PUBLIC DIAGNOSTIC EXEMPLAR · 02 IN READING ORDER

OUTLIER

/ AI · PLATFORM RELIABILITY

STRUCTURAL DIAGNOSIS

Recurring Platform Failure — from publicly observable evidence.

> *"Outlier is recruiting experts to do compliance work.*
> *That is why nothing else has worked."*

ARTIFACT FAMILY

#	ARTIFACT	ACCESS
01	Translation Artifact	**Public**
02	**Diagnostic Artifact**	← ***This document***
03	Redesign Executive Summary	**Public**
04	Redesign Artifact	**Restricted**

PREPARED BY Jamie Forrester · Independent Systems Architect · Edinburgh, UK
CONTACT hello@jamieforrester.com · jamieforrester.com
DATE April 2026
CLASSIFICATION Public · 02 in reading order · derived from public evidence

EVIDENTIARY STANDARD	JUDGMENT BOUNDARY
This artifact distinguishes directly evidenced facts **[A][B]**, convergent public symptom patterns **[C]**, consequence evidence **[D]**, and structural inference **[E]**. The governing diagnosis is a high-confidence structural inference from convergent public evidence — not a claim about internal intent. Full source documentation is in Section 14.	This artifact makes no claim about intentional deception, fraud, or bad faith by Outlier or Scale AI. The mismatch described is consistent with rapid scaling under conflicting commercial pressures, not with malice or misconduct. No inference is made about internal dashboard logic, undisclosed client instructions, or adjudicated legal liability.
Evidence-class discipline: [C] sources are convergent symptom evidence — not proof of any individual event. [D] sources are consequence evidence and, where legal proceedings remain unresolved, allegation-level evidence only. [E] marks structural inference derived from convergence across public evidence classes. No single complaint, lawsuit, review, or forum post is treated as sufficient proof.	**What this artifact does not say:** It does not claim Outlier or Scale AI acted in bad faith. It does not claim adjudicated liability. It does not claim access to internal communications or proprietary systems. It does not claim all contributors are misclassified. It does not prescribe a technology solution.

SECTION 0

WHAT THIS DOCUMENT IS

This is a structural diagnosis from public evidence. It is not a legal brief, an operational audit, a moral judgment, a technology assessment, or a redesign specification. It identifies the governing primitive, the first structural break, the function collapse condition, the compensatory burden layer, and the dependency order for correction — derived entirely from publicly observable system behaviour.

The **Redesign Artifact** (Artifact 04 in this family) specifies the target-state architecture, transition sequence, and implementation logic. This document defines what must be corrected and in what order. It does not specify how. For role-level implementation, see Redesign Artifact, Section 16.

RECOMMENDED READING ORDER	
Start here	Translation Artifact (Artifact 01) — plain-language explanation for all audiences
Full evidence	Diagnostic Artifact (Artifact 02 — this document
Replacement logic	Redesign Executive Summary (Artifact 03 — public board summary
Builder spec	Full Redesign Artifact (Artifact 04) — restricted; not published publicly

SECTION 0.2

PLAIN-LANGUAGE GLOSSARY

Technical terms used in this artifact, defined for non-specialist readers.

TERM	PLAIN MEANING
Ghost structure	Hidden human work — Queue Managers interpreting queue logic, contributors maintaining their own tracking tools, community members building open-source patches — that the platform architecture should carry but does not.
Collapse node	The single point where too many system failures converge. Here: the one-frame contributor architecture that compresses two incompatible work types into the same gate, scoring layer, and interface.
Sensing-authority fusion	When the system's quality detection (sensing) is directly coupled to irreversible consequence (deactivation, pay reduction) with no human review layer in between. A mis-specified sensor with executive authority produces systematic error at scale.
Generative reframing	The capability to detect that a rubric, task definition, or boundary condition is itself wrong — and surface that signal to a layer that can act on it. Currently absent from Outlier's architecture at every contributor tier.
Asset depreciation	The economic framing of recruitment waste: the platform spends intake cost to acquire credentialed specialist capability, then places it in a system that depreciates that capability to zero by treating specialist variance as a failure signal.
Signal switch	The moment a contributor who entered through expertise-coded recruitment receives their first reliability-coded rubric. The question shifts from ‘use your expertise’ to ‘match the manual’.

DIAGNOSTIC SPINE

The complete diagnostic finding in governed form. Every section of this artifact traces back to one or more rows of this table.

DIAGNOSTIC ELEMENT	FINDING
Governing primitive	Expertise-coded reliability pipeline
Stated promise (paraphrased)	Expert human judgment improves advanced AI systems
Actual operating object	Variance-controlled instruction adherence at scale
First structural break	Qualification-to-execution handoff — where expertise-coded selection meets reliability-coded execution for the first time
Function collapse	Sensing fused with authority: automated scoring detects quality and triggers consequence (deactivation, pay reduction) with no human review between detection and action
Missing function	Portable cross-project contributor memory — system resets to zero at every project boundary
Hidden burden carriers	Queue Managers (ambiguity absorption); contributors (queue-state inference, unpaid onboarding); community tooling (EmptyQueue-Extension, GitHub)
Correction prerequisite	Name the primitive correctly at leadership level before redesigning any gate, stream, or interface

DIAGNOSTIC ELEMENT	FINDING
Proof condition	**Same pressure → different response.** The redesign is real only when the failure pattern does not return under identical structural conditions.

SECTION 1

EXECUTIVE SUMMARY

Outlier has a recurring platform failure that its current correction loops cannot resolve. Credentialed contributors fail production. Quality scores do not predict output quality. Qualified contributors sit idle in Empty Queue states. High-signal specialists churn fastest while reporting the work is not what they were recruited to do. Successive interventions — stricter credentials, longer guidelines, Marketplace visibility, dashboard redesigns, Queue Manager support, webinars — have not reduced the core friction. They have only moved it.

GOVERNING DIAGNOSIS

Outlier recruits through the grammar of an expertise market while operating through the mechanics of a reliability pipeline.

Directional cost anchor [C][D][E]: With 100,000+ active contributors [A] and documented high churn among high-signal cohorts [C][D], the structural mismatch generates compounding costs across recruitment waste, QM support overhead, legal tail-risk, and upstream signal degradation. Each month of deferral deepens the mismatch because new credentialed contributors continue entering the wrong execution frame. See Section 11 for the cost-class breakdown.

An expertise market purchases open-ended specialist judgment and values correct variance. A reliability pipeline purchases repeatable instruction adherence under constraint and suppresses variance through qualification gates, unitised tasks, rubrics, and continuous quality control. Outlier's recruitment surface signals the first. Its execution layer runs the second.

The first structural break occurs at the qualification-to-execution handoff. Contributors are selected through expertise-coded signals — degrees, LinkedIn profiles, domain screenings, identity verification — then moved into success conditions dominated by rubric adherence, timing, formatting, and bounded outputs. The work shifts from *use your expertise* to *match the manual*. The expertise is not consumed. Compliance is consumed. More precisely: **the scoring layer treats specialist variance as a failure signal. A contributor whose domain knowledge leads them to diverge from the rubric will be algorithmically downgraded. The system does not need to intend this outcome for the structure to produce it: where specialist judgment conflicts with rubric conformity, the visible scoring architecture appears to reward conformity over correct domain variance.**

This is why every prior fix has failed. Stricter credentials tighten a gate already measuring the wrong thing. Longer guidelines intensify the signal-switch, converting expert labour into compliance labour earlier and more completely. Each intervention addresses its own layer. None touch the primitive.

A second-order consequence is compensatory human labour. Queue Managers — contributors, not employees — function as human bridges between opaque routing, shifting guidelines, and contributor confusion. Their role is load-bearing, not supplementary. A third-order consequence:

community members built the EmptyQueue-Extension specifically to surface the 19 EQ reason codes not surfaced in the contributor-facing interface. When the system's opacity is corrected by open-source tooling, the structural failure has exceeded its own carrier layer.

BOARD VERDICT

Outlier is an expertise-coded reliability pipeline. Its recurring failure is not a contributor problem. It is a category error embedded in the system's primitive. Correct the primitive, or every future fix will fail in the same way.

SECTION 2

ENGAGEMENT SCOPE

System under analysis

The Outlier contributor platform — the full contributor lifecycle from recruitment surface through qualification, onboarding, task routing, quality-scoring, queue-state, support layers, payment, and the external consequence layer where legal and reputational pressure have appeared.

Framing note

The structural mismatch diagnosed here is not a moral failure. It is the predictable output of a system scaled rapidly under commercial pressure — Scale AI's obligation to deliver reliable, high-throughput training data to frontier AI clients — while its public recruitment surface borrowed the legitimacy grammar of expert sourcing. The two requirements point in opposite directions. The mismatch is structural, not intentional.

Parent system context

Outlier's governing constraint is Scale AI's upstream commercial mandate: reliable, scalable human-generated inputs for AI clients including OpenAI, Meta, Microsoft, Google, and Anthropic [A]. This constraint is treated as fixed.

Evidence basis

This diagnosis is based entirely on publicly available materials. Primary sources [A]: Outlier's official website, FAQ, blog posts, specialist job pages, Marketplace announcements, Community Guidelines, Terms of Service; Scale AI official materials; Oxford Economics commissioned report (December 2025). Symptom evidence [C]: Glassdoor, Indeed, Trustpilot, Quora, Reddit (r/outlier_ai), BBB complaints, GitHub (EmptyQueue-Extension). Consequence evidence [D]: McKinney v. Scale AI (December 2024), Rogowicz PAGA (January 2025), psychological harm class action (January 2025), Bloomberg WARN Act (October 2024), TechCrunch DOL investigation (March 2025), The Guardian (April 2026), Oxford Internet Institute labour standards assessment (2023).

Fixed constraints

Contractor classification model; client NDA and codename structure; demand-driven project availability; algorithmic quality scoring backbone; core task types (rating/ranking, prompt creation, editing, annotation). Internal systems, proprietary routing logic, client contracts, and internal quality

dashboards are excluded. All legal references are allegation-level evidence, not adjudicated findings.

SECTION 3

FAILURE PATTERN RECONSTRUCTED FROM PUBLIC EVIDENCE

The following represents the operating problem as the public record reconstructs it — in the platform's functional language, before structural translation. These statements are derived from convergent public sources [A][C], not from undisclosed internal communications.

CLAIM TYPE	HOW THIS ARTIFACT TREATS IT
First-party platform statements [A][B]	Direct evidence of stated promise or operating surface. Treated as the most reliable source class.
Contributor reports [C]	Convergent symptom evidence. No individual report is treated as proof. Pattern across many sources is treated as structural signal.
Legal filings [D]	Allegation-level consequence evidence. Legal proceedings are unresolved. Claims are treated as consistent with the structural diagnosis, not as adjudicated proof of it.
Media and investigation reports [D]	Public consequence evidence. Treated as corroborating the consequence layer, not as primary structural proof.
Structural diagnosis [E]	Inference from convergence across all source classes. Clearly marked [E]. Not a claim about internal intent or undisclosed systems.

→ Contributor success at intake does not reliably predict success inside the production workflow.

→ High-signal contributors — especially those with advanced specialist backgrounds — either fail early assessment gates, enter production and encounter work materially different from what they expected, or churn rapidly.

→ People who pass the QA process are not delivering the depth of response clients actually need.

→ Quality variance persists despite increasingly strict compliance layers.

→ The Empty Queue is endemic. Contributors qualified for specific work cannot access that work.

→ Support functions — Queue Managers, webinars, office hours, war rooms — have become permanent compensatory infrastructure.

These are not isolated complaints. The pattern is identical across qualification, production, routing, and retention — the first structural signal that this is not an execution problem.

SECTION 4

EARLY STRUCTURAL SIGNALS

Five structural signals are visible from the public record before the formal diagnosis is applied.

1. The two grammars diverge at the handoff.

The recruitment surface is uniformly expertise-coded: degrees, LinkedIn profiles, academic verification, specialist domain selection [A]. The execution surface is uniformly reliability-coded: standardised tasks, time limits, 50+ page rubrics, algorithmic scoring by formatting compliance and time-on-task [A][B]. Two different grammars describing two different systems, operating on the same contributor population with no reconciliation mechanism between them.

2. Every correction has been interface-level.

Every public response to the recurring failure has addressed a surface: routing interface, navigation interface, compliance interface, intake filter [A]. Because the system has not explicitly declared its true production primitive, correction efforts remain confined to interface-level adjustments rather than primitive-level redesign.

3. Compensatory human structures have become load-bearing.

Queue Managers exist to absorb system ambiguity that the platform's own information architecture cannot carry [C]. When an organisation requires a permanent human layer to interpret its own operational conditions to the people doing its core work, the system's design has failed at a level below the interface.

4. The external pressure is patterned, not random.

The lawsuits, the DOL investigation, the investigative journalism, and the labour standards score all cluster around the same tension: the gap between what the platform signals it is and what contributors experience when they are inside it [D]. Patterned external pressure is a consequence signal, not a cause.

5. The system's opacity has been externalised into third-party tooling.

The EmptyQueue-Extension (GitHub: andreytakhtamirov) was built specifically to surface the 19 EQ reason codes visible in the platform's developer tools but not surfaced in the contributor-facing interface [C]. When contributors must build open-source tooling to access information that exists in the platform's own developer environment, the system's coordination costs have been externalised beyond its own human carrier layer. This is strong evidence that queue-state opacity is an information-architecture condition rather than a pure technical limitation: the data exists in the platform environment; the contributor-facing interface does not surface it.

SECTION 4.1

COMMON MISDIAGNOSES

The following explanations are frequently proposed for Outlier's recurring failure. Each is downstream of the actual break and will not resolve it.

MISDIAGNOSIS	WHY IT IS INSUFFICIENT
Contributor quality problem	The gate selects for expertise signals while the production system rewards compliance. High-quality contributors are failing a test designed for the wrong work class.
Guideline clarity problem	Clearer guidelines intensify the signal-switch — they convert expert labour into compliance labour earlier and more completely. More pages deepens the mismatch.
Routing inefficiency	Routing is downstream of the qualification gate. Routing cannot compensate for a gate that admits contributors into the wrong work class.
Support problem	QM support load is generated by structural ambiguity. Adding more support absorbs symptoms; it does not remove the source.
Dashboard / interface problem	Interface visibility cannot correct a wrong primitive. Making the Empty Queue more visible does not correct the mismatch that produces it.
Scaling problem	Scale exposes the mismatch; it does not create it. More contributors through a broken gate produce more failure, not different failure.
Legal / compliance problem	Legal pressure is a consequence of the structural mismatch, not its cause. Changing arbitration clauses or disclosure terms does not stop the system from penalising expertise. It changes the venue of the failure, not the source.

WHY LONGER GUIDELINES MADE IT WORSE — [C]

A concrete example of failed correction: When quality variance persisted, the platform responded by lengthening guidelines from 20 to 50+ pages. The intention was clarity. The visible effect: contributors report spending more unpaid time learning to suppress specialist judgment to match the rubric. The signal-switch intensified. This is not an opinion — it is the pattern visible in convergent symptom evidence [C]. Longer guidelines are the wrong treatment for a primitive mismatch.

SECTION 5

SYSTEM BOUNDARY MAP

Figure 1 — Category Divergence Map

Two incompatible grammars operating simultaneously on the same contributor population, with the qualification gate as the point where the mismatch becomes operational.

FEATURE	RECRUITMENT GRAMMAR (expertise-coded)	EXECUTION GRAMMAR (reliability-coded)
What is signalled	Specialist judgment, credentialed access, expert contribution	Rubric adherence, task completion, compliance

FEATURE	RECRUITMENT GRAMMAR (expertise-coded)	EXECUTION GRAMMAR (reliability-coded)
Success metric	Depth of domain expertise	Formatting, time-on-task, peer reviewer agreement
Feedback loop	Specialist validation	Algorithmic scoring, EQ state
Unit of work	Open-ended problem, judgment call	Bounded task with defined output format
Variance	Valued — correct variance is the product	Suppressed — variance is the failure signal
Memory	Implied portfolio, cumulative expertise	Per-project only, no portable record

Queue Managers appear as a compensatory loop outside the main pipeline, absorbing the ambiguity generated by the grammar mismatch. The missing Upstream Signal Channel is a structural absence — no interface exists for contributors to surface reframing insight to a layer where it can be acted on.

Primary actors

ACTOR	ROLE	POSITION IN SYSTEM
Prospective contributor	Self-selects based on recruitment grammar	Pre-intake
Contributor (Attempter)	Completes entry-level tasks under rubric	Production layer
Contributor (Reviewer)	Quality-assesses other contributors' work	QA layer
Contributor (Oracle)	Senior review and escalation	QA escalation
Queue Manager (QM)	Interprets system ambiguity; hosts webinars and war rooms	Coherence layer (undesigned)
Algorithmic scoring system	Detects quality; triggers access, payment, removal	Authority layer
Project routing system	Matches contributors to available tasks	Routing layer
Client (codename-protected)	Sets task definitions, rubrics, project scope	Upstream
Scale AI	Sets commercial mandate, owns client relationships	Parent

Five points where structural failure concentrates

■ FIVE POINTS OF STRUCTURAL FAILURE

1. What constitutes a passing assessment — currently measures the wrong input class [A][B][C]

2. How quality scores are calculated — currently measures compliance, not expertise [B][C]

3. Why a contributor's queue is empty — currently not surfaced to contributors [B][C]

4. Whether a contributor can appeal a score or deactivation — currently: no mechanism [B][D]

5. Whether upstream reframing feedback is collected — currently: no such interface exists [A][B]

SECTION 6

WORKFLOW AND PIPELINE TRACE

The documented operating sequence:

Diagram key: <-- = signal feeds into this layer · x = structurally absent (no mechanism exists)

```
EXPERTISE-CODED RECRUITMENT
         |
         v
  QUALIFICATION GATE  <--- [expertise signals: degrees, domain, LinkedIn]
         |
         |  <-- SIGNAL SWITCH  (First Structural Break)
         v
RELIABILITY-CODED EXECUTION
         |
         v
 ALGORITHMIC SCORING  <--- [measures: rubric adherence, time-on-task]
         |
   [no human review layer between signal and consequence]
         v
  ROUTING / PAY / REMOVAL CONSEQUENCE
         |
         v
CONTRIBUTOR CHURN  +  QM BURDEN  +  PUBLIC PRESSURE

Structural absences:
  x  Portable cross-project contributor memory
  x  Upstream expert signal channel
  x  Separated review authority (sensing fused with consequence)
  x  Contributor-visible queue state (19 EQ codes not surfaced in contributor interface)
```

Detailed trace

1. Contributor creates account, uploads resume, selects skills, completes ID verification via Persona [B]
2. System cross-references LinkedIn, validates academic email, verifies publications [B]
3. Contributor is offered project-specific onboarding — or hits Empty Queue immediately [A][C]
4. Contributor reads guidelines (50+ pages), may attend webinars — inconsistently compensated [B][C][D]
5. Contributor completes qualification task graded by AI or peer reviewers [A][B]
6. If pass: contributor enters production queue. If fail: no explanation provided [C]
7. Contributor receives unitised task with time limit, rubric, and output format [A][B]
8. Response scored algorithmically or by peer reviewers [A][B]
9. Score affects future task access, project eligibility, tier status [B][C]
10. When demand drops or project ends: blank dashboard, no explanation, no timeline [C]
11. Contributor seeks help from Queue Managers — webinars, Slack, war rooms [C]
12. Contributor either continues through cycles or churns [C][D]

Critical structural observations

→ **No step for upstream reframing exists [A][B].** No mechanism allows a contributor to report that a rubric is wrong, a time limit is unrealistic, or a task definition is malformed. The workflow accepts output only.

→ **Memory is per-project only [A][B].** A contributor who excelled on Project A cannot carry that signal to Project B. Every project begins assessment from zero.

→ **The Empty Queue is not a workflow step. It is a silent capacity management mechanism [B][C].** 19 EQ reason codes are visible in the platform's developer tools (browser network tab) but are not surfaced in the contributor-facing interface. This is an information-architecture condition: the codes exist in the platform environment; contributors cannot access them without external tooling.

→ **Queue Managers are inserted after the fact — a structural patch, not a designed component [C].**

SECTION 7

FUNCTIONAL DIAGNOSIS

Applying the Four-Function Law [E]: institutions fail under scale when sensing, interpretation, authority, and memory remain fused or misaligned at the point of consequence.

FUNCTION	FORMAL LOCATION	CONDITION	STRUCTURAL CONSEQUENCE
Sensing	Algorithmic scoring + peer review	Mis-specified — calibrated to detect rubric adherence proxies rather than the recruited input class	Quality detection operates on the wrong signal. Specialist variance is flagged as failure. [A][B][C]
Interpretation	QM layer + guideline documents	Fragmented — displaced into informal, non-authoritative human support	Meaning of quality varies by project and reviewer. Consistent contributors receive inconsistent scores — no stable signal for routing or contributor trust. QMs interpret but cannot correct the frame. [C]
Authority	Automated system	Fused with sensing — automated consequence without review separation	Publicly visible terms and contributor reports are consistent with a structure in which automated or score-mediated quality signals can trigger access, pay, routing, or deactivation consequences without a sufficiently legible, contributor-accessible review layer. Direct structural pathway to legal exposure. [B][D][E]
Memory	Per-project only	Local and non-portable — no cross-project contributor record	System cannot learn. Routing cannot improve. Every project begins from zero. [A][B][C]

CRITICAL FUSION — STRUCTURAL ORIGIN OF LEGAL EXPOSURE [B][D][E]

The critical fusion: sensing and authority are collapsed in the algorithmic scoring system. A mis-specified sensor operating across 100,000+ contributors and 3.4 million completed tasks produces systematic misclassification as a structural output — not as an occasional error. Publicly visible terms and contributor reports are consistent with a structure in which score-mediated signals can trigger deactivation or pay consequences without a sufficiently legible, contributor-accessible review layer. This sensing-authority fusion is structurally consistent with the legal and regulatory pressure documented in Sections 11 and 14: when a mis-specified sensor has executive authority, its errors are not correctable by the people they affect.

Interpretation fragmentation: Queue Managers are the de facto interpretation layer but have no design authority and cannot correct upstream errors [C]. Glassdoor: *"Directions change every minute and are inconsistent amongst reviewers."*

SECTION 8

STRUCTURAL FAILURE POINT

DIAGNOSTIC LAW

A contributor system fails at the handoff when the signal used to admit contributors differs from the signal used to evaluate them. Outlier is a direct instance of this law.

The first structural break: the qualification-to-execution handoff

This is the moment a contributor who entered through expertise-coded recruitment receives their first project-specific rubric. It is where the system's two grammars meet and where the mismatch becomes operational.

The contributor was selected using expertise signals — degree, domain knowledge, specialist identity [A][B]. At the handoff, they receive a 50+ page document specifying exactly how tasks must be completed: formatting requirements, length limits, time constraints, response structures, scoring rubrics [A][B][C]. The question shifts from *what is the correct answer given your expertise?* to *what answer conforms to this rubric?* When they diverge, the scoring layer sides with the rubric every time.

SHARPEST STRUCTURAL CLAIM — PUBLICLY SUPPORTED [C][E]

Where specialist judgment diverges from the rubric, the visible scoring architecture appears to treat that variance as error rather than value. In that condition, expertise becomes operationally unsafe: the contributor is rewarded for conformity to the rubric, not for correct domain variance. The system does not need to intend this outcome for the structure to produce it. The visible architecture appears to create structural incentives against exercising expertise when expert variance conflicts with rubric conformity. — Glassdoor [C]: "Marked as incorrect for giving the technically accurate answer because it didn't match the rubric wording."

What the break is not

→ **Not a guideline clarity problem** — clearer guidelines intensify the signal-switch [E]

→ **Not a contributor skill problem** — the system is not primarily designed to consume open-ended specialist judgment [A][B][E]

→ **Not a routing inefficiency** — routing is downstream of the handoff [E]

The three capability modes

MODE	DESCRIPTION	HOW OUTLIER HANDLES IT
Execution reliability	Stable instruction adherence under monitoring	Actively selected for. This is the actual operating need. [A][B]
Diagnostic capability	Error detection within defined criteria	Partially present via Reviewer and Oracle tiers. [B][C]
Generative reframing	Detecting that the criteria or boundary are wrong	No interface exists for purchasing or routing this capability. The third mode is structurally absent. [A][B][E]

What generative reframing would look like in practice [E]: A contributor notices that a rubric for a mathematics task asks for 'the correct answer' but the rubric's expected output is wrong for a specific edge case. The contributor submits a structured defect report through a designated interface. The report is routed to a human review node with authority to change the rubric. The rubric is corrected. All future tasks of that type use the corrected version. This capability — surfacing and acting on structural errors upstream — is currently absent from Outlier's architecture at every contributor tier, including the Queue Manager tier.

Evidence of the break

Trustpilot [C]

> *"I took a test for my area of expertise, with no indication of which skills were required. The test asked me about technologies I did not have experience with. Then it removed me."*

Glassdoor [C]

> *"There are often very long tasks that require more time than the paid time. So you are basically working as free labour after that. You then get a warning for using a long time for those extremely long tasks."*

BBB complaint, November 2024 [C]

> *"Since joining Outlier, I have dedicated hundreds of unpaid hours to onboarding processes, training tutorials, assessment tasks, and project-specific modules."*

McKinney v. Scale AI [D] — allegation-level evidence

> *Workers not compensated for time reviewing instructions, attending mandatory webinars, or completing onboarding. Effective hourly rate alleged to be below California minimum wage when unpaid time is included.*

SECTION 9

COMPENSATION AND HIDDEN BURDEN

The platform appears to function. Tasks are completed. Scores are issued. The strongest structural reading is that it functions because a set of informal human carriers is performing the coherence functions the architecture does not formally support.

Primary carrier — Queue Managers

Queue Managers (Attempter → Reviewer → Oracle → Queue Manager) are contributors, not employees [B]. Their documented functions: hosting webinars, running war rooms, moderating Slack, answering questions the platform's documentation cannot answer [C].

Their actual function: absorbing structural ambiguity. When guidelines change without notice, QMs translate the change. When scoring is opaque, QMs provide informal interpretation. When contributors cannot understand why their queue is empty, QMs offer the closest available explanation [C]. This is load-bearing coherence work. If QMs disappeared, the platform would become functionally incoherent at the contributor level.

Critically: QMs cannot correct the frame. They can answer *"how do I comply with this guideline?"* They cannot answer *"is this guideline correctly framed?"* The structural problem generating the

ambiguity they absorb remains invisible to the only human layer with system-adjacent access [C][E].

Secondary carrier — contributors themselves

Contributors carry significant unpaid coherence burden: repeated project-specific onboarding, queue-state inference assembled from community forums and GitHub tools rather than platform information, and guideline interpretation effort not compensated as task time [C][D].

THE EMPTYQUEUE-EXTENSION — THIRD STRUCTURAL CARRIER

The EmptyQueue-Extension (GitHub) was built specifically to surface the 19 EQ reason codes not surfaced in the contributor-facing interface. The 19 codes are visible in the browser network tab. Withholding them is not a technical limitation — the effect is to externalise capacity management costs onto contributors' unpaid time. This creates what the cost record implies is a Shadow Payroll: contributor inference, interpretation, and queue-management labour that does not appear in operating budgets but generates material legal tail-risk [D]. The Empty Queue functions as a silent capacity-management mechanism: contributor availability is preserved while the cost of uncertainty is carried largely by contributors rather than by the platform. When that burden extends into open-source infrastructure, the system has exceeded its own carrier layer.

Economic consequence (allegation-level evidence [D]): Per the Rogowicz PAGA suit: contributors alleged to be working 10-hour days while compensated for approximately 5. The system externalises its coordination costs onto contributors.

SECTION 10

GOVERNING DIAGNOSIS

Outlier's recurring friction is not best explained as a contributor capability problem, a guideline clarity problem, or a routing inefficiency. The structural reading — derived from convergent public evidence across the recruitment surface, operating mechanics, symptom pattern, and legal record [A][B][C][D][E] — is this:

THE GOVERNING DIAGNOSIS — COMPLETE STATEMENT

The platform recruits through the legitimacy grammar of expertise while operating through the control grammar of a reliability pipeline. Because the intake architecture and the operating architecture describe different inputs, the system predictably generates expectation mismatch at the qualification gate, routing opacity at the queue, compensatory support burden at the QM layer, quality variance at the scoring layer, and legal and reputational pressure at the consequence surface. Interface-level corrections cannot resolve this because the failure originates upstream of every interface — at the primitive. The architecture derives legitimacy benefits from expert-coded recruitment while operationally consuming compliance-governed output. These have never been the same thing, and no amount of interface improvement will make them converge until the primitive is named and corrected.

SECTION 10.1

PRIMITIVE MISMATCH TABLE

DIMENSION	STATED	ACTUAL
Primitive	Expert human judgment improves advanced AI systems	Variance-controlled instruction adherence at scale
What is purchased	Open-ended specialist reasoning	Repeatable rubric-conformant output
What is measured	Depth of domain judgment	Formatting, time adherence, rubric alignment
Who succeeds	Domain experts (stated)	Rubric-compliant performers (actual)
Who is penalised	—	High-signal specialists whose expertise produces rubric variance
Who churns	—	Highest-credentialed contributors — those for whom the expert-market promise was most credible

SECTION 10.2

WHY THE MISMATCH IS STRUCTURAL, NOT INCIDENTAL

The mismatch is structurally produced by the collision between Scale AI's obligation to deliver reliable, scalable, quality-controlled inputs to AI clients [A], and the legitimacy grammar required to attract credentialed contributors and reassure those same clients [A][B]. Reliability requires variance suppression. Genuine expertise requires that correct variance be valued. The platform cannot satisfy both requirements inside the same system with the same gate and the same scoring layer [E].

This is not a transitional state that will resolve itself. The commercial mandate produces it continuously. Until the primitive is named and the two work classes are formally separated, the

mismatch will be regenerated by the same forces that created it.

SECTION 10.3

VALIDATION TEST

If Outlier were to implement the Vanish List in Section 12 and internally declare its production primitive as “reliability pipeline with an optional expert judgment stream,” would the remaining friction pattern be explained by execution quality rather than structural mismatch? If yes, the diagnosis holds.

Measurable signals that would confirm the diagnosis without full redesign:

→ Expose EQ reason codes (19 exist in browser developer tools; no new data collection required). If churn among high-signal contributors declines materially after exposure, queue opacity was a primary driver — diagnosis supported. If QM interpretive load falls but specialist churn is unchanged, opacity was a symptom rather than the main driver of churn — diagnosis still holds but the mechanism requires refinement.

→ Pause intake and run a pilot with a reliability-only gate (no credential-density requirement). If churn among new hires drops materially and QA variance decreases, the gate was admitting the wrong input class.

→ Run a single project with an explicit rubric-challenge interface. If contributors surface structural rubric errors that improve task quality when corrected, generative reframing capability exists in the contributor pool but has no structural route.

Falsification condition: If specialist contributor churn persists even after the primitive is correctly named and the gate is rebuilt, revisit whether the fixed constraints in Section 2 prevent full primitive correction. The diagnosis is structurally sound if the symptom pattern maps precisely to the primitive mismatch — which the convergent evidence currently supports [A][B][C][D][E].

STRONGEST FALSIFIER — SECTION 10.3

Strongest falsifier: if Outlier separates its recruitment promise, qualification gate, scoring logic, consequence review, queue-state visibility, and cross-project contributor memory — yet the same specialist-churn, QM-burden, queue-opacity, and quality-variance pattern persists under comparable demand — the diagnosis must be revised. The diagnosis holds if and only if correcting the primitive changes the symptom pattern. If it does not, the fixed constraints in Section 2 prevent correction and the redesign scope must be renegotiated.

SECTION 11

COST OF CONTINUATION

A directional assessment of costs generated by the unresolved primitive mismatch. These are structural inferences from public evidence, not internal financial data.

Recruitment waste — Toxic Asset Accumulation

The platform recruits, screens, verifies, and onboards contributors whose full capability the production layer is not primarily designed to consume [A][B][C]. Every high-credential contributor who churns within one project cycle represents full intake cost with zero retained production value. At 100,000+ active contributors [A] with documented high churn rates among high-signal cohorts [C][D], this is a recurring structural drain.

ASSET DEPRECIATION — NOT RECRUITMENT WASTE

This is not recruitment waste in the ordinary sense. It is asset depreciation. The platform spends intake cost to acquire credentialed specialist capability, then immediately depreciates that capability to zero by placing it inside a reliability stream that treats specialist variance as a failure signal. Structurally implied [E]: every advanced-degree contributor onboarded into the current frame behaves as a high-cost asset depreciated in under 30 days. The platform is not merely wasting money. It is systematically destroying the capital it worked hardest to attract.

Support overhead — structurally recurring

QM infrastructure has ceased to be temporary scaffolding and become permanent compensatory architecture [C]. It cannot be reduced without correcting the underlying ambiguity that creates demand for it.

QA inflation — non-linear under scale

Each compliance layer added in response to quality variance increases QA cost without reducing the variance, because the variance originates at the primitive, not at the execution layer [E]. Execution-layer fixes relocate the variance temporarily.

Upstream client risk — signal mismatch [E]

Because no public contributor-facing mechanism is visible for contributors to surface structural problems with rubrics or task definitions [A][B], malformed task definitions may be scaled rather than corrected. If a client commissions expert judgment but the operating layer rewards rubric conformity, the delivered signal may differ from the capability the recruitment surface implies. This is a structural risk, not a claim about any specific client contract, model outcome, or internal quality standard. The upstream clients listed in public sources [A] may be unaware that the contributor-facing scoring architecture treats specialist variance as a failure signal.

Legal and regulatory exposure — tail-risk severe

Three class action and PAGA lawsuits, a federal DOL investigation, and a pattern of legal action consistent with the structural gap between the platform's expert-facing promise and its reliability-pipeline operating reality [D]. The sensing-authority fusion identified in Section 7 is the direct structural origin of this exposure: score-mediated signals appear capable of triggering deactivation without a sufficiently legible review layer, and that architecture is now the subject of active legal scrutiny. Tail exposure is severe; timing of crystallisation is uncertain.

Reputational damage — directional and observable

National press coverage, investigative journalism, and a 1/10 Oxford Internet Institute labour standards score [D] are consistent with rising contributor acquisition friction, with coverage accelerating between 2024 and 2026.

The compounding dynamic — no internal exit

STAGE	MECHANISM	STRUCTURAL ORIGIN
1	Recruitment waste increases cost pressure	Wrong gate admits wrong population
2	Cost pressure → more algorithmic quality management	No primitive correction; execution-layer response
3	More QA → more scoring opacity	Sensing-authority fusion deepens
4	More opacity → more contributor frustration	No queue-state visibility; no appeal mechanism
5	More frustration → more public complaints + legal action	Ghost structure visibility rises; legal exposure crystallises
6	More compliance → more signal-switch intensity	Guidelines lengthen; expert-compliance gap widens
7	More churn among high-signal contributors	Most credentialed cohort exits fastest
8	More recruitment waste → return to Stage 1	**No internal exit without primitive correction.**

Structural upside if corrected: A correctly scoped reliability pipeline — with honest recruitment, a gate that selects for actual production need, and transparent queue logic — would materially reduce recruitment waste, QM overhead, legal exposure, and contributor churn simultaneously. The redesign is not a cost. It is a recovery mechanism.

SECTION 12

CORRECTION DEPENDENCY ORDER

This is a dependency order, not a list of recommendations. Right fixes in the wrong sequence still fail. This section identifies what must stop and the order in which correction must proceed. The Redesign Artifact (Artifact 02, Section 10, Moves 1–6) specifies the full target-state architecture, transition sequence, and interface changes.

Vanish List — Non-Negotiable Refusal Conditions

What stops before redesign can hold. These are not negotiable starting points for redesign conversations. They are preconditions. Redesign attempted while any of these persist will replicate the mismatch in its first phase.

STOP	WHY	FAILURE IF NOT STOPPED
Using credential density as a proxy for production quality	Credentials predict expertise. The system is not primarily designed to consume expertise. The gate selects the wrong population. [A][B][E]	Wrong population continues entering the wrong stream
Treating longer guideline documents as a response to quality variance	More pages intensify the signal-switch. They convert expert labour to compliance labour earlier and more completely. [C][E]	Expert-market promise continues collapsing into compliance burden
Treating the Empty Queue as a communication problem	19 EQ reason codes exist in developer tools [B][C]. The information exists. The information exists in the platform environment but is not surfaced to contributors.	Ambiguity remains socially carried; support burden stays load-bearing
Continued volume growth before primitive correction	Every additional credentialed contributor added before the primitive is corrected deepens the mismatch.	Redesign begins while old intake compounds structural waste

Correction sequence

Move 1 — prerequisite for everything else

Name the primitive correctly.

The system must decide, at the leadership level, what it is actually purchasing: reliable repeatable instruction adherence at scale, genuine open-ended specialist judgment, or both separated into distinct streams. Every subsequent design decision depends on this. Without it, every downstream fix addresses a symptom while the cause regenerates [E].

Owner

Platform leadership and Scale AI executive layer

Risk if skipped

All subsequent moves will be interpreted through the old frame and will partially replicate the mismatch

Redesign reference

Redesign Artifact, Section 10, Move 1

Move 2 — follows from Move 1

Redesign the qualification gate to select for the actual production need.

The current gate selects for expertise signals. The platform is not primarily structured to consume expertise. The gate must be rebuilt to select for what the system actually requires.

Owner

Product management and contributor operations

Risk if skipped

Any work-stream separation replicates the mismatch in two streams instead of one

Redesign reference

Redesign Artifact, Section 10, Move 2

Move 3 — follows from Move 2

Separate the two legitimate work classes currently collapsed into one broken frame.

The public evidence is consistent with two structurally different kinds of work: constrained reliability execution, and genuine expert judgment [A][B][C][E]. Collapsing them into one frame with one gate and one scoring system is the source of the mismatch.

Owner

Product management

Risk if skipped

Separation becomes cosmetic; mismatch persists in two streams

Redesign reference

Redesign Artifact, Section 10, Move 3

Move 4 — follows from Move 3

Install an upstream signal channel.

No public contributor-facing mechanism is visible for a contributor who identifies a structural problem to surface that signal to a layer where it can be acted on [A][B][E]. This channel does not exist at any tier, including the QM tier.

Owner

Product and engineering

Risk if skipped

Generative reframing capability remains structurally absent; rubric errors scale uncorrected

Redesign reference

Redesign Artifact, Section 10, Moves 4–5

Move 5 — follows from Move 4

Rebuild cross-project memory.

The system currently starts every contributor’s project assessment from zero [A][B]. Demonstrated performance does not carry across project cycles. The memory architecture is in the Redesign Artifact.

Owner

Engineering

Risk if skipped

Routing cannot improve; system cannot learn from contributor track record

Redesign reference

Redesign Artifact, Section 10, Move 3

DEPENDENCY SUMMARY

Move 5 cannot precede Move 4. Move 4 cannot precede Move 3. Move 3 cannot precede Move 2. Move 2 cannot precede Move 1. The current pattern of interface corrections fails precisely because it attempts Move 3 and Move 4 interventions without Move 1 in place. That is why every prior fix has held temporarily and then collapsed.

ACCEPTANCE CRITERIA

WHAT MUST BE IN PLACE BEFORE THE REDESIGN ARTIFACT BEGINS

The Redesign Artifact cannot proceed until the following are confirmed. Without explicit acceptance of the governing diagnosis, every downstream redesign move will be interpreted through the old frame and will partially replicate the mismatch.

1. Leadership formally acknowledges the primitive mismatch: Outlier is a reliability pipeline that has been recruiting through an expertise-market grammar.
2. The Vanish List above is agreed. The four items on it are stopped.
3. The five-move dependency order is accepted as the sequence for correction. No move is attempted before the move that precedes it is complete.
4. Intake volume growth is paused until the new qualification gate (Move 2) is live and tested.
5. The Redesign Artifact (Artifact 04) is commissioned as the next engagement.

SECTION 13

ROLE-LEVEL DIAGNOSTIC IMPLICATIONS

Structural translation of the diagnosis into diagnostic implications by role. These follow the dependency sequence in Section 12. For full system object model, interface specifications, and implementation sequence, see the Redesign Artifact (Artifact 04) .

ROLE	DIAGNOSTIC IMPLICATION	MUST STOP ASSUMING	IMMEDIATE DIAGNOSTIC CHECK	REDESIGN REFERENCE
Engineering	Queue opacity and sensing-authority fusion are structurally material. The 19 EQ codes exist in developer tools; surfacing them is the highest-ratio immediate action. Automated scoring must be separated from deactivation and pay consequence.	That queue opacity is a UI problem.	Can contributors see their actual EQ reason code without external tooling?	Redesign Artifact § 16 — Engineering Moves 1, 4

ROLE	DIAGNOSTIC IMPLICATION	MUST STOP ASSUMING	IMMEDIATE DIAGNOSTIC CHECK	REDESIGN REFERENCE
Product management	The qualification gate is measuring the wrong production input. Redesigning it is Move 2 and cannot proceed until Move 1 (primitive declaration) is complete. Guidelines that change mid-project are a primary generator of QM support load.	That credentials predict production fit in the current frame.	Does the qualification gate test the work actually performed in production?	Redesign Artifact § 16 — Product Moves 2, 3
Operations / QM layer	Queue Managers are carrying system coherence. Their interpretive function is load-bearing, not supplementary. Reducing QM burden requires correcting the system, not cutting QMs.	That more support infrastructure will resolve the ambiguity.	What percentage of daily QM work explains system ambiguity rather than genuine edge cases?	Redesign Artifact § 16 — Operations
Founders and leadership	The primitive must be named before any redesign can hold. Internally accept that Outlier is a reliability pipeline. Pause intake volume growth until the new gate is live. Commission the Redesign Artifact.	That the failure is in execution quality.	Has the platform formally named which work class each stream purchases?	Redesign Artifact § 16 — Leadership Move 1

SECTION 14

APPENDIX — EVIDENCE MAP

This map traces each major diagnostic claim to its evidence class and representative sources. It is designed to make the evidentiary basis of each structural claim impossible to dismiss as reliant on a single source class.

Evidence tier key

TIER	CLASS	DESCRIPTION
[A]	First-party claims	Outlier/Scale AI official surfaces — website, FAQ, blog, job pages, announcements, Community Guidelines, Terms of Service
[B]	First-party operating signals	Platform mechanics and governance documents — onboarding docs, pay structures, tier hierarchy, client codename structure
[C]	Convergent symptom evidence	Independent public platforms — Glassdoor, Indeed, Trustpilot, Quora, Reddit, BBB, GitHub (EmptyQueue-Extension). Treated as pattern evidence, not individual fact.
[D]	External consequence evidence	Legal filings (allegation-level), regulatory action, investigative press — McKinney, Rogowicz, psychological harm class action, Bloomberg, TechCrunch, The Guardian, Oxford Internet Institute
[E]	Structural inference	Frameworks from which the diagnostic logic is derived — The Expertise Illusion (Forrester 2026a), The Four-Function Law (Forrester 2026b)

Claim-to-source map

DIAGNOSTIC CLAIM	EVIDENCE TIER	REPRESENTATIVE SOURCES
Recruitment is expertise-coded	**[A]**	Outlier specialist job pages; Oxford Economics report (84% advanced degrees); LinkedIn cross-referencing in onboarding
Execution is reliability-coded	**[A][B]**	Onboarding documentation; 50+ page rubrics; algorithmic time-limit scoring; pay reduction for time overruns; task type definitions
Empty Queue opacity is endemic and designed	**[B][C]**	19 EQ codes confirmed in developer tools (first-party platform); Reddit r/outlier_ai; Trustpilot; EmptyQueue-Extension (GitHub)
Queue Managers are load-bearing coherence carriers	**[C]**	Glassdoor, Indeed, Quora (tier progression confirmed); Reddit (QM function described); community-acknowledged webinars and war rooms

DIAGNOSTIC CLAIM	EVIDENCE TIER	REPRESENTATIVE SOURCES
Sensing-authority fusion is structural	**[B][E]**	Terms of Service (arbitration; no appeal mechanism); Community Guidelines (deactivation process); Four-Function Law (Forrester 2026b)
Consequence pressure is patterned, not random	**[D]**	McKinney v. Scale AI (Dec 2024); Rogowicz PAGA (Jan 2025); Bloomberg WARN Act (Oct 2024); TechCrunch DOL (Mar 2025); The Guardian (Apr 2026); OII 1/10 labour score (2023)
The governing diagnosis — expertise-coded reliability pipeline	**[A][B][C][D][E]**	Convergent across all five tiers. No single tier is sufficient alone. The structural inference holds if and only if all four lower tiers are consistent with it — which they are.

SECTION 14.1

SOURCE REGISTER

Representative primary sources by evidence tier. This register names specific items rather than source categories. For live engagements, a full source log with archive dates is available on request.

TIER	SOURCE	SPECIFIC ITEM	USED FOR
[A]	Outlier official website	Specialist job pages, FAQ, blog posts, Marketplace announcements	Stated promise, recruitment grammar, official operating claims
[A]	Scale AI official materials	Client list, company description, AI client relationships (OpenAI, Meta, Microsoft, Google, Anthropic)	Parent system mandate and commercial constraint
[A]	Oxford Economics commissioned report	"The Economic Impact of the Data Annotation Industry", Oxford Economics, commissioned by Scale AI, December 2025. Figures used: contributor count (100,000+), completed tasks (3.4m), degree holders (84%). Note: Scale AI-commissioned report; treated as first-party evidence of stated platform composition.	First-party scale and composition claims
[B]	Outlier Community Guidelines	Tier progression (Attempter/Reviewer/Oracle/QM); deactivation policy; appeal absence	Operating signal: consequence structure, tier model

TIER	SOURCE	SPECIFIC ITEM	USED FOR
[B]	Outlier Terms of Service	Arbitration clause; deactivation terms; contractor classification	Operating signal: legal architecture, consequence authority
[C]	GitHub: andreytakhtamirov/emptyqueue-extension	EmptyQueue-Extension; 19 EQ reason codes surfaced from developer tools	Queue-state opacity as information-architecture condition
[C]	Glassdoor / Indeed / Trustpilot / Quora	Convergent contributor reports on signal-switch, rubric adherence, scoring opacity	Convergent symptom evidence; not treated as individual proof
[C]	Reddit r/outlier_ai	Queue Manager function, EQ code discussion, unpaid onboarding reports	Convergent community symptom evidence
[C]	BBB complaint database	Unpaid onboarding hours; November 2024 complaint cited in Section 8	Convergent symptom evidence
[D]	McKinney v. Scale AI	Filed December 2024; unpaid onboarding, below-minimum wage allegations	Allegation-level consequence evidence; unresolved
[D]	Rogowicz PAGA	Filed January 2025; PAGA (California) labour violation allegations	Allegation-level consequence evidence; unresolved
[D]	Bloomberg / TechCrunch	WARN Act (October 2024); DOL investigation (March 2025)	Public consequence evidence: regulatory pressure pattern
[D]	The Guardian	April 2026; investigative coverage of Outlier platform conditions	Public consequence evidence: reputational pressure
[D]	Oxford Internet Institute	Labour standards assessment (2023); 1/10 score on fair pay and worker voice	Independent institutional consequence evidence
[E]	Four-Function Law (Forrester 2026b)	Structural inference framework applied in Section 7	Structural diagnosis [E]; not independently evidenced

SECTION 15

APPENDIX — DIAGNOSTIC ACCEPTANCE PROTOCOL

ACCEPTANCE PROTOCOL

DIAGNOSTIC ACCEPTANCE PROTOCOL (BINDING) This is the protocol used in live engagements. Within 14 calendar days of receiving this Diagnostic Artifact, the commissioning leadership team must return a signed copy of the acceptance checklist below. In live engagements, the Redesign Artifact is only offered after explicit acceptance of the governing diagnosis or a substantive written structural counter-argument. Without that acceptance, the diagnostic engagement closes and no redesign sequence begins. Fees paid to date are non-refundable. In this public exemplar, the client and organisation fields are left unpopulated. The checklist, consequence clause, and signature blocks are reproduced in full. For live engagements, this document is completed, signed, and returned within 14 days. The protocol is identical to what any commissioning client signs. This document was not commissioned by Outlier or Scale AI.

In a live engagement, a diagnostic is only operative if leadership formally accepts it. The following protocol is required before the Redesign Artifact engagement begins. Without formal acceptance, the governing diagnosis can be “noted” and the redesign never commissioned.

Acceptance checklist

■ **Primitive mismatch acknowledged**
Leadership confirms: Outlier is a reliability pipeline that has been recruiting through an expertise-market grammar. This acknowledgment is the prerequisite for every subsequent move.

■ **Vanish List agreed**
The Non-Negotiable Refusal Conditions in Section 12 are stopped. No new volume growth; no new credential-density intake; EQ codes surfaced from the UI; guidelines frozen mid-project.

■ **Dependency order accepted**
The five-move sequence in Section 12 is accepted as the correction sequence. No move is attempted before the move that precedes it is complete. No cherry-picking.

■ **Intake pause confirmed**
Volume growth is paused until the new qualification gate (Move 2) is live, tested, and validated against the production primitive.

■ **Redesign Artifact commissioned**
Artifact 02 (Redesign Artifact) is formally commissioned as the next engagement. The diagnostic engagement is complete at this point.

■ **Expert signal channel acknowledged**
Leadership confirms that no mechanism currently exists for contributors to surface structural errors in rubrics or task definitions, and that this is a required component of the Redesign Artifact (Move 4).

■ **Build boundary accepted**
Stakeholders agree not to claim redesign success from partial implementation. The full minimum replacement requires all five components in the Redesign Artifact.

■ Platform capture resistance noted
Leadership notes that the redesign must not create a new form of institutional capture: expert stream exists but has no upstream task-authority; continuity record exists but is not used for routing. These are named recapture risks in the Redesign Artifact.

Acceptance signatures

ROLE	NAME	SIGNATURE	DATE
Chief Executive Officer			
Chief Product Officer			
Chief Technology Officer			
Platform Operations Lead			

Once signed, this document and the signed acceptance page are delivered to Jamie Forrester as the formal trigger for the Redesign Artifact engagement. Contact: **hello@jamieforrester.com** · **jamieforrester.com**

CLOSING NOTE

CLOSING STATEMENT

This artifact names the system condition that public evidence makes visible: a single contributor frame is being used to coordinate two incompatible work classes. Until that frame is corrected, the same failure will continue to reappear under new interfaces, stricter guidelines, and heavier support layers. The redesign is real only when the same pressure produces a different response.

WHAT THIS DEMONSTRATES

This public diagnostic exemplar demonstrates the diagnostic logic, structural evidence tiering, governing diagnosis, and correction dependency order. It is not sufficient for implementation. The restricted Redesign Artifact (Artifact 04)*** contains the replacement architecture, transition sequence, refusal invariants, recapture gates, admissibility conditions, and builder handoff. This document demonstrates the method. The Redesign Artifact specifies the implementation.***

This artifact was produced by Jamie Forrester, Independent Systems Architect, Edinburgh, UK. It is based entirely on publicly available information and was not commissioned by Outlier or Scale AI.

The diagnostic logic, structural interpretation, and governing diagnosis are the author's own.

For the Redesign Artifact (Artifact 04) or the Translation Artifact (Artifact 01) , contact:
hello@jamieforrester.com · **jamieforrester.com**