



AI Readiness: How Companies Move from AI Pilots to Production in 2026

Research Report

Katerina Merzlova & Kirill Funtikov, SumatoSoft

Contents

TL;DR	2
Key statistics from this research	3
Abstract	4
What is AI readiness?	5
AI readiness research demographics and key findings	7
Respondent profile	7
Survey respondents by country	7
Survey respondents by industry	8
How we validate responses	9
Key findings	9
What enabled production deployment	10
The #1 factor: Workflow redesign (61%)	10
Data readiness (22%)	10
Executive sponsorship (14%)	10
Other enablers (3%)	11
Biggest AI readiness gaps	12
Data quality and consistency (58%)	12
Integration with legacy systems (24%)	12
Security, privacy, and compliance (11%)	13
Other gaps (7%)	13
What changed after AI deployment	14
Cycle time compression	14
Error rate reduction	15
Capacity and throughput expansion	17
Cost and financial impact	17
How companies measure value and manage risk	18
Value metrics	18
Risk controls	18
What made companies scale AI or stop the implementation	20
Scaling triggers	20
Stopping triggers	20
How our findings align with industry research	21
The SumatoSoft AI Readiness Checklist	22
Conclusion	23
References	24
How to cite this research	27

TL;DR

- **Workflow redesign drives production deployment.** 61% of respondents cited rebuilding processes around AI as the #1 factor that moved them from pilot to production. Data readiness came second at 22%, executive sponsorship at 14%, and MLOps or cross-functional factors at 3%.
- **Data quality is the most common blocker.** 58% of respondents named messy, fragmented, poorly labeled, or inconsistently structured data as their biggest readiness gap. Every company that skipped data standardization reported unreliable outputs and delayed deployment.
- **Measured results cluster in a consistent range.** Cycle time reductions of 35–40% are typical within the first 90 days. Error rates in structured extraction and classification tasks dropped 50–90%. Small teams regularly operated at 2–3 times their previous capacity. Direct cost savings of 20–35% were reported across logistics, content production, construction, and healthcare operations.
- **Human oversight is universal.** 96% of respondents maintain human-in-the-loop review for any customer-facing, compliance-sensitive, financially significant, or legally binding AI output. No respondent reported fully autonomous AI in these workflows.

Key statistics from this research

- According to SumatoSoft's 2026 survey of 72 executives, 61% named workflow redesign as the #1 factor that moved AI from pilot to production.
- 58% of respondents in SumatoSoft's 2026 AI readiness research identified data quality and consistency as their biggest readiness gap.
- 96% of the 72 leaders surveyed by SumatoSoft maintain human-in-the-loop review for customer-facing, compliance-sensitive, financially significant, or legally binding AI output.
- The median efficiency gain in SumatoSoft's 2026 study is a 35–40% reduction in cycle time, task duration, or error rate within the first 90 days of production deployment.
- Companies in the SumatoSoft dataset that standardized data inputs before deploying AI reported 2–3× higher output reliability than those that cleaned data during deployment.
- In structured extraction, triage, and classification tasks, respondents to SumatoSoft's survey reported error-rate reductions of 50–90% and cycle-time compression of 60–80%.

Abstract

88% of organizations now use AI in at least one business function, according to [McKinsey's 2025 State of AI survey](#). Yet most never move beyond experimentation. Only about one-third have begun scaling AI across their enterprises, and MIT researchers found that [95% of generative AI pilots fail to deliver measurable P&L impact](#).

The gap between piloting AI and running it in production is wide, but the reasons behind it remain poorly understood at the operational level. To build a ground-level view of the transition, we surveyed executives and functional leaders involved in implementing AI. The dataset includes 72 validated responses from leaders across 30+ industries who shared their workflows, described improvements, explained the organizational changes that enabled deployment, and quantified results.

Technology was rarely the bottleneck. Workflow redesign was the primary enabler of production deployment, cited by 61% of respondents, followed by data readiness (22%), executive sponsorship (14%), and MLOps or cross-functional factors (3%). The most common readiness gap was data quality and consistency (58%), followed by integration with legacy systems (24%), security/privacy concerns (11%), and other factors (7%).

96% of respondents maintain a human-in-the-loop control on AI outputs, particularly for customer-facing or compliance-sensitive work. Reported efficiency gains cluster in a consistent range: 30–40% reductions in cycle time, task duration, or error rates within the first quarter of production deployment.

This report gives executives, product owners, operational leaders, and technology teams an evidence-based picture of what AI readiness looks like: what to fix first, what results to expect, what risks to control, and what organizational conditions separate a pilot that ships from one that fails.

The report closes with the SumatoSoft AI Readiness Checklist: ten conditions that separated pilots that shipped from pilots that stalled.

What is AI readiness?

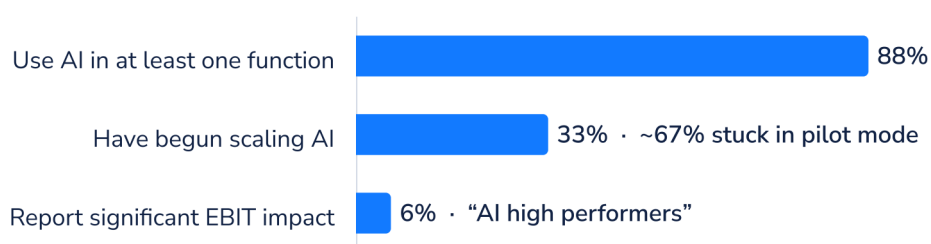
AI readiness is a company's ability to move AI from pilot to production within specific processes and keep it there without degrading quality or increasing risk.

Readiness consists of five interrelated components:

1. **Data readiness.** A company is data-ready when its information sources are defined and managed: it's clear where the "truth" lies, who owns the data, how documents are updated, and how versions are processed. There is a defined entry contract: format, required fields, validation rules, and restrictions on certain data types. Access is limited by roles, and sensitive domains are segregated. Without this, AI begins to hallucinate on junk input or outdated sources. If a company isn't sure whether its data is ready for AI implementation, it can start with [AI consulting services](#) for a data audit.
2. **Workflow readiness.** A company is process-ready when AI occupies a specific place in the workflow. There's a unit of work (ticket, request, report, case, call), and AI's actions are defined: extracting fields, generating a draft, classifying, and suggesting the next step. A human's actions are also defined: approving, correcting, rejecting, or escalating. A key part is handling uncertain cases: what happens when confidence is low, sources conflict, or non-standard input arrives.
3. **Governance readiness** is management and control readiness. A company is governance-ready when risk is controlled by operational rules. It's clear where a human is needed, which responses or actions require confirmation, what should be logged, and how traceability is ensured. There's a policy for accessing AI tools and prohibiting "shadow AI," where employees transfer sensitive data to public services. Change management is in place: who approves prompt and model updates, how quality assurance is conducted, how rollbacks are handled, and how changes are documented. Governance is a set of operational constraints, not a policy document on a shelf.
4. **Organizational readiness.** A company is organizationally ready when process owners are appointed, and lines of responsibility between product, IT, security, and business are clear. Teams know how to work with probabilistic answers: they know what to check, where not to trust, how to log errors, and how to improve source data or rules. Standards for inputs and outputs ensure AI results don't disrupt team collaboration.
5. **Infrastructure readiness.** A company is infrastructure-ready when AI can be integrated into record systems and monitored as a production component. There's secure key and secret storage, network boundaries, environment isolation, and leak control. There's quality and cost monitoring, logs, alerts, and metrics for errors and fixes. There's a release framework: a test environment, gradual rollouts, and rollback capabilities.

When any of these dimensions is missing, the result is what the industry calls "pilot purgatory": AI initiatives show initial promise in testing but never scale to production or deliver business impact. According to IDC research, for every 33 AI pilots a company launches, only 4 make it to production, which translates to an 88% failure rate.

Figure 1. The pilot purgatory funnel



Source: McKinsey, *The State of AI 2025* (n=1,993).

Stage	Share	Note
Use AI in at least one function	88%	McKinsey 2025 (n=1,993)
Have begun scaling AI	33%	~67% stuck in pilot mode
Report significant EBIT impact	6%	"AI high performers"

AI readiness research demographics and key findings

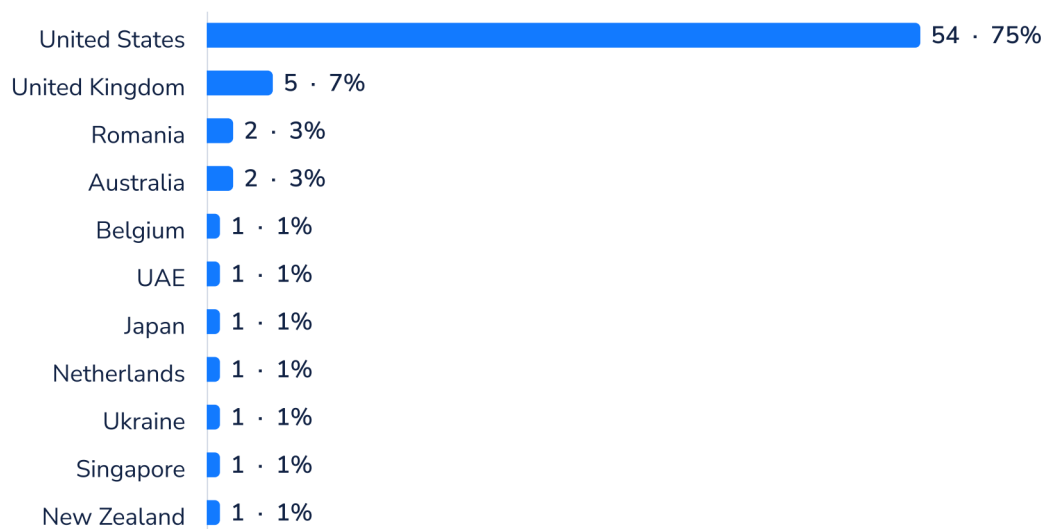
We compiled our database from 72 verified responses from executives and functional leaders directly involved in AI implementation, collected over 21 days via MentionMatch and Featured. Only responses with specific, measurable results and a clear description of the production workflow were included in quantitative analysis.

Respondent profile

- 54% were founders, CEOs, presidents, or sole operators
- 26% were directors, VPs, or heads of product/operations/marketing
- 20% were CTOs, engineering leads, ML engineers, or data scientists
- Company sizes ranged from solo operators to enterprises with 10,000+ employees
- 30+ industries represented, including SaaS, fintech, healthcare, logistics, construction, legal tech, real estate, manufacturing, digital marketing, and consumer products
- Geography: primarily North America, with representation from Europe, Asia-Pacific, and the Middle East

Survey respondents by country

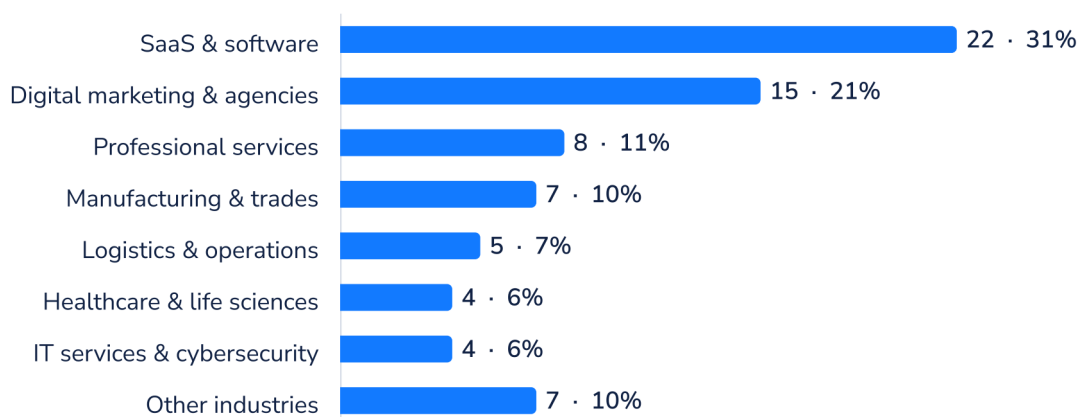
Figure 2. Survey respondents by country (n=72)



Country	Respondents	Share
United States	54	75%
United Kingdom	5	7%
Romania	2	3%
Australia	2	3%
Belgium	1	1%
UAE	1	1%
Japan	1	1%
Netherlands	1	1%
Ukraine	1	1%
Singapore	1	1%
New Zealand	1	1%
Total shown	70	97%

Survey respondents by industry

Figure 3. Survey respondents by industry (n=72)



Industry	Respondents	Share
SaaS & software	22	31%
Digital marketing & agencies	15	21%
Professional services	8	11%
Manufacturing & trades	7	10%
Logistics & operations	5	7%
Healthcare & life sciences	4	6%
IT services & cybersecurity	4	6%
Other industries	7	10%
Total	72	100%

How we validate responses

We collected responses through MentionMatch and Featured over a 21-day window. Each response was screened against four criteria before inclusion: a named company or a credible anonymized context, a defined production workflow, at least one measurable result with a baseline, and a description of the barrier the team addressed. Responses that offered general enthusiasm, unverifiable claims, or numbers detached from the described workflow were excluded or used as background only.

All figures are self-reported by the named respondents and were not independently audited. We treat them as directional evidence rather than controlled experimental proof. Some companies have participated in more than one SumatoSoft study; in every case, each study uses only the responses submitted to its own questions.

Key findings

1. Workflow redesign was the #1 cited enabler of production deployment (61%).
2. Data quality and consistency were the #1 cited readiness gap (58%).
3. Human-in-the-loop review is the dominant risk control, maintained by 96% of respondents for customer-facing or compliance-sensitive output.
4. The median reported efficiency gain is a 35–40% reduction in cycle time, task duration, or error rate within the first 90 days.
5. Companies that standardized data inputs before deploying AI reported 2–3× higher output reliability than those that cleaned data during deployment.
6. The largest improvements occurred in structured extraction, triage, classification, and first-draft generation tasks, where error rates dropped by 50–90% and cycle times compressed by 60–80%.
7. No respondent reported fully autonomous AI without human oversight in any customer-facing, compliance-sensitive, financially significant, or legally binding workflow.

What enabled production deployment

We asked every respondent to name the single most important factor that moved their AI from pilot to production. The responses converged.

The #1 factor: Workflow redesign (61%)

Across industries and company sizes, the dominant answer was redesigning the business process so that AI had a defined role within it. Respondents described this in similar terms: they stopped trying to “add AI” to existing processes and instead rebuilt the workflow with AI as a participant from the start.

Roman Milyushkevich, CTO at HasData: “Our early pilots failed because we tried to layer AI on top of existing processes. After several failures, we decided to redesign the entire workflow so AI could have a defined role, but never make the final decision without human oversight.”

David Kemmerer, CEO of CoinLedger, described the same pattern in customer support: “AI isn’t something you turn on. We had to move from a human-first queue to an AI-first triage model. It meant redesigning the hand-off triggers and letting AI become the filter.”

This aligns with McKinsey’s State of AI data: high-performing organizations are nearly 3x more likely to have redesigned workflows as part of their AI efforts. Winners don’t bolt on models. They rebuild processes.

Data readiness (22%)

The second most cited enabler was having clean, structured data before attempting production deployment. Companies in data-intensive environments such as marketing analytics, financial services, IoT telemetry, and logistics consistently named this as the prerequisite that made everything else possible.

Ilya Telegin at Improvado described how standardizing data from dozens of advertising platforms made AI-generated dashboards trustworthy: once the ETL layer was built to clean, map, and standardize incoming data, dashboard generation time dropped from 6 hours to 15 minutes.

Executive sponsorship (14%)

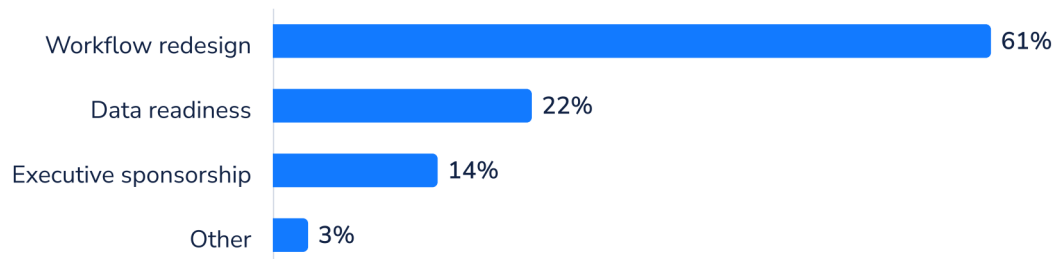
A smaller group identified executive support as the critical unlock, particularly in regulated industries and legacy-heavy environments where AI required significant budget for integration work.

Clayton Eidson, CEO of AZ Health Insurance Agents, noted that without his direct intervention, the team “would have remained using spreadsheets.” In small organizations, where the CEO is often the only person with authority to force workflow changes, sponsorship and workflow redesign become inseparable.

Other enablers (3%)

A small number of respondents cited [MLOps/LLMOps infrastructure](#), cross-functional collaboration, vertical integration, or domain-specific model tuning as their primary enabler.

Figure 4. #1 factor that enabled production deployment



Factor	Share
Workflow redesign	61%
Data readiness	22%
Executive sponsorship	14%
Other	3%
Total	100%

Biggest AI readiness gaps

We asked respondents to identify the single largest gap they had to close before AI could operate reliably in production.

Data quality and consistency (58%)

The most common readiness gap was that the data was too messy, fragmented, poorly labeled, or inconsistently structured for AI to produce reliable outputs. This manifested differently across industries, but the pattern was the same: organizations had the information AI needed, but not in a form AI could consume.

Support tickets were “historically unstructured and inconsistently tagged” (HasData). Help center articles were “full of outdated articles from previous tax years” (CoinLedger). Design templates and user behavior data were “scattered across multiple systems” (Brizy). Product catalogs had been “entered by more than one person over a number of years” with “inconsistent naming, lack of specs, duplicate records” (Medmart).

The fix: standardizing input formats, cleaning legacy records, consolidating data into centralized repositories, and introducing validation rules before AI processing. Multiple respondents described weeks or months of manual data cleanup as the prerequisite to production deployment.

Specific improvements after data standardization:

- HasData: first-response drafting time dropped 38% within one month
- Brizy: landing page design time from 4 hours to under 45 minutes
- Improvado: marketing dashboard generation from 6 hours to 15 minutes
- ITECH Recycling: destruction certificate accuracy from ~80% to 98%+
- 8Rental: dispatch preparation time from 18 minutes to 4 minutes per booking (78% reduction)
- Proximity Plumbing: booking error rate from 12.4% to 3.1%

Integration with legacy systems (24%)

The second most common gap involved connecting AI tools to existing software infrastructure: CRMs, ERPs, ticketing systems, and proprietary platforms never designed for AI integration.

Ketch built “API bridges and lightweight middleware” to connect AI pipelines to older data storage and CRM tools. Alliance InfoSystems virtualized legacy server workloads and deployed data connectors to stream data from on-premises systems to cloud AI. Sundance Networks created “custom hybrid APIs” to bridge legacy on-premises systems with AI tools, reducing integration time from 6 weeks to 2.

Security, privacy, and compliance (11%)

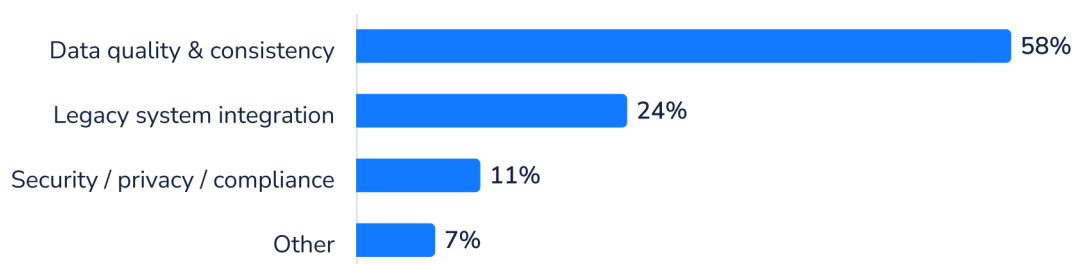
Organizations handling sensitive data (healthcare records, financial information, personal data, and law enforcement case files) cited security and compliance as their primary blocker. The challenge was proving that AI systems could handle sensitive information within regulatory boundaries.

Repello AI addressed the “who watches the watchmen” problem by isolating their AI red-teaming engine’s inference pipeline from customer model outputs at the architecture level, reducing false-negative rates from 3.2% to 0.04%. Tracker Products built AI features with strict RBAC and MFA, logging every prompt and output as immutable audit events to satisfy CJIS requirements for law enforcement data.

Other gaps (7%)

A smaller number of respondents cited team adoption, trust, output reliability, or context management as their primary readiness gap.

Figure 5. Biggest AI readiness gap



Gap	Share
Data quality & consistency	58%
Legacy system integration	24%
Security / privacy / compliance	11%
Other	7%
Total	100%

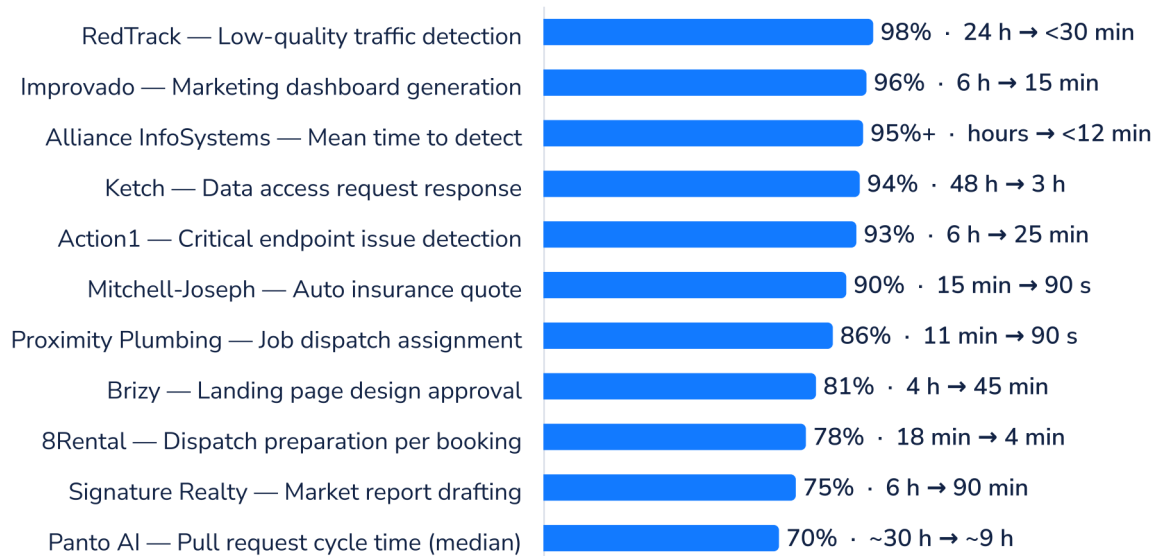
What changed after AI deployment

The reported improvements fall into four categories: cycle time compression, error rate reduction, capacity expansion, and cost savings.

Cycle time compression

The most consistently reported improvement was a reduction in how long specific tasks take. AI compressed research, triage, classification, and first-draft generation from hours to minutes.

Figure 6. Cycle time compression before vs after AI (reported improvement per workflow)

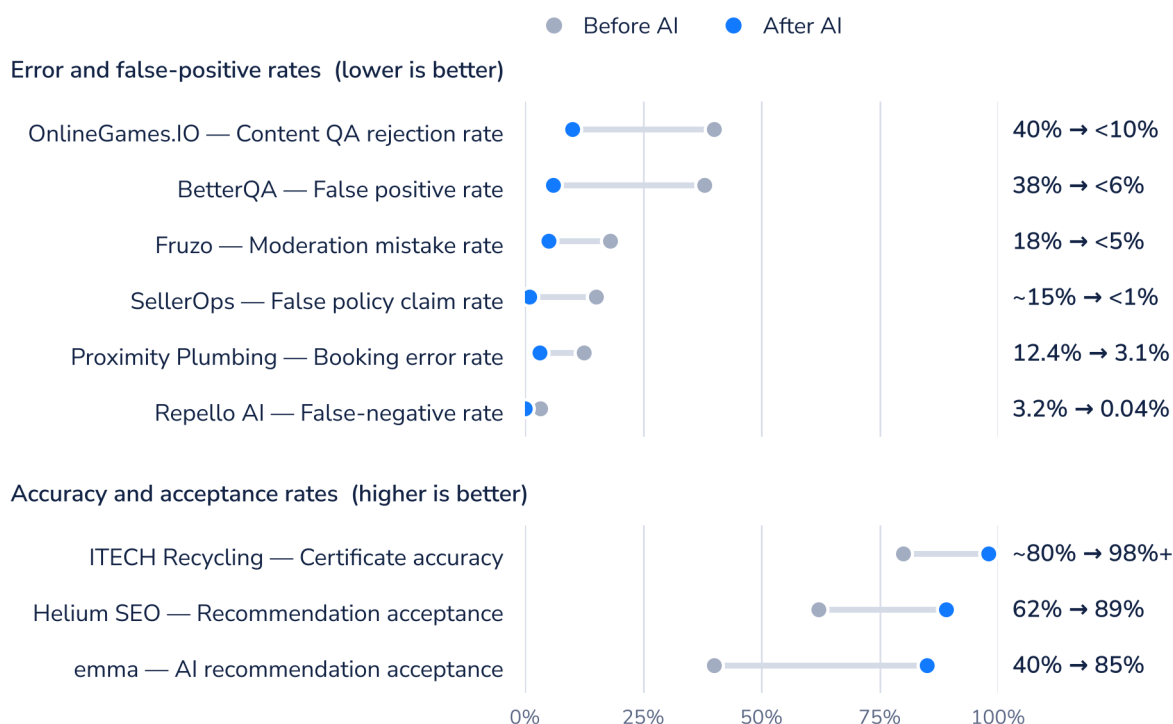


Company	Task	Before	After	Improvement
Improvado	Marketing dashboard generation	6 hours	15 minutes	96% faster
Mitchell-Joseph Insurance	Auto insurance quote generation	15 minutes	90 seconds	90% faster
Brizy	Landing page design approval	4 hours	45 minutes	81% faster
8Rental	Dispatch preparation per booking	18 minutes	4 minutes	78% faster
Panto AI	Pull request cycle time (median)	~30 hours	~9 hours	70% faster
Signature Realty	Market report drafting	6 hours	90 minutes	75% faster
Action1	Critical endpoint issue detection	6 hours	25 minutes	93% faster
Ketch	Data access request response	48 hours	3 hours	94% faster
Alliance InfoSystems	Mean time to detect (MTTD)	Several hours	Under 12 minutes	95%+ faster
RedTrack	Low-quality traffic detection	24 hours	Under 30 minutes	98% faster
Proximity Plumbing	Job dispatch assignment	11 minutes	90 seconds	86% faster

Error rate reduction

AI-assisted classification, extraction, triage, and moderation tasks showed consistent accuracy improvements once data inputs were standardized.

Figure 7. Error rate reduction before vs after AI



Mobile Vision Technologies (90% fewer false-positive security alerts) is not plotted because no baseline rate was reported.

Company	Metric	Before	After	Improvement
ITECH Recycling	Destruction certificate accuracy	~80%	98%+	+18 pp
Repello AI	False-negative rate (vulnerability detection)	3.2%	0.04%	80× better
SellerOps	False policy claim rate	~15%	Under 1%	15× better
BetterQA	False positive rate (productivity detection)	38%	Under 6%	6× better
Mobile Vision Technologies	False-positive security alerts	Baseline	90% reduction	10× better
Proximity Plumbing	Booking error rate	12.4%	3.1%	75% reduction
Fruzo	AI moderation mistake rate	18%	Under 5%	72% reduction
OnlineGames.IO	AI-generated content QA rejection rate	40%	Under 10%	75% reduction
Helium SEO	Recommendation acceptance rate	62%	89%	+27 pp
emma (cloud management)	AI recommendation acceptance rate	40%	85%	+45 pp

Capacity and throughput expansion

Several respondents reported that AI lets small teams operate at multiples of their previous capacity by eliminating administrative overhead and repetitive tasks.

Numonic, a 4-person startup, shipped a production platform with 122 API endpoints, 100+ database migrations, and 1,700+ automated tests while simultaneously closing a pre-seed funding round. Their CEO attributed this to AI being embedded as infrastructure in their development and content workflows.

CoinLedger's Fin AI increased automated resolution of customer queries from 14% to 43%, absorbing a 10× seasonal volume spike without additional hiring. Proximity Plumbing increased dispatcher throughput from 4.2 jobs per hour to 6.8 jobs per hour. Blackbelt Commerce helped a client achieve 700% organic sales growth in 3 months through AI-assisted SEO operations.

Cost and financial impact

Direct cost reductions were reported across multiple domains:

- Hanzo Logistics: transportation costs down 25% after AI route optimization, overstocking incidents down 35%
- Gunita Restoration: material project overages from 12% to 3%
- LocalSEO: cost per published content asset from \$125 to \$75
- Sundance Networks: client SLA from 99.2% to 99.8%
- ProMD Health Bethesda: no-show rate from ~14% to ~9%, consult-to-book rate improved 18%

How companies measure value and manage risk

Value metrics

Respondents track a wide range of metrics, but they cluster around one principle: measure the outcome AI is supposed to improve. The most frequently cited categories:

Time-based metrics (most common): cycle time reduction, time-to-resolution, time-to-first-draft, response time, time-from-pickup-to-certificate. These are favored because they're easy to measure and directly tied to operational cost.

Quality-based metrics: error rates, acceptance rates, rework frequency, false positive/negative rates. These matter most in regulated industries and compliance-sensitive environments.

Business outcome metrics: conversion rates, revenue per customer, cost per acquisition, and client retention. A smaller but growing group of respondents ties AI value directly to financial outcomes rather than operational proxies.

Risk controls

The risk management picture is uniform. Across 72 responses spanning 30+ industries, the controls converge on one principle: AI drafts, humans decide.

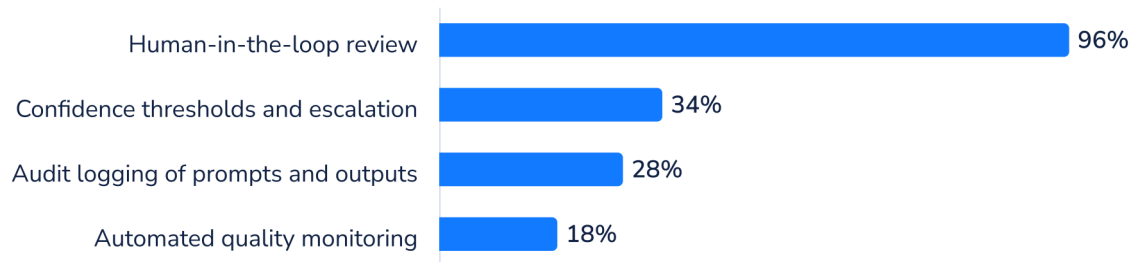
Human-in-the-loop review (96%): Nearly every respondent requires human review before any AI output reaches a customer, patient, regulator, or external stakeholder. Jessica-Lee Tingley at The JobBridge: "A misrepresented disability disclosure in a cover letter isn't a UX bug, it's a legal and human harm issue."

Confidence thresholds and escalation (34%): AI outputs with a confidence score below a defined threshold are automatically routed to human review. Nicky Zhu at Dymesty described setting a 0.74 confidence threshold gate, with predictions below that score suppressed and the system reverting to its original behavior.

Audit logging of prompts and outputs (28%): Common in regulated industries, financial services, and legal tech, where traceability is a compliance requirement.

Automated quality monitoring (18%): Continuous monitoring for model drift, accuracy degradation, or anomalous outputs, with alerts when performance drops below defined thresholds.

Figure 8. Risk controls used in production AI deployments (% of respondents, n=72)



Control	Share
Human-in-the-loop review	96%
Confidence thresholds and escalation	34%
Audit logging of prompts and outputs	28%
Automated quality monitoring	18%

What made companies scale AI or stop the implementation

Scaling triggers

Respondents described consistent patterns for expanding AI from initial deployment to broader use:

Sustained accuracy above defined thresholds. HasData scaled once structured extraction accuracy held above 90%. Repello AI expanded after vulnerability detection rates stabilized at 94.6%. The common thread: a quantitative bar, held for a defined period, before expanding scope.

Consistent savings without increased rework. Multiple respondents described a “prove it works, then expand” pattern: running AI on a contained scope for 60–90 days, measuring savings, and expanding only after confirming that efficiency gains didn’t create downstream problems.

Volume pressure that made manual processes untenable. CoinLedger scaled because tax season created a 10× volume spike that couldn’t be staffed manually. ITECH Recycling scaled after a single data center decommission brought in 400+ devices in one day.

Executive confidence secured for expansion budget. In several organizations, the first deployment’s measured results became the business case for broader rollout. Once leadership could see the numbers from a contained scope, they approved the budget and headcount for the next phase.

Stopping triggers

Respondents were equally specific about when they paused or killed AI initiatives:

Error rates exceeding defined thresholds. Several respondents described hard stops triggered by accuracy degradation: false positive rates climbing above 5%, recommendation quality dropping, rework increasing, or customer complaints spiking.

Compliance or governance gaps. HasData paused expansion in areas where “AI outputs introduced ambiguity or compliance concerns” until governance mechanisms were in place. Repello AI would “pause or roll back deployment if correction rates exceeded defined risk thresholds.”

No measurable financial impact. Bully Max stopped experiments that “produced insights but no measurable financial impact.” Venture Smarter advises clients to abandon projects that fail to meet financial benchmarks within two consecutive quarters.

AI creating more work than it saved. Lighthouse Energy stopped using AI for free-form code interpretation after it “confidently suggested an approach that would’ve failed local inspector preference.” Bruce Kemp now restricts AI to drafting within approved templates only.

How our findings align with industry research

Our data from 72 practitioners aligns with, and adds operational detail to, findings from major industry surveys.

On the pilot-to-production gap: McKinsey's State of AI survey found that roughly two-thirds of organizations have not begun scaling AI across the enterprise. IDC shows that for every 33 AI pilots launched, only 4 reach production. Our data confirms this but adds a useful detail: respondents who reached production consistently identified the same organizational prerequisites (workflow redesign, data standardization, clear ownership, and defined success metrics) rather than technology improvements.

On workflow redesign as the critical enabler: McKinsey found that high-performing organizations are nearly 3× more likely to have redesigned workflows as part of their AI efforts. In our survey, 61% named workflow redesign as the #1 factor. BCG's 10-20-70 principle (10% algorithms, 20% data and technology, 70% people, processes, and cultural transformation) maps directly onto what our respondents described.

On the data quality bottleneck: Our finding that 58% cited data quality as their biggest readiness gap is consistent with Gartner's emphasis on "AI-ready data" as a distinct organizational capability. Multiple respondents spent weeks or months on data cleanup before AI could function reliably. This echoes S&P Global Market Intelligence's finding that data quality is a primary reason AI projects stall after proof-of-concept.

On human oversight as standard practice: 96% of our respondents maintain human-in-the-loop controls. Even McKinsey's high performers, the 6% seeing significant EBIT impact from AI, maintain human review gates. The difference is where in the workflow the review occurs and how much AI handles before that gate.

On the value measurement challenge: McKinsey found that only 39% of organizations report EBIT impact at the enterprise level from AI. Our respondents generally track value at the workflow level (time saved, errors reduced, throughput increased, and cost avoided) rather than at the enterprise P&L level, suggesting that even successful deployments often lack the measurement infrastructure to show organization-wide financial impact.

The SumatoSoft AI Readiness Checklist

Based on patterns across 72 validated responses, the following conditions appear to be prerequisites for moving AI from pilot to production.

Before deployment:

1. **Map the target workflow end-to-end.** Identify where human judgment is irreplaceable and where AI handles execution. Multiple respondents described this mapping exercise as the most important pre-deployment step.
2. **Standardize data inputs.** Clean, consolidate, validate, and structure the data AI will consume. Every respondent who skipped this step reported unreliable outputs and delayed deployment.
3. **Define success metrics before launching.** Tie AI to a specific, measurable business outcome. Respondents who defined metrics upfront moved to production faster than those who tried to demonstrate value after the fact.
4. **Assign a single owner.** One person accountable for the workflow, the metrics, the decision to ship or kill, and the escalation path when things break. Multiple respondents cited clear ownership as the factor that separated their successful deployment from earlier failures.

During deployment:

5. **Start with a contained scope.** Deploy AI on a single workflow or a single customer segment. Expand only after sustained performance above defined thresholds.
6. **Maintain human review gates.** No AI output should reach a customer, a patient, a regulator, or an external stakeholder without human verification until accuracy and reliability are proven over a meaningful period.
7. **Log everything.** Prompts, outputs, confidence scores, human overrides. This data is needed for debugging, compliance, and continuous improvement.
8. **Set confidence thresholds.** Define the score below which AI outputs are suppressed or routed to human review rather than delivered to end users. Multiple respondents cited this as the control that caught errors before they reached customers.

After deployment:

9. **Monitor for drift.** AI performance degrades as data patterns change. Set up automated monitoring with defined thresholds for when to retrain, pause, or roll back.
10. **Scale based on evidence.** Expand AI to new workflows only when the current deployment shows sustained, measurable improvement without increased rework or error rates.

Conclusion

Across 72 case studies from leaders who have deployed AI in production, the pattern is consistent: the organizational prerequisites matter more than the technology choices.

Companies that moved from pilot to production share a common set of practices: they redesigned workflows before deploying AI, standardized their data before asking AI to consume it, assigned clear ownership, and maintained human oversight throughout.

The reported results are consistent. Cycle time reductions of 35–40% are typical. Error rate improvements of 50–90% are common in structured extraction and classification tasks. Small teams regularly operate at 2–3× their previous capacity. These gains appear within the first 90 days when the organizational prerequisites are in place.

The gains are also conditional. Respondents who skipped data standardization, bolted AI onto broken processes, deployed without clear ownership, or launched without defined metrics reported the same outcome: unreliable outputs, low adoption, stalled momentum, and eventual abandonment. This matches what McKinsey, BCG, Gartner, and MIT have documented at the macro level, but our data adds the operational detail that macro research lacks.

For executives evaluating AI readiness, the evidence points to a clear sequence: fix the workflow first, clean the data second, set the controls third, deploy AI fourth. Companies that reverse this order are the ones that remain in pilot purgatory.

References

We would like to thank all companies and individuals who participated in this research.

SaaS and software

- Roman Milyushkevich, CTO, HasData — hasdata.com
- David Kemmerer, Co-Founder & CEO, CoinLedger — coinledger.io
- Dimi Baitanciuc, CEO, Brizy — brizy.io
- Ilya Telegin, Head of Content, Improvado — improvado.io
- Dirk Alshuth, CMO, emma — emma.ms
- Colleen Barry, Head of Marketing, Ketch — ketch.com
- Peter Barnett, VP Product Strategy, Action1 — action1.com
- Vlad Zhovtenko, CEO, RedTrack.io — redtrack.io
- Steve Derezinski, Advisor, Riff Analytics — riffanalytics.ai
- Casey Milone, Co-Founder & CEO, Numonic — numonic.ai
- Aryaman Behera, Founder & CEO, Repello AI — repello.ai
- Ritwick Dey, Co-Founder & CTO, Panto AI — getpanto.ai
- Tudor Brad, Managing Director, BetterQA — betterqa.co
- Kris, Founder, SellerOps — sellerops.io
- Nicolas Moreno, Founder, Reddinbox — reddinbox.com
- Amy Bos, Co-Founder & COO, Mediumchat Group — mediumchat.co.uk
- Raghav Sharma, Machine Learning Engineer, Workday
- Lily Stoyanov, CEO, TFY (Transformify) — transformify.org
- Meryll Dindin, VP Product & Engineering, Parallel Learning — parallelllearning.com
- Tiago Strammiello, Founder & CEO, ClaireAI — theclaireai.com
- Yanis Mellata, CEO & Founder, NextPhone — getnextphone.com
- Sylvain Kalache, Co-Founder, Rootly — rootly.com
- Marin Cristian-Ovidiu, CEO, OnlineGames.IO — onlinegames.io

Digital marketing and SEO

- Sean Markey, Founder, LocalSEO.net — localseo.net
- Andrew Swiler, Founder, AnswerManiac — answermaniac.ai
- Karen Cleaver, COO, Underground Marketing — undergroundmarketing.com
- Jennifer Bagley, CEO, CI Web Group — ciwebgroup.com
- Adam Bocik, Founder, Evergreen Results — evergreenresults.com
- Scott Kasun, Founder, ForeFront Web — forefrontweb.com
- Cesar Beltran, Blackbelt Commerce — blackbeltcommerce.com
- Stephen Taormino, CEO, CC&A Strategic Media — ccastrategicmedia.com
- Cyrus Kennedy, Chairman, The Ad Firm — theadfirm.net
- Aaron Whittaker, VP Demand Generation, Thrive Internet Marketing Agency — thriveagency.com
- Chris Kirksey, Founder & CEO, Direction.com — direction.com
- Danyon Togia, Founder, Expert SEO — expertseo.co.nz
- Rachita Chettri, CEO, Linkible
- Gareth Hoyle, Managing Director, Marketing Signals — marketingsignals.com

Healthcare and wellness

- Scott Melamed, President & COO, ProMD Health — promdbethesda.com
- Ryan Pittillo, Franchise Owner, ProMD Health Bel Air — promdbelair.com
- Dr. Seth J. Crapp, Founder, South Florida Radiology — sflrad.com
- Sara Cemin, Head of Customer Relations, Helio Cure

IT services and cybersecurity

- Ryan Miller, Founder, Sundance Networks — sundancenetworks.com
- Sara Szot, President, Alliance InfoSystems — ainfosys.com
- Roland Parker, Founder & CEO, Impress Computers — impresscomputers.com
- Marc Umstead, President, [Plus 1 Technology](#)

Logistics, fleet, and field services

- Cole Russell, Operations Director, Hanzo Logistics — hanzologistics.com
- Hani Alkayali, President, Blue Diamond Towing — bluediamondtowingllc.com
- Eliot Vancil, CEO, Fuel Logic — fuellogic.net
- Robert Novak, Operations Manager, 8Rental — 8rental.com
- Emily Demirdonder, Director of Operations & Marketing, [Proximity Plumbing](#)

Manufacturing, construction, and trades

- Bill Dudley, Gunitite Restoration — gunititerestoration.com
- Bruce Kemp, President/CEO, Lighthouse Energy Services — lighthouseenergyco.com
- Felix Bagr, Owner, ITECH Recycling — itechrecycling.com
- Delbert Lee, President, [Wynbert Soapmasters Inc.](#)
- Jill Frattini, Service Coordinator, Ohio Heating — ohheating.com

Real estate and financial services

- Brett Sherman, Signature Realty — thesignaturerealty.com
- Jeffrey Joseph, Owner, Mitchell-Joseph Insurance — mitchelljoseph.com
- Einar Vollset, Founder, Discretion Capital — discretioncapital.com
- Tricia Watts, Founder, MaxNet Homes — maxnethomes.com

Security and surveillance

- Chris Edens, Founder, Mobile Vision Technologies — mobilevisiontechnologies.com
- Dan Wright, Founder, Duck View Systems — duckviewsystems.com

Education and workforce

- Jessica-Lee Tingley, Founder, The JobBridge — thejobbridge.com
- Don Poh, Group CEO, [Lorna Whiston Schools](#)
- Anatolii Kasianov, Co-Founder & Co-CEO, Holywater / My Passion — my-passion.com

Professional services and consulting

- Jon Siegal, Co-Founder, Venture Smarter — venturesmarter.com
- Filip Pesek, CEO, DonnaPro — donnaapro.com
- Syed Asif Ali, Founder, Point Media — point-media.link
- Sanjay, Founder, Acquisales — acquisales.com
- Ryoji Morii, Founder, Insynergy Inc. — insynergy.io
- Ganesh Kompella, Managing Director, Kompella Technologies — kompella.io
- Raj Baruah, SupaFunnel — supafunnel.com
- Lukas Kruger, Klarus — klarus.com
- Rahul Gupta, Head of AI, [Insight Global Consulting](#)

Other industries

- Nicky Zhu, AI Interaction Product Manager, Dymesty — dymesty.com
- Matthew Kinneman, Founder & CEO, Bully Max — shop.bullymax.com
- Myles Schepetin, CEO, New York Custom Labels — newyorkcustomlabels.com
- Ben Townsend, Founder & CEO, Tracker Products — trackerproducts.com

- Ken Herron, Co-Founder, VCONify — vconify.com
- Abhishek Bhatia, CEO, [ShadowGPS](https://shadowgps.com)
- John DeMarchi, CEO, Social Czars — socialczars.com
- Jay Ellenby, President, Safe Harbors Travel Group — safeharbors.com
- Athena Kavis, Founder, Quix Sites — webdesignlasvegas.com
- Alex Jones, Founder, FundForge — fundforge.net
- Joseph Riviello, CEO, Zen Agency — zen.agency
- Dawn McGrath, Marketing Director, Keller Heartt — kellerheartt.com
- Bill Joseph, Founder, Frontier Blades — frontierblades.com
- Naz Mitchell, Founder, MoveMetal — movemetalcrm.com
- Carla, General Manager, Wardnasse — wardnasse.org
- Daniel Reynolds, Managing Director, Dynamo LED Displays — dynamo-led-displays.co.uk
- David Hunt, COO, Versys Media — versysmedia.com

Supporting industry research

- McKinsey & Company, “The State of AI: Global Survey 2025” (November 2025)
- McKinsey & Company, “AI in the Workplace: A Report for 2025” (January 2025)
- MIT NANDA Institute, findings on generative AI pilot ROI failure rates (2025)
- S&P Global Market Intelligence, AI initiative abandonment data (2025)
- Deloitte, “State of AI 2026” report
- TechRepublic, “AI Adoption Trends in the Enterprise 2026” (January 2026)

How to cite this research

Merzlova, K., & Funtikov, K. (2026). *AI Readiness: How Companies Move from AI Pilots to Production*. SumatoSoft Research. <https://sumatosoft.com/research/business-ai-readiness>

Dataset: 72 validated responses from leaders across 30+ industries, collected over 21 days and screened against the inclusion criteria described in the methodology.

Statistics and charts from this study may be reused with attribution and a link to the online version of this study.



Thank you for your time!

Any questions? Drop us a line!

Headquarters

One Boston Place, Suite 2602
Boston, MA 02108, United States

Other ways to get in touch

info@sumatosoft.com
sumatosoft.com