Seven Verbs: A Theory of Human Judgment in AI-Augmented Production Systems

Rajesh Iyer

iyer70@gmail.com

January 2026

Abstract

This paper proposes a seven-verb taxonomy for characterizing human work in AI-augmented environments. Through dual independent derivation—from organizational outputs backward and from irreducible human capacities forward—we identify seven verbs that describe the complete set of human contributions that persist when artificial intelligence handles execution: WANT (originate intent), INTERROGATE (pressure test output), TASTE (judge quality), OWN (bear consequences), ATTEST (stake reputation), WITNESS (provide presence), and CONVENE (align humans). We validate this taxonomy against 1,000 occupations in the U.S. Department of Labor's O*NET database across 23 job families. All seven verbs load across all job families; none are orphaned. The distribution reveals ATTEST (85%) and TASTE (68%) as near-universal, INTERROGATE (47%) and OWN (44%) as core to decision roles, and WITNESS (27%), WANT (25%), and CONVENE (19%) as specialized human-presence functions. We argue that the primary economic value of AI augmentation lies not in labor substitution but in decision quality improvement: same people, same hours, much better decisions. We further identify failure modes when each verb degrades and derive implications for organizational design, governance, and workforce development.

1 Introduction

The dominant discourse on artificial intelligence and labor markets focuses on job displacement: how many roles will be automated, which tasks are susceptible, what percentage of the workforce is "at risk." This framing assumes the economic value of work resides primarily in labor—in the execution of tasks. If machines execute tasks, humans become redundant.

This paper proposes an alternative framing. The economic value of work has never resided primarily in labor. It resides in outcomes: decisions made, risks assessed, quality achieved, accountability borne. Labor was the mechanism by which humans produced those outcomes. When AI handles execution, the mechanism changes; the outcomes remain.

The question is not *how many jobs will AI take?* but *what will humans actually do when AI handles the doing?*

We propose a taxonomy of seven verbs that describe the irreducible human contributions to work in AI-augmented environments. This taxonomy is derived through two independent methods, validated against comprehensive occupational data, and designed to be immediately actionable for organizational design, governance, and workforce development.

This taxonomy is not a task taxonomy, a skills framework, or a job classification. It is a theory of *irreducible human acts* in socio-technical production systems. Tasks, skills, and roles may change; these verbs do not. They describe the minimum set of human functions required for accountable outcomes in any AI-augmented organization.

This paper contributes a **theory of human judgment in AI-augmented production systems**, rather than a theory of automation or employment. The central claim is that organizational performance in AI regimes is bounded by human judgment capacity, not execution throughput—and that this bound has formal, testable, and designable properties. This theory is expected to be strongest in decision-intensive, accountability-bearing domains such as financial services, healthcare, legal, and regulated industries; its applicability to low-stakes or purely creative production systems remains an open empirical question.

2 Related Work: Positioning Against LLM-Maker Research

Recent empirical work by AI developers has documented the potential scope of AI's labor market impact. OpenAI's "GPTs are GPTs" study [5] found that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by large language models, with 19% of workers potentially seeing 50% or more of their tasks impacted. Anthropic's Economic Index research [6] reports that 77% of enterprise API usage now reflects automation patterns, and projects that AI could increase U.S. labor productivity growth by 1.8% annually.

This paper complements rather than contradicts these findings. OpenAI and Anthropic measure AI capability and adoption—what AI *can do* and *is doing* to tasks. This paper measures what humans *must do* regardless of AI capability—the irreducible judgment, accountability, and coordination functions that persist when AI handles execution.

The distinction is not semantic. Consider Anthropic's own acknowledgment of a measurement gap:

> "Our analysis has limits. Most notably, we can't account for additional time humans spend on tasks outside of their conversations with Claude, including validating the quality or accuracy of Claude's work." [6]

The seven verbs describe precisely what Anthropic cannot measure: the human acts of INTERROGATE (validating), TASTE (assessing quality), OWN (bearing consequences), and ATTEST (staking reputation on accuracy). These are not residual tasks awaiting automation; they are the control layer that makes AI output organizationally usable.

Three specific gaps in existing LLM-maker research motivate this work:

Governance architecture. Neither OpenAI nor Anthropic address how AI-augmented decisions should be governed in regulated environments. Their frameworks do not distinguish between who originates a decision, who owns its consequences, and who attests to its validity. In financial services, healthcare, and legal domains, these distinctions are not optional—they are regulatory requirements. The separation of OWN and ATTEST across control functions (first line, second line, third line) is absent from existing research.

Failure mode taxonomy. Existing research measures capability and adoption but does not explain how organizations fail with AI. Productivity gains and automation rates do not predict when INTERROGATE degrades into abdication, when TASTE atrophies into rubber-stamping, or when ATTEST becomes signature theater. This paper provides a verb degradation framework that explains systemic failure in terms that risk committees and regulators can operationalize.

Economic framing. OpenAI and Anthropic implicitly accept a labor substitution narrative: tasks get automated, productivity increases, fewer workers may be needed. This paper proposes an alternative primary effect: *same people, same hours, much better decisions*. The economic value of AI augmentation lies predominantly in decision quality improvement, not headcount reduction. Where substitution is pursued at the expense of judgment capacity, outcome quality degrades and gains erode.

In summary, LLM-maker research documents what AI contributes; this paper documents what humans contribute. Both are necessary for a complete theory of AI-augmented work. Neither is sufficient alone.

3 Methodology

To avoid confirmation bias and ensure robustness, we derived the verb taxonomy through two independent paths that did not reference each other, then compared results for convergence.

3.1 Path 1: From Organizational Outputs

We began with the question: *What outputs must exist after AI completes its work, and what human acts are required to produce them?*

We enumerated output categories that organizations require:

- Consequential decisions with risk exposure

- Trusted attestations with liability surfaces

- Quality artifacts meeting standards

- Novel directions and strategic intent

- Human connection and care

- Coordinated action across multiple humans

- Verified compliance with regulations

For each output, we identified the minimal human act required to complete it.

3.2 Path 2: From Human Capacities

Independently, we asked: *What can humans do that AI cannot or should not do?*

We enumerated irreducible human capacities:

- Bear legal and moral accountability

- Stake personal reputation

- Exercise qualitative discernment

- Generate novel desire or intent

- Provide embodied presence

- Negotiate between human selves

- Accept personal liability

For each capacity, we identified how it manifests in work.

3.3 Convergence

The two paths produced verb lists that mapped one-to-one with minor synonymy (Table 1).

Initial derivation produced seven verbs from Path 1 and seven from Path 2. Upon examination, CERTIFY and ATTEST collapsed into a single verb (ATTEST), and we added INTERROGATE to capture the pressure-testing function that emerged as central in role mapping. The final taxonomy comprises seven verbs.

Path 1	Path 2	Canonical
COMMIT	OWN	OWN
VOUCH	ATTEST	ATTEST
TASTE	JUDGE	TASTE
WANT	ENVISION	WANT
WITNESS	ATTEND	WITNESS
CONVENE	ALIGN	CONVENE
CERTIFY	WARRANT	ATTEST

Table 1: Convergence of independently derived verb lists

4 The Seven Verbs

4.1 WANT

Originate intent. Generate what should exist. This is the creative act, the strategic vision, the entrepreneurial impulse. AI can generate infinite options; humans decide which options matter. WANT is upstream of all other verbs—it sets direction.

4.2 INTERROGATE

Pressure test AI output. Find it wanting. This is the adversarial function: not "does this look right?" but "walk me through your reasoning—where are you uncertain? What assumptions are you making?" INTERROGATE is the new core skill for knowledge workers. Every human in an AI-augmented workflow must be capable of interrogating the AI's work.

4.3 TASTE

Discern quality. Accept or reject. AI can produce; only humans can be satisfied or dissatisfied. TASTE is the sommelier function—the judgment that says "good enough" or "not yet." It cannot be automated because it is constitutively human: preferences, standards, and discernment are what make humans the customer of AI's output.

4.4 OWN

Bear the consequences. Make the call. Sign the loan, approve the surgery, ship the code. When decisions carry risk, someone must be accountable. That someone is always human. OWN is the verb of consequence-bearing.

4.5 ATTEST

Stake reputation. Put your name on it. ATTEST is the notary function generalized: I certify that this meets required standards and controls. In regulated systems, attestation does not imply origination or ownership of the underlying decision, only formal accountability for its validity. ATTEST may be distributed across control functions (see Section 7).

4.6 WITNESS

Provide human presence. Be there. The nurse at the bedside, the therapist in the room, the teacher who sees the student struggling. Some work requires a human body in the space—not for what it does, but for what it is.

4.7 CONVENE

Align multiple humans. Coordinate wanting. Before AI can act on behalf of a group, the group must agree on what it wants. CONVENE is the work of leadership, negotiation, facilitation, and sales. It is human-to-human, not human-to-AI.

5 Empirical Validation

We validated the taxonomy against the U.S. Department of Labor's O*NET database, which contains 1,016 occupational titles covering 923 data-level occupations across 23 job families and approximately 55,000 job titles in the U.S. economy.

For each occupation, we assigned applicable verbs based on the occupation's task statements and work activities. An occupation could have multiple verbs; most had 2-4.

5.1 Frequency Distribution

Table 2 presents the frequency of each verb across the 1,000 occupations mapped.

Verb	Count	%	Pattern
ATTEST	849	85%	Near-universal
TASTE	675	68%	Broad
INTERROGATE	473	47%	Core decision skill
OWN	435	44%	Consequence roles
WITNESS	274	27%	Presence/care
WANT	249	25%	Origination
CONVENE	187	19%	Coordination

Table 2: Verb frequency across 1,000 O*NET occupations

5.2 Job Family Coverage

All seven verbs loaded across all 23 O*NET job families. No verb was orphaned (appearing in zero families), and no job family was unmappable (requiring verbs outside the taxonomy).

The distribution of primary verbs varied by job family as expected:

- **Management**: OWN, CONVENE, WANT, ATTEST

- **Healthcare**: INTERROGATE, TASTE, OWN, WITNESS

- **Legal**: OWN, ATTEST, INTERROGATE

- **Creative**: WANT, TASTE

- **Production**: TASTE, ATTEST

- **Care**: WITNESS, ATTEST

The taxonomy is comprehensive: every occupation in the U.S. economy can be described as a combination of these seven verbs.

6 Economic Implications

6.1 Decision Quality, Not Labor Substitution

The conventional AI-and-labor narrative models:

Fewer people $\rightarrow$ Same output $\rightarrow$ Labor savings

Our framework suggests a different primary effect:

Same people $\rightarrow$ Same hours $\rightarrow$ Much better decisions

Consider an underwriter. Pre-AI, they analyze 10 applications per day, achieving a loss ratio of X%. Post-AI, they interrogate AI-generated analyses of 100 applications per day, achieving a loss ratio of (X-30)%. The headcount is unchanged; the economic value created is dramatically higher.

The primary value of AI augmentation is not labor arbitrage. It is decision quality improvement at scale.

This has significant implications for how organizations should measure AI ROI. Headcount reduction is a second-order effect and often counterproductive if it degrades the human judgment layer that makes AI output valuable.

This framework does not assert that labor substitution will not occur. Rather, it asserts that labor substitution is not the primary or most durable source of economic value from AI in decision-intensive industries. Where substitution dominates, outcome quality often degrades, eroding the gains. Where decision quality improves, economic surplus compounds.

Existing labor economics models decompose work into tasks and study the substitution or complementarity of machines at the task level [2, 3]. This paper argues that task-level decomposition is no longer sufficient once AI systems generate outputs that humans did not explicitly specify. In such systems, the binding constraint is not task execution but human judgment, accountability, and coordination. The seven verbs constitute a higher-level primitive that subsumes tasks and explains organizational failure modes that task-based models cannot predict.

6.2 Volume Economics by Verb

Different verbs have different volume economics:

- **INTERROGATE, TASTE, ATTEST**: Scale with AI. One human can pressure test, judge, and sign off on many more AI outputs than they could produce themselves.

- **OWN**: Scales modestly. Consequence-bearing has cognitive limits. One person can own more decisions with AI support, but not infinitely more.

- **WITNESS**: Does not scale. Presence is one-to-one or one-to-few. AI does not change this.

- **CONVENE**: Does not scale. Aligning humans is human-limited, not AI-limited.

- **WANT**: Does not scale. Origination is scarce by nature.

Organizations should expect productivity gains concentrated in INTERROGATE/TASTE/ATTEST roles, with WITNESS/CONVENE/WANT roles remaining human-throughput-limited.

6.3 Verb Stacks

Most roles are not defined by a single verb but by a **stack** executed in sequence (e.g., INTERROGATE $\rightarrow$ TASTE $\rightarrow$ OWN $\rightarrow$ ATTEST). AI primarily compresses the execution layer beneath the stack, increasing the throughput of the stack without eliminating it. Organizational design failures often occur when stacks are broken across too many roles (diffusing accountability) or collapsed into too few without governance separation (conflating OWN and ATTEST). The verb stack is the unit of role design in AI-augmented organizations.

6.4 The Seven Verbs as a Human Control System

AI-augmented organizations function as control systems. AI performs execution at machine speed; humans provide sensing, judgment, and accountability. INTERROGATE and TASTE operate as the *sensing layer*, detecting error, uncertainty, and quality drift. OWN and ATTEST form the *actuation and accountability layer*, converting judgment into binding decisions. WANT, CONVENE, and WITNESS define *system boundaries*: what objectives are pursued, who must agree, and where human presence is non-substitutable.

This framing yields the central proposition of this paper:

Proposition 1 (Judgment-Bounded Scaling). In AI-augmented production systems, organizational performance is upper-bounded by the capacity of human judgment verbs (INTERROGATE, TASTE, OWN, ATTEST), not by AI execution throughput. Scaling execution beyond this bound produces increasing error variance and latent risk accumulation, even when average performance metrics improve.

Let E represent execution throughput and J represent aggregate human judgment capacity (the combined bandwidth of INTERROGATE, TASTE, OWN, and ATTEST). For stable system performance, $dE/dt \leq dJ/dt$. When $dE/dt \gg dJ/dt$, systems exhibit delayed failure modes rather than immediate error signals—a condition we term *judgment debt*.

Judgment capacity (J) is a latent construct but admits observable proxies, including: decision throughput per accountable owner, review depth per AI output, override and reversal rates, exception handling frequency, audit finding density, time-to-attestation, and post-hoc correction costs. Future empirical work may test Proposition 1 by examining how these proxies behave as execution throughput increases. The prediction is specific: as E scales faster than J, proxies will show compression (shallower review, faster attestation) followed by lagging degradation (higher reversal rates, increased audit findings, rising correction costs).

The practical implication: **never scale execution faster than judgment**. Violating this principle produces the illusion of productivity while silently accumulating systemic risk.

Failures emerge not from model error but from broken control loops—when sensing degrades, when actuation diffuses, or when boundaries blur.

7 Governance Implications

7.1 OWN and ATTEST Separation

In regulated industries, OWN and ATTEST are deliberately separated across control functions:

- **First line**: Decision-maker OWNs the decision (e.g., underwriter owns the risk decision)

- **Second line**: Control function ATTESTS to process (e.g., risk committee attests model validity)

- **Third line**: Audit ATTESTS to control framework (e.g., internal audit attests governance soundness)

This separation is a design principle: *the person who OWNs should not ATTEST their own work*. With AI, this principle becomes more important, not less. AI can scale bad decisions faster than humans can catch them; robust control separation is essential.

7.2 Verb Degradation and Systemic Failure

AI failures in organizations rarely originate in model performance; they originate in degradation of the human verbs that surround the model.

Each verb can degrade. Organizations should monitor for:

- **INTERROGATE → abdication**: "The model said." Humans stop pressure testing.

- **TASTE → rubber-stamping**: Speed pressure kills judgment. Approval rates approach 100%.

- **ATTEST → signature theater**: Signatures without examination. Audit trails without accountability.

- **OWN → committee fog**: No individual bears consequences. Diffused accountability.

- **CONVENE → scaled misalignment**: AI amplifies execution of misaligned strategy.

- **WITNESS → presence deficit**: Efficiency pressure removes humans from spaces where presence matters.

- **WANT → origination monoculture**: Power concentrates with the few who set direction.

The CONVENE failure mode is particularly dangerous: low-frequency (19%) but catastrophic when absent. AI does not fix misalignment—it scales it.

8 Workforce Development Implications

If the taxonomy is valid, training programs should develop:

- **INTERROGATE**: Adversarial questioning, assumption surfacing, uncertainty localization

- **TASTE**: Quality discernment, standards calibration, exposure to excellence

- **OWN**: Decision confidence, consequence tolerance, accountability identity

- **ATTEST**: Reputational awareness, certification rigor, liability understanding

These are teachable skills, distinct from domain expertise. An INTERROGATE curriculum would train people to probe AI reasoning regardless of domain; domain context is learnable on the job.

9 Limitations and Future Work

This taxonomy is a working framework, not a finished theory. Limitations include:

1. **Binary verb assignment**: Occupations were assigned verbs as present/absent. Future work should model verb intensity.

2. **Task-level granularity**: We mapped at the occupation level. Task-level mapping would increase precision.

3. **Cross-cultural validity**: The O*NET database covers U.S. occupations. International validation is needed.

4. **Dynamic effects**: The taxonomy describes steady-state work. Transition dynamics deserve separate treatment.

Verb Load and Cognitive Saturation

The most significant extension of this framework would model verb *intensity* rather than presence. In practice, some roles carry heavy OWN (a surgeon mid-operation), while others carry thin ATTEST (a routine sign-off). OWN and ATTEST exhibit cognitive and reputational saturation limits; one person can only bear so many consequential decisions per day before judgment degrades. INTERROGATE and TASTE scale further but degrade under time compression.

Many organizational failures attributed to "AI error" are better explained as *verb overload*: humans asked to own or attest beyond sustainable limits. This suggests that safe AI scaling requires not just technical guardrails but *verb load balancing*—explicit constraints on how many decisions a human can meaningfully own, how many outputs they can genuinely taste, how many attestations carry real weight.

Falsifiability and Disconfirming Evidence

This framework would be falsified if one or more of the following were observed empirically:

1. Sustained AI-driven production systems operating at scale with no human INTERROGATE, TASTE, OWN, or ATTEST functions and no degradation in outcome quality over time;

2. Organizations in which decision quality continues to improve as execution scales while human judgment capacity remains constant or decreases;

3. Occupations emerging at scale whose core function cannot be expressed as a combination of the seven verbs.

To date, we observe none of these conditions in regulated or high-stakes domains. The framework's predictive claim—that scaling execution beyond judgment capacity produces delayed failure—remains untested at sufficient time horizons but is consistent with observed patterns in financial services, healthcare, and other accountability-intensive sectors.

10 Conclusion

We propose that human work in AI-augmented environments can be comprehensively described by seven verbs: WANT, INTERROGATE, TASTE, OWN, ATTEST, WITNESS, and CONVENE. This taxonomy was derived through dual independent methods, validated against 1,000 occupations across 23 job families, and is immediately actionable for organizational design, governance, and workforce development.

The economic implication is significant: the primary value of AI augmentation is decision quality improvement, not labor substitution. Same people, same hours, much better decisions.

Organizations that understand this will design AI-augmented work differently: protecting the human judgment layer, separating OWN from ATTEST, monitoring verb degradation, and investing in INTERROGATE and TASTE as core competencies.

As formalized in Section 6.4, the seven verbs form a control envelope around AI execution: INTERROGATE and TASTE constitute the sensing layer; OWN and ATTEST form the accountability layer; WANT, WITNESS, and CONVENE anchor the irreducibly human functions that AI cannot perform. The envelope persists; only the execution inside it changes.

If this theory is correct, then organizations that pursue AI scale primarily through labor substitution will exhibit fragile performance gains that erode under stress, while organizations that invest in judgment capacity will compound advantage over time. AI progress will increase, not decrease, the strategic value of human accountability and discernment. The economic frontier will be set not by faster models, but by better judgment architectures—and the organizations that thrive will be those that invested most deliberately in their human control layer.

This theory is descriptive of system behavior, not prescriptive of moral value; its design implications follow from stability constraints rather than ethical claims.

The verbs are simple. The implications are not.

Acknowledgments

Data Availability

The complete occupation-verb mapping (1,000 O*NET occupations mapped to all seven verbs) is available as supplementary material at:

https://github.com/corpXiv/corpXiv/blob/
main/data/occupation_verb_mapping.xlsx

The dataset includes mapping rationale and verb frequencies by job family.

References

[1] National Center for O*NET Development. *O*NET OnLine*. U.S. Department of Labor, Employment and Training Administration. https://www.onetonline.org/

[2] Autor, David H. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives*, 29(3): 3-30, 2015.

[3] Acemoglu, Daron and Pascual Restrepo. "Tasks, Automation, and the Rise in U.S. Wage Inequality." *Econometrica*, 90(5): 1973-2016, 2022.

[4] Board of Governors of the Federal Reserve System. "Supervisory Guidance on Model Risk Management." SR Letter 11-7, April 4, 2011.

[5] Eloundou, T., Manning, S., Mishkin, P., & Rock, D. "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." *Science*, 384(6702), 1306-1308, 2024. arXiv preprint arXiv:2303.10130, 2023.

[6] Anthropic. "The Anthropic Economic Index." Anthropic Research, September 2025. https://www.anthropic.com/research/anthropic-economic-index-september-2025-report