partial; gaicc AIMS artefact: AI risk assessment: supports; nistAiRmf MEASURE 4: addresses; euAiAct Article 9: supports; oecd Principle 1.3: supports

Decision scale–renew–stop

CHECK	CROSSWALK (STANDARD, CLAUSE ID: RELATION)
Outcome evidence relevant to the decision is available. outcome–evidence–available	iso42001 Clause 9: supports; gaicc AIMS artefact: Evidence of implementation: supports; nistAiRmf MEASURE 1: supports; euAiAct Article 72: supports; oecd Principle 1.5: supports
The full relevant cost boundary is accounted for. full–cost–accounted	iso42001 Annex A.4.2: partial; gaicc AIMS artefact: Management review: supports; nistAiRmf MAP 3: partial; euAiAct Article 17: supports; oecd Principle 1.5: supports
Material risks and trade-offs are assessed. risk–tradeoffs–assessed	iso42001 Annex A.5.2: addresses; gaicc AIMS artefact: AI risk assessment: supports; nistAiRmf MAP 5: addresses; euAiAct Article 9: addresses; oecd Principle 1.4: supports
Reasonable alternatives or the status quo are compared. alternatives–compared	iso42001 Annex A.5.2: partial; gaicc AIMS artefact: Risk treatment: supports; nistAiRmf MANAGE 2: supports; euAiAct Article 9: supports; oecd Principle 1.5: supports
The rule for scale, renew, expand, or stop is explicit. value–rule–explicit	iso42001 Clause 9: supports; gaicc AIMS artefact: Management review: supports; nistAiRmf GOVERN 1: supports; euAiAct Article 17: supports; oecd Principle 1.5: supports

Appendix B: figure tables

Figure 1. How launches were selected

BOX	KIND	RECORDS
Records identified in the sampling frame	Selection stage	40,806
Records screened	Selection stage	1,675
Assessed on the text of their own page	Selection stage	963
Included	Selection stage	247
Included with too little page text to judge	Selection stage	53
Excluded at stage 2: not-ai	Excluded	246
Excluded at stage 1: not-ai	Excluded	615
Excluded at stage 2: ai-adjacent	Excluded	178
Excluded at stage 2: unreachable-at-screening	Excluded	82
Excluded at stage 1: duplicate	Excluded	42
Excluded at stage 1: model-or-paper	Excluded	18
Excluded at stage 1: library-no-product-claim	Excluded	29
Excluded at stage 2: library-no-product-claim	Excluded	19
Other reasons (each under ten, merged)	Excluded	12
Not collectable: robots-blocked	Excluded	99
Not collectable: redirected-off-boundary	Excluded	35

Figure 2. Checks for the decision: Keep exploring

ROW	LAUNCHES	SHARE	95% CI
The artefact or action can be reproduced or inspected.	withheld: withheld under the pre-registered shortfall rule		
Important supplied context, assumptions, and inferences are visible.	withheld: withheld under the pre-registered shortfall rule		

Figure 3. Checks for the decision: Permit reliance for the named use

ROW	LAUNCHES	SHARE	95% CI
The intended use and boundaries are explicit.	169 of 300	56%	51–62%
Acceptance criteria are explicit and testable enough for the use.	withheld: withheld under the pre-registered shortfall rule		
Evidence shows the acceptance criteria were met.	withheld: withheld under the pre-registered shortfall rule		
Relevant failure modes and fallback behaviour were evaluated.	withheld: withheld under the pre-registered shortfall rule		
Known limitations and out-of-scope conditions are stated.	162 of 300	54%	48–60%

Figure 4. Checks for the decision: Sell, operate, support, or staff repeatedly

ROW	LAUNCHES	SHARE	95% CI
The Deliverable gate for the intended use has passed.	<10		
Operational ownership is explicit.	withheld: withheld under the pre-registered shortfall rule		
The assurance mechanism is defined for the important failure modes.	withheld: withheld under the pre-registered shortfall rule		
Fallback or escalation is defined.	44 of 300	15%	11–19%
Recovery or rollback has been tested where failure consequence warrants it.	<10		
Support, maintenance, debugging, and change responsibility are defined.	withheld: withheld under the pre-registered shortfall rule		
The relevant operating-cost boundary is known well enough for the decision.	104 of 300	35%	30–40%

Figure 5. Checks for the decision: Determine whether the initiative changed the result

ROW	LAUNCHES	SHARE	95% CI
The outcome measure is defined.	withheld: withheld under the pre-registered shortfall rule		
A comparable baseline or counterfactual is defined.	withheld: withheld under the pre-registered shortfall rule		
The outcome was measured after the intervention.	withheld: withheld under the pre-registered shortfall rule		
Before/after measurements use the same relevant definition.	<10		
Material confounds, uncertainty, or attribution limits are named.	withheld: withheld under the pre-registered shortfall rule		

Figure 6. Checks for the decision: Scale, renew, expand, or stop

ROW	LAUNCHES	SHARE	95% CI
Outcome evidence relevant to the decision is available.	withheld: withheld under the pre-registered shortfall rule		
The full relevant cost boundary is accounted for.	withheld: withheld under the pre-registered shortfall rule		
Material risks and trade-offs are assessed.	withheld: withheld under the pre-registered shortfall rule		
Reasonable alternatives or the status quo are compared.	withheld: withheld under the pre-registered shortfall rule		
The rule for scale, renew, expand, or stop is explicit.	withheld: withheld under the pre-registered shortfall rule		

Figure 7. Evidence by decision

ROW	LAUNCHES	SHARE	95% CI
Keep exploring	<10		
Permit reliance for the named use	<10		
Sell, operate, support, or staff repeatedly	<10		
Determine whether the initiative changed the result	<10		
Scale, renew, expand, or stop	<10		

Figure 8. Discovery sources: Keep exploring

ROW	LAUNCHES	SHARE	95% CI
Show HN	<10		
Reddit	not in this release		
GitHub-only	not in this release		

Figure 9. Discovery sources: Permit reliance for the named use

ROW	LAUNCHES	SHARE	95% CI
Show HN	<10		
Reddit	not in this release		
GitHub-only	not in this release		

Figure 10. Discovery sources: Sell, operate, support, or staff repeatedly

ROW	LAUNCHES	SHARE	95% CI
Show HN	<10		
Reddit	not in this release		
GitHub-only	not in this release		

Figure 11. Discovery sources: Determine whether the initiative changed the result

ROW	LAUNCHES	SHARE	95% CI
Show HN	<10		
Reddit	not in this release		
GitHub-only	not in this release		

Figure 12. Discovery sources: Scale, renew, expand, or stop

ROW	LAUNCHES	SHARE	95% CI
Show HN	<10		
Reddit	not in this release		
GitHub-only	not in this release		

Figure 13. Launch cohorts: Keep exploring

ROW	LAUNCHES	SHARE	95% CI
quarter 1	<10		
quarter 2	<10		
quarter 3	<10		
quarter 4	<10		

Figure 14. Launch cohorts: Permit reliance for the named use

ROW	LAUNCHES	SHARE	95% CI
quarter 1	<10		
quarter 2	<10		
quarter 3	<10		
quarter 4	<10		

Figure 15. Launch cohorts: Sell, operate, support, or staff repeatedly

ROW	LAUNCHES	SHARE	95% CI
quarter 1	<10		
quarter 2	<10		
quarter 3	<10		
quarter 4	<10		

Figure 16. Launch cohorts: Determine whether the initiative changed the result

ROW	LAUNCHES	SHARE	95% CI
quarter 1	<10		
quarter 2	<10		
quarter 3	<10		
quarter 4	<10		

Figure 17. Launch cohorts: Scale, renew, expand, or stop

ROW	LAUNCHES	SHARE	95% CI
quarter 1	<10		
quarter 2	<10		
quarter 3	<10		
quarter 4	<10		

Figure 18. Claims and evidence (part 1 of 5: explore)

ROW	LAUNCHES	SHARE	95% CI
The artefact or action can be reproduced or inspected. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
The artefact or action can be reproduced or inspected. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Important supplied context, assumptions, and inferences are visible. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Important supplied context, assumptions, and inferences are visible. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		

Figure 18. Claims and evidence (part 2 of 5: rely)

ROW	LAUNCHES	SHARE	95% CI
The intended use and boundaries are explicit. (claimed, no artefact linked)	62 of 300	21%	16–26%
The intended use and boundaries are explicit. (publicly evidenced)	169 of 300	56%	51–62%
Acceptance criteria are explicit and testable enough for the use. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Acceptance criteria are explicit and testable enough for the use. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Evidence shows the acceptance criteria were met. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Evidence shows the acceptance criteria were met. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Relevant failure modes and fallback behaviour were evaluated. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Relevant failure modes and fallback behaviour were evaluated. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Known limitations and out-of-scope conditions are stated. (claimed, no artefact linked)	<10		
Known limitations and out-of-scope conditions are stated. (publicly evidenced)	162 of 300	54%	48–60%

Figure 18. Claims and evidence (part 3 of 5: operate)

ROW	LAUNCHES	SHARE	95% CI
The Deliverable gate for the intended use has passed. (claimed, no artefact linked)	<10		
The Deliverable gate for the intended use has passed. (publicly evidenced)	<10		
Operational ownership is explicit. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Operational ownership is explicit. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
The assurance mechanism is defined for the important failure modes. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
The assurance mechanism is defined for the important failure modes. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Fallback or escalation is defined. (claimed, no artefact linked)	<10		
Fallback or escalation is defined. (publicly evidenced)	44 of 300	15%	11–19%
Recovery or rollback has been tested where failure consequence warrants it. (claimed, no artefact linked)	<10		
Recovery or rollback has been tested where failure consequence warrants it. (publicly evidenced)	<10		
Support, maintenance, debugging, and change responsibility are defined. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Support, maintenance, debugging, and change responsibility are defined. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
The relevant operating-cost boundary is known well enough for the decision. (claimed, no artefact linked)	27 of 300	9%	6–13%
The relevant operating-cost boundary is known well enough for the decision. (publicly evidenced)	104 of 300	35%	30–40%

Figure 18. Claims and evidence (part 4 of 5: measure-outcome)

ROW	LAUNCHES	SHARE	95% CI
The outcome measure is defined. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
The outcome measure is defined. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
A comparable baseline or counterfactual is defined. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
A comparable baseline or counterfactual is defined. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
The outcome was measured after the intervention. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
The outcome was measured after the intervention. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Before/after measurements use the same relevant definition. (claimed, no artefact linked)	<10		
Before/after measurements use the same relevant definition. (publicly evidenced)	<10		
Material confounds, uncertainty, or attribution limits are named. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Material confounds, uncertainty, or attribution limits are named. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		

Figure 18. Claims and evidence (part 5 of 5: scale-renew-stop)

ROW	LAUNCHES	SHARE	95% CI
Outcome evidence relevant to the decision is available. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Outcome evidence relevant to the decision is available. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
The full relevant cost boundary is accounted for. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
The full relevant cost boundary is accounted for. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Material risks and trade-offs are assessed. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Material risks and trade-offs are assessed. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
Reasonable alternatives or the status quo are compared. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
Reasonable alternatives or the status quo are compared. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		
The rule for scale, renew, expand, or stop is explicit. (claimed, no artefact linked)	withheld: withheld under the pre-registered shortfall rule		
The rule for scale, renew, expand, or stop is explicit. (publicly evidenced)	withheld: withheld under the pre-registered shortfall rule		

Figure 19. Where launches put their evidence

ROW	LAUNCHES	SHARE	95% CI
homepage	139 of 300	46%	41–52%
repository-readme	95 of 300	32%	27–37%
policy-page	52 of 300	17%	13–22%
repository-file	48 of 300	16%	12–21%
docs	37 of 300	12%	9–17%
other source types	126 of 300	42%	37–48%

Appendix C: release and disclosure summary

Release

v1, 2026-10-25, 300 included launches

Disclosure ledger

v1

Tuples audited

192

Suppressed tuples

117 (117 primary, 0 complementary)

Exact audit

passed: every suppressed count has at least 5 possible values

Wording lint

41 statements, 0 errors

Identifier scan

3 files, 0 hits

Process figures outside $k = 10$ (addendum A14)

- **screening–audit**: screening error of the audited exclusions: a property of the screener, not of any launch (addendum A14).
- **collection–recall**: supported and raw collection recall: a property of the collector, not of any launch (addendum A14).
- **person–edits**: the count of labels edited by a person: a property of the process, not of any launch (addendum A14).
- **confirmation–total**: the total share of proposed positives changed at confirmation: a property of the coder and confirmer, not of any launch (addendum A14).

Release v1 · 2026-10-25 · aggregate figures only · Data: public-data/ · Licence: Apache-2.0