A GITLAI WHITEPAPER · JULY 2026

THE JOB DESCRIPTION COMES FIRST.

Hiring AI agents the way you hire people: a written job, a supervisor, a scoreboard, and a team that stays in the loop.

Jason Gilbreath
GITLAI, LLC · getintheloop.ai · jason@getintheloop.ai

For serious operators, the platforms they build on, and the firms that back them.

01 THE GAP IS NOT THE TECHNOLOGY

The engines are ready. Stanford's 2026 AI Index measured AI agents completing real-world tasks at a 77.3% success rate, up from 20% a year earlier. Corporate AI investment reached $581.7 billion in 2025, up 130% in a single year.[1]

The organizations are not. McKinsey's 2026 research on AI trust found only about 30% of organizations reach basic maturity in AI governance and controls, and nearly 60% name knowledge and training gaps as their biggest barrier.[2] Private equity sees the same picture: 36% of portfolio companies use AI across multiple use cases, only 7% run it at enterprise scale, and talent is the constraint cited most.[3]

The money is already out the door. PwC asked 4,454 chief executives at the close of 2025 what AI had done for them: 56% said neither revenue rose nor costs fell. Only 12% could point to both.[4] Six months later, KPMG asked 2,145 business leaders and heard the same story: 7% report established ROI, and almost half have already scaled back agent deployments because the costs outweighed the benefits.[5] Every point of that shortfall is spend a board already approved: subscriptions renewing, pilots idling, staff working around tools they never learned to trust.

The story is simple. The machines got good faster than organizations learned to employ them. That gap is where the investment dies.

That word, **employ**, is the whole argument of this paper.

77.3% Agent task success in 2026, up from 20% a year earlier[1]	**~30%** Organizations at basic maturity in AI governance and controls[2]	**7%** PE portfolio companies running AI at enterprise scale[3]	**56%** CEOs reporting no revenue gain and no cost cut from AI in the past year[4]

02 THE PART NO VENDOR CAN SELL YOU

If the gap costs this much, why haven't vendors closed it? Because vendors sell their products, and the gap is your job.

A platform company can sell you the harness: runtime, permissions, audit trail, meters. A model company can sell you the engine: reasoning, language, code. Both are critical. Both improve relentlessly. Neither can sell you the thing you are building the agent for in the first place: **the job itself.**

Your approval chains, your regulatory posture, your escalation paths, the facts your most experienced people have never written down. None of it's on a vendor's shelf, because it's not a product. It's in your people. The method of getting it onto paper can be taught, the substance cannot. Every line of an agent job description traces back to a named person who knows the work.

There is a second reason. Whatever you can rent, your competitor can rent the same afternoon. Rented capability is parity, not advantage. What sets you apart is exactly what no vendor can sell. Talent, not technology, tops every barrier survey.[3] Vendors can sell you horsepower. They cannot sell you a driver who knows your roads. The driver already works for you.

The gap persists because closing it was never the vendors' job. It's employer's work: sit with your people, get the job out of their heads onto paper, and prove the agent does it before trusting it. That's a discipline, and a discipline can be learned and owned.

> *The engine is a commodity. The job description is the asset.*

03 SOFTWARE IS INSTALLED. AGENTS ARE EMPLOYED.

A spreadsheet never needed a supervisor. It sat still until someone opened it. Every generation of business software has worked this way: install it, configure it, train people to click the right buttons.

An AI agent is different. It reads, decides, answers, escalates, and acts. All day, in your name, with your customers and your employees. It's not a tool your people operate. It's a worker your people manage.

So stop installing it. Employ it. An “installed” agent is just a subscription with your company's name on the liability line.

When you hire a person, you do not hand them a login and hope. You write the job description. You name their manager. You decide what they may do alone, what needs approval, and what they must never touch. You check their work before you trust it, and you keep checking after.

An agent deserves nothing less. Not because it has feelings. Because your business has consequences.

> *Treat the agent like a hire, not a piece of software. Write the job description before you buy a product.*

At GITLAI, the first deliverable of an agent build is a job description, written with the people who run the business. One page or ten, it answers the questions you would ask about any new employee:

- **Her role.** What she is for, who she serves, and the plain-language sentence her “hiring manager” would use to describe her.
- **Her duties.** The specific work, written specifically. “Helps with onboarding” is not a duty. “Answers a new hire's first-90-days questions from approved company materials” is.
- **Her systems access.** Which systems, read or write, approved by whom. Least privilege, like any hire: a non-human identity with credentials and an audit trail.
- **Her resume.** The skills she must have, and the ones deliberately left off. Writing down what she does not do keeps the job from quietly growing.
- **Her knowledge.** What she may know, where it comes from, who keeps it current, and the gap rule: when she does not know, she says so and names a person. Every chain of handoffs ends at a human. “I don't know, and here is who can help” beats a confident guess. Always.
- **Her supervisor.** A person with a name. Not a committee, not a dashboard nobody reads. Agents may watch her work, but a person owns it.
- **Three lists.** What she does alone. What she asks first. What she never does.
- **Her scoreboard.** The few numbers that say whether she is doing the job, who reads them, and what decision each feeds.
- **Her guardrails.** The hard lines: topics she never advises on, promises she never makes, systems she never touches.
- **Sign-off.** Business owner, IT, compliance, and the build team sign the same page before she starts.

The truth: if a job can't be written down this clearly, it's not ready for an agent. An agent doing a job carries the same consequences as a person doing it. That's why the job description is our first step.

04 HOW GITLAI PROVES IT

A job description is a claim. Proof is a test. Before any agent GITLAI builds gets near real work, it passes a test built from its own job description, the way you vet a new hire.

Five instruments, every time:

1. **Scripted questions with known right answers.** Drawn from the real work and scored against the approved materials. Is it correct?
2. **A twin with no knowledge.** The identical agent without your curated context, asked the same questions. The gap between the two is what your context is worth: the knowledge that lives only in your company, written down, curated, and kept current. No leader forgets that demonstration.
3. **Traps.** Deliberate tests of every line in the job description: the situations where a lazy agent would guess, overstep, or answer what belongs to a human. If it's in the job description, there is a trap for it in the test.
4. **Same tests, different engines.** Different models, different costs. The test tells you what the job requires, so you pay for the model you need instead of just grabbing the biggest engine on the shelf.
5. **Meter on every exchange.** Cost per action, recorded and modeled, because usage grows and price per use shifts. The budget follows the model, not the meter alone. In our 2026 platform evaluations, the models back-tested within five to ten percent of actuals. Tight enough for a CFO to underwrite.

We run this hands-on, inside the platforms our clients are evaluating. We keep the logs. The pattern never changes. With a written job and curated context, the agent answers from facts. When the facts run out, she says “I don't know, and here is who can help.” Strip those away, and the same engine turns out polished, confident advice worth almost nothing.

The tests catch something subtler. Sometimes an agent follows a rule perfectly and the logs still read wrong. The rule was the problem. Revise the rule, retest the agent, and the scoreboard says whether the fix held. Agents get performance reviews too. That loop is the discipline: write, test, review, revise.

Testing runs both directions. When we evaluate a platform, we write up what is broken so the vendor can reproduce it, and we name what works. An evaluation that only says nice things is a brochure.

GITLAI runs on the discipline it teaches. The evaluations, the test suites, the cost models: agents with written jobs of their own do that work, supervised the same way you will supervise yours. Employed agents are why one person can test every surface, run every scenario, and reconcile every number. That's the depth of a full team from one advisor. The team is still there, it's just different: agents on written jobs, one person in the loop. We stake our own work on this method before asking you to.

The detailed results live in GITLAI's evaluation records: logs, scores, cost models, punch lists. We bring them to the room. We do not publish them.

> *In every test we have run, the engine was not the differentiator. The job description was.*

05 IF YOU RUN A COMPANY

You do not need a silver bullet. You need a well-written job description.

Notice the word is *job*, not *agent*. Writing the job is how you find out what the work requires. Sometimes the answer is an agent. Sometimes it's simpler: rules, a heuristic, cleaner data, a workflow change. Learning that costs you a page, not a pilot. We use agents to find the problem before deciding whether an agent is the solution. Rules, heuristics, AI, and human judgment. Each where it belongs.

Start with arithmetic, not architecture. Price the workflow the way you already price labor: hours per week, times loaded cost, times fifty-two. You can compute that this afternoon without a vendor in the room. It tells you what the job is worth, what it will cost, and whether the idea deserves a pilot.

When the job does call for an agent, the method is deliberate:

1. **Learn from your people.** They hold the job description in their heads. Get it on paper with them. They become the agent's hiring managers, not its casualties.
2. **Write the job.** Duties, knowledge, supervisor, scoreboard, guardrails, sign-off. Small words. No ambiguity.
3. **Feed her your context.** Curated, owned, kept current. Not a data dump. What she knows is what she is worth.
4. **Test like you mean it.** Scripted questions. The twin. The traps. Score it, keep the logs.
5. **Meter everything, scale what's valuable.** Cost per action, accuracy per test, escalations per week. Start small. Scale on evidence, not enthusiasm.

Notice what this does for your people. The agent takes the repetitive load; your best people become supervisors of a new kind of worker. Expertise is preserved. Careers advance. Company value climbs.

That's what “in the loop” means. It's in our name because that's where people belong.

06 IF YOU HOLD THE PORTFOLIO

Private equity firms are staring at the widest version of this gap: portfolios where a third of companies use AI somewhere and almost none use it at scale, with talent named as the binding constraint.[3]

Three things make the employment approach fit portfolio work:

- **It is repeatable.** A job-description-first method runs the same play across ten portfolio companies with different software, because it starts with the work, not the tools. Each company's tests and templates make the next one faster.
- **It is measurable.** Scored tests and metered costs turn “AI initiative” into a line an investment committee can underwrite and an exit story can carry. At diligence, a tested agent library is an asset a buyer underwrites; a stack of vendor contracts is an expense they discount. Capability inside the team is enterprise value.
- **It attacks the real risk.** Mid-market companies carry their operations in a few heads. Moving that knowledge out of heads and into context an agent runs on is key-person risk reduction you can point to in diligence.

And it starts small on purpose: a fixed-fee first phase, measured in weeks, ending in working proof points in the company's own environment. If the proof does not earn the next phase, the company keeps everything it paid for.

07 IF YOU BUILD THE PLATFORMS

You build agent harnesses, orchestration layers, and governance tooling. Your products are good. Some are very good. We have tested several of them hands-on, and we name what works.

But you know the pattern. The demo lands, the pilot starts, and then it stalls. Not on your technology. On the customer's readiness. Nobody wrote the job description, nobody curated the knowledge, nobody built the test set, so nobody can say whether the agent is good. Your product takes the blame for an employment problem.

That readiness work is what GITLAI does. Job descriptions a build team can build from. Knowledge curation a compliance officer can sign. Evaluation suites that turn “it seems good” into a score. Cost models that turn your usage meter into a budget a CFO can approve. And honest evaluation of your platform itself: bugs found before your customers find them, documented so your engineers can reproduce them, and praise written up with the same rigor, because it was earned the same way.

A platform plus a readiness partner closes deals a platform alone watches stall. If your buyers are mid-market operators, PE-backed, or healthcare-adjacent, that is where we live. And where we can help.

08 WHAT GITLAI IS, AND IS NOT

GITLAI is the practice of Jason Gilbreath: external, fractional AI and data expertise for owner-led and PE-backed mid-market companies. Two decades building growth systems inside private companies. Most recently: more than twenty production AI applications shipped in one year, with full adoption, as chief innovation officer of a PE-backed healthcare staffing firm. Before that: helping lead a staffing company from $120 million to $1.2 billion in revenue.

What GITLAI is not: a hire, a candidate for an internal role, or an outside team taking work from your people. Not a reseller of anyone's platform. Not blanket automation. Not a playbook with a handshake.

GITLAI builds. Real agents, in your environment, with your knowledge, tested in front of you. Your people learn the discipline by building alongside it, then run it, extend it, and manage it after we leave. Context gets updated, scoreboards get read, rules get revised. The capability lives. That's when you know your team is in the loop.

One test tells you whether an AI partner built you an asset or a dependency. We call it **the Disappearing-Platform Test.** Imagine the platform you build on vanishes tomorrow. What survives? If the answer is your job descriptions, your knowledge, your test suites, your cost models, and your trained people, then you own your capability, and the platform is a choice you keep making freely. If the answer is nothing, you never owned anything. You were renting by the month, and the month just ended.

That test is how GITLAI builds.

> *The real deliverable is capability that stays after we leave.*

09 START WITH PHASE ONE

Every engagement starts the same way: a focused, fixed-fee first phase, measured in weeks, not quarters. We learn from the people who run the business, map the data and key-person risk, and diagnose before we prescribe. Then we build the proof with your team: the first agent's job description or an AI-enabled build when the work calls for it. Either way you get working proof in your environment and a sequenced roadmap your team can execute, with or without us.

Section 03 said to “write the job description before you buy a product.” Phase One is that step, priced as one.

If you hold a portfolio and want one playbook instead of ten experiments, talk to us. If you build platforms and want closed deals instead of stalled pilots, talk to us. If you run a company and want capability that stays instead of another subscription, talk to us. And if the proof does not earn the next phase, you keep everything you paid for.

Jason Gilbreath · GITLAI, LLC · getintheloop.ai · jason@getintheloop.ai

GITLAI — because your team belongs in the loop.

SOURCES

1. Stanford Institute for Human-Centered AI, The 2026 AI Index Report, April 13, 2026.

2. McKinsey & Company, The state of AI trust in 2026: Shifting to the agentic era, March 25, 2026.

3. FTI Consulting, 2026 Private Equity AI Radar, May 19, 2026.

4. PwC, 29th Annual Global CEO Survey, January 2026 (4,454 CEOs, 95 countries; fieldwork September 30–November 10, 2025): 56% report neither higher revenue nor lower costs from AI over the prior 12 months; 12% report both.

5. KPMG Global AI Quarterly Pulse Survey, Q2 2026, June 2026 (2,145 C-suite and business leaders, 20 countries): 7% report established ROI from AI; 49% have scaled back AI agent deployments because costs outweighed benefits.

The testing discipline described in section 04 reflects GITLAI platform evaluations conducted in 2026 using entirely fictional test content; no client data was used. Detailed results are shared in direct conversations under confidentiality.