AI Readiness 2025

Benchmarking Infrastructure, Strategy, and Scale

Survey Insights on Enterprise AI Maturity and the Document Infrastructure Powering It

apryse

Contents

Introduction:

The Criticality of the AI Input Problem

Today's Artificial Intelligence (AI) is becoming smarter, more flexible, and able to work on its own. The different types of AI (predictive, generative, and agentic) are quickly moving out of research labs and into the core operations of enterprises. These systems can now analyze information, create new content, and even plan and complete complex, multi-step tasks by themselves. They are transforming how businesses operate and innovate.

AI may be the future, but it has a fatal flaw in that it's only as good as the data that powers it. Before an AI model can perform a complex task, the raw, unstructured data, hiding in your PDFs, images, emails, and forms, must be converted into clean, structured, and context-aware data. This report reveals the disconnect between ambitious AI deployment and the maturity of the underlying document infrastructure. It explores the readiness gap that is currently limiting AI scalability, focusing on the hidden infrastructure layer (intelligent pre-processing) that serves as the ground level for intelligent automation.

Based on a survey of technical and product leadership across North America (NA), Europe, and Oceania, this report analyzes current adoption trends, core infrastructural deficiencies, and strategic investment priorities. It aims to guide leaders toward building a secure, robust document-to-data foundation capable of supporting the next generation of AI-driven enterprise.

Firmographics:

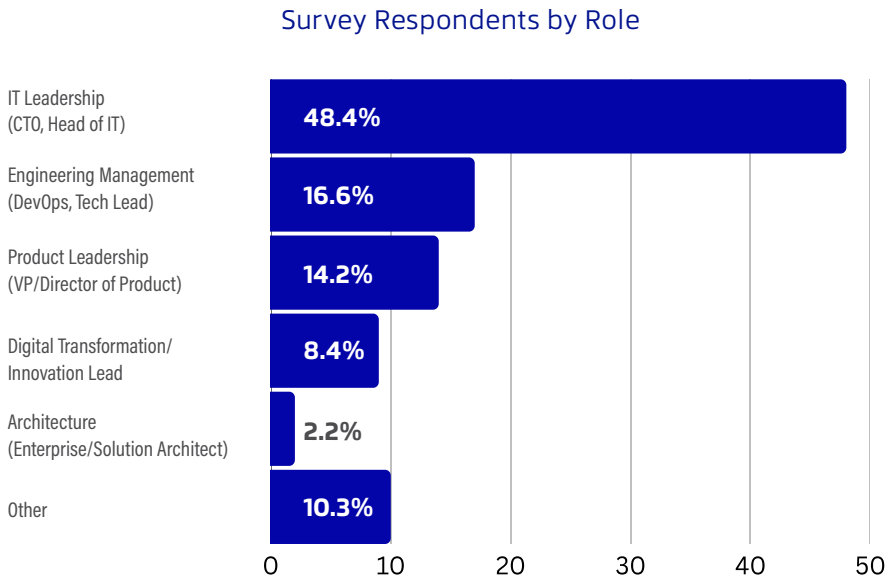
The Builders and Their Environment

Understanding the audience is key to interpreting the findings. The survey was intentionally focused on a technically sophisticated group of leaders who are not just observing AI but actively engineering its deployment and integration.

Defining the Audience

The survey audience highlights that the challenge of AI infrastructure is fundamentally an engineering and product problem, addressed at the highest levels of technical management.

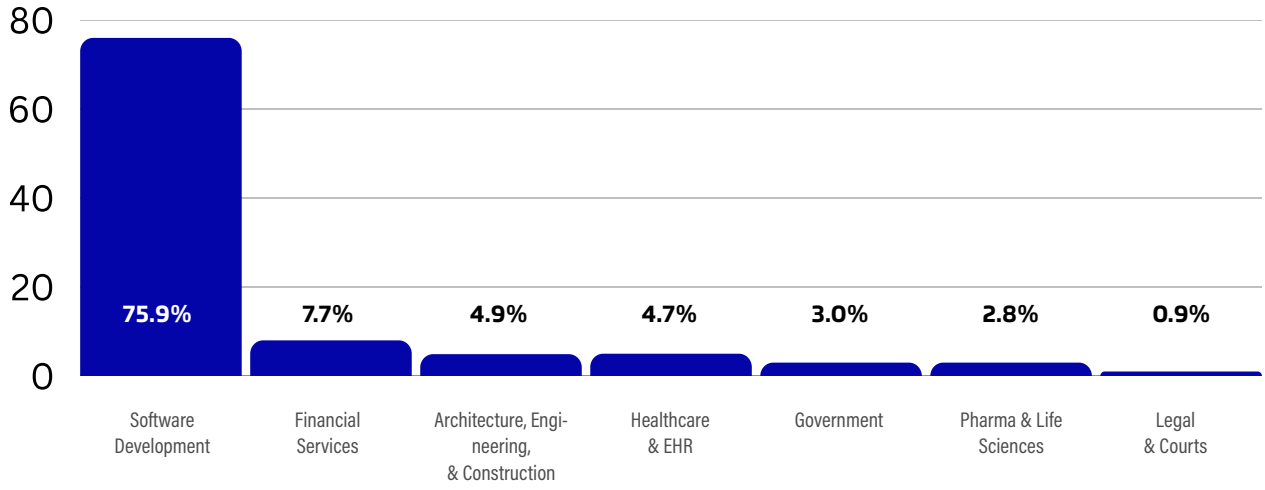
Role and Function: The audience consists mostly of IT Leadership (**48.4%**), demonstrating that AI strategy is now a C-suite agenda item, especially in concern to infrastructure and security. This is strongly supported by the presence of Engineering Management (**16.6%**) and Product Leadership (**14.2%**), who are responsible for building, integrating, and maintaining the scalable systems required for AI. Their insights are invaluable as they deal with the operational friction points such as integration complexities, development cycles, and data quality issues that are often encountered in projects moving from pilot to production.



Source: Apryse, September 2025, AI Readiness Survey

Industry Concentration: The Software Development (**75.9%**) sector forms the majority, indicating a strong developer-centric focus. This suggests that the pain points identified are deeply technical and show that there is a need for robust SDKs, APIs, and embeddable components that allow for custom application development. Other critical, document-heavy sectors such as Financial Services (**7.7%**) and Healthcare & EHR (**4.7%**) show that the findings are relevant to highly regulated, data-intensive environments.

Survey Respondents by Industry

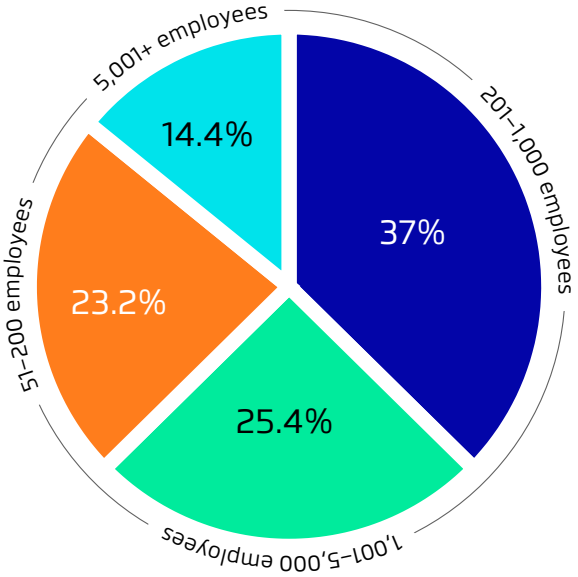


Source: Apryse, September 2025, AI Readiness Survey

Scale of Operations:
Most participants represent mid-sized to large organizations, with over two-thirds (**62.4%**) coming from companies with 201 to over 5,000 employees.

This confirms that AI readiness is centered mainly on organizations that have the resources and the complexity of data and legacy systems that require robust, enterprise-grade solutions. The size and complexity of the organization, from the sheer volume of data to the number of stakeholders, make the document-to-data challenge particularly difficult.

Survey Respondents by Company Size



Source: Apryse, September 2025, AI Readiness Survey

The Global View:

Regional Priorities and Pacing

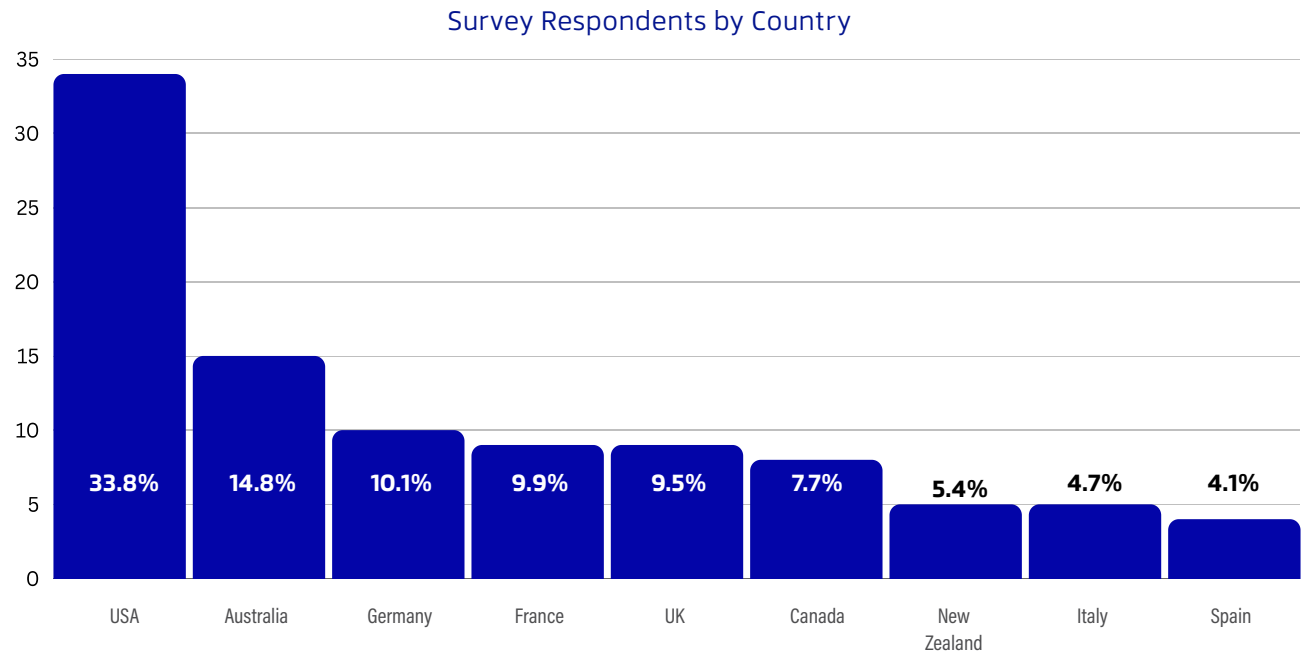
The regional focus across North America, Europe, and Oceania provides a critical perspective into how geography and regulatory environments shape AI strategy.

North America (NA): NA (USA and Canada) respondents (**41.5%**) show the highest rate of AI in production, indicating a more aggressive approach to commercializing AI applications. But this rapid implementation does come with a higher reported need for sophisticated data governance to keep pace.

Europe: European countries (UK, Germany, France, Spain, and Italy) are heavily engaged in the pilot and testing phase, suggesting a methodical, potentially regulatory-influenced approach. Their findings highlight the barriers that must be overcome before their significant pipeline of projects can transition to scale.

Oceania: Oceania (Australia and New Zealand) consistently registers the highest levels of concern regarding Data Quality and Privacy/Security. This reflects both stringent regional data residency requirements and a highly analytical approach to AI deployment.

The following sections include the survey data combined with these regional insights to provide a full picture of the state of global AI readiness.



Source: Apryse, September 2025, AI Readiness Survey

The AI Boom Meets a Bottleneck:

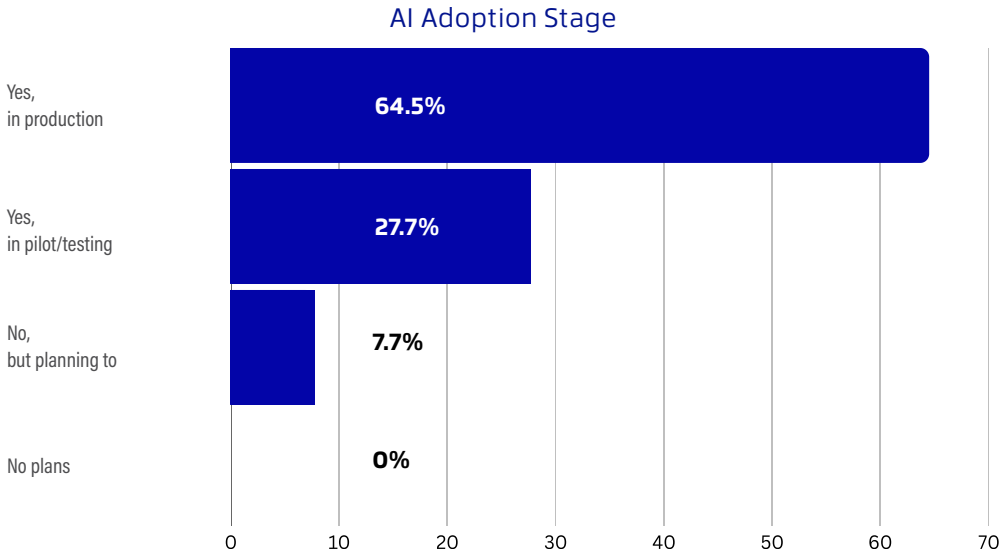
Deployment, Goals, and the Hard Walls of Scale

AI is no longer exploratory; it is deeply woven into enterprise operations. This section examines the current state of deployment and analyzes the primary barriers organizations face with scaling their AI ambitions to their full potential.

The Mainstream Reality of AI Deployment

The survey's findings put an end to the debate about AI adoption: the technology is firmly established in the operational toolkit.

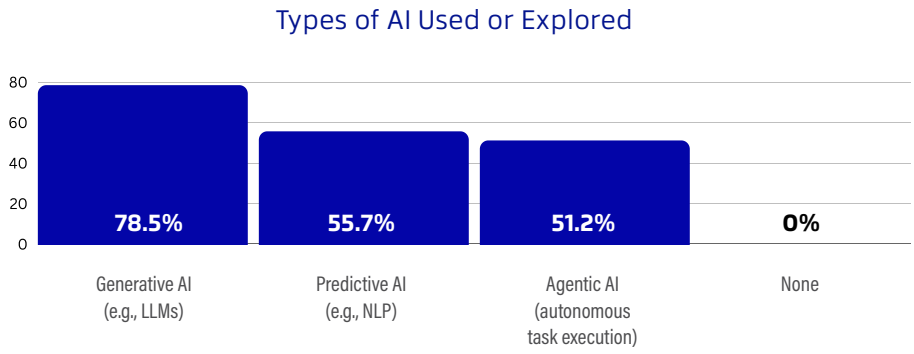
From "Pilot" to "Product": A substantial **64.5%** of organizations report having AI in production. This means AI systems are currently handling live customer data, executing financial tasks, and driving mission-critical workflows. An additional **27.7%** are in the pilot/testing stage, which represents a robust pipeline for future scaling. With only **7.7%** still in the planning phase, we can see that the market is mature and actively implementing AI.



Source: Apryse, September 2025, AI Readiness Survey

Regional Pacing: This maturity level is not uniform. North America leads significantly, with 77.7% in production, 27.7 percentage points higher than Europe (50.0%). This suggests NA organizations are focused on speed to market and are quicker to push successful pilots live. Europe leads in the pilot/testing stage (38.2%), indicating a more measured approach. This potentially is driven by the need to navigate complex regulations before fully deploying at an enterprise level.

Layered Intelligence: Deployment is also multimodal. While Generative AI (**78.5%**) provides the headline capabilities for content generation and summarization, it is being paired with foundational technologies: Predictive AI (**55.7%**) and Agentic AI (**51.2%**). With over half of the respondents adopting Agentic AI, there is a clear and strategic move toward full automation where systems must be capable of autonomous task execution. This then places a greater demand on the reliability and structure of input data.



Source: Apryse, September 2025, AI Readiness Survey

Goals and Strategic Drivers

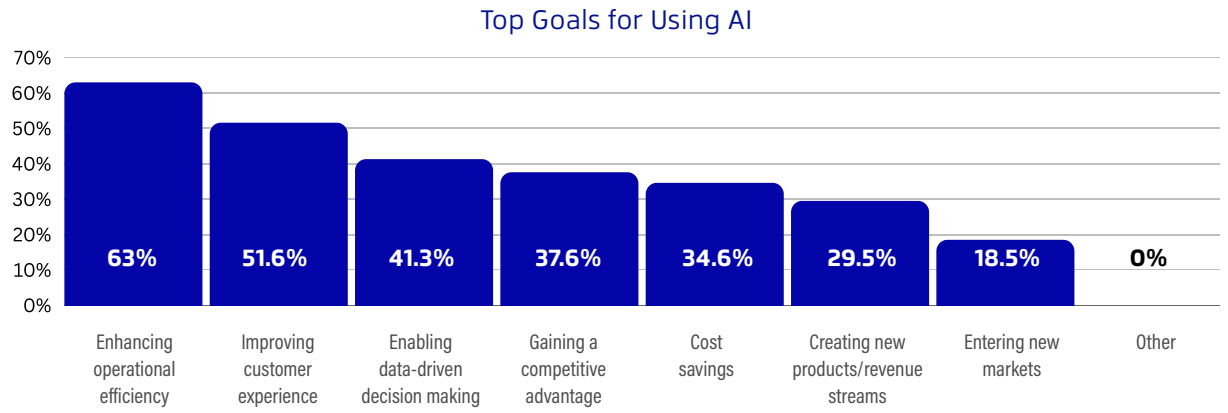
AI is overwhelmingly viewed as a performance enhancer, rather than purely a cost-reduction tool. The top three strategic goals are clear:

Enhancing Operational Efficiency (63.0%): The clear primary driver, focused on optimizing core business processes, reducing processing times, and streamlining internal workflows.

Improving Customer Experience (51.6%): AI, particularly generative and conversational models, can transform customer interactions, from automated support to personalized service.

Enabling Data-Driven Decision Making (41.3%): Focusing on extracting rapid, actionable insights from complex data sets, which can be severely hindered by poor data quality.

Regional Goal Interpretation: While efficiency is a shared global goal, North America leads in prioritizing Customer Experience (**55.4%**), suggesting a higher investment in customer-facing AI applications. Oceania shows a strong lead in prioritizing Operational Efficiency (**66.0%**) and Data-Driven Decision Making (**45.7%**), reinforcing their earlier lead in predictive and generative AI adoption which indicates a more internal, optimization-focused strategy.



Source: Apryse, September 2025, AI Readiness Survey

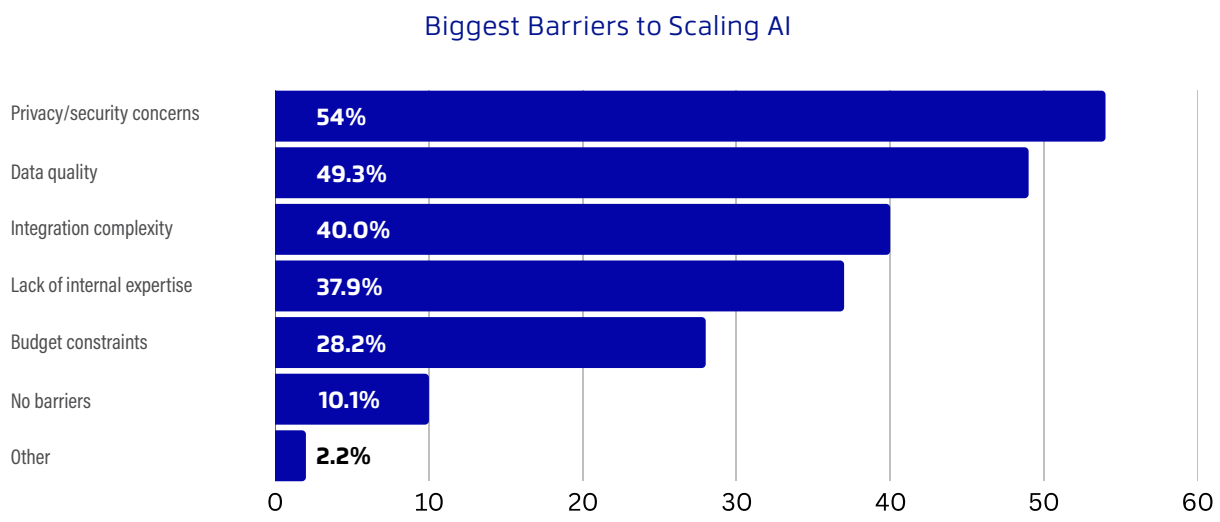
The Hard Walls of Scalability

Despite the clear goals and production readiness, foundational hurdles can still obstruct the path to true enterprise-wide scale. When organizations were asked about their biggest barriers to scaling AI, the top two were clear indicators of a trust and data challenge.

Privacy/Security Concerns (54.0%): This is the top barrier globally, reflecting the high sensitivity of enterprise data. This is particularly evident in highly regulated industries where solutions require data to be transmitted to external, non-secure environments, driving a fundamental demand for control.

Data Quality (49.3%): Nearly half of organizations recognize that their data is simply not clean, consistent, or structured enough to sustain high-accuracy AI at scale. This is an issue with the preparation of the source data.

These two barriers significantly outweigh more traditional concerns like Integration Complexity (**40.0%**) and Lack of Internal Expertise (**37.9%**). It reveals the critical strategic insight that an organization can hire the best staff and tackle integration, but if the underlying data is flawed and security is compromised, the AI implementation will fail.



Source: Apryse, September 2025, AI Readiness Survey

Regional Barrier Insights: The urgency of these issues is most apparent in Oceania, which leads in concern for both Data Quality (**64.9%**) and Privacy/Security (**58.5%**). The sharp focus on clean data and security governance highlights the complex regulatory landscape and the technical maturity of Oceania organizations, who are more sensitive to the need for pristine inputs. The high rates of both generative/predictive AI adoption and data quality concerns in Oceania suggest that these organizations are both pushing the boundaries of AI deployment and recognizing the vulnerability of their data at the same time.

The Infrastructure No One Talks About:

The Document-to-Data Pipeline

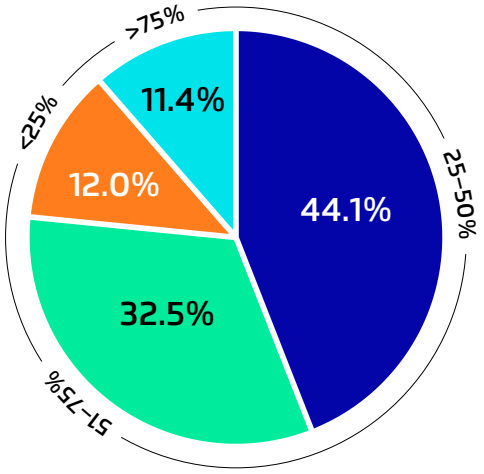
Data quality leads as a barrier due to the unstructured nature of enterprise information. Before AI can act, documents must be transformed into interpretable data.

The Unstructured Data Mountain

Documents are the primary repository of enterprise knowledge, validating the mission of document AI and automation:

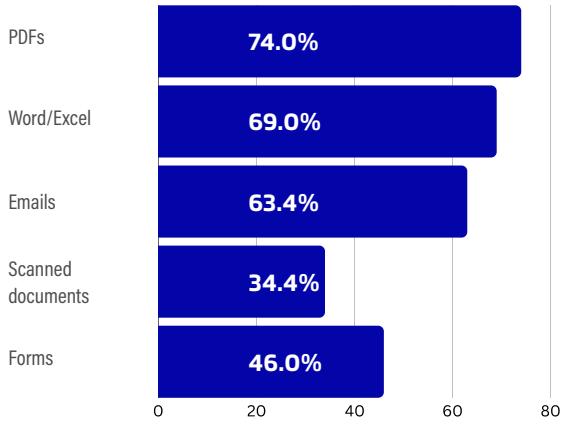
Core Asset Reliance: A massive **76.6%** of organizations report that between **25.0%** and **75.0%** of their total data lies in documents such as PDFs, scans, and forms. This confirms that critical information, from contracts and invoices to patient records and engineering drawings, is contained in formats designed for human eyes, not machine consumption.

Percentage of Data Stored in Documents



Source: Apryse, September 2025, AI Readiness Survey

Common Document Types



Source: Apryse, September 2025, AI Readiness Survey

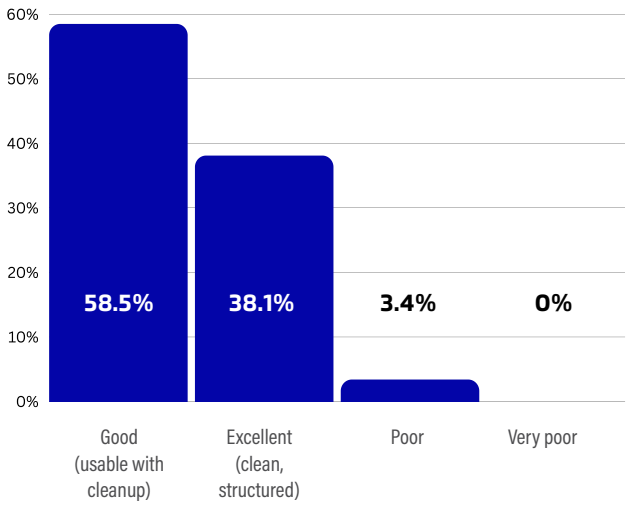
Document Type Domination: The document format is key to understanding the technical challenge: PDFs (**74.0%**) are the most common, followed closely by Word/Excel files (**69.0%**). These are known as semi-structured formats which contain data organized in tables, headers, and forms but without machine-readable metadata or consistent structure. This requires intelligent tools that can interpret layout and context, which is beyond simple text scraping. The **34.4%** who still rely on scanned documents also highlight the ongoing need for robust Optical Character Recognition (OCR).

The Quality Deficit and the Need for Pre-Processing

Despite the heavy reliance on document-based knowledge, quality remains a major hurdle:

The "Good Enough" Trap:
Only **38.1%** of organizations rate their document data as "Excellent" (clean, structured, and labeled). The majority (58.5%) describe it as merely "Good," requiring cleanup before being fit for AI. Labelling data "Good Enough" results in a systemic risk. While marginal data quality may be tolerated in human-centric workflows, it introduces compounding errors and unpredictable outputs when fed into large, autonomous AI models, which directly contributes to the Data Quality barrier.

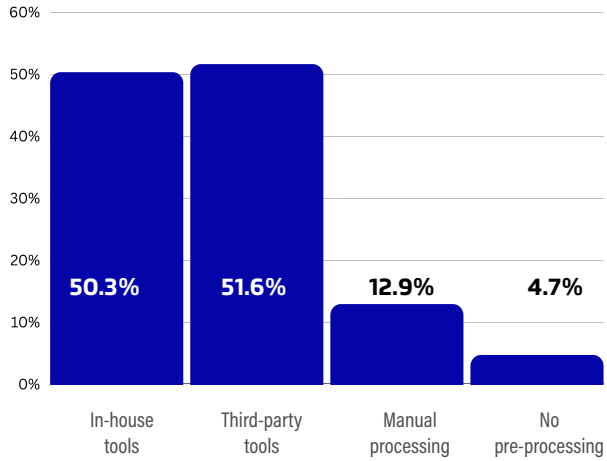
Quality of Document Data for AI



Source: Apryse, September 2025, AI Readiness Survey

Tooling Fragmentation:
Organizations are split on their approach to infrastructure. A near-even distribution exists between in-house (**50.3%**) and third-party tools (**51.6%**) for document pre-processing. This suggests that no single solution dominates, and many companies are forced to stitch together a custom solution. With **17.6%** still relying on manual processing or no pre-processing at all, these organizations are exposed to high labor costs, slow turnaround times, and severe risks of data omission or error.

Use of Pre-processing Tools



Source: Apryse, September 2025, AI Readiness Survey

Regional View of Quality and Tooling: The North American lead in organizations reporting "Excellent" data quality (**43.0%**) matches with their higher AI-in-production rates, meaning a better data foundation makes for faster deployment. Oceania's strong lead in the use of third-party tools (**63.8%**) suggests a reliance on the expertise of vendors to quickly deploy ready-made components that address their high data quality concerns. This emphasizes speed and efficiency over custom development.

From Chaos to Context:

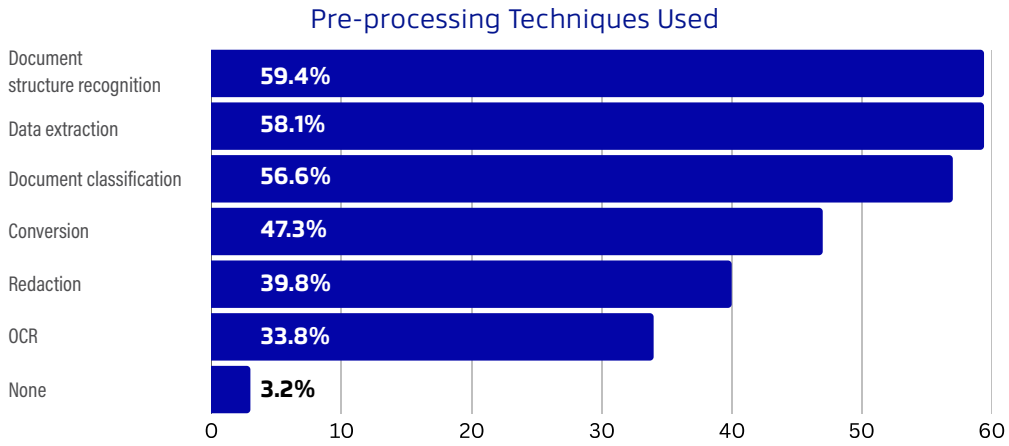
Intelligent Pre-Processing and the Readiness Gap

The true value of document automation lies in its ability to extract context from chaos. This clearly shows the need for intelligent pre-processing techniques that understand a document's layout, identifying tables, forms, and key entities to feed structured, contextualized data to AI models.

The New Priorities in Capability

Organizations are making strategic investments in capabilities that go far beyond simple digitization.

Focus on Structure: Document Structure Recognition (**59.4%**) and Data Extraction (**58.1%**) are the most widely used techniques. This is a clear sign that the priority is no longer just making text selectable but understanding where that text resides, such as in a table, a header, or a key-value field to add semantic context for the AI.



Source: Apryse, September 2025, AI Readiness Survey

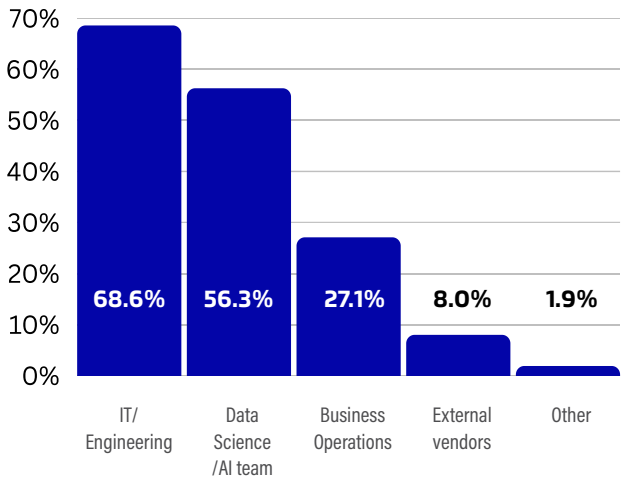
The OCR Gap: The fact that OCR is used by only **33.8%** of respondents suggests that, for many enterprises, the primary challenge has moved past scanning paper and into extracting value from digitally native, semi-structured formats like complex PDFs and legacy digital forms.

This focus on advanced structuring is directly related to the real-world impact of poor data: **62.8%** of companies report that document quality issues "Occasionally" or "Frequently" have a negative impact on their AI or automation outcomes. This high failure rate reinforces the idea that the document-to-data layer is currently the most prominent point of failure in the AI pipeline.

The Confidence/Reality Disconnect

While technical teams are actively involved with IT/Engineering (**68.6%**) and Data Science (**56.3%**) owning the process, there is still a surprising disconnect between stated confidence and documented failure rates.

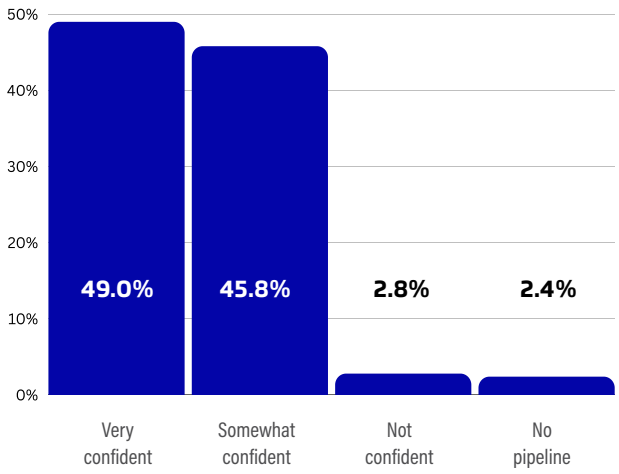
Who Prepares Documents for AI



Source: Apryse, September 2025, AI Readiness Survey

Fragile Confidence:
Almost **95.0%** of organizations express at least "Somewhat Confident" in their document processing pipeline, with **49.0%** claiming to be "Very confident." But this high confidence comes mainly from the **62.8%** who report frequent or occasional pipeline failures. The market is highly optimistic but lacks the technical certainty needed for the deployment of mission-critical, high-volume AI.

Confidence in Document Processing Pipeline



Source: Apryse, September 2025, AI Readiness Survey

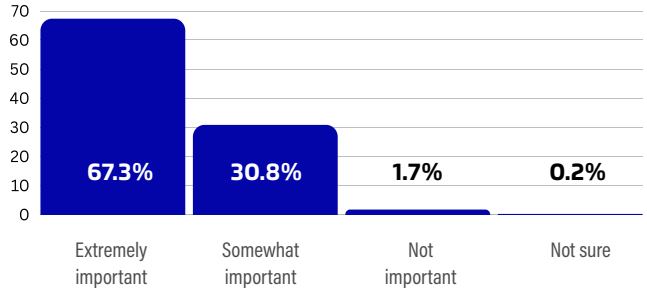
Regional View of Confidence: This gap is most evident when comparing deployment leaders. North America reports the highest level of "Very Confident" respondents (**59.1%**), a full 21.5 percentage points higher than Europe (**37.6%**). This higher NA confidence aligns with their better data quality and faster production deployments, which suggests that investing in a robust foundation directly pays off in perceived technical stability.

Security, Privacy, and Control

The security and privacy concerns identified earlier (**54.0%** of all barriers) translate directly into strategic requirements for document processing tools, creating a non-negotiable demand for data security.

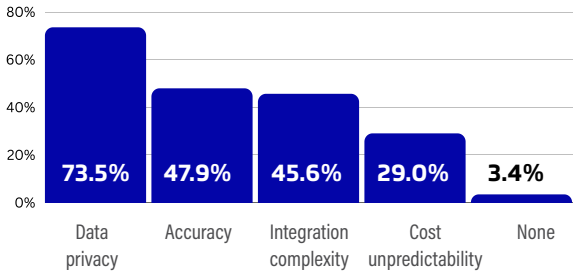
Control Over Cost: 67.3% of organizations report that it is "Extremely important" to keep document data processing within their own infrastructure. This demand for on-premises, private cloud, or embeddable processing is directly related to the primary concern with third-party tools: Data Privacy (**73.5%**). Other major concerns include Accuracy (**47.9%**) and Integration Complexity (**45.6%**). The market is willing to tolerate complexity if it means retaining control over sensitive data.

Importance of Keeping Document Data Processing In-House



Source: Apryse, September 2025, AI Readiness Survey

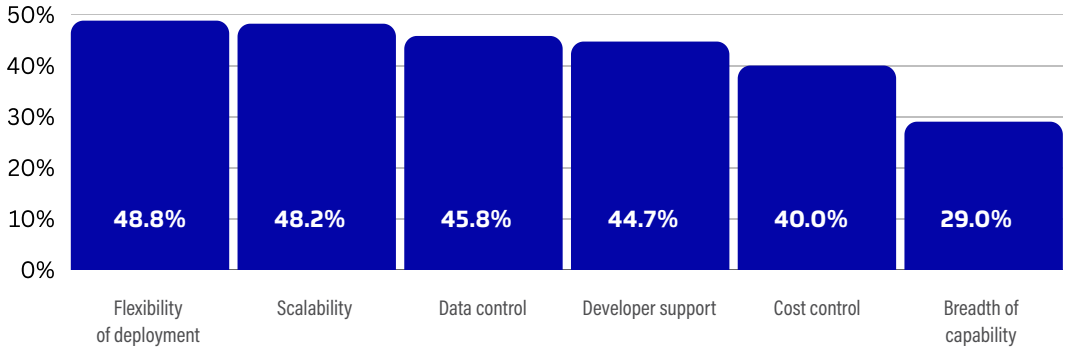
Top Concerns with Third-Party Tools



Source: Apryse, September 2025, AI Readiness Survey

Vendor Priority Alignment: When selecting an SDK vendor, the top priorities reflect the need for control and engineering enablement. Flexibility of Deployment (**48.8%**), Scalability (**48.2%**), and Data Control (**45.8%**) are extremely important. This confirms that the audience requires embeddable, developer-friendly solutions that allow them to integrate best-in-class capability while maintaining full data governance and avoiding the unpredictable cost of usage-based cloud APIs.

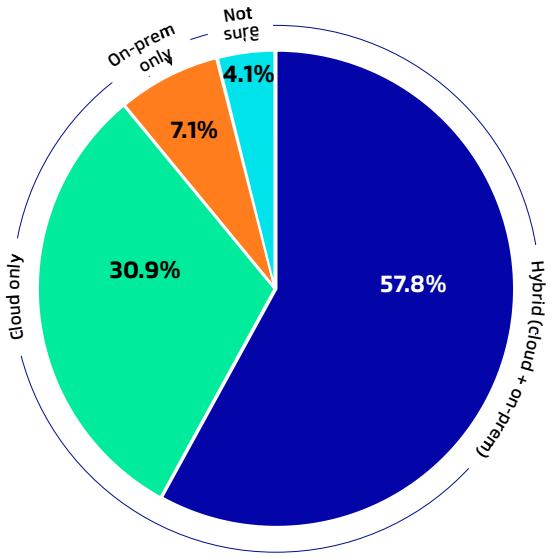
Priorities When Selecting SDK Vendor



Source: Apryse, September 2025, AI Readiness Survey

Regional View of Security:
North America is the global leader in demanding in-house processing (**78.8%**), a striking 24.3 percentage points higher than Europe. This indicates a strong aversion to moving data outside internal perimeters. On the other hand, Oceania shows the highest usage of the Hybrid cloud model (**67.0%**) and the greatest concern for third-party data privacy (**80.9%**), showing a sophisticated, risk-aware strategy that seeks deployment flexibility with strict enforcement of data residency rules.

Use of Cloud-Based APIs for Document Processing



Source: Apryse, September 2025, AI Readiness Survey

Building the Infrastructure for Intelligent Workflows:

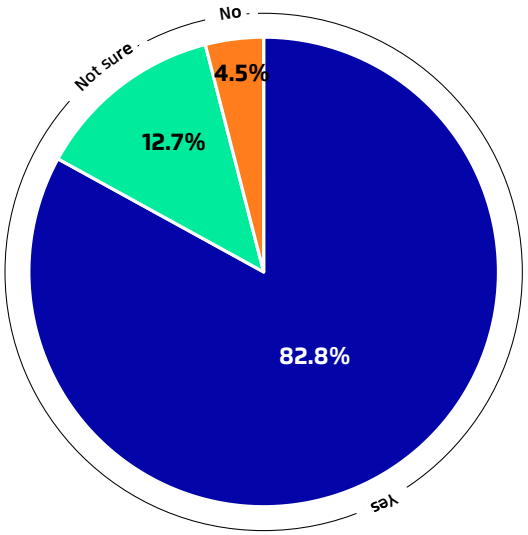
Closing the Readiness Gap

The final phase of the report shifts from identifying the problem to defining the solution. With investment intent high and clear priorities established, organizations must now make strategic, focused investments in the right document infrastructure to close the AI Readiness Gap.

The Momentum of Investment

The market is set for a massive investment in the document layer with **82.8%** of organizations planning to invest in document automation or AI-ready data infrastructure in the next 12 months. This high intent is consistent across all regions (NA: **83.9%**, Europe: **82.0%**, Oceania: **81.9%**), confirming that the challenges identified are recognized as urgent, and that immediate investment is warranted.

Investment Plans in Next 12 Months



Source: Apryse, September 2025, AI Readiness Survey

The Strategic Capabilities for Structured Data

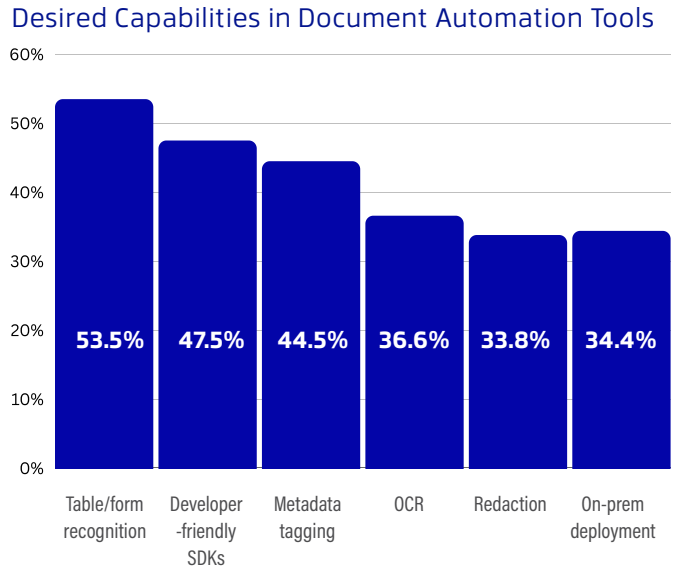
The capabilities required in a document automation tool appear to be centered primarily on unlocking structured data for feeding into AI.

The top three requested capabilities are:

Table/Form Recognition (53.5%): The leading priority, underlining the shift from simple character recognition to sophisticated spatial and structural understanding. AI needs to know not just the numbers on an invoice, but that those numbers belong to the "Total Due" column of a table.

Developer-Friendly SDKs (47.5%): This is a functional requirement highlighting the fact that organizations prefer to integrate powerful components into their existing custom applications, rather than adopting rigid, vendor-locked systems.

Metadata Tagging (44.5%): Essential for contextualizing data, tagging allows AI to classify, index, and apply appropriate governance (like redaction) to specific extracted data points, making the data searchable and compliant.



Source: Apryse, September 2025, AI Readiness Survey

Regional Capability Demand: The demand for structured data is highest in Oceania, which leads in prioritizing Table/Form Recognition (**59.6%**) and SDK Deployment Flexibility (**53.2%**). This reinforces their need for high-accuracy tools that can handle diverse, complex documents across varied regulatory requirements.

A Holistic Approach to Readiness

While infrastructure is a technical challenge, achieving true AI readiness requires a holistic approach that includes people and strategy. Respondents pointed out four interconnected areas for improvement:

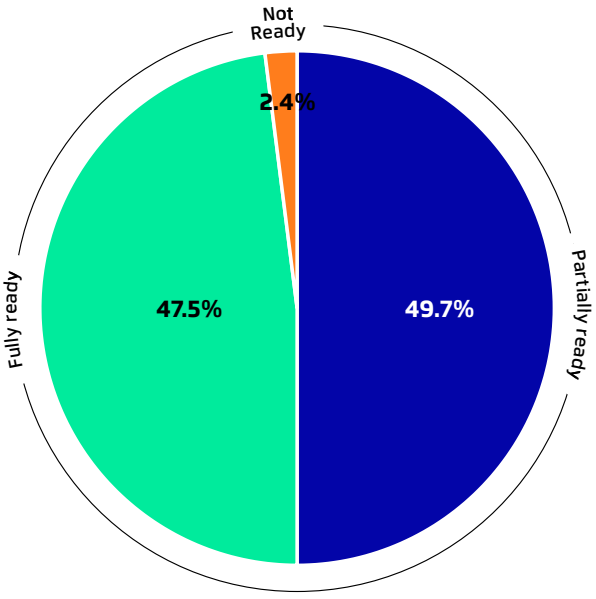
More Skilled Staff (32.5%):
Addressing the need for engineers and data scientists capable of building and maintaining these complex AI pipelines.

Clearer Strategy (22.8%):
Ensuring AI initiatives are aligned with measurable business outcomes and regulatory requirements.

Better Document Infrastructure (21.9%):
The technical tools that convert unstructured content into AI-ready data.

Better Tools (21.5%):
The specific, high-quality components and SDKs necessary for execution.

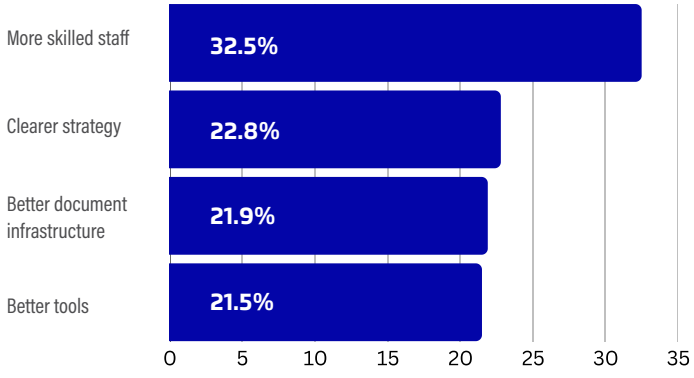
Organizations' AI Readiness



Source: Apryse, September 2025, AI Readiness Survey

This confirms that technical and functional solutions (Infrastructure and Tools) must be combined with the right organizational context (Skills and Strategy) to be truly effective.

What Would Improve AI Readiness



Source: Apryse, September 2025, AI Readiness Survey

Conclusion:

Apryse and the Foundation for AI Success

The survey establishes that AI ambition is far outpacing the maturity of enterprise document infrastructure. The success of different AI models (predictive, generative, agentic) relies on resolving the common barriers of Data Quality and Data Privacy. To scale successfully, organizations must recognize the document-to-data layer as the non-negotiable, foundational base for AI.

Apryse's strategic role is to solve this foundation problem. We are the infrastructure provider that ensures documents are clean, structured, and secure, transforming unstructured content into AI-ready data.

Apryse's solutions directly address the five most pressing issues identified in this report:

Data Privacy and Sovereignty: By providing embeddable, developer-friendly SDKs for processing documents, Apryse enables the entire workflow, from parsing to extraction to redaction, to run entirely within the customer's environment (on-premises, private cloud, or hybrid). This meets the high demand for data control (**67.3%** "Extremely Important") and removes the top barrier of Privacy/Security Concerns (**54.0%**).

Data Quality and Structure: Apryse's core technology delivers advanced intelligent pre-processing capabilities, including best-in-class Table/Form Recognition (**53.5%** most desired capability) and context-aware extraction, converting the "Good" data (**58.5%** of the market) into "Excellent," structured data, overcoming the **49.3%** Data Quality barrier.

Integration Complexity: By offering clean, developer-friendly SDKs (**47.5%** most desired), we simplify the integration process for IT/Engineering teams (the primary owners), reducing the high barrier of Integration Complexity (**40.0%**) and speeding up time-to-production.

Operational Failures: Our robust, mature platform provides high accuracy in document parsing and extraction, directly addressing the **62.8%** of companies experiencing frequent or occasional quality issues. This also builds the technical confidence required for mission-critical deployment.

Flexibility and Scale: The architecture supports the desired Flexibility of Deployment and Scalability (both high vendor priorities). This provides a cost-predictable model that scales with organizational growth, unlike unpredictable usage-based cloud models.

The time for simple AI adoption is over. The time for foundational investment is now. The future of enterprise AI, whether predictive, generative, or agentic, depends on the quality of the infrastructure built today, and Apryse is positioned to provide that essential, secure, and intelligent foundation.

Interested in learning how Apryse can help provide the foundation for your AI implementation?

Contact Us Today

apryse

Connect with Apryse

Stay updated with all things Apryse by following us on: