

From AI Adoption to AI Optimization

Why Every Organization Needs Workforce Intelligence to Maximize AI Business Value

Thomas Giannulli, MD, MS
Founder, Inteltivity • Executive White Paper • July 2026

Executive Summary

Enterprise artificial intelligence has reached an inflection point. The first phase of adoption rewarded experimentation: organizations purchased licenses, encouraged prompting, deployed copilots, and celebrated rising numbers of active users, interactions, agents, and consumed tokens. That phase established familiarity with AI, but it also encouraged a dangerous assumption—that more AI activity necessarily produces more business value.

The brief corporate enthusiasm for “tokenmaxxing” exposed the weakness of that assumption. Some organizations treated token consumption as a proxy for innovation and productivity, even creating internal leaderboards. The metric was easy to collect but difficult to connect to customer value, revenue, quality, cycle time, or margin. Once token consumption became a target, employees could increase the metric without improving the business. Rising model costs and uneven productivity results have since pushed enterprises toward model routing, spend controls, workflow measurement, and outcome-based management.

Market evidence supports this transition. Stanford’s 2026 AI Index reports that organizational AI use reached 88% in 2025, yet the use of autonomous agents remained in the single digits across individual business functions. Gartner has found that estimating and demonstrating value is a leading barrier to AI adoption, while McKinsey reports that workflow redesign—not merely deployment—has the strongest relationship with enterprise-level financial impact. Microsoft’s 2026 Work Trend Index similarly found that organizational factors were associated with roughly twice as much reported AI impact as individual factors.

This white paper introduces **Workforce Intelligence**—a management capability that connects people, roles, AI tools, workflows, operational systems, costs, quality measures, and business outcomes through a contextual data model. Its purpose is not to maximize or suppress AI use. Its purpose is to **right-size AI utilization**: increasing it where AI creates incremental value, improving it where quality or workflow design is deficient, and constraining it where cost, risk, duplication, or rework exceeds the benefit.

The first era of enterprise AI asked, “Are people using AI?” The optimization era asks, “Where, when, and under what conditions does AI improve the economics and performance of work?”

This shift has direct implications for CEOs, CFOs, CIOs, and private equity operating partners. AI can no longer be managed as a collection of licenses or isolated pilots. It must be managed as a portfolio of workforce capabilities, each with an explicit use case, economic hypothesis, quality threshold, risk owner, and measurable outcome.

The Strategic Inflection: From Adoption to Right-Sized Utilization

AI adoption is a necessary precondition for value, but it is not itself value. A purchased license is access. An active user is adoption. A completed prompt is activity. Tokens measure computational consumption. None of these, in isolation, establishes that work became faster, better, less expensive, less risky, or more valuable.

Evidence from controlled research shows why management must look beyond averages. In a study of 5,179 customer-support agents, access to a generative AI assistant increased productivity by approximately 14% on average, with gains of roughly 34% for novice and lower-skilled workers. A separate experiment involving 758 consultants found that AI improved speed and quality for tasks within the technology’s capability frontier, but performance could deteriorate when users applied it to tasks outside that frontier.

Other workplace studies demonstrate that time savings do not automatically become organizational output. An NBER field experiment across 66 firms and 7,137 workers found that users saved approximately two hours per week on email and reduced after-hours work, but the study did not detect a significant change in the overall quantity or composition of tasks performed. Gartner has likewise observed that employees can report substantial time savings while organizations fail to recapture that capacity as revenue, throughput, innovation, or cost reduction.

The implication is not that AI fails to create value. It is that value is **conditional**. It depends on the suitability of the task, the capabilities of the model, the quality of the context supplied to it, the user’s expertise, the workflow surrounding the model, the review burden, and what the organization does with any capacity released.

The AI Utilization Curve

A useful executive model is an AI utilization curve with three operating zones:

Zone	Characteristics & Management Response
Under-utilization	Employees have access to appropriate tools but do not use them in eligible workflows. Causes: weak training, poor integration, unclear policies, fear of mistakes, lack of approved data access, or tool-work mismatch. Cost includes unused licenses, foregone capacity, and slower organizational learning.

Optimal utilization	Appropriate AI capability applied to suitable tasks within redesigned workflows, with acceptable cost, quality, human review, and risk. Optimal levels differ by role—engineers, support reps, attorneys, analysts, and operators will have different economically optimal patterns.
Over-utilization	Additional AI activity creates little or negative incremental value: premium models for simple tasks, agents on poorly specified work, content requiring extensive correction, overlapping subscriptions, or token consumption treated as a performance target.

This framework reframes AI management as an optimization problem. The objective is not maximum use or minimum cost. It is the point at which the **marginal business benefit of additional AI-assisted work remains greater than its marginal financial, quality, operational, and risk cost.**

Workforce Intelligence and the Context Graph

The central obstacle to AI optimization is not a lack of telemetry. Enterprises have vendor administration consoles, model gateways, cloud billing records, security logs, application events, employee-experience platforms, customer systems, project-management systems, and financial reports. The obstacle is that these sources rarely share enough context to explain business value.

An AI administration console might show that a user was active. A model gateway might show tokens and cost. An observability platform might show latency and errors. A SaaS management platform might show license utilization. A workforce analytics system might show time allocation. An operational application might show that a case closed faster. Without a shared contextual layer, leadership cannot reliably determine whether these events describe the same person, workflow, customer, project, or business outcome.

Workforce Intelligence is the management capability that closes this gap. It combines AI, workforce, operational, financial, and governance signals to answer five questions:

- 1. Who or what used AI?
- 2. In which workflow and business context?
- 3. Which model, tool, agent, or capability was used?
- 4. What changed in cost, speed, quality, capacity, or risk?
- 5. Should the organization expand, modify, route, govern, or retire that use?

The Context Graph

The enabling concept is a Context Graph: a governed set of relationships connecting workforce entities, AI events, workflows, and outcomes. Its purpose is to preserve the relationships necessary for attribution—for example, connecting an AI-assisted support interaction to the employee's role, approved model, customer case, resolution time, quality review, satisfaction result, and fully loaded cost. A single event does not prove causation, but repeated patterns across cohorts and time periods can reveal where AI use is associated with better or worse outcomes.

This model also enables organizations to compare apparently similar AI use cases. Two groups may have comparable active-user rates while differing materially in workflow penetration, rework, model cost, or quality. One team may use AI to accelerate a well-defined document workflow; another may use it primarily for ad hoc experimentation. A seat-utilization dashboard would classify both as successful adoption. A Context Graph can show that only one is generating repeatable economic value.

The graph should be designed around purpose limitation. It does not require collecting every prompt or keystroke. In many cases, metadata and outcome relationships are sufficient: approved tool, model, timestamp, workflow, token consumption, completion status, quality result, and business identifier. Aggregation and privacy thresholds should be the default for executive reporting.

A Workforce Intelligence program should therefore be positioned as **workflow and investment governance**, not covert employee monitoring. Transparency, data minimization, role-based access, retention limits, and employee participation are design requirements rather than secondary controls.

Measurement Architecture and the Executive Scorecard

Accounting Telemetry vs. Performance Telemetry

A rigorous measurement system separates two complementary telemetry families:

Telemetry Type	What It Answers & Key Measures
Accounting Telemetry	What was provisioned, consumed, and paid for? Licenses, entitlements, active users, models, requests, tokens, agent runs, credits, unit prices, infrastructure, and vendor spend. Essential for inventory, chargeback, forecasting, and model routing—but consumption is an input, not a result.
Performance Telemetry	How well did the system and workflow perform? Latency, errors, cycle time, throughput, first-pass quality, rework, escalation, customer satisfaction, revenue conversion, capacity created, and compliance outcomes.

Neither is sufficient alone. Accounting without performance produces spend reports and activity leaderboards. Performance without accounting can show improvement but not whether it is economical. Both without context produce correlations that cannot be mapped to workflows, customers, or accountable owners.

A Layered KPI and ROI Framework

The executive scorecard should follow a value chain rather than a flat list of metrics:

Layer	Core KPI	Executive Interpretation
Access	Approved AI coverage; eligible users	Can the workforce use appropriate, governed AI?
Adoption	Activation rate; weekly active users	Is access becoming sustained capability?
Workflow Penetration	AI-assisted eligible tasks + total eligible	Is AI embedded in work, not used sporadically?
Productivity	Cycle time, throughput, capacity created	Does work move faster or can more be completed?
Quality	First-pass acceptance, rework, exceptions	Is speed achieved without hidden quality loss?
Economics	Cost per accepted outcome or case	Is consumption efficient at the unit of work?
Enterprise Value	Incremental revenue, avoided cost, margin	Do operational gains affect org performance?
Governance	Policy exceptions, audit completeness	Is value captured within acceptable risk?

Useful pilot defaults (treat as hypotheses until validated): activation above 60% after 60 days, workflow penetration above 30% in selected workflows, cycle-time improvement of 10–25% for well-scoped use cases, quality at or above pre-AI baseline, and training completion above 80%.

A defensible ROI calculation must use fully loaded costs (licenses, tokens, infrastructure, integration, security, training, review, and rework) and validated benefits. “Hours saved” should not automatically be valued at hourly compensation. Saved time creates economic value only when eliminated from cost, converted into throughput, redeployed to higher-value work, or used to improve service or risk outcomes. For many use cases, **cost per accepted outcome** is a better optimization metric than cost per token.

The Category Landscape and Executive Visibility Gap

No single established technology category consistently connects AI consumption, workforce context, operational performance, financial outcomes, and governance:

Category	Primary Object	Typical Strength	Key Limitation
AI Observability	Model, agent, trace	Technical behavior & evals	Limited workforce & business context
FinOps	Technology cost	Unit economics & budgets	Weak workflow attribution
SaaS Management	App, license, entitlement	Discovery & license optimization	No quality or outcome measures
Workforce Analytics	Employee, activity pattern	Time, focus, collaboration	Indirect outcome linkage
Workforce Intelligence	Person + AI + workflow + outcome	Cross-system attribution	Emerging; requires integration

The resulting visibility gap leaves critical questions unanswered:

- Which workflows create the greatest net benefit after model cost, review, and rework?
- Which groups are underutilizing suitable AI capabilities, and why?
- Where is heavy AI activity failing to improve throughput or quality?
- Which use cases should move to smaller models or deterministic automation?
- Which licenses should be expanded, reassigned, consolidated, or retired?
- How much released capacity is actually converted into growth, service, or cost advantage?

Maturity Model and 180-Day Implementation Roadmap

Workforce Intelligence should be implemented as a management system, not an analytics installation. A practical maturity model progresses through four phases:

Phase	Days	Key Activities & Deliverables
Discover	1–30	Establish cross-functional owner. Select 2–4 high-frequency workflows. Inventory approved and shadow AI tools. Establish pre-AI baselines for cycle time, quality, cost, and exceptions. Define a testable value thesis for each workflow.
Correlate	31–60	Activate native vendor analytics, cost feeds, and operational events. Map to a minimum viable Context Graph. Prefer metadata over prompt content. Incorporate privacy, security, legal, and labor requirements before instrumentation.
Measure	61–120	Compare pilots with credible baselines using phased rollouts or matched cohorts. Track adoption alongside workflow penetration, quality, rework, cost per accepted outcome, and business results. Sample human review explicitly.
Optimize	121–180	Make portfolio decisions: standardize high-value workflows; redesign or retire weak ones; reassign licenses by role and eligibility. Convert the scorecard into a monthly operating mechanism spanning CIO, CFO, operating leaders, risk, and HR.

Governance Principles

NIST’s AI Risk Management Framework provides an appropriate foundation. A Workforce Intelligence implementation should add controls specific to workforce data and performance attribution:

Governance Area	Required Practice
Purpose & Transparency	State what is measured, why, and how decisions will be made
Data Minimization	Prefer metadata and workflow outcomes over prompt or content capture
Aggregation	Default executive reporting to team, workflow, and cohort levels
Access & Retention	Restrict individual-level data; retain only as long as necessary
Quality Assurance	Track accepted output, human rework, exceptions, and downstream effects
Human Accountability	Preserve named responsibility for consequential decisions and outputs
Spend Governance	Apply budgets, alerts, model routing, and approval thresholds
Security & IP	Enforce approved tools, data classifications, audit logs, and vendor safeguards
Fairness & Workforce Impact	Test whether measurement or automation disadvantages roles or groups
Change Management	Train employees, involve managers, and explain how released capacity will be used

Governance must also address cultural cost. Research has found that work perceived as AI-assisted can affect judgments of worker competence, creating incentives to hide legitimate AI use. Management systems that reward raw activity or punish transparent experimentation may degrade both trust and data quality.

What Leaders Should Do Next

For CEOs, CFOs, CIOs, and private equity operating partners, the immediate agenda is practical:

- **Stop treating adoption metrics as success metrics.** Active-user rates and token volumes are necessary but insufficient.
- **Select a small number of high-frequency workflows** where outcomes can be measured and a value thesis can be stated.
- **Instrument for context, not surveillance.** Connect AI events to roles, workflows, costs, quality, and business identifiers while minimizing prompt and content capture.
- **Optimize the portfolio.** Expand, redesign, route, or retire use cases based on net economic contribution—not on how loudly they consume tokens.
- **Govern for trust.** Transparent purpose, data minimization, and clear rules about how released capacity will be used are prerequisites for durable measurement.

Conclusion

Enterprise AI has moved beyond the point at which adoption statistics can carry the business case. Leaders now need to distinguish access from activity, activity from workflow penetration, workflow penetration from operational improvement, and operational improvement from enterprise value.

The organizations that win the next phase of AI will not necessarily consume the most tokens, deploy the most agents, or purchase the most licenses. They will be the organizations that know where AI changes the economics of work—and that can continuously direct people, models, capital, and governance toward those opportunities.

AI adoption made the technology available.
Workforce Intelligence makes it manageable.
AI optimization makes it valuable.

Selected Sources

1. Kahn, J. Fortune. “Tokenmaxxing Is Over.” May 28, 2026.
2. Stanford HAI. AI Index Report 2026 — Economy & Organizational Adoption.
3. McKinsey & Company. “The State of AI: How Organizations Are Rewiring to Capture Value.” 2025.
4. Brynjolfsson, Li & Raymond. “Generative AI at Work.” NBER Working Paper.
5. Dell’Acqua et al. “Navigating the Jagged Technological Frontier.” Harvard Business School.
6. NBER. “Shifting Work Patterns with Generative AI.” Field evidence on time savings vs. output.
7. NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0) and Generative AI Profile.
8. Microsoft WorkLab. 2026 Work Trend Index: Agents, Human Agency, and Opportunity.
9. Gartner. Surveys on generative AI deployment barriers and value challenges (2024–2025).
10. Harvard Business Review. “The Hidden Penalty of Using AI at Work.” 2025.
11. Associated Press. “Workplaces Look for Cheaper AI as ‘Tokenmaxxing’ Fades.” July 2026.

12. FinOps Foundation. Unit Economics capability and FOCUS specification guidance.