2026

AI Roadmap & Predictions

THE GREAT CONVERGENCE

What small and mid-sized businesses should build, skip, measure, and protect as AI moves from chat to work.

THE THESIS AI capability is becoming abundant. Reliable, permissioned, measurable execution is not.

Contents

HOW TO READ

Evidence labels used throughout: **FACT** = observed data or published research. **INTERPRETATION** = our synthesis. **FORECAST** = a probability-weighted bet. **UNKNOWN** = a variable that can still break the story.

The useful answer, before the long answer

The AI market is not settling down. It is becoming more competitive, more affordable per unit, and more operationally demanding at the same time.

Keep it a buck: most smaller companies do not need a frontier-model strategy. They need a **work strategy**. The winning question is not "Which model is smartest?" It is "Which recurring job can we install, supervise, and prove?"

88%	70%	3.3%	16%
of surveyed organizations report using AI[3]	use generative AI in at least one function[3]	gap between top closed and open models, March 2026[2]	relative employment decline for ages 22–25 in highly exposed roles[6]

FIVE CALLS FOR OPERATORS

01 **The model race is converging. The system race is opening.** Capability differences still matter on the hardest tasks, but model access alone is no longer a moat. Context, workflow ownership, evaluation, permissions, and distribution are.[2]

02 **Adoption is high; operational maturity is not.** AI is present in most organizations, while agent deployment remains in the single digits across nearly every business function. Translation: a lot of logins, not yet a lot of installed labor.[3]

03 **Small businesses can move faster than large enterprises, but they have less room for expensive mistakes.** A narrow workflow with a clear owner and measurable "done" beats a broad transformation program with a beautiful deck.[5]

04 **Human approval becomes the new operating system.** As execution gets cheaper, judgment, exception handling, and accountability become the scarce layer. The approval queue can become the bottleneck just as easily as the old inbox did.

05 **The losses are real but uneven:** entry-level pathways, privacy, trust, attention, and local infrastructure carry more near-term risk than a cinematic jobs apocalypse.[6]

2026 POSTURE Aggressive about learning. Conservative about permissions. Ruthless about measurement. Unimpressed by vibes.

The market is converging technically—and diverging politically

The same underlying technology is being organized into different economic and regulatory systems. That matters even when your company has twelve employees.

> FACT As of March 2026, the top U.S. model led the top Chinese model by 2.7%, while the top closed model led the top open model by 3.3%. Those are gaps, not coronations.

MARKET POLE	WHAT IT IS OPTIMIZING	SMB TRANSLATION
UNITED STATES	Frontier labs, cloud distribution, venture capital, and premium managed services. The advantage is ecosystem depth; the risk is concentration, lock-in, and expensive scaling.	Buy capability, but insist on portability and usage visibility.
CHINA	Open-weight scale, aggressive pricing, domestic hardware adaptation, and strong industrial deployment. Chinese model families are now part of the global default set, not a side market.	Benchmark on your own language, data, and security requirements—not on nationality alone.
EUROPE	Risk-based regulation, content transparency, data sovereignty, and public investment in regional infrastructure. AI-generated public-interest content and deepfakes face new labeling duties from August 2026.	If you publish or serve EU users, make provenance and disclosure part of the workflow now.
THE REST	India, the Gulf, Southeast Asia, Africa, and Latin America are not waiting for a two-country verdict. Adoption is shaped by language, labor costs, mobile-first distribution, public infrastructure, and local regulation.	Global reach requires local evaluation. English benchmark leadership does not guarantee dialect, culture, or jurisdiction fit.

What the map means for a smaller business

DO THIS

- Keep data, prompts, evals, and workflow logic exportable.
- Test at least two model families on the same real cases.
- Separate provider choice from the customer-facing experience.
- Document where data is stored and which tools can act on it.

DO NOT DO THIS

- Sign a long exclusive contract because one leaderboard moved.
- Assume U.S.-built means automatically safer—or foreign-built means automatically risky.
- Call every released-weight model “open source.”
- Let geopolitics become an excuse for skipping technical due diligence.

What the numbers are not telling us

Benchmark gaps do not price the full system. They do not reveal uptime, real inference margins, provider survival, tool-use reliability, data retention, or the cost of fixing a wrong answer after it has touched a customer. They also hide language variance: a globally strong model can be locally mediocre.

OPERATOR RULE Design for a multi-model world. The model market may consolidate financially while remaining plural technically.

Open-weight is outside in a hoodie. Closed still owns the bouncer.

The binary is fake. The practical answer is a routing layer that uses the least expensive model able to clear the task.

Most systems marketed as “open source AI” are better described as **open-weight**: the model parameters are available, but the training data, full code, or usage freedoms may not be. The Open Source Initiative’s definition requires the ability to use, study, modify, and share the system, plus the preferred form needed to make modifications.[10]

OPTION	WHY IT WINS	WHAT YOU INHERIT
FRONTIER MANAGED	Highest capability on the hard tail; strong multimodal and safety tooling; low infrastructure burden.	Higher variable cost, policy changes, provider dependence, and opaque internals.
OPEN-WEIGHT / PRIVATE	Control, customization, data residency, predictable high-volume economics, and the ability to run inside your environment.	You inherit deployment, evaluation, patching, security, and model-operations work.
SMALL / LOCAL	Fast, cheap, private, and often excellent at classification, extraction, routing, or narrow repetitive tasks.	Lower ceiling and more brittle behavior outside the lane.

THE COST PARADOX

GPT-3.5-level inference became more than **280× cheaper** from late 2022 to late 2024. The token got cheaper. The workflow got hungrier.

SOURCE: STANFORD AI INDEX 2025 COST ANALYSIS[12]

Agentic systems call models repeatedly: plan, search, read, draft, check, revise, and log. A lower unit price can coexist with a larger total bill. Your cost model therefore needs **cost per accepted outcome**—not cost per token, seat, or prompt.

The demo is not the business case. Tuesday is.

AI pilots usually stall at the seam between an impressive answer and a repeatable operation. Your wallet is not imagining it.

The viral “95% of AI pilots fail” line came from a preliminary Project NANDA report focused on measurable P&L impact from integrated generative-AI initiatives. It was not a census of every AI project, and it should not be treated as a universal failure rate. Its useful point survives the hype: generic tools can help individuals while custom systems fail to learn the workflow, retain context, or reach production.[13]

The seven ways a good demo becomes expensive wallpaper

01 **No baseline.** The team cannot say how long, costly, or error-prone the current process is.

02 **No unit of “done.”** Output volume is celebrated while accepted completions, rework, and customer outcomes stay invisible.

03 **No system connection.** The chatbot writes; a human still copies, pastes, updates, schedules, and follows up.

04 **No memory.** Every session starts over, so the same corrections are paid for repeatedly.

05 **No exception loop.** The happy path works. The weird invoice, angry customer, or missing field sends the process into the ditch.

06 **No owner.** Everyone experiments; nobody owns accuracy, permissions, adoption, or results.

07 **No stop rule.** The project survives because it is interesting, not because it beats the old process.

AI ROI **ACCEPTED OUTCOMES ÷ FULL COST.** The vibes are free. Rework is not.

The one-page business case

ECONOMIC QUESTIONS

- What measurable outcome changes—and what is today's baseline?
- How often does the work occur, and what does delay cost?
- What data and permissions are required?

CONTROL QUESTIONS

- What quality threshold must the system clear?
- What happens when it is wrong, unavailable, or manipulated?
- Who owns the outcome, and when do we scale, redesign, or stop?

OECD research finds that SMEs still lag larger firms in AI adoption, with connectivity, data and compute, skills, and finance acting as core enablers. That is a polite way of saying the model is often the easiest part.[5]

From chatbot to supervised operator

The product is not the chat. The product is the completed, reviewable unit of work.

A chatbot waits for a prompt. An agent watches for a trigger, carries context, uses tools, moves a workflow forward, requests approval when needed, and records the result. The “agent” label is cheap; the operating loop is the actual asset.

Six pieces that separate an operation from a demo

LAYER	WHY IT EXISTS
1 **SHARED MEMORY**	Customers, voice, policies, decisions, and prior outcomes persist across sessions.
2 **WORKFLOW LOOPS**	Triggers, steps, retries, handoffs, and finish conditions are explicit.
3 **TOOL ACCESS**	The agent can act in approved systems with least-privilege credentials.
4 **APPROVAL GATES**	Risk determines what drafts, what batches for review, and what can run automatically.
5 **AGENTIC HQ**	A dashboard exposes queues, exceptions, cost, outcomes, and the work waiting on a human.
6 **EVALUATION**	Real cases score quality, speed, cost, and failure patterns before the lane expands.

THE APPROVAL ECONOMY When execution gets cheap, judgment gets expensive. “What is allowed to happen without me?” becomes an operating-design question.

NIST’s risk frameworks emphasize governance, measurement, documentation, pre-deployment testing, and incident management. For a smaller company, the translation is simple: explicit owners, narrow permissions, logs, tests, and a way to stop the system.[8]

The same Tuesday, twice

One workday, run both ways. This is the whole argument without the keynote music.

TIME	THE OLD WAY	WITH A SUPERVISED AGENT
7:40 AM	The owner opens 41 messages. Six matter; nobody knows which six yet.	The owner opens six decisions. The other 35 were triaged overnight, with drafts and reasons attached.
9:05 AM	A no-show is discovered after the slot is already empty.	An unconfirmed appointment was flagged at 7:00; the waitlist received an approved offer.
12:20 PM	A restock email is started, interrupted, and forgotten.	Sell-through crossed a rule. The agent drafted the order; a manager approved it from the dashboard.
3:40 PM	A lead from Monday finally gets a reply. The buyer already booked elsewhere.	The lead was enriched, routed, and answered in the owner's voice within the approved lane.
5:55 PM	The CRM, books, and project board disagree. Tomorrow begins with cleanup.	Systems were updated as the work happened; one exception is waiting for a human.
6:15 PM	The owner writes the same three follow-ups as last Tuesday.	The owner reviews three outcomes, changes one rule, and goes home.

Nothing on the right requires science fiction. The difference is not intelligence. **It is installation.**

Work moves first when it is…

AGENT-READY

- Recurring and deadline-shaped
- Driven by digital inputs
- Rule-bounded, with a checkable “done”
- Reversible or easy to escalate
- Frequent enough to justify wiring

KEEP HUMAN

- Hiring, firing, and hard conversations
- Negotiation endgames and pricing judgment
- Legal, safety, or money commitments
- Relationship moments where showing up is the message
- Taste decisions that define the brand

A probabilistic system can draft the apology. A person owns the apology. That boundary is not anti-AI; it is what makes sustained adoption possible.

AI takes tasks first. Careers still feel it.

The near-term labor story is not “everybody is fired.” It is fewer rungs, thinner teams, and more pressure to prove judgment earlier.

Stanford researchers using payroll data found a **16% relative employment decline** for workers ages 22 to 25 in the most AI-exposed occupations, while older workers in those occupations and workers in less-exposed fields remained stable or grew. The finding is concentrated, not economy-wide—but it is a real warning about the entry-level ladder.[6]

The true loss: apprenticeship

Junior work has always contained a lot of low-status repetition: first drafts, data cleaning, basic analysis, support tickets, test cases, and research. That work was inefficient, but it was also how people learned what good looked like. Remove the task without rebuilding the learning path and a company can become more productive this quarter while becoming less capable next year.

FOR OWNERS AND MANAGERS

- Automate the first draft; keep the review as a teaching surface.
- Track junior hiring and promotion, not only labor savings.
- Rotate people through exception handling—the place where judgment grows.
- Use saved time for customer work, training, and process improvement, not automatic workload inflation.

FOR WORKERS AND STUDENTS

- Move from formatting information to owning outcomes.
- Learn to verify sources, calculations, code, and claims.
- Build a portfolio that shows the problem, your reasoning, AI use, and final result.
- Keep unaided reps in writing, memory, analysis, and domain craft.

Three durable career moats

MOAT	WHAT IT MEANS
TRUST	Customers, teams, and partners choose people whose judgment and motives they understand.
CONSEQUENCE	The closer work gets to money, safety, legal responsibility, or a relationship, the more accountability matters.
CONTEXT	Physical reality, institutional history, taste, and tacit knowledge are difficult to compress into a prompt.

CAREER RULE Do not become the prompt clerk. Become the person who defines the workflow, judges the edge case, and owns the result.

For everyday people: establish a family verification phrase, confirm urgent financial requests through a second channel, and decide which personal data never enters a general-purpose AI tool.

The goods, the bads, and the losses that hide between them

AI can make a small firm feel larger. It can also make a fragile process fail at machine speed. Both are true.

THE GOODS

- Faster lead response and customer follow-up
- Lower cost of research, prototyping, and first drafts
- More accessible interfaces for writing, translation, vision, and speech
- Consistent queue maintenance after humans log off
- Earlier detection of missing data, delays, and anomalies
- A small team's ability to operate with big-company cadence

THE BADS

- Confident errors entering customer or financial systems
- Impersonation, phishing, and synthetic-media fraud
- Private data leaking through poorly governed tools
- Vendor lock-in and invisible usage costs
- Content slop overwhelming channels and degrading trust
- Work intensification when "time saved" becomes "more expected"

The true losses are quieter

01 **Approval debt.** A hundred decent drafts wait behind one owner who never redesigned the sign-off process.

02 **Skill atrophy.** People stop practicing the exact cognitive work they later need to audit.

03 **Accountability blur.** The vendor blames the user, the user blames the model, and the customer still has the problem.

04 **Data exhaust.** Every prompt, edit, and workflow decision can become valuable business memory—or a liability someone else controls.

05 **Local infrastructure stress.** Global percentages can look modest while a specific grid, water system, or community absorbs a concentrated burden.

ENERGY REALITY

Data centers used about **415 TWh** of electricity in 2024; the IEA projects roughly **945 TWh** by 2030. That is a small global share with very large local consequences.

The per-query guilt meme misses the point. Efficiency per task can improve while total demand rises because more people run more capable, multi-step systems. The planning problem is aggregate load and location.[9]

A fair test for any deployment

Does the system increase agency—or only output? Does it improve the customer outcome—or only shorten the employee's timer? Who gets the gain, who absorbs the error, and can the affected person appeal to a qualified human?

Demystify the fake news before it enters the budget

The biggest AI myths usually contain a useful signal wrapped in a bad conclusion.

“Open source AI is unsafe.”

Open-weight systems can be deployed responsibly or recklessly, just like managed systems. Openness changes the control surface; it does not replace governance.

ACTION Use the right term, evaluate licenses and data, and apply the same tests, permissions, and monitoring.

“95% of AI pilots fail.”

A preliminary report found little measurable P&L impact in a specific set of integrated initiatives. It did not prove that 95% of all AI use is worthless.

ACTION Take the integration warning seriously; do not use the headline as either gospel or an excuse to do nothing.

“Agents are autonomous employees.”

On structured computer-use benchmarks, leading agents still fail roughly one in three attempts. METR also warns that time-horizon estimates above 16 hours are unreliable with its current suite.

ACTION Deploy bounded operators with permissions, checkpoints, audit logs, and human escalation.

“AI will take 30% of jobs next year.”

Exposure is not displacement. Measured harm is currently concentrated in some entry-level, AI-exposed roles—not a broad labor-market collapse.

ACTION Watch hiring ladders and tasks, not one viral percentage.

“AI is basically free now.”

Model prices fell sharply. Integration, review, rework, security, drift, and agentic token volume did not disappear.

ACTION Measure cost per accepted outcome, including human time and exceptions.

“The best benchmark model wins.”

Benchmarks saturate, contain invalid items, and reward test-specific adaptation. A model can win elite math tests and still fail mundane perception.

ACTION Build a private eval set from your own work. Switch only when the lift is material.

Evidence: Stanford AI Index technical and responsible-AI chapters; METR time-horizon research; Stanford labor study; OSI definition; Project NANDA preliminary report.[2,4,6,7,10,13]

REALITY CHECK A model can be brilliant and brittle in the same minute. That is not a philosophical puzzle. It is a permissions problem.

The safest way to talk about AI is at the workflow level: this system, on these cases, with these tools, under these permissions, at this acceptance rate and full cost. Everything else is branding until proven otherwise.

Ten bets for the back half of 2026 through 2027

Each prediction includes a confidence level and a tripwire—the evidence that would make us change our mind.

1 Seats give way to supervised work

85%

The default AI purchase for smaller firms shifts from “another login” to a recurring unit of work: follow-up, reporting, reconciliation, content operations, scheduling, or support.

TRIPWIRE → seat-based copilots keep producing clearer ROI than installed workflows across 2027.

2 Model routing becomes standard

85%

Routine tasks flow to small or open-weight systems; the hard tail reaches frontier models. Customers increasingly buy the outcome without knowing—or caring—which model handled each step.

TRIPWIRE → one provider sustains a decisive quality-cost lead across most business tasks.

3 Open-weight commoditizes routine intelligence

80%

The best closed model keeps an edge on the hardest work, while open-weight families remain close enough to pressure price, improve sovereignty, and win high-volume lanes.

TRIPWIRE → a sustained gap above roughly 15 points appears on decontaminated, real-world evaluations.

4 Approval debt gets a name

80%

Companies deploy more agents than they design approval surfaces. The next bottleneck is not generation; it is one distracted manager holding 200 drafts hostage.

TRIPWIRE → low-risk autonomy becomes reliable enough that review queues shrink by default.

5 Shared business memory becomes the moat

75%

The durable asset is not the prompt. It is the permissioned record of customers, decisions, corrections, outcomes, and what the business learned.

TRIPWIRE → general models become so context-aware and persistent that proprietary memory adds little.

6 Autonomous “digital employees” remain rare

80%

Agents expand, but production systems stay bounded, supervised, and reversible through 2027—especially around money, customers, law, and safety.

TRIPWIRE → audited multi-day workflows exceed 90% reliability with minimal intervention across multiple firms.

7 The entry-level squeeze persists

65%

Hiring remains soft in exposed junior white-collar roles even without mass layoffs. Companies that do not rebuild apprenticeship create a mid-career talent problem later.

TRIPWIRE → junior postings and employment recover to trend while AI use continues rising.

8 Provenance becomes a product feature

70%

Content labels, source trails, and human-made signals become part of brand trust—especially in public-interest, commerce, and reputationally sensitive content.

TRIPWIRE → platforms and regulators fail to enforce disclosure and users show no preference for provenance.

9 An over-permissioned agent causes a reset

55%

A visible security or financial incident shifts the market from “give the agent tools” to least privilege, transaction limits, and stronger identity verification.

TRIPWIRE → agent adoption scales materially without a major permissions-related incident through 2027.

10 The killer apps stay boring

80%

The most valuable deployments remain lead response, service, reporting, reconciliation, scheduling, coding, and content operations—not a brand-new consumer universe.

TRIPWIRE → a new AI-native consumer category reaches durable mass-market retention and profit.

UNKNOWN WATCHLIST — Frontier inference economics · Taiwan / chip supply · energy interconnection · liability · consumer trust · the apprenticeship replacement rate.

Autoresearch loop: monthly, track model price and quality on your evals; exception and approval rates; labor and hiring signals; security incidents; regulatory changes; and customer trust. A prediction is only useful when it can lose.

Build one system, not ten experiments

A 90-day roadmap for owners who want evidence before a transformation budget.

Choose the first workflow with a readiness score

FACTOR	QUESTION	SCORE
FREQUENCY	Does it happen daily or weekly?	`0 rare · 1 monthly · 2 weekly+`
VERIFIABLE DONE	Can a human check completion quickly?	`0 subjective · 1 mixed · 2 clear`
REVERSIBILITY	Can mistakes be caught and undone?	`0 irreversible · 1 costly · 2 easy`
DATA ACCESS	Are inputs digital, available, and permitted?	`0 blocked · 1 partial · 2 ready`
ECONOMIC VALUE	Does delay, labor, or error materially cost money?	`0 low · 1 moderate · 2 high`
PROCESS STABILITY	Do people agree on the current rules?	`0 chaos · 1 exceptions · 2 stable`
OWNER	Is one person accountable for the result?	`0 none · 1 shared · 2 named`

SCORE INTERPRETATION

- **11–14:** install now with controlled permissions.
- **7–10:** redesign the process, then pilot.
- **0–6:** wait. The workflow—not the model—is the problem.

BEST FIRST LANES

- Lead response and follow-up
- Inbox and customer-service triage
- Recurring reporting and data reconciliation
- Content drafts, repurposing, and approvals
- Scheduling, reminders, and exception alerts

STOP RULE If the pilot cannot beat the baseline on accepted completion rate, cycle time, and full cost after a fair test, redesign it or kill it. Curiosity is not a renewal metric.

Build, buy, partner, or wait?

The right answer depends on whether AI is the product, the workflow, or just a useful tool inside the work.

PATH	USE IT WHEN	TRADE-OFF
USE A GENERAL TOOL	Individual drafting, research, brainstorming, and analysis where the user remains the operator.	Low setup; weak persistence and follow-through.
BUY A POINT SOLUTION	A standardized function such as meeting notes, support triage, or document extraction.	Fast time-to-value; limited cross-system customization.
PARTNER / FORWARD-DEPLOY	Recurring work spans multiple tools, needs business-specific memory, approvals, and operating support.	Best when the last mile is the hard part and internal capacity is thin.
BUILD INTERNALLY	AI behavior is core IP, differentiation is strategic, and the company can staff engineering, evaluation, security, and operations.	Maximum control; highest execution and maintenance burden.
WAIT	The process is unstable, data access is unresolved, failures are irreversible, or success cannot be measured.	Waiting is rational when it prevents expensive theater.

The operator scorecard

METRIC	WHAT IT ANSWERS
ACCEPTED COMPLETION RATE	What share of tasks finish at the required quality without redo?
CYCLE TIME	How long from trigger to accepted outcome?
HUMAN REVIEW TIME	Did the system remove work—or create a faster pile to inspect?
EXCEPTION RATE	How often does the case leave the lane, and why?
COST PER ACCEPTED OUTCOME	Model + infrastructure + integration + human review + rework.
BUSINESS RESULT	Revenue recovered, cost removed, response time, customer retention, or risk reduced.

BUYER WARNING "Unlimited AI" is not a financial model. Ask what happens when volume grows, models change, tokens spike, or the vendor disappears.

Contract terms worth fighting for

- Data export and deletion rights
- Access to logs and usage reporting
- Notice of material model or policy changes
- Clear retention, training, and subprocessor terms
- An exit plan for prompts, memory, workflows, and integrations
- Service levels tied to the workflow—not only platform uptime

The FlipTech Pro operating thesis

We do not think smaller companies need another AI tab. They need a supervised operating layer installed around the work that already costs them.

This report's slant is transparent: FlipTech Pro builds and licenses a configurable agent workspace—an **Agentic HQ**—around a company's recurring pain points. The agent is customized to the business, the dashboard shows the work and the receipts, and workflow loops connect the chat to real systems. Agents draft; humans approve. Autonomy expands only inside lanes the owner chooses.

What the platform is designed to operationalize

PRINCIPLE	WHAT IT MEANS
ONE BUSINESS BRAIN	Shared memory across content, sales, service, and operations—so the system does not reset every Monday.
WORKFLOW LOOPS	Triggers, schedules, retries, handoffs, approvals, and measurable finish conditions around the highest-friction work.
MODEL ROUTING	Frontier capability where judgment pays for it; efficient open-weight or small models where routine volume dominates.
AGENTIC HQ	One surface for chat, queues, approvals, metrics, exceptions, and the company's growing memory.
HUMAN + AGENT POD	Forward-deployed people handle the last mile, tune the system, and own incidents instead of wishing the software good luck.
PORTABILITY	The client's site, content, and data remain exportable; the underlying engine is licensed, not disguised as bespoke magic.

THE COMMERCIAL MODEL

A **$10,000** build, **fourteen days** to keys, then a partnership from **$500/month** with AI usage included.

A qualified business can receive a password-protected working demo of its own company before a sales call. Terms and pricing can change; the operating principle should not: **proof before pitch**.

Where we deliberately keep humans

Nothing posts, sends, or ships without approval until the owner loosens a lane on purpose. We do not promise outcomes that cannot be measured, and regulated-data workflows requiring controls the standard stack cannot meet run on different rails—or we pass.

BRING THE JOB DESCRIPTION

Point us at the expensive, repetitive process nobody wants to own. We will show the work before we sell the work. Explore the platform and request a working demo at **fliptechpro.com**.

Sources, method, and limits

A forecast is useful only when it can be wrong. The sources below support the facts; the interpretations and bets are ours.

Method. We prioritized primary research, public institutions, and official policy sources available through July 17, 2026. Provider and preliminary research is used when it contributes a useful signal, but it is labeled. Benchmarks are treated as capability indicators—not deployment guarantees. Predictions are probability-weighted judgments, not certainties.

1 **Stanford HAI — 2026 AI Index Report.** Overview of capability, investment, adoption, labor, science, policy, and public opinion.

2 **Stanford HAI — Technical Performance, 2026 AI Index.** Closed/open and U.S./China gaps, agent benchmarks, benchmark quality, and jagged capability.

3 **Stanford HAI — Economy, 2026 AI Index.** Organizational adoption, early agent use, investment, productivity, and labor expectations.

4 **Stanford HAI — Responsible AI, 2026 AI Index.** Hallucination, governance, transparency, multilingual performance, and safety trade-offs.

5 **OECD — AI Adoption by Small and Medium-Sized Enterprises.** SME adoption gaps and the role of skills, data, compute, connectivity, and finance.

6 **Stanford Digital Economy Lab — Canaries in the Coal Mine?** Administrative payroll evidence on younger workers in AI-exposed occupations.

7 **METR — Task-Completion Time Horizons of Frontier AI Models.** Agent reliability over task duration and measurement limits.

8 **NIST — Generative AI Profile for the AI Risk Management Framework.** Cross-sector guidance on governance, testing, measurement, incident disclosure, and risk.

9 **International Energy Agency — Energy and AI.** Global data-center electricity use, projections, local concentration, and uncertainty.

10 **Open Source Initiative — Open Source AI Definition 1.0.** Definition of use, study, modify, and share—and the distinction from open weights.

11 **European Commission — AI Act implementation and transparency.** Official timeline and AI-generated content transparency obligations.

12 **Stanford HAI — Research and Development, 2025 AI Index.** Inference cost declines and hardware efficiency trends.

13 **Project NANDA — The GenAI Divide: State of AI in Business 2025.** Preliminary report behind the viral 95% claim; used with explicit methodological caution.

14 **NIST — AI Risk Management Framework 1.0.** Use-case-agnostic framework for governing and managing AI risk.

15 **Stanford HAI — Public Opinion, 2026 AI Index.** Global optimism, anxiety, workplace use, and trust in regulation.

Publication note: this report is educational and strategic, not legal, financial, employment, security, or regulatory advice. Re-check time-sensitive claims and jurisdiction-specific obligations before acting.

BRING THE JOB DESCRIPTION

We will show the work before we sell the work.

Point us at the expensive, repetitive process nobody wants to own.
A working demo of your business, before we ever get on a call.

Get my demo fliptechpro.com