r·Potential

METHODOLOGY FOR SUPPLIER UOP LIBRARIES · V1.2

The Alignment Layer

Three Gates. Seven Primitives. The Conditions for Human + AI Co-Creation.

Supersedes v1.1 · Emergence Conditions · Gate 0 canonical · L5 three-axis re-architecture · see Changelog v1.1 → v1.2

PREPARED FOR

OpenAI Executive Leadership
Peter Doolan · Nick Tresser
Google Cloud AI leadership

PREPARED BY

Paul Joly
r.Potential · Product

VERSION & DATE

v1.2 · Apr 24, 2026

01

THE ANSWER FIRST

Executive Summary

Here is how to measure, honestly and independently, where supplier capabilities produce real enterprise outcomes — with human accountability intact, multi-vendor credibility architected in, and the conditions for human + AI co-creation made explicit so that Realized Potential can exceed what any pre-designed plan could predict.

01

Substrate + unit of analysis

L1 Global Labor Graph (verified today: 1.28B postings · 594M profiles · 28.8K canonical skills). L2 UoP primitive — now instantiated within an **Organization**'s **System of Work** in pursuit of redefining a **Coordination Mechanism** (3 new v1.2 Primitives, from Choudary *Reshuffle*). Work-first, human-first, constraint-first.

02

Three gates with configuration between

Gate 0 · Agent Capability (cross-cutting): can the vendor frontier actually do this work today? **Gate 1 · L3 Absorption** (macro, outside-in, 9 dims across OR × SR × EC sub-blocks anchored to the 3 new Primitives): can this Organization metabolize anything? L4 Neutral Shelf + L5 Workforce Plan (three axes: Value Clarity × Agency × Substance) configure the UoP. **Gate 2 · L6 Stakeholder Alignment** (micro, inside-out, three-stage depth): will the coalition commit to adopting the new Coordination Mechanism? `AlignmentScore = min(Gate 0, Gate 1, Gate 2)`.

03

Reality + Realized Potential

L7 Live Telemetry — including human telemetry (actual redeployment vs. plan). **Realized Potential** (Realized Value / Potential Ceiling) is the governing metric tested against ex-ante Alignment Score. v1.2's falsifiable claim: UoPs with high Alignment Score produce Realized Potential *exceeding* Estimated Potential (non-linear human + AI co-creation unlock); low Alignment Score produces decay. Every Anchor Participant deployment tests this.

WHAT'S NEW IN V1.2

Four substantive additions that tighten the methodology from *scoring apparatus* into *human + AI co-creation apparatus*:

1. **Emergence Conditions doctrine** made canonical. UoPs at Estimated/Modeled tier are *hunches-as-artifacts* for human reaction, not ground-truth plans. **Realized Potential** (what actually emerges post-deployment) is the governing metric; the **Alignment Score** is the operational expression of whether Emergence Conditions are met.
2. **Three new Primitives** — *Organization*, *System of Work*, *Coordination Mechanism* (per Choudary *Reshuffle*). Methodology Primitives expand from 4 to 7. Makes explicit what L3 Absorption always implicitly scored. Ratified via workchart-architect v1.2 fold-in (2026-04-24).
3. **Gate 0 (Agent Capability) formally promoted.** The three-gate `min()` composite is now canonical, not a candidate: `AlignmentScore = min(Gate 0, Gate 1, Gate 2)`. Full operationalization in the companion *Alignment Score Deep-Dive*.
4. **L5 Workforce Impact re-architected with three workforce-level Emergence Conditions** — *Value Clarity* × *Agency* × *Substance*. Not a distributional mandate; a condition-setting discipline. Produces the falsifiable claim that non-linear productivity unlock requires all three conditions simultaneously present.

Plus: 2 new Lenses (Strategic Reframe · Adoption Constraint), canonical density-conditional Lens applicability matrix, signed-off sentiment calibration contract for D2 Workforce Readiness. Full list in the **Changelog v1.1 → v1.2**.

Why this methodology, for this audience, now

Doolan and Tresser are operator-class skeptics. Both ex-Salesforce. Both lived the gap between "deployed" and "generating value." A methodology whose weights cannot be defended in real time is dead on arrival. A shelf that shows OpenAI winning every slice is propaganda and they will register it as such within thirty seconds. A workforce story without TAG-grade redeployment substrate is ESG window dressing.

This document is built to survive all three filters. Every number has a provenance. Every weight has a review cadence. Every layer has a named opinionated stance that rejects the received-wisdom alternative. The uncomfortable slice where OpenAI visibly wins less than half is not a concession — it is the structural credibility proof that makes every other number in the document survive its next conversation.

v1.2 sharpens this specifically: the Emergence Conditions doctrine gives every skeptic a single falsifiable claim to test — *do UoPs with high Alignment Score actually produce Realized Potential exceeding Estimated Potential, or not?* — and the methodology commits to publishing that data quarterly across Anchor Participant deployments. If the claim is false, the methodology fails publicly and gets revised. That's what makes it a methodology and not a marketing artifact.

> *"The Stateful Runtime only executes. We help you decide what it should be executing, and we measure whether it worked."*

How to read this document

Section 02 states the root question and the MECE decomposition into seven layers — **journey-ordered in v1.1** (data → unit → gate → configure → configure → gate → reality). Sections 03 through 09 walk each layer in order: L1 Global Labor Graph · L2 UoP Primitive + Archetypes · L3 Absorption Filter (Gate 1) · L4 Neutral Shelf · L5 Workforce Impact · L6 Stakeholder Alignment Feasibility (Gate 2) · L7 Live Telemetry + Flywheel. Section 10 maps the seven layers to the three commercial value propositions landing differently on each executive. Section 11 covers governance and the confidence-tier system. Section 12 names the anti-patterns the loop is designed against. The Uncomfortable Number supplement follows as Section 13 with the empirical shelf grounding and demo-script draft. Appendix C carries the v1 → v1.1 Changelog.

---

02 THE ARCHITECTURE

One Question · Seven Layers

For this supplier's capabilities, in this market, how much will actually get absorbed into real enterprise work that produces real human and economic outcomes — and where?

This is one question, not seven. The seven layers are a MECE decomposition — mutually exclusive so no number is double-counted, collectively exhaustive so no credible objection points at a gap. They are **journey-ordered**: *data substrate → unit of analysis → capability check → macro gate → configure the UoP → configure the human plan → micro gate → reality + feedback.*

The **three-gate shape** (v1.2 canonical) is deliberate. **Gate 0 (Agent Capability)** — cross-cutting — kills UoPs where the vendor frontier can't do the work today at acceptable reliability, regardless of organizational readiness. **Gate 1 (L3 Absorption)** kills the ~80% of pursuits that lack Organizational capacity regardless of UoP design. **Gate 2 (L6 Stakeholder Alignment)** kills the ~20% of those passing both Gate 0 and Gate 1 where the specific UoP will never achieve coalition consensus on the proposed new Coordination Mechanism. All three gates are necessary because they fail orthogonally — an Organization can have a ready frontier with broken absorption, strong absorption with no frontier, or strong frontier AND absorption but a fractured coalition. Any of the three kills deployment.

Alignment Score formal composite (v1.2 canonical):

```
AlignmentScore = min(Gate0 AgentCap, Gate1 OrgAbsorb, Gate2 StakeAlign)
Confidence     = min over all contributing signals (Estimated → Modeled → Measured)
```

The `min()` rule is the MECE-safe composite. A UoP is a chain; the weakest link breaks it; weighted-sums hide chain-breakers. Every score decision becomes "*raise the binding constraint,*" not "*optimize the average.*" Operationally, the Alignment Score is the quantitative expression of whether **Emergence Conditions** are met — the philosophical framing that makes the methodology's product thesis falsifiable (high score → Realized Potential exceeds projections; low → decay).

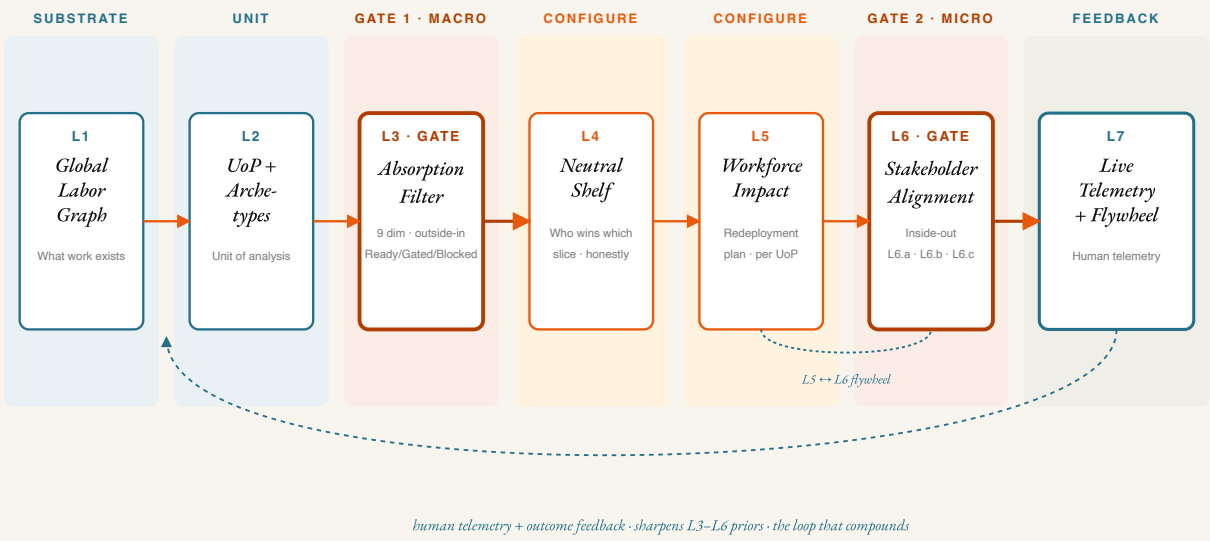

Figure 1 · The 7-Layer Journey (v1.1) — **Gate 1** at L3 (macro absorption) kills ~80% of pursuits; **Gate 2** at L6 (micro alignment) kills the ~20% where stakeholder consensus will never close. L5 ↔ L6 is a bidirectional flywheel; L7 feedback sharpens every prior layer's priors.

> *The received-wisdom alternative we reject. The AI industry builds methodologies tech-first ("here's a capability, find use cases, measure adoption"). We build it work-first, human-first, constraint-first.* ***A UoP is the latent optimal configuration of resources — human and AI — to achieve a business outcome, grounded in business constraints.*** *Start from the outcome and the constraints. Find the latent reconfiguration that achieves it. Flag the projects that sound amazing but will never align internally — those are traps, not opportunities. Automating the past is the Task Fallacy we explicitly reject.*

#	Layer	Sub-question	Role
1	**Global Labor Graph**	What work exists in the market, at what volume, in what shape?	Substrate
2	**UoP Primitive + Work Archetype Classification**	What is the unit of coordination we score, and what are the cross-industry patterns we score it against?	Unit of analysis
3	**Absorption Filter** (outside-in)	Can this enterprise, in general, mechanically absorb AI transformation?	**Gate 1 · macro**
4	**Neutral Shelf** (vendor allocation)	When work gets done, which vendor wins which slice, honestly?	Configuration
5	**Workforce Impact + Redeployment Plan**	What happens to the humans — which roles, which geographies, on what redeployment paths?	Configuration
6	**Stakeholder Alignment Feasibility** (inside-out)	Will the internal coalition actually commit to deploy this specific UoP?	**Gate 2 · micro**

#	Layer	Sub-question	Role
7	**Live Telemetry + Flywheel**	What really happened — outcome + human telemetry — and how does it sharpen L3–L6 priors?	Reality + feedback

01

SUBSTRATE · DEMAND

Global Labor Graph

What work exists in the market, at what volume, in what shape?

The shape of the work that actually exists in the enterprise market — by occupation, skill cluster, industry, geography, standardization, volume — joined from Lightcast, TAG proprietary depth, public stats, and our own UoP Ontology, at a granularity no single participant in the AI supply chain can reproduce internally.

The labor market is a joined dataset from many authoritative sources — it is not any single vendor's product. We aggregate Lightcast (bulk postings + profiles + taxonomies), TAG HR worldwide (proprietary placement + interaction telemetry), public stats (BLS, ILOSTAT, OECD, Eurostat, World Bank, UN), and the r.P UoP Ontology layered on top. No single source is the spine; the joined graph is.

Ambition vs. verified current (April 22, 2026)

Every number carries its confidence tier. Ambition numbers are our target steady-state when Lightcast full ingestion + TAG ingestion + Lattice Q3 are live and canonical dedup is complete. Verified current numbers are what is live in supply-side Supabase + Databricks Delta Share today.

Tier	Source	Scale · ambition	Scale · verified current
Labor market data	Lightcast Delta Share + public stats	3.5B postings · 1.65B profiles · 2.4M enterprises firmographic · 41K skills · 175 countries	**1.28B postings** (2010 → Apr 22, 2026, 5,949 distinct days) · **594M pseudonymized profiles** (192K cities) · **~7.3M raw / ~2M canonical enterprises** (dedup in progress) · **28.8K canonical skills** (narrative says 41K; delta is variant-inclusion)
Proprietary depth	TAG HR worldwide	100K enterprise HR clients · 104M placements/yr · 300M candidate interactions/yr	TAG-operational at stated scale; ingestion into r.P platform in progress
UoP Ontology	r.P-curated	14,200 absorption-scored · 7 archetypes · 2,850 AI-agency providers · flywheel priors	30 seeded · 1 archetype fully scored (Legal Drafting) · 2,850 provider catalogue in progress · derivation from lcdata pending

Tier	Source	Scale · ambition	Scale · verified current
Q3 add	Lattice HRIS	~5M managers, real-time performance + skill telemetry	Planned integration

The honest line. Credibility wins over overclaiming. The ambition-vs-current framing is visible everywhere in the product — every number carries Estimated / Modeled / Measured tier in the UI. Earlier r.Potential artifacts cited GLG at "3.5B / 1.65B / 41K" without the reconciliation. v1.1 retires the unreconciled framing.

WHY OPENAI SHOULD CARE

Frontier sells horizontal capability. Enterprises buy vertical outcomes. The GLG lets a Frontier conversation begin with "*here is the work pattern, here are the 1,200 enterprises where it repeats, here are the three where you already win.*"

EXECUTIVE PRAGMATISM NOTE

A labor graph without placement data is a postings index — what employers *said* they wanted, not what happened. TAG's 104M annual placements are the empirical ground truth that separates Lightcast-plus-something from Lightcast alone. That "something" is the moat.

Role in the cascade: L1 is the ground; it does not filter. Its output is a queryable substrate every downstream layer can pull from.

02

UNIT OF ANALYSIS

UoP Primitive + Work Archetype Classification

What is the unit of coordination we reason about, and what are the cross-industry patterns we score it against?

A **Unit of Potential (UoP)** is a scored, governed decision about how to configure human + AI resources to achieve a business outcome under stated constraints. An **archetype** is a cross-industry work pattern that generates specific UoPs when instantiated for a specific enterprise × slice.

The contrarian inversion. A UoP starts from the business outcome — not from a task, not from an API, not from a model. Outcome-anchored · constraint-aware · governed. What the primitive rejects is the Task Fallacy: automating the past. When stakeholders align on the outcome, the new tasks that achieve it get derived from the configuration — we do not inherit the old task list and automate it slot by slot.

The UoP primitive — three non-negotiable properties

OUTCOME-ANCHORED

A UoP starts from the business outcome — not a task, not an API, not a model. The outcome defines the success criterion the whole stack is measured against.

CONSTRAINT-AWARE

A UoP carries the enterprise's actual constraints — budget, timing, workforce composition, regulatory context, stakeholder map — as first-class inputs, not afterthoughts.

GOVERNED

A UoP is a decision artifact. Every score traces to signals; every signal carries provenance; every ceiling rule is legible; every stakeholder named.

SLICE DECOMPOSITION · THE KEY METHODOLOGICAL ACT

Each archetype decomposes into 4–8 discrete slices (e.g. Legal Drafting: Clause Generation / Precedent Research / Compliance Review / Redlining / Summarization / Adversarial Judgment). Slices are the unit at which vendor allocation (L4) and workforce impact (L5) are measured.

The v1.1 archetype library — seven archetypes

Inclusion threshold: >50 cross-enterprise deployment instances with observable slice decomposition. Below that, it is a use case wearing a framework hat.

Archetype	Role in the library	Status
Legal Drafting & Review	Measured proof anchor. $65M identified, $30M realized, 27% attorney time reallocated (TAG × Harvey). Multi-vendor split visible.	MEASURED
Software Engineering Agents	Neutrality test — Anthropic wins most of this today. Showing it honestly is what earns trust for every other archetype's allocation.	MODELED
Multi-Agent Orchestration	Sells Frontier's TAM — Frontier's addressable market is gated by absorption, not by model quality. r.P unblocks.	MODELED
Customer Support Triage	Volume anchor. Highest token-per-enterprise of any archetype. AgentForce vs. reasoning slices split.	MODELED
Financial Close	Back-office core, high-structure workflows.	ESTIMATED
Knowledge Synthesis (Regulated)	Pharma / financial services — audit-grade, human-decisive fallbacks at every slice.	MODELED
Clinical Documentation	Mission-sensitive, ships behind independent clinical review board.	ESTIMATED

Software Engineering — archetype anatomy (worked in full)

NAMED SLICES

Code generation · code review · multi-file refactoring · test generation · documentation synthesis · debugging / trace reasoning · long-context codebase reasoning · agentic coding loops

SKILL CLUSTER (LIGHTCAST)

Programming languages × framework familiarity × system-design reasoning × toolchain fluency

ENTERPRISE SIGNATURES

Monthly SWE postings >40 as "high activity." LLM-fluency appearing in JD requirements as leading indicator. Wage-inflation signal as absorption proxy.

INTERFACE MODES PER SLICE

Code gen = Draft→Refine · review = Analyze→Decide · refactoring = Execute→Supervise · debugging = Analyze→Decide or Synthesize→Validate (by trace depth)

Pattern taxonomy underneath (canonical, 6 patterns)

Synthesis · Analysis · Judgment · Orchestration · Monitoring · Execution. Every slice of every archetype resolves to one or more of these patterns. SW Engineering spans Analysis (debugging), Execution (code gen),

Synthesis (docs), and Orchestration (agentic loops).

Why 7 and not 9. The inclusion threshold is discipline, not convenience. An archetype no one can point at across multiple enterprises is a use case in costume. Extensible beyond 7 with evidence — today the library is 7, and the absence of an eighth is itself evidence of methodological discipline.

Role in the cascade: L2 defines the unit (UoP) and the patterns (archetypes). Every downstream layer operates on (enterprise × archetype) or (enterprise × archetype × slice) tuples.

03

GATE 1 · MACRO · OUTSIDE-IN

Absorption Filter

Can this enterprise, in general, mechanically absorb AI transformation?

Whether a given enterprise can metabolize AI transformation at all — scored across nine dimensions from outside-in signal, with three ceiling rules each grounded in mechanism, provenance, and override condition. Blocked enterprises are filtered out of the Hot Roadmap before any UoP is configured.

Absorption is knowable from outside data. 79% of enterprises are failing at AI adoption; 54% of C-suite admit AI is "tearing the company apart." The failure mode is not intelligence — it's absorption. You do not need to be inside an enterprise to know whether hiring velocity, tech-stack signals, governance posture, economic capacity, workflow specificity, data readiness, regulatory context, competitive intensity, and stakeholder-tolerance posture support AI deployment. The Global Labor Graph contains enough signal.

Nine dimensions · v1.1 global baseline

Weights are hand-set priors. Each Anchor Participant deployment shifts them via regression on outcome variance. Weights total 100%.

#	Dimension	Weight	What it measures	Signal source
1	**Org Readiness**	**15%**	Change-management maturity · executive sponsorship · prior AI-deployment track record · change-absorption rate	Posting patterns for transformation roles · public announcements · press density
2	**Workforce Capability**	**15%**	Skill-role adjacency · retraining capacity · union/regulatory friction · workforce diversity	Profile histories · posting velocity · ILOSTAT regional data
3	**Tech Infrastructure**	**10%**	API readiness · integration complexity · security posture · cloud-stack maturity	Software-skill signals · cloud/data-stack mentions

#	Dimension	Weight	What it measures	Signal source
4	**Economic Capacity**	**10%**	Budget allocation · ROI threshold · cost-per-work-unit · J-curve tolerance	Salary bands · firmographic revenue · BLS wage anchors
5	**Workflow Specificity**	**10%**	Decomposability of target work · title-cleanliness patterns · process standardization	Title-normalization residuals · NAICS/O*NET mapping precision
6	**Data Readiness** (new)	**15%**	Data quality · governance · access patterns · lineage maturity. Split from Tech Infra because data readiness predicts AI-scalability independently of integration readiness.	Data-engineering posting density · cloud-data-platform mentions · published governance audits
7	**Regulatory Context** (new)	**10%**	Sector-specific clock: pharma / financial services / clinical / regulated utilities move at different cadences than retail / SaaS / advertising.	Industry classification · regulatory-compliance posting signals · published sector-audit posture
8	**Competitive Intensity** (new)	**10%**	Is this a market where AI-laggards get killed (accelerant) or a regulated utility where nobody adopts (decelerant)? Urgency is its own dimension.	Industry margin compression · public AI-adoption announcements per sector · competitor deployment density
9	**External Stakeholder Tolerance** (new)	**5%**	Customer / partner / union / regulator tolerance for visible AI in the workflow. Distinct from internal Workforce — this is the *outside-the-firm* friction.	Public sentiment · union-communication corpus · customer survey data where available

Archetype weight deltas

Weights above are a **global baseline**. Each archetype can override dimensions where a measured deployment has revealed variance. In v1.1 only one archetype has earned a published delta.

Archetype	Published delta vs. global baseline	Source
Legal Drafting & Review	Workflow Specificity +5% (→ 15%) · Economic –3% (→ 7%) · Regulatory +2% (→ 12%)	TAG × Harvey measured deployment, 4 quarters
All others	*Using global baseline; delta pending first Measured deployment.*	—

The point: weights are honest where we've measured, conservative where we haven't. Every Anchor Participant adds a delta; every delta is published with provenance.

Ceiling rules — mechanism · provenance · override

The composite is capped by three empirical ceilings. Each ceiling is a prior, not an opinion — it says "*until we see disconfirming evidence, absorption cannot exceed this level because the underlying condition structurally prevents deployment at scale.*"

Ceiling	Mechanism (why it caps)	Provenance	Override condition
No executive sponsor → 40	Sponsorless transformations collapse at the first budget cycle. Absorption requires sustained political capital the organization will not issue without a named owner who holds budget, outcome, and narrative.	Change-management literature (Kotter 1996; HBR change-failure corpus 2018–2024); TAG's 100K-client placement dataset indicates sponsorless AI pilots survive past month 6 in under 15% of cases. 40 is the empirical upper bound observed in this data.	Named C-suite champion issuing budget → ceiling lifts to composite.
External signals only → 60	External data predicts mechanical capacity, not internal will. 60 is the operationalization of the confidence system's rule that *Modeled confidence is structurally lower than Measured*.	Directly derived from the confidence system. 60 is the maximum composite a score built entirely from outside-in signal can hold without becoming dishonest.	Internal engagement (sponsor interview + stakeholder sit-downs, i.e. L6.a engagement) → score tier becomes Modeled → cap lifts to 85. Full Procensus deliberation (L6.c) → Measured → cap dissolves.
Low governance maturity → 50	Undocumented governance means AI deployments cannot be audited. No audit trail → no board approval → no production deployment at scale. Reflects a structural operational block, not a capability gap.	Posture inferred from GDPR / SOC 2 / ISO-27001 / HIPAA signals in firmographic data plus published regulatory-audit patterns. 50 is the "in principle possible, operationally capped" bound.	Governance uplift plan (named CISO commitment, data-governance review scheduled, AI-audit framework adopted) → conditional lift to 70 with deployment-gated release of full ceiling.

Ceilings are **empirical priors, revised quarterly** as Anchor Participant data accumulates. v2.0 will publish with 95% confidence-interval bands on each ceiling number.

> **v1.1 explicit drop.** v1 contained a fourth rule — a *Nadler-Tushman congruence penalty* deducting 5% from the composite for any dimension pair with >30-point gap. **Dropped in v1.1.** The Nadler-Tushman framework does not operationalize as point-gaps in a scoring composite; the borrowed citation weakened the methodology more than the penalty strengthened it. Congruence variance, where it matters, is surfaced as a named risk signal in L6.b rather than as a numeric penalty at L3.

Absorption tiers — after all adjustments

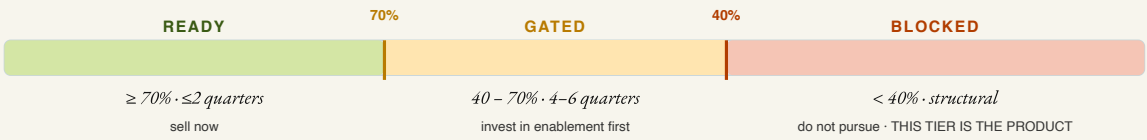

Figure 2 · Absorption tiers. Tiers are *archetype-specific, not company-specific*. The same enterprise can be Ready for Legal Drafting and Blocked for Multi-Agent Orchestration.

Role in the cascade. L3 is **the first stop-here**. Blocked enterprises are filtered out of the Hot Roadmap. The remaining set (~15% of the universe, ~2,180 of 14,200) flows into L4. For hyperscalers, DeployCo and Frontier Alliance use L3 tiers for weekly account prioritization: Ready accounts get immediate motion, Gated accounts get enablement-investment plans, Blocked accounts get deferred. The Demand Map globe is literally an L3 visualization.

04

CONFIGURATION · VENDOR ALLOCATION

Neutral Shelf *(internally: Category Captain)*

When work actually gets done for a given archetype, which vendor wins which slice — honestly?

When work actually gets done, which vendor's capability does which slice at what quality — allocated honestly from Anchor Participant telemetry · market signal · benchmarks, each cell labeled with its evidence strand, with OpenAI visibly winning less than half of at least one slice as the structural credibility proof. Capability-per-vendor-per-slice is what the Shelf measures.

A shelf where every slice routes 100% to one vendor is advertising, not intelligence. Honest allocation is the leader's commercial asset, not a constraint. Markets are not winner-take-all at the slice level — they are winner-take-most. The methodology that publishes the slice-level truth compounds trust with buyers, and the leader benefits disproportionately, because a trustworthy shelf attracts buyer attention and a rigged shelf repels it. Received wisdom says "never show yourself losing." Received wisdom is wrong for the leader, and mostly right for the challenger. OpenAI is the leader.

Role change from v1. This layer was called "Category Captain Shelf" in v1 — retained as internal shorthand, but renamed "Neutral Shelf" for external surfaces. The antitrust literature on retail Category Captains (FTC 2003, Gundlach-Loff 2018) is largely critical; any hostile reviewer lands on "*r.Potential is replicating the anti-competitive structure the FTC flagged, funded by the dominant vendor*" in one paragraph. The rename neutralizes the metaphor without changing the mechanism. UL / Moody's / independent-scoring framings are the defensible external analog.

Three evidence strands · each with its own confidence tier

Each (vendor × slice) cell in the Shelf carries its strand label. As Anchor Participant telemetry accumulates, Strand A expands and displaces B/C for those slices.

Strand	Confidence	Mechanism
A. Anchor Participant telemetry	MEASURED	Direct deployment outcome data from signed Anchor Participants (TAG × Harvey today, more as the program grows). Gold standard; scarce.
B. Market signal triangulation	MODELED	Lightcast postings scanned for vendor-specific skill demand per slice · case-study corpus · announcement density. Cross-referenced.
C. Benchmark extrapolation	ESTIMATED	Public leaderboards (MT-bench, HumanEval, legal-bench, SWE-bench, etc.) mapped to slices. Honest but fragile — task-weighted to avoid single-number inflation.

The math — allocation fitness

```
F(p, s) = 0.35 · CapabilityFit + 0.25 · DeploymentReadiness + 0.20 · EconomicFit + 0.20 · TrustAndGovernanceFit A(p, s) = F(p, s) / Σ_p' F(p', s) # normalized allocation share
UoP-level allocation = volume-weighted roll-up of slice allocations.
```

Three gating rules (applied before allocation)

1. **Hard-zero gate.** If TrustAndGovernanceFit < archetype-specific threshold (e.g. <40 for HIPAA-scoped data), provider is zeroed regardless of other factors.
2. **Minimum viable floor.** Any provider with fitness <30% of the top provider is floored to zero — avoids the cosmetic "every vendor gets 2%" shelf.
3. **Human-retention gate.** Certain slices are marked as mandatory human-decisive regardless of vendor scoring: adversarial judgment in Legal Drafting ≥75% human · architecture decisions in Software Engineering ≥65% human · clinical triage escalation ≥80% human. Governance decisions, not capability measurements.

> **The principle behind the retention gate (v1.1 · Paul).** *The retention gate is a principle we aim to honor by helping enterprises redeploy people toward higher-value tasks that deliver financial outcomes. In a capitalist system, any publicly listed company retains valuable people precisely because the system rewards that — people who produce measurable financial outcomes are assets the company's own market logic protects. Our benchmark is a retention gate above 75%, but the company decides for itself. Our role is to provide clarity — so that leaders making absorption decisions can see the redeployment path and do not default to the easier, shorter-horizon route of cost-cutting. The gate is a prompt for that clarity, not a constraint imposed from outside.*

Worked example — Legal Drafting & Review shelf (measured, TAG × Harvey)

Slice	Volume	OpenAI	Anthropic	Harvey	Human	Confidence
Clause generation	42%	**52%**	26%	12%	10%	MEASURED

Slice	Volume	OpenAI	Anthropic	Harvey	Human	Confidence
Precedent research	16%	18%	8%	**68%**	6%	MEASURED
Compliance review	14%	12%	28%	**46%**	14%	MEASURED
Redlining & markup	12%	22%	**48%**	12%	18%	MEASURED
Summarization	8%	**58%**	22%	10%	10%	MEASURED
Adversarial judgment	8% (gate)	6%	12%	4%	**78%**	MEASURED

OpenAI aggregate share of Legal Drafting: ~35%. Anthropic: ~22%. Harvey: ~22%. Human: ~21%. Every number defensible from TAG × Harvey deployment telemetry + public validation.

Worked example — Software Engineering shelf (April 2026)

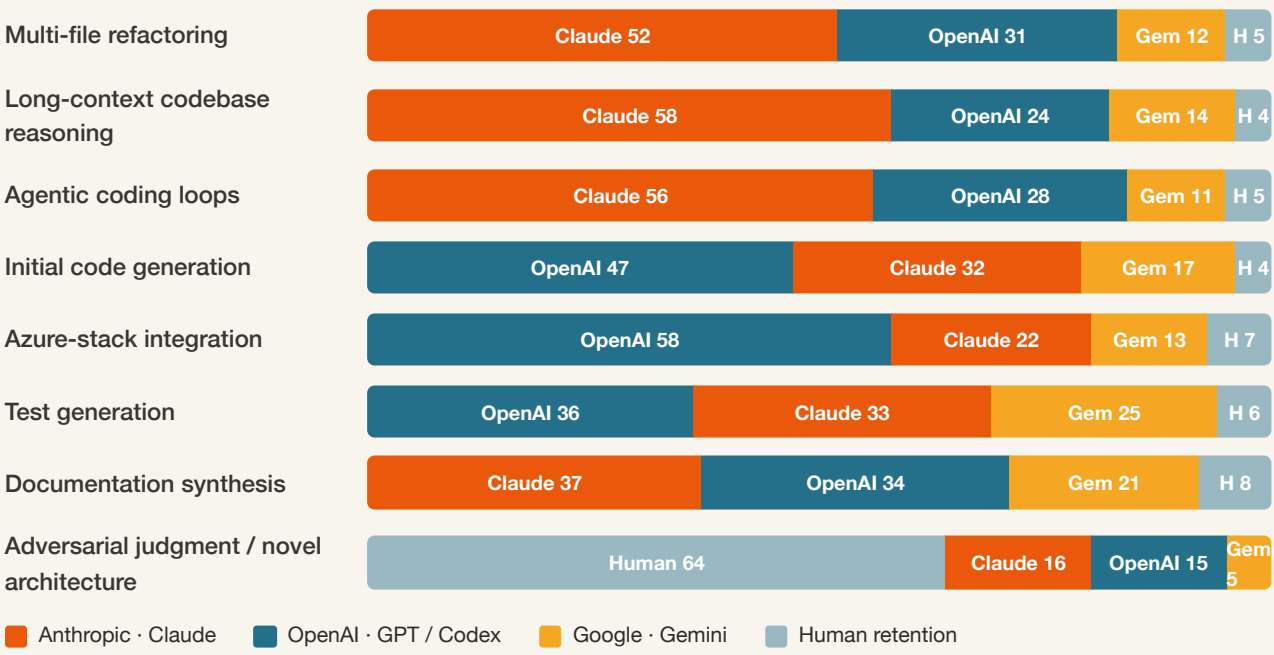

Figure 3 · Software Engineering archetype · illustrative slice allocations (April 2026 · see Section 12 for source-level evidence).

Note on weights. The 35/25/20/20 weighting for allocation (L4) differs from the 9-dimension composite for absorption (L3). This is correct: L3 measures *can-the-org-absorb-anything*; L4 measures *which-vendor-wins-the-slice*. Orthogonal questions, different weight structures. Both are in the weight changelog, both reviewed quarterly, both are v1.1 hypotheses.

Role in the cascade: L4 produces the per-slice competitive map that feeds both L5 (for the slices AI takes, what's the human impact?) and L6 (are stakeholders aligned on this particular vendor split?).

05

CONFIGURATION · HUMAN-CENTERED

Workforce Impact + Redeployment Plan

For this specific UoP, what happens to the humans — which roles, which geographies, on what redeployment paths?

The honest, role-level, geographically-resolved consequence to humans when a UoP deploys — macro (fed back into the Workforce dimension of L3) and micro (as the UoP-level augmented / reskilled / retained / displaced math), anchored in TAG's 104M annual placements as the redeployment substrate.

Augmentation, not replacement. Reskilling, not displacement. Receipts, not claims. The cost-cutting trap — AI frees up time, time goes to layoffs or productivity-capture drift — is the default failure mode. The redeployment plan is the prevention mechanism. Without it, a UoP is an optimization artifact dressed as a transformation. AI favorability among US registered voters is net −20 (worse than either political party). Every AI lab has promised workforce-positive outcomes; none have the measurement infrastructure to prove it. ***r.P's unique right-to-play — and v1.2's sharpening of it — is that we start with the humans AND we name what "human-centered" operationally means: setting the conditions for human + AI co-creation to produce emergent value that no pre-designed plan could predict, while delivering the factual clarity both leaders and workers need to navigate the new system of work.*** *Dignity-with-productivity: productivity gains → EBITDA → salary share for workers is the target loop. Not Marxist no-displacement absolutism; not paternalism; not anti-productivity. Condition-setting, not distributional mandate.*

The three workforce Emergence Conditions (v1.2 canonical)

Inside D2 Workforce Readiness of Gate 1 Absorption, L5 decomposes into three sub-dimensions that score whether the Emergence Conditions for non-linear productivity unlock are met. These are not distributional mandates — they are the factual conditions under which humans + AI actually co-create beyond what the ex-ante plan specified. **Where all three hold, Realized Potential exceeds Estimated Potential. Where they don't, cost-cutting decay.**

VALUE CLARITY

Bidirectional factual transparency on value-creation. Leaders see what value is created where; workers see the map of where value is shifting, so they can unbundle / re-bundle their roles toward higher-value tasks. NOT a distributional mandate — empowering through clarity, not protection. Informed by the calibrated public-sentiment consumption contract (deep-dive §3.9) + Anchor Participant engagement data.

AGENCY

Workers are **participants in reinventing their own roles** — co-creation workshops · role-design input rights · binding feedback on L4 allocation and L5 redeployment. Distinct from works-council representation. Where non-linear productivity comes from: human productivity is non-linear when engaged (vs. linear for agents/robots). Agency is the *condition*, not the goal.

SUBSTANCE

The new work **holds, alleviates, or fills constraints** — per the workchart 3-type × 2-sign Constraint taxonomy (scarcity · risk · coordination × negative/positive, Choudary Ch 5). Where economic value and meaning converge. Higher-value human work absorbs scarcity (specialized judgment), holds positive risk (accountability, safety), or composes coordination. *Substance axis at Modeled/Measured tier depends on workchart Constraint primitive schema-completion; Value Clarity and Agency are publishable at Estimated tier today.*

Falsifiable claim. UoPs with high workforce-axis scores (all three Conditions met) produce Realized Potential that *exceeds* Estimated Potential — because humans in co-creation mode invent/adapt beyond what the ex-ante plan specified. UoPs with low workforce-axis scores produce Realized Potential *below* Estimated Potential — cost-cutting decay mode. L7 telemetry tests this across Anchor Participant deployments. r.P commits to publishing this data quarterly — if the claim is false, the methodology fails publicly and gets revised.

Two altitudes

MACRO (ENTERPRISE-LEVEL)

Feeds back into the Workforce Capability dimension of L3 (15% weight). Enterprise policies, reskilling infrastructure, budget allocation for retraining, geographic concentration of affected workers, union and regulatory friction. The macro view tells you which enterprises can *plan* for redeployment, period.

MICRO (PER-UOP)

For this specific UoP, with this specific vendor allocation from L4, what happens to role X in geography Y? Role-level and slice-level impact profiles. See the TAG × Harvey Legal Drafting template below.

Measured template — TAG × Harvey Legal Drafting (4 quarters)

Role	Roles in scope	Augmented	Reskilled	Retained	Displaced
Junior associate	210	72%	18%	10%	**0%**
Paralegal	340	54%	32%	14%	**0%**
Contract specialist	180	48%	24%	28%	**0%**
Compliance counsel	260	38%	18%	44%	**0%**
Litigation counsel	160	16%	8%	76%	**0%**
Partner	90	12%	4%	84%	**0%**

Aggregate outcome: 1,240 roles in scope · 340 augmented · 120 reskilled into higher-value practice areas · **0 involuntary displacements** · 27% attorney time reallocated from repetitive drafting to adversarial judgment.

The economic decomposition — why alignment beats layoff

I

Miss the Wave

$0 net · competitors reconfigure, underprice, outmaneuver.

II

Automate + Layoff (Klarna Trap)

+$2T savings, –$3T institutional knowledge destroyed, –$2.5T engagement drag.

III

Automate into Disengagement

$0 net · +$2T automation offset by $2.5T engagement drag.

IV

Align

+$9.7B / 19% revenue gain · Automate +$1.95B · Engagement +$2.7B · Redeployment Premium +$3.9B · Adoption Synergy +$1.2B.

×

Redeployment Premium

2.3× vs. cost savings alone. Freed humans move to higher-value work.

×

Engagement Amplifier

3.5× adoption differential. Humans who choose their AI augmentation adopt faster.

Upstream gating at micro level

The workforce plan can *veto* a specific UoP deployment even when L3 and L4 clear it. Example: if the plan's reskilling component requires geographic concentration of affected workers in an already-fragile labor market (industrial closure town, regulated jurisdiction with transition protections), the UoP is gated at L5 until a redistributive plan exists.

> **Executive pragmatism note · Doolan's moment.** Doolan spent 20 years at Salesforce watching customer-success teams argue about attribution. He knows the real-world difficulty of saying "*this specific job at this specific site changed this specific way because of this specific deployment.*" L5 is the attribution engine he never had. TAG is what makes it operational — not a dashboard over a dataset, but a dataset that hires, places, and reskills humans at 104M/yr cadence.

WHY OPENAI SHOULD CARE

AI favorability is net –20. Data-center protests went from 5 in 2023 to 100 projected in 2026 with a 40% cancellation rate. 54% of C-suite call AI "tearing my company apart." 60% plan layoffs for non-adopters. IPO roadshow will face "*how do you measure harm?*" and today the answer is silence. L5 is the answer — independently produced, with TAG as the redeployment engine, visible at role-and-geography level, built into every UoP by default.

HUMAN TELEMETRY — THE NAMED DATA STREAM

For every UoP deployed, quarterly observations of *where the people actually went* — augmented role patterns, reskilling completion rates, retention patterns, geographic landing points. Not capital-deployment data (everyone tracks that). **Human-deployment data.** This is what Congressional hearings need. See L7 for how the telemetry compounds.

Role in the cascade: L5 must be complete before L6 can evaluate feasibility — stakeholders can only align on something concrete, which requires the workforce plan in hand. ***But L5 is NOT a one-shot*** *— it's half of the L5 ↔ L6 flywheel.*

5↔6 BIDIRECTIONAL · V1.1 ADDITION

The L5 ↔ L6 Flywheel

L5 and L6 are not sequenced as a linear handoff. They are bidirectional. The human telemetry from L7 improves both L5 plan accuracy AND L6 alignment-decay prediction. Each deployment sharpens the flywheel.

Pre

Pre-deployment

L5's workforce plan is **Estimated** (industry priors + academic literature). L6's stakeholder alignment is **Estimated** (external signals about the coalition).

Eng

Engagement

L5 plan sharpens to **Modeled** (enterprise-specific workforce data + TAG HR patterns for this industry). L6 alignment sharpens to **Modeled → Measured** (Procensus deliberation, sponsor sign-off).

Post

Post-deployment

L7's **human telemetry** observes reality — did people actually land where the plan said? Did reskilling pathways activate at the modeled velocity? Did stakeholders sustain their commitment or did alignment decay in month three? That observed reality feeds back to sharpen L5 priors AND L6 priors.

Human telemetry as a named concept. This is the specific data stream r.P commits to collecting and publishing: for every UoP deployed, quarterly observations of *where the people actually went* — augmented role patterns, reskilling completion rates, retention patterns, geographic landing points. Not capital-deployment data (everyone tracks that). Human-deployment data. This is what Congressional hearings need, what IPO-roadshow CFOs need, what r.P's neutrality architecture requires us to publish.

06

GATE 2 · MICRO · INSIDE-OUT

Stakeholder Alignment Feasibility

For this specific UoP — with its vendor allocation (L4) and workforce plan (L5) — will the internal coalition actually commit to deploy it?

The second stop-here. L3 told us the enterprise has the mechanical capacity. L4 told us who wins which slice. L5 told us what happens to the humans. But transformations do not fail on capacity or allocation or planning — they fail on **stakeholder alignment**. L6 measures whether the internal coalition of stakeholders will actually commit to deploying *this specific UoP*.

Stakeholder alignment is measurable, not hand-waved. The CFO doesn't sign, the COO diverts budget, the Head of HR blocks reskilling, the General Counsel raises a risk, the Head of Engineering insists on a different vendor. An enterprise can pass every prior layer and still not deploy if internal consensus doesn't exist. Measurement requires structured deliberation — interviews, sponsor assessment, Procensus-style consensus measurement with faction

> *detection and audit trails. External data cannot tell you whether the CFO and the Head of HR agree on the workforce plan; only engagement can.*

Three depth stages · mirror how real deals qualify → discover → commit

Stakeholder alignment is not a single gate applied at one point in time — it is a progressive commitment that deepens as engagement widens. Each stage publishes its own confidence tier; each is a valid stopping point for different go/no-go decisions. No stage is skippable for progression.

Stage	Activity	Confidence tier	Decision it supports
L6.a · Directional Alignment	1–2 executive conversations (sponsor candidate + one key stakeholder). Tests whether a credible sponsor exists and whether the workforce plan is directionally acceptable.	MODELED · lifts L3's 60 cap to 85	Early Go / No-Go for the account. Permits UoP to progress from Generated to Configured in the library; not yet committed for deployment.
L6.b · Coalition Assessment	Structured interviews with 3–5 named stakeholders (Head of People, Head of Function, CIO, GC where relevant, prior-transformation PMO). Named-risk identification; sponsor deep-dive; budget-cycle alignment.	MODELED → early MEASURED	Deployment planning viable. Surfaces congruence variance as a named risk (replaces the dropped Nadler-Tushman penalty from L3). Enables pricing and contract conversations.
L6.c · Procensus Deliberation	Full structured consensus protocol — Semantic-RBO algorithm, faction detection, audit trail, three-component consensus score.	MEASURED	Committed deployment authorization. Produces the composite feasibility number and the alignment-trap / ready-to-deploy / needs-more-alignment label.

What's measured across all three stages (accumulating depth)

Signal	Captured at	Confidence contribution
Sponsor posture	L6.a (surfaced) → L6.b (validated) → L6.c (quantified)	Binary ceiling: no sponsor → Feasibility ≤40
Org Readiness (internal)	L6.b primarily — interviews with Head of People, Head of Change, prior-transformation PMO data	Modeled → Measured
System Readiness (internal)	L6.b — actual integration review, security posture audit, data-governance review	Modeled → Measured
Stakeholder Consensus	L6.c only — **Procensus deliberation** — structured multi-agent consensus protocol, Semantic-RBO algorithm, faction detection, audit trail. The only dimension that cannot be inferred externally or from lightweight conversations.	Measured

The stakeholder consensus score — three components

Produced by Procensus (L6.c) as a single number with three components:

ALIGNMENT ON PURPOSE

Do all named stakeholders agree on what outcome this UoP serves?

ALIGNMENT ON AUTHORITY

Do they agree on who decides, who is consulted, who is informed — standard RACI?

ALIGNMENT ON ACCOUNTABILITY

Do they agree on what changes for whom, and who carries the consequences?

COMPOSITE

Each component scored; the composite is the L6 feasibility contribution for the consensus dimension.

The decision this layer produces

✓

Ready-to-deploy

All three stages clear. Sponsor committed. Coalition aligned. Procensus consensus above threshold.

~

Needs-more-alignment

Go-with-conditions. Specific stakeholder-conversation action items. Return to L6.a or L6.b.

✗

Misaligned-trap

Kill this UoP; it will fail. The alignment traps are the UoPs r.P warns enterprises away from — projects that look great on paper, that pass every macro gate, but that no internal coalition will actually sustain.

Role in the cascade. L6 is **the second stop-here**. If misaligned, the UoP does not deploy — no matter how good L3/L4/L5 looked. The output that enables L7: once stakeholders align on the outcome, the *new tasks* that achieve it get derived from the alignment — not inherited from the legacy task list. This is the concrete moment the Task Fallacy gets rejected. The derived task list is what gets deployed; L7 observes whether reality matches.

07

REALITY + FEEDBACK

Live Telemetry + Flywheel

What actually happened — and how does that learning compound?

The closed loop where every production UoP's measured outcome refines every other layer's calibration — the mechanism by which the methodology compounds, stays current, and earns the "Measured" confidence tier over time. Includes **human telemetry** as a first-class observation: where the people actually went, quarterly.

Every deployment sharpens every future prediction, across all six prior layers. The feedback loop is not a nice-to-have; it is what turns the methodology from a one-shot scoring exercise into a learning system. Without the loop, the methodology is an estimate factory. With it, the methodology converges toward measured truth with every Anchor

> *Participant deployment. Received wisdom buys the framework document. We sell the feedback loop — the Anchor Participant hires the methodology for its ongoing calibration, not for its current snapshot. The snapshot is the demo. The compounding is the product.*

What's observed

Observation	Feeds back to	Confidence uplift
Realized value vs. identified potential	L4 vendor allocation (did the vendors deliver on their scored slices?)	Strand A telemetry for the Shelf
Token economics — realized-value-per-token	L4 + L3 Economic dimension	CFO-grade metric
Time reallocated per role (e.g., 27% attorney time)	L5 workforce plan accuracy	Plan priors sharpen
Human telemetry — actual redeployment vs. plan	L5 + L6	**v1.1 headline addition**
Alignment persistence — did stakeholders stay committed?	L6 feasibility priors	Decay-rate modeling
Slice-level vendor performance	L4 Strand A expansion	Shelf confidence increases
UoP lifecycle state (Generated → Configured → In Production → Compounding)	L2 archetype library	Measured archetypes replace Modeled ones

Three states, one progression

01

Theoretical

UoP exists as proposed alignment decision from external data only. Score caps at 60 (L3 ceiling rule). No enterprise-specific validation.

02

Observed

Enterprise context present: CPO signals, firmographic enrichment, procurement conversations. L6.a or L6.b engagement lifts caps. Allocation refines. Workforce moves from Estimated to Modeled.

03

Business Outcome

UoP in production with telemetry: adoption rates, redeployment outcomes, RoP (Realized Potential / Potential Ceiling). All layers upgrade to Measured for that instance and — via anonymized compounding — for the adjacent profile cluster.

The compounding claim

After 10 measured deployments across diverse enterprises and archetypes, the weights in L3 can be regressed against outcome variance, replacing hand-set priors with data-driven ones. Strand A evidence in L4 expands from ~3 slices (today, Legal Drafting × Harvey) to dozens. L5 plan-accuracy priors shift from Estimated to Modeled. L6 decay-rate predictions become enterprise-specific. The methodology version migrates: v1.0 → v1.1 (this doc) → v2.0 (first data-driven revision).

Anchor Participant program — the mechanism

The validation path is not academic simulation. It is paid pilots at strategic partners (OpenAI, Salesforce, TAG, Google, Anthropic...) where Anchor Participants get first access to the Shelf + methodology + roadmap in exchange for contributing telemetry into the flywheel. Pricing is against IPO/credibility value, not intelligence value. "*If they flinch, they've told us how real they believe their own numbers are.*"

> **Rollup rule across all layers.** Lowest tier wins for any composite view of a UoP. A Measured workforce outcome composed with an Estimated L1 substrate signal yields an Estimated composite. The progression from Estimated → Modeled → Measured is the explicit upgrade arc — and it is what every Anchor Participant relationship is priced against.

08

THE LOOP'S KNOWN FAILURE MODES

Anti-patterns the Loop is Designed Against

These are the failure modes the methodology treats as first-class risks — not theoretical hazards, but patterns observed in real transformation corpora. v1 had five; v1.1 adds two new ones specific to the v1.1 architecture.

#	Anti-pattern	What goes wrong	Methodology's response
1	**Context Starvation**	A UoP gets scored on thin data and that score gets treated as authoritative. Most damaging in L3 (outside-in absorption) where Estimated-tier numbers can be mistaken for Measured.	**Data Sources Graph** makes data thinness transparent in the product UI; every number carries its confidence tier (Estimated / Modeled / Measured); external-only signals are capped at 60 in L3. Lowest-signal-wins rule prevents confidence inflation.
2	**Output Accumulation**	Stale UoPs pile up and never get revisited; the library rots.	**Staleness protocol** — every parameter and every UoP has a refresh deadline. Expired parameters revert to conservative defaults with visible flags. Measured tier re-verifies quarterly; Modeled tier re-verifies bi-annually.
3	**Model Staleness**	Organizational reality evolves (exec turnover, M&A, budget cycle, strategy pivot) but absorption scoring doesn't. The score becomes fiction 6 months after it was last run.	**Re-assessment triggers** on named organizational change events. v1.1 names seven triggers: executive sponsor departure, head-of-function change, M&A event, budget-cycle shift, strategy pivot announcement, regulatory-posture change, reorganization at L2 or above. Any trigger flags the enterprise's score for re-run within 30 days.

#	Anti-pattern	What goes wrong	Methodology's response
4	**Fragmented Tooling**	Different departments produce conflicting UoP scores because they run different versions of the methodology, different data sources, or different interpretations of the same archetype.	**Portfolio-level coherence check** at the UoP library level. Every archetype has one canonical scoring spec; every enterprise has one canonical cohort profile; conflicts surface on a coherence dashboard rather than silently coexisting.
5	**Decision Amnesia**	Scores change but nobody knows why. Six months later, the methodology board can't defend a ranking change under pressure.	**Audit trail** on every parameter change, every methodology-board revision, every weight changelog entry. Changelog entries (like the v1 → v1.1 one in Appendix C of this doc) are the public surface of this audit trail.
6	**Flywheel Capture** (new in v1.1)	The Anchor Participant program is supposed to be a validation mechanism, but if one partner's telemetry dominates early, the flywheel biases toward their deployment pattern and the library becomes an accidental monoculture.	**Diversity constraint on Anchor Participants.** The methodology board sets a minimum diversity threshold: no single Anchor Participant's data can contribute more than 40% of the Measured observations for any given archetype's weight regression. Enforcement is structural (published in the governance charter), not advisory.
7	**Gate Collapse** (new in v1.1)	The two-gate architecture (L3 macro + L6 micro) degenerates into a single gate when teams skip L6 because L3 already said "Ready." Salespeople optimize for closing, not alignment.	**Deployment cannot be logged without an L6 output.** L6.a minimum (directional alignment from at least two exec conversations) is a hard prerequisite for UoP progression from Generated → Configured state. No L6, no progression — regardless of what L3 said.

Why these matter for v1.1 specifically. Anti-patterns #6 (Flywheel Capture) and #7 (Gate Collapse) are new in v1.1 because the architecture changes that created them are new: the Anchor Participant flywheel itself exists only in v1.1, and the two-gate architecture (L3 Gate 1 + L6 Gate 2) was consolidated only in v1.1. Each structural innovation in the methodology creates new failure modes; the anti-pattern list grows in lockstep.

09

CROSS-CUTTING

One Machine · Three Outputs

The seven layers are not three independent outputs — they are one machine producing three outputs simultaneously.

This is the structural answer to the inevitable "why don't we build this internally?" objection: your internal team would need to replace three separate capabilities, none of which they can produce honestly alone, all of which emerge from the same methodology running the same loop.

VALUE PROP · 01

GTM Efficiency Edge

"Stop wasting quarters on absorption-blocked accounts." Spend enterprise sales energy on accounts that will actually close and adopt. Skip the ones that look attractive on face value but are absorption-blocked. A quieter pipeline is a higher-converting pipeline.

Produced by L1 + L2 + L3 (Gate 1)
Demo moment Beat 1 · Total → Cold → Hot globe shrinkage
Lands hardest on Doolan

VALUE PROP · 02

Congress & IPO Shield

"The answer when the Senate asks how you measure harm." A structurally independent, auditable record of responsible deployment and responsible human redeployment that survives adversarial reading by congressional staff, journalists, and IPO regulators. Human telemetry is the named artifact.

Produced by L3 + L5 + L6 (Gate 2) + L7
Demo moment Beat 5 · role-level UoP · zero-displacement math
Lands hardest on Dresser

VALUE PROP · 03

Consulting Amplifier

"The trust layer that makes DeployCo and Frontier deliverables believable." Independent measurement that sits upstream of execution. Neutral Shelf posture — honest multi-vendor allocation that preempts the internal-build objection.

Produced by L4 + L7
Demo moment Beat 3 · SW Eng shelf · OpenAI <50% in at least one slice
Lands on Both executives

> *"Efficiency → Responsibility → Credibility.*
>
> *Operational → Existential → Structural."*

Each prop raises the stakes. Each prop is produced by the same honest absorption measurement. Each prop compounds only if the methodology is a living loop — which is why the leave-behind methodology document is the conversion asset, not the demo.

10 GOVERNANCE

What Makes the Methodology Defensible

PRODUCT GOVERNANCE

Every score traces to specific inputs and weightings. No black-box rankings. Every number carries a confidence tier. Every weight change is versioned and published.

CONTRACTUAL GOVERNANCE

Holland & Knight as independent legal counsel for methodology-board independence. Anchor Participant fees grant zero scoring influence and no pre-publication review rights over archetype or allocation outputs.

METHODOLOGY BOARD

Independent review board for weight revisions, threshold adjustments, archetype additions, data-source integrations. Quarterly reviews with published changelogs. No single Anchor Participant holds more than one board seat.

VERSIONING

Major (v1 → v2) for structural changes. Minor (v1 → v1.1) for parameter refreshes. Every parameter has its own version history with source, date, decay clock.

SOURCE HIERARCHY (ENFORCED)

Peer-reviewed benchmark > independent deployment study > multi-vendor analyst report > vendor demo. Vendor claims flagged "vendor-sourced, unvalidated" with minimum 50% discount if no higher-tier source exists.

STALENESS PROTOCOL

Every parameter has a half-life; crosses staleness thresholds at 0.5, 1.0, 2.0 half-lives. Crossing triggers visible flags, refresh searches, and — at 2.0 — reversion to conservative defaults.

11

SUPPLEMENT · THE STRUCTURAL CREDIBILITY PROOF

The Uncomfortable Number

In the Software Engineering archetype, current-state allocation shows Claude models winning three specific slices empirically. OpenAI leads others. The honest allocation is slice-level, not vendor-level.

This moment — Beat 3 of the demo — is the structural credibility proof of the whole methodology. A shelf that shows OpenAI winning every slice is registered as advertising in under 30 seconds; the whole framework is downgraded to marketing. A shelf that shows OpenAI winning substantial slices honestly, and losing specific slices that you already know you lose, earns every other number the right to survive its next conversation.

> **Framing for the room.** "*Our job isn't to predict who wins. It's to map who wins what, honestly, and update the map quarterly. Today Claude leads three slices; OpenAI leads others. Tomorrow a new frontier release ships and the map updates. The allocation framework is what's durable — the allocation itself is live.*"

Three-signal composite

A single claim is tearable in 30 seconds. Three independent signals are structurally harder to dismiss — the composite only falls if all three are simultaneously wrong, and they come from different evidence categories.

SIGNAL 01 · INDEPENDENT BENCHMARK

Multi-file Refactoring

SWE-bench Verified (Apr 2026)	Claude Opus 4.7: 87.6% · GPT-5.3-Codex: 85.0% · gap: 2.6pp Claude
SWE-bench Pro	Claude Opus 4.7: 64.3% · GPT-5.4: 59.1% · gap: 5.2pp Claude
Cross-file dependency tracking	Claude correctly tracks imports/types/cascades across 10+ files ~23% more often (MorphLlm)

Teardown test. Defensible under pressure. SWE-bench is peer-reviewed, independent, not Anthropic-commissioned. The gap widens on the harder variant — structurally the right direction.

SIGNAL 02 · PUBLIC PRICING + INDEPENDENT BENCHMARK

Long-Context Codebase Reasoning

Needle-in-Haystack 1M tokens	Claude Opus 4.7: 98.5% accuracy
Claude 1M context pricing	Standard rates, no surcharge (since Mar 13, 2026)
GPT-5.4	400K token max; 2× input / 1.5× output surcharge beyond 272K tokens
Chroma Research	Claude decays slowest with context length; GPT erratic at long context

Teardown test. Highly defensible. Accuracy claim from independent benchmark. Pricing structure is public and un-spinnable. Strongest of the three signals.

SIGNAL 03 · REVEALED-PREFERENCE MARKET BEHAVIOR

Agentic Coding Loops

Claude Code ARR	$2.5B (Indagari credit-card transaction analysis via TechCrunch, Mar 2026)
YoY growth	300%+
Token efficiency	Claude Code ~33K tokens vs. Cursor (GPT-5) ~188K on identical multi-file refactor (MorphLlm)
Success without human edits	Claude Code 78% vs. Aider 71% · Codex ~27% less efficient refactor planning (Render.com)

Teardown test. Defensible on adoption/revenue (market voting). Partially defensible on token efficiency (methodology debate). Revenue is from third-party credit-card transaction analysis, not an Anthropic press release.

The 60-second demo script — Beat 3

> *"Software Engineering is the neutrality test, and here it is. If we showed OpenAI winning every slice on this shelf, the methodology is broken — and your engineers would know it before I finished the sentence.*
>
> *So: today, April 2026, Claude leads three specific slices. Multi-file refactoring — SWE-bench Verified, 87.6 against 85. Long-context reasoning — 98.5% at 1M tokens at standard pricing where the alternative is a 2× surcharge past 272K. Sustained agentic loops — $2.5B ARR on Claude Code, 300% year-on-year.*
>
> *On the other side: OpenAI leads greenfield code generation. OpenAI leads Azure-first integration. Test generation is roughly parity.*
>
> *This shelf doesn't stay fixed. We re-run it quarterly. The next Frontier release lands, the map updates. The durable thing isn't the current allocation — it's the methodology that updates it honestly. That methodology is the commercial asset. The honest shelf is the commercial weapon, because only the leader can afford to carry one."*

A

APPENDIX

What's Measured vs. Modeled vs. Estimated Today

A methodology that hides its frontier is not a methodology. Honest assessment of v1.1 evidence today:

Layer	Best evidence today	Confidence class
L1	Lightcast Delta Share live (lcdata): 1.28B postings, 594M profiles, 28.8K canonical skills.	Measured (labor-market data), Modeled (r.P aggregation)
L2	Legal Drafting archetype fully defined · 6 others at Modeled.	Measured for 1, Modeled for 6
L3	TAG (Ready, 82) Measured; 29 other enterprises scored at Modeled. Weights are hand-set priors at global baseline; Legal Drafting delta published.	1 Measured, 29 Modeled, thousands Estimated
L4	Legal Drafting Shelf Measured from TAG × Harvey; 6 other archetype shelves Modeled.	Measured on Legal Drafting, Modeled elsewhere
L5	TAG × Harvey Legal Drafting role-level plan Measured.	Measured
L6	TAG × Harvey Procensus consensus score Measured.	Measured for TAG; Estimated for all unengaged enterprises
L7	4 quarters of TAG × Harvey deployment telemetry.	Measured for 1 deployment; feedback loop architected but not yet compounding

Honest one-liner: r.P has one fully Measured end-to-end UoP (TAG × Harvey Legal Drafting) and a methodology architected for compounding from that anchor. v2.0 publishes after 10 measured deployments. Every Estimated number in the product carries its label.

Known open frontier (v1.2 candidates)

- **Archetype weight deltas.** Only Legal Drafting has a published delta. The Anchor Participant program is the validation mechanism — every new measured deployment publishes its delta. Target: v2.0 after 10 deployments, all archetypes carrying measured deltas.
- **v1.2 dimension candidates still open.** Change-Absorption Rate (orthogonal to Org Readiness — static maturity vs. dynamic frequency of successful change); Cultural Affinity (culture's specific stance on AI, distinct from Workforce Capability). See `v1.1-gaps-and-next-steps.md`.
- **Human telemetry publication grain.** Enterprise level, anonymized-aggregate level, or both? Anchor Participant contracts may specify; governance review board decides.
- **L6 consensus score shape.** Three-component (purpose/authority/accountability) or single composite? Currently three; some enterprises prefer single for roadmap reporting. Defer to Procensus + board.
- **Reliability threshold on L4 slice allocations.** When a vendor's measured reliability drops below a load-bearing threshold on a slice, does the shelf dynamically reallocate? Current answer: yes, quarterly. Open: should this be live?
- **The UoP primitive's minimum viable definition.** v1.1 = outcome-anchored + constraint-aware + governed. Is there a fourth non-negotiable property? Candidates: adversarially-reviewed, workforce-impact-bound, telemetry-committed.
- **Procensus engineering.** Whitepaper-complete but engineering-pending as of April 2026. Once operational, feeds L6.c directly and enriches L7 feedback via real stakeholder positions.

B

APPENDIX

Data Source Cadence

Source	Refresh	Used in layers (v1.1)
Lightcast postings + profiles + firmographic	Weekly	L1, L2, L3, L4
TAG placements + interactions	Daily (per enterprise)	L1, L3, L5
O*NET + BLS time-use	Annual	L1, L2
ILOSTAT + OECD + Eurostat + World Bank + UN	Annual	L1, L3
Public benchmarks (SWE-bench · GDPval · Aider · Terminal-Bench · legal-bench)	Quarterly	L4 (Strand C)
Public governance artifacts (AI principles, policies, audit posture)	Continuous monitoring	L3, L6
Anchor Participant production UoP telemetry	Continuous (post-contract)	L4 (Strand A), L5, L6, L7
Procensus deliberation (L6.c)	Per-engagement	L6
Methodology board reviews	Quarterly	All (governance)

C′

APPENDIX · V1.2 ADDITIONS

Changelog · v1.1 → v1.2

Sixteen material changes between v1.1 and v1.2. The arc: v1.2 folds in the Emergence Conditions doctrine (philosophical spine), three new methodology Primitives from Choudary *Reshuffle* (Organization · System of Work · Coordination Mechanism), the formal promotion of Gate 0 Agent Capability (making the `min()` three-gate composite canonical), L5's three-axis re-architecture (Value Clarity × Agency × Substance), and two new Lenses with a canonical density-conditional applicability matrix — ratified via workchart-architect v1.2 fold-in and substrate-architect sentiment-deweighting sign-off on 2026-04-24.

#	Change	Rationale
1	**Emergence Conditions doctrine** made canonical	v1.1 implicit; v1.2 explicit. The methodology's product thesis — high Alignment Score → Realized Potential exceeds Estimated Potential; low → decay — is now named and falsifiable. Articulates why the primitive is called *Unit of Potential* and not *Unit of Plan*.
2	**UoP gains 4th property: Emergence-oriented**	v1.1 had 3 properties (Outcome-anchored · Constraint-aware · Governed). v1.2 adds: at Estimated/Modeled tier the UoP is a hunch-as-artifact for human reaction, not a ground-truth plan. Ground truth emerges via Realized Potential post-deployment.
3	**3 new methodology Primitives** — Organization · System of Work · Coordination Mechanism	Per Choudary *Reshuffle*. Expands Primitives from 4 to 7. Makes explicit what L3 Absorption always implicitly scored; gives Archetype its proper parent; resolves "what exactly is being aligned-on in L6" ambiguity. Ratified 2026-04-24 via workchart-architect v1.2 fold-in.
4	**Gate 0 (Agent Capability) formally promoted**	v1.1.1 candidate → v1.2 canonical. Three-gate `min()` composite (`min(Gate 0, Gate 1, Gate 2)`) is canonical architecture. Deep-dive §2 carries the 5-dim capability composite; this whitepaper references.
5	**L3 canonically decomposed into OR × SR × EC** with Primitive anchors	v1.1.1 deep-dive proposal → v1.2 canonical. Each sub-block scores a named Primitive: OR = Organization · SR = System of Work · EC = Environmental modifier. Block weights 0.45 / 0.35 / 0.20 (with D9 as veto-ceiling: OR 50 / SR 40 / EC 10).
6	**L5 re-architected with three Emergence Conditions**: Value Clarity × Agency × Substance	v1.1 had generic augmented/reskilled/retained/displaced math. v1.2 adds the three-axis Condition framework — not a distributional mandate, but the conditions under which non-linear human + AI co-creation actually produces Realized Potential above projection. Substance axis depends on workchart Constraint primitive (3 types × 2 signs).
7	**2 new Lenses registered canonically** — Strategic Reframe · Adoption Constraint	Both Paul-approved 2026-04-23 via workchart-ontology-v2 ratification. Strategic Reframe HARD at L3 (4 Choudary where-to-play questions); Adoption Constraint HARD at L2/L3 (5-dim geometric-mean gate). Lenses inventory expands from 5 to 7.
8	**Density-conditional Lens applicability matrix canonical**	7 Lenses × 3 densities (L1/L2/L3) with HARD / RECOMMENDED / LIGHT / N/A. L3 UoPs pass through most Lens scrutiny; L1 pass through fewest.

#	Change	Rationale
		Canonical: `_meta/lenses/density-conditional-applicability.md`.
9	**D9 External Stakeholder Tolerance converted from scalar to veto-ceiling**	v1.1 had D9 as scalar inside EC (5% weight). v1.1.1 deep-dive proposed veto conversion; v1.2 canonical. A lethal stakeholder class sinks the deployment regardless of other-dimension strength. Sub-block weights adjust to OR 50 / SR 40 / EC 10 (with D7, D8 only as scalars).
10	**3 new L3 ceiling rules** — data platform · DORA CFR · regulatory intervention	No primary data platform → SR ≤ 50 · DORA CFR > 30% → SR ≤ 55 · active regulatory intervention → EC ≤ 30. Each with mechanism · provenance · override documented in deep-dive §3.
11	**Governance ceiling made density-conditional** (L2/L3 only; L1 exempt)	v1.1 governance ceiling (50) applied universally. v1.2: L1 role-level UoPs have contained scope and don't hit this cap. Matches empirical pattern from TAG placement dataset.
12	**SR decomposes into 5 operational sub-signals**	Process standardization · Exception-tail mass · Reversibility surface · Feedback-loop latency · HITL infrastructure — measurable before any agent is built via process-mining. Product-visible diagnostic; deep-dive §3.8.
13	**Sentiment calibration contract signed off** — D2 Workforce Readiness consumes `raw_sentiment.*` under calibrated de-weighting	Substrate exposes raw + `bias_tag`; methodology applies v1 calibration multipliers (7 bias-tag clusters × multiplier from 0.65 to 1.00) + 70% single-source cap. **Calibration is informative to D2 only — does NOT gate UoP scoring, render decisions, or stage promotion.** Contract: `_meta/sentiment-deweighting-contract.md`; canon: deep-dive §3.9.
14	**5th taxonomy category: Substrate sub-primitives** (workchart-architect's territory)	5 graph-schema objects (WorkEvent · Flow · Participant · Constraint · ValueAttribution) sit one level below methodology Primitives. Render the methodology's Primitives onto navigable substrate (Supabase + Neo4j). Documented in `METHODOLOGY-TAXONOMY.md`.
15	**Constraint taxonomy locked at 3 types × 2 signs** = 6 classes	Per Choudary Ch 5 p.167 canonical. Supersedes Paul's earlier 5-type working framework. Scarcity / Risk / Coordination × Negative / Positive. Workchart constraint-classifier v0.3 implements with 8 positive-sign hints covering AI-oversight contexts (oversight · agent-handoff · HITL · kill-switch · accountable owner · escalation path · prompt engineering · review/audit AI output).
16	**5 Coordination Levers become first-class L6 design variables**	Per Choudary Ch 1 p.56: representation · decision · execution · composition · governance. Every UoP's L6 output now names how each lever changes from as-is to proposed Coordination Mechanism. Makes "what stakeholders are aligning on" operationally precise.

What stays the same from v1.1: MECE decomposition logic + 7-layer journey ordering · Confidence system (Estimated → Modeled → Measured, lowest-signal wins) · Neutrality architecture (Product · Contract · Governance) · Anchor Participant program as the validation mechanism · L5 ↔ L6 flywheel (bidirectional) · 4 existing Primitives (UoP · Archetype · Slice · Alignment Score) — definitions retained, only UoP extended with 4th property · TAG × Harvey Legal Drafting as first Measured end-to-end UoP.

C

APPENDIX · V1.1 HISTORY

Changelog · v1 → v1.1

Fifteen material changes between v1 and v1.1 (retained as reference; the full v1.1 → v1.2 deltas live in Appendix C′ above). Every one is a response either to red-team pressure (Les, Patrick), to Paul's product direction, or to reconciled reality (Apr 22, 2026 verified counts).

#	Change	Rationale
1	**Journey ordering** replaces family-grouping	v1 grouped by family (Substrate / Intelligence / Accountability / Compounding). v1.1 sequences by the causal funnel the methodology enacts — data → unit → gate → configure → configure → gate → reality → feedback. Stronger narratively; aligns with how the methodology is actually executed.
2	**L1 = Global Labor Graph** (was "H2A Capability Ceiling" in v1)	v1's L1 asked "what can AI do?" before grounding in "what work exists?" — backwards. Capability ceiling is now embedded in L4 vendor scoring (where it belongs — capability is vendor-per-slice). Reordering the substrate restores logical precedence.
3	**L3 = Absorption Filter** (was L4 in v1)	v1 had Absorption after Matching. v1.1 promotes Absorption to Gate 1 — it's a macro filter applied BEFORE configuration work (L4, L5), not after. Matches how hyperscalers use it (DeployCo prioritization).
4	**L5 = Workforce Impact** (was L6 in v1)	v1 had Workforce as the 6th layer after Allocation. v1.1 moves it earlier — the workforce plan must be built before stakeholders can align, because alignment requires concrete content.
5	**L6 = Stakeholder Alignment Feasibility** (new as its own layer)	v1 folded stakeholder alignment into the L4 Absorption composite. v1.1 separates: L3 is the external/macro absorption signal, L6 is the internal/micro alignment gate. Different data sources, different confidence levels, different stages.
6	**L5 ↔ L6 bidirectional flywheel** (new)	v1 treated the layer cascade as one-way. v1.1 names the flywheel: human telemetry from L7 improves both L5 plan accuracy and L6 alignment-persistence priors. Every deployment sharpens the predictions.
7	**"Neutral Shelf"** replaces "Category Captain Shelf" on external surfaces	Antitrust literature on retail Category Captains (FTC 2003, Gundlach-Loff 2018) is largely critical. "Category Captain" retained as internal shorthand; external-facing artifacts use Neutral Shelf with UL/Moody's framing as the defensible B2B analog.
8	**L3 expanded from 5 → 9 dimensions**	v1's 5 dims (Org Readiness · Workforce · Tech Infra · Economic · Workflow Specificity) missed four predictors that materially drive absorption: **Data Readiness**, **Regulatory Context**, **Competitive Intensity**, **External Stakeholder Tolerance**. v1.1 adds all four with explicit weights. Total still sums to 100%.
9	**Archetype weight deltas** published as measured cases earn	Rather than per-archetype weight sets (overkill) or a single global set (dishonest), v1.1 publishes a global baseline + archetype-specific delta

#	Change	Rationale
	them	tables. Legal Drafting has its published delta (from TAG × Harvey measured data, 4 quarters); other 6 archetypes carry the baseline until their first Measured deployment.
10	**Ceiling rules** rebuilt with mechanism · provenance · override triad	v1's ceiling rules (no sponsor → 40, etc.) were asserted, not defended. v1.1 gives each ceiling a *mechanism* (why it caps), a *provenance* (where the number comes from — published literature + TAG dataset), and an *override condition* (how an enterprise unblocks). The **Nadler-Tushman congruence penalty** introduced in an earlier v1.1 draft was **dropped** — the framework does not operationalize as point-gaps in a composite; the citation was borrowed credibility.
11	**Human telemetry** named as a first-class concept in L7	v1 had "measured outcomes" generically. v1.1 names the specific data stream r.P commits to publishing — where the people actually went, quarterly. This is what Congressional hearings need.
12	**Numbers reconciled** to real verified counts (Apr 22, 2026)	Prototype + v1 claimed 3.5B postings / 1.65B profiles / 41K skills. Verified counts: **1.28B postings · 594M profiles · 28.8K canonical skills.** Ambition-vs-current framing added everywhere. Credibility wins over overclaiming.
13	**Anti-patterns section** ported forward from v1 and expanded	v1 had five anti-patterns. v1.1 ports them forward and adds two new ones specific to the v1.1 architecture: **Flywheel Capture** (single-partner dominance biasing the weight regression) and **Gate Collapse** (L6 skipped when L3 already said Ready). Each architectural innovation creates new failure modes.
14	**Human retention gate** stated as a principle, not a cap	v1 had no explicit treatment of how the ≥75% human-retention gates on certain slices should be thought of. v1.1 states the principle: in a capitalist system, valuable people are retained because the system rewards that; the gate is a prompt to redeploy toward higher-value work, not a constraint; companies decide their own gate; our role is to provide clarity so they don't default to cost-cutting.
15	**L6 three-stage depth** (L6.a · L6.b · L6.c)	v1 + early-v1.1 drafts treated L6 as a single binary gate. v1.1 publishes three depth stages mirroring real deal progression: **L6.a Directional Alignment** (1–2 exec conversations, Modeled, lifts L3 cap to 85), **L6.b Coalition Assessment** (3–5 stakeholder interviews, Modeled → early Measured, surfaces congruence variance as named risk), **L6.c Procensus Deliberation** (full consensus protocol, Measured). Closes the methodology's gap between outside-in scraping and full structured deliberation — where most real enterprise validation actually happens.

What stays the same from v1

- MECE decomposition logic
- Confidence system (Estimated → Modeled → Measured, lowest-signal wins)
- Neutrality architecture (Product · Contract · Governance)
- Anchor Participant program as the validation mechanism
- The contrarian r.P stance (work-first, human-first, constraint-first)

v1.1 supersedes v1. Next revision expected after the first Anchor Participant deployment telemetry lands, or after Doolan-meeting feedback — whichever comes first. v2.0 publishes after 10 measured deployments, when L3 weights are regressed against outcome variance and hand-set priors are replaced with data-driven ones.