BARC

Harnessing Unstructured Data for AI Innovation

Problems, Practices, and Principles for Success

BARC Research Study

Research sponsored by:

DataHub ohalo | Data X-Ray

Authors

Kevin Petrie | VP of Research, BARC US

Kevin Petrie is the VP of Research at BARC, where he leads the data management practice and writes about topics such as AI, data integration and data governance. For 30 years Kevin has deciphered what technology means to practitioners, as an industry analyst, instructor, marketer, services leader, and tech journalist.

Kevin built a data analytics services team for EMC Pivotal in the Americas and EMEA, and ran field training at the data integration software provider Attunity (now part of Qlik). A frequent public speaker and co-author of two books about data management, Kevin most loves teaching data and AI leaders about evolving strategies, tools and techniques to capitalize on the value of data.

Contact

Mail: kpetrie@barc.com

www.barc.com

Social Media:
LinkedIn

Merv Adrian | Senior Analyst Data & AI, BARC Fellow

With over 40 years in IT, Merv Adrian has experience as a programmer, marketing and strategy director, IT market analyst, and research executive. As a BARC Fellow, his focus includes mapping the DBMS, Big Data, and NoSQL landscape. Previously, he managed all technology research at Forrester and served as a vendor lead at Gartner, roles that provided him with deep insights into software and cloud-based architectures, which he continues to leverage as a BARC Fellow.

Contact

Mail: madrian@barc.com

www.barc.com

Social Media:
LinkedIn

Table of Contents

Foreword

This research, *Harnessing Unstructured Data for AI Innovation*, surveys global respondents and builds upon three earlier BARC surveys. In *Optimizing Your Architecture for AI Innovation* (335 global respondents), published in March 2024, the authors define characteristics of artificial intelligence (AI) "Leaders" and identify that only 20.5% of respondents met all seven leadership criteria. In *Lessons From the Leading Edge* (December 2025, 421 respondents), the authors analyze what these AI Leaders have accomplished—and in some cases, what they have not. In March 2025, BARC surveyed 370 firms for *Preparing and Delivering Data for AI*, narrowing the lens to focus on data-related issues.

Here, we focus on unstructured data while taking the pulse of overall AI adoption. The pulse remains strong. AI adoption has moved into high gear across industries, bringing the issues of production, performance, and governance into sharper focus. We find that accelerating innovation within platforms, models, and tools exposes longstanding issues with unstructured data that represent potential roadblocks if not addressed now. The stakes are high, the lessons are critical, and as in earlier research, the Leaders remain worth emulating.

The findings draw on an online survey of 225 data, analytics and AI leaders conducted by BARC in February and March 2026, covering a broad range of industries, company sizes and regions.

Kevin Petrie and Merv Adrian
June 2026

Executive Summary

Executive Summary

As organizations deploy AI across business functions, they show renewed interest in the value of unstructured data. Our research explores the critical attributes of mature adopters, projects, and initiatives.

1 Production AI exposes familiar challenges

Organizations have moved beyond experimentation into production. Their survey responses reveal concerns in areas such as optimization, data quality, and security. We see increasing involvement by data engineers and machine learning (ML) engineers for governance tasks and a shift in data scientists' focus from theory to "getting it done." In a similar spirit, classification and validation now play a leading role in data preparation. Data quality has risen to the top of overall AI obstacles over the last two years as organizations recognize the risks of faulty model inputs.

As respondents move beyond foundational engineering, business units such as marketing, operations and security/infosec, sales, regulatory & compliance, and research & development (R&D) all play a larger role than in past surveys. As they consider Responsible AI, data privacy remains a top priority along with security, transparency/auditability, and accuracy, all near the top. These issues come into sharper focus when moving into business-facing production.

2 Uneven governance and misplaced confidence

AI adopters recognize the imperative for accurate inputs, and they prioritize Responsible AI objectives such as data privacy, security, and human-AI collaboration. They also show surprising progress in governing unstructured data: across categories, a higher portion of respondents implement, and even optimize, data governance controls compared with one year ago. Leaders and North American respondents show higher maturity than others.

However, much work remains. Data quality, bias, and lineage still require urgent attention at many organizations, as do the risks of bad decisions or damaging behavior by misguided AI agents. More troubling still, some AI adopters still "don't know what they don't know." Four out of five respondents express confidence that they can extract value from unstructured data without compromising governance controls, even though one third report deficiencies in lineage and bias. About one third of respondents cannot consistently enforce data access and usage policies. Automated decisions and actions by agents only compound the potential damage of poorly governed unstructured data.

Executive Summary

3 Unstructured data remains obscure, and clarity depends on people… for now

For many organizations, unstructured data—critical in adding the context that drives value for models—is the next frontier. Less than half of their unstructured data is discoverable and usable for analytics/AI, according to 70% of respondents. Digging deeper, when asked how to define and measure the AI readiness of unstructured data, respondents point to accuracy, business relevance, consistency, completeness, and timeliness. However, these measures require probabilistic logic and judgment rather than deterministic quality rules, marking a sharp contrast with the traditional world of structured data. As a result, the case for humans in the mix remains strong, and human validation ranks high in responses as well. Until a semantic layer can meaningfully supplement this process, their role seems secure… for now, at least.

4 The need for a cross-platform semantic layer

Unstructured data, like all data, dwells in far-flung, diverse environments. Most Amazon Web Services (AWS) users also have unstructured data on Azure and more than half of databases containing unstructured data reside on-premises or in hybrid environments. The distributed nature of these environments will persist thanks to migration complexity, data gravity, and rising sovereignty concerns. Given this reality, organizations will need an independent semantic layer that can access, refine, and consume all data types across platforms. This requires federated querying, aggregated data views, and integrated metadata. While hyperscalers and lakehouse platforms aim to address this need, we should expect AI adopters to favor platform-independent tools that support open ecosystems and eliminate lock-in risk.

The semantic layer gives data teams the map they need to begin their AI voyage. It's critical for unstructured data, because these documents, emails, images, and so on contain the necessary context for model training and inference.

Survey Results

Survey Results

Readiness

AI Leaders, Still a Minority, Can Teach Others How to Harness Unstructured Data

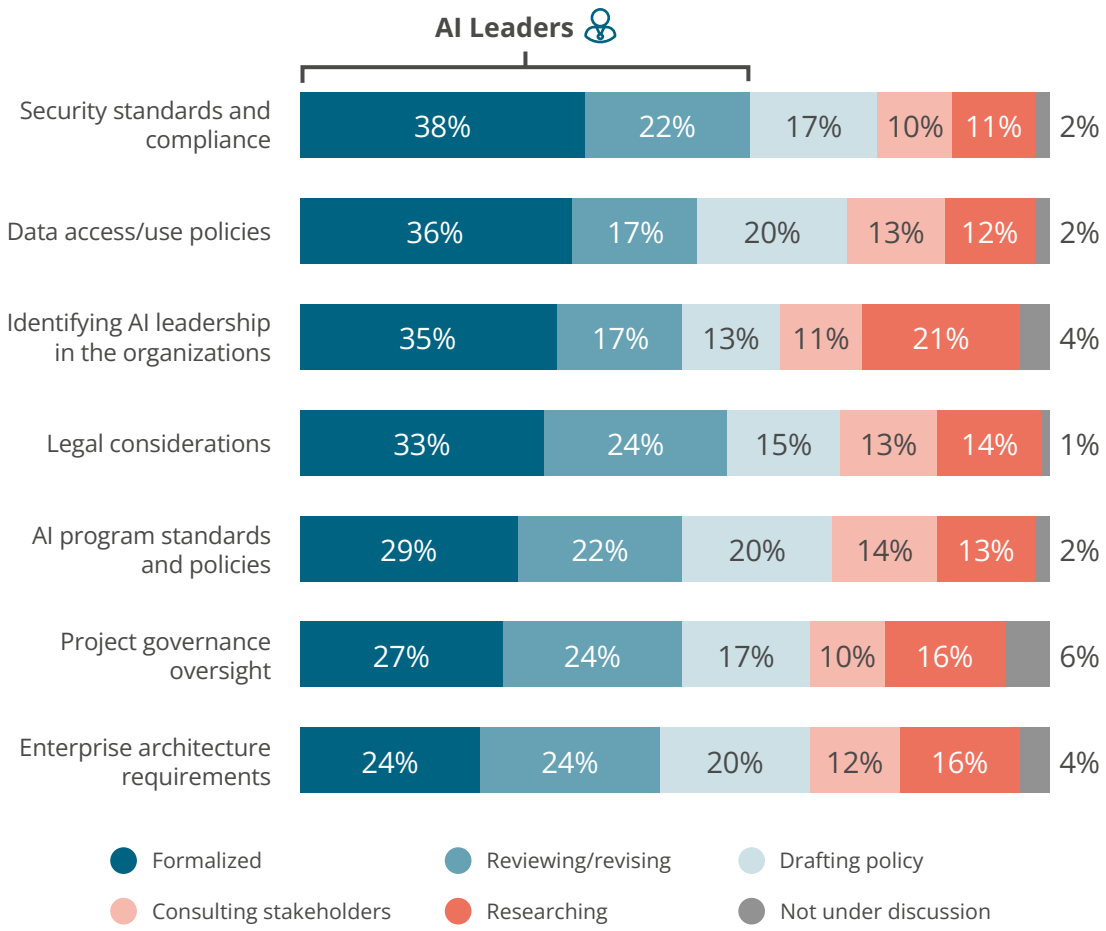


Figure 1: At what stage is your organization regarding these critical AI planning/process initiatives? (n=216)

Viewpoint

BARC identifies a cadre of AI "Leaders," whose mature, cross-functional programs drive innovation while controlling risk. They have formalized or are reviewing/revising all seven of our program elements: executive leadership, project governance oversight, enterprise architecture requirements, legal considerations, AI program standards and policies, data access/use policies, and security standards and compliance. Leaders tackle more sophisticated projects, reach production faster, and achieve better business results.

These Leaders remain a small segment of the market. Since March 2024, our research consistently finds that Leaders comprise just 20-21% of the overall market. This survey shows 23%, a slight uptick, with North American respondents once again outpacing Europeans 26% to 20%. Maturity levels for each dimension have not changed significantly since our *Lessons From the Leading Edge* report in December 2025.

Security standards and compliance have the highest maturity level, with 60% of respondents in the formalized or reviewing/revising stages. This holds true in all four surveys. Legal considerations, at 57%, have risen with each survey. Leaders come in above 50% in all categories except enterprise architecture requirements (48%).

This report will highlight where Leaders can teach the rest of the market how to harness unstructured data for AI. Other adopters must pay close attention and move fast to catch up, because unstructured data provides the deep context that AI agents need to take smart, safe actions.

Unstructured Data Remains Uncharted Territory

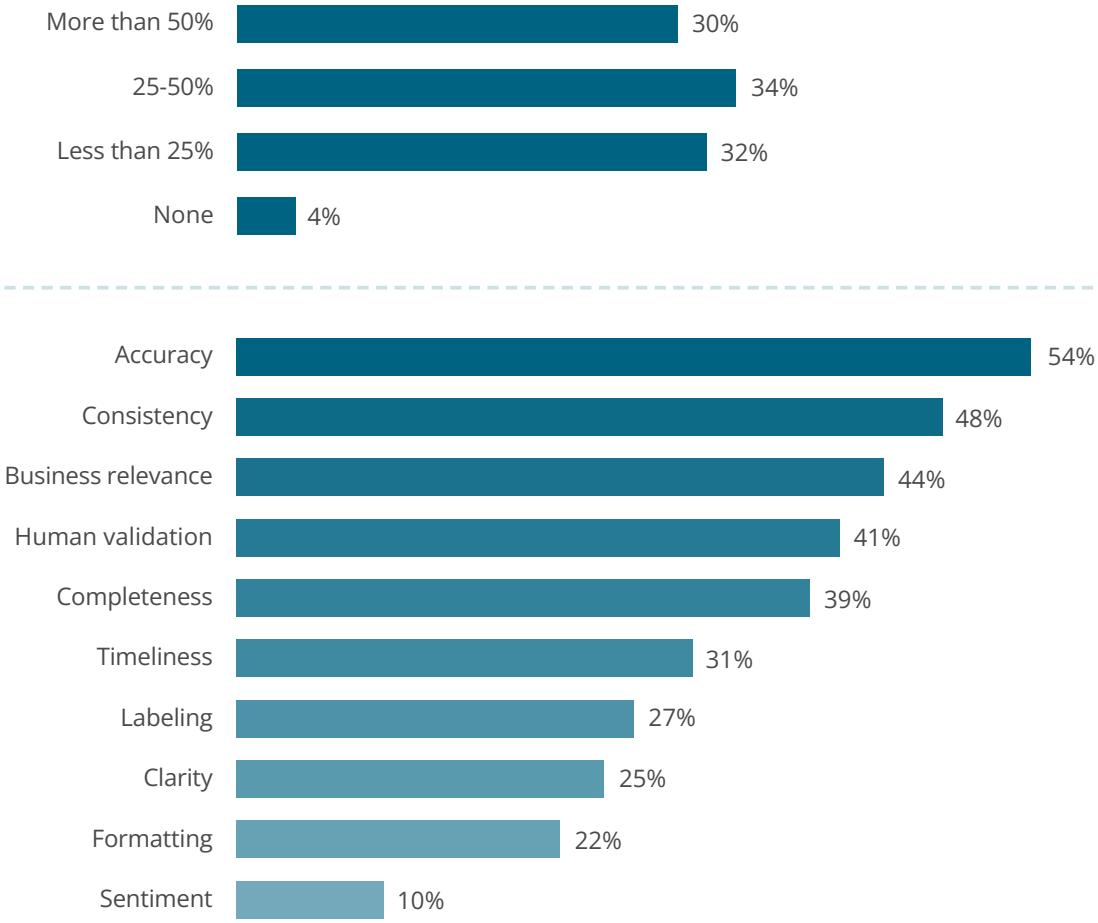


Figure 2: How much of your organization's unstructured data is currently discoverable and usable for analytics or AI initiatives? (n=196)

Figure 3: How do you measure the AI readiness of unstructured data? (n=195)

Viewpoint

Most unstructured data is not ready for AI. In fact, 70% of respondents say that less than half their unstructured data is discoverable and usable for analytics/AI. The problem is less widespread for Leaders, North Americans, and Information Technology (IT) industry respondents, whose portions are 55%, 61%, and 67%, respectively. Still, most adopters, in most cases, have significant work to do. They will fail to provide contextual understanding to agentic AI until they discover most or all of their unstructured data and make it usable.

All this raises the question: how does one define and measure the AI readiness of unstructured data? The answers reinforce longstanding yardsticks for data quality, including accuracy (54%), consistency (48%), completeness (39%), and timeliness (31%). Respondents also prioritize business relevance (44%). However, all these measures require probabilistic logic and judgment rather than deterministic quality rules, which marks a sharp contrast with the traditional world of structured data. This helps explain why human validation ranks high at 41%.

Unstructured Data Environments Are Diverse and Distributed

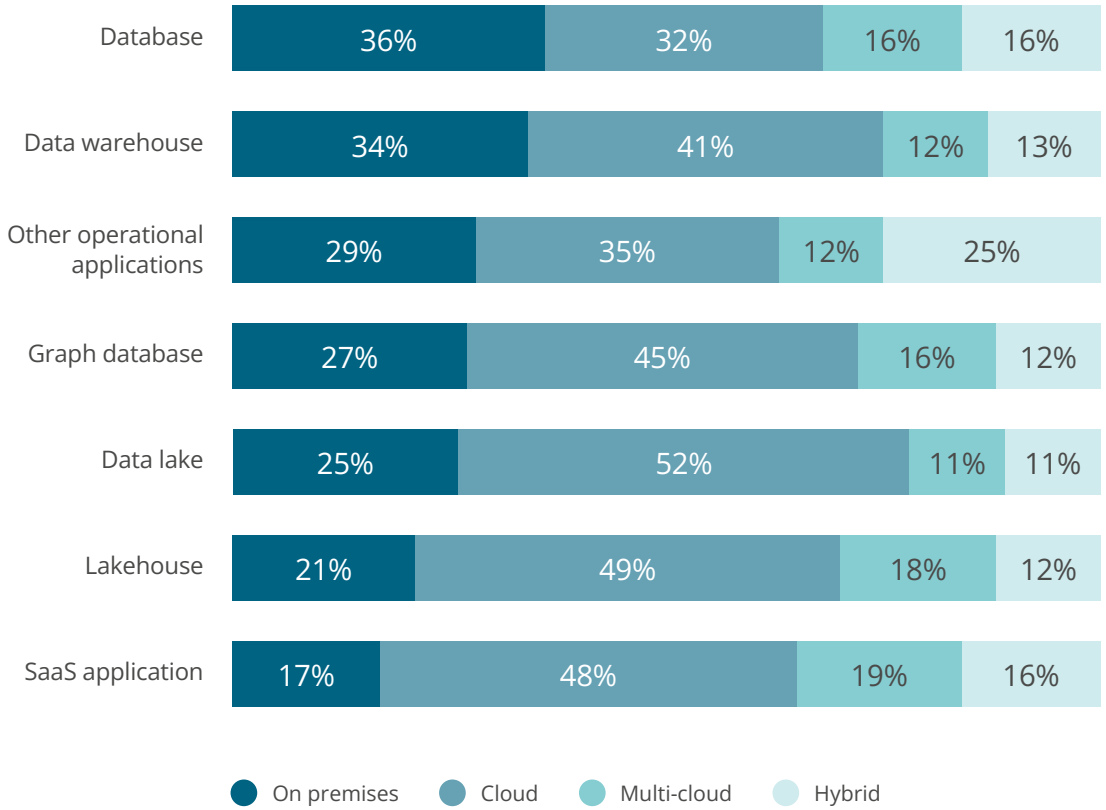


Figure 4: Where does your unstructured data reside? (n=208)

Viewpoint

Unstructured data, which comprises about one third of enterprise data according to our first two AI surveys, spans numerous environments. As with many AI inputs, unstructured data lives in diverse, far-flung sources on multiple platforms. On-premises and hybrid infrastructure continue to play a major role, ranging from 33% of Software-as-a-Service (SaaS) applications to 52% of database environments. We can expect on-premises installations to persist given data gravity, migration complexity, and rising sovereignty concerns. The distributed nature of unstructured data will restrict enterprise-wide AI projects for companies that lack an independent semantic layer.

Platform and segment differences reflect broader industry trends. Data lakes, lakehouses, and SaaS applications have a stronger cloud and multi-cloud presence, while databases and data warehouses are more common on premises. Leaders, always cloud-friendly, tend to host unstructured data in cloud-based data lakes and databases as well as hybrid applications. These hybrid and cloud ecosystems provide access to advanced tooling, especially from hyperscalers, enabling Leaders to push more AI projects into production faster.

Survey Results

State of Adoption

Application Gorillas Join Hyperscalers as Preferred Platforms

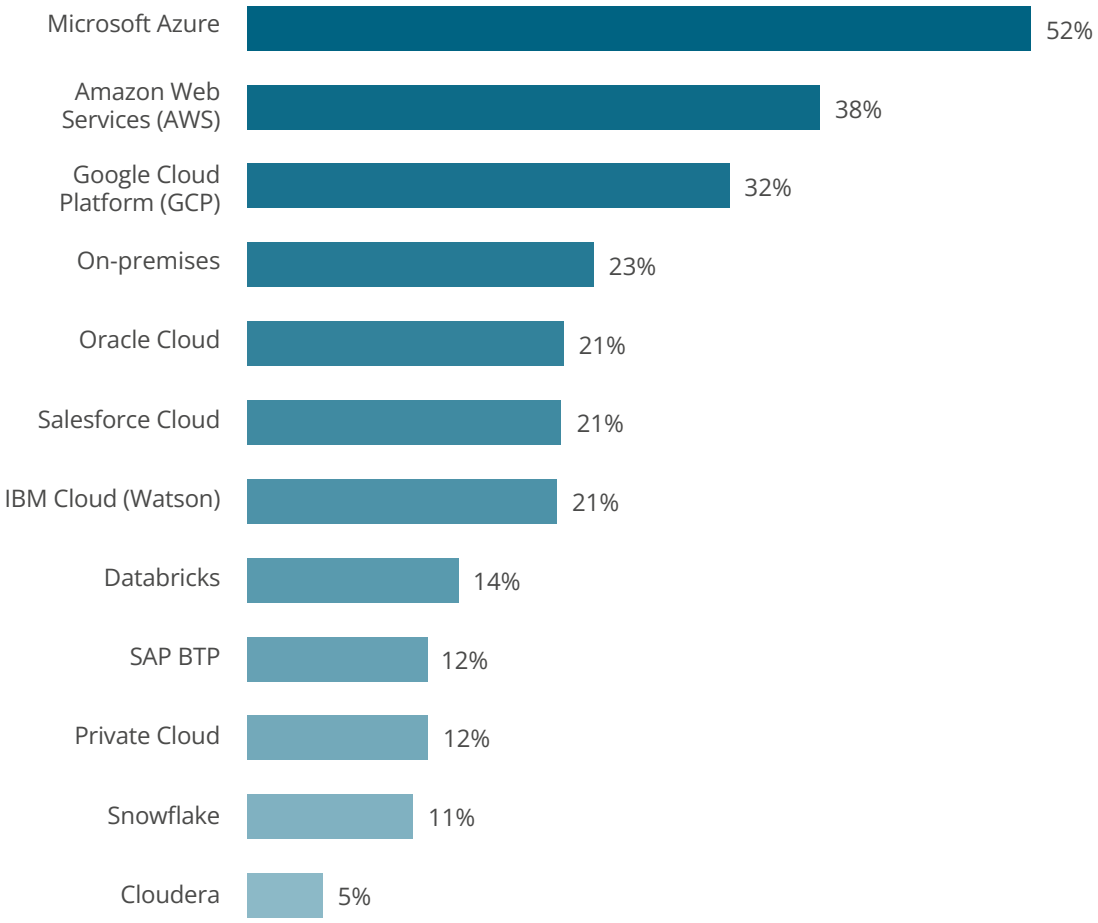


Figure 5: Which of the following platforms does your company use for storing and managing unstructured data in support of AI? (n=224)

Viewpoint

The choice of platform for deployment remains relatively stable. For storage and use of unstructured data for AI, Microsoft Azure continues to hold a substantial lead over its competitors: 52% compared to 38% for AWS and 32% for Google Cloud. These figures echo last year's aggregate survey findings for data, analytics, and AI. And they reflect the reality that AWS' commanding market share lead for overall cloud computing derives from operational workloads rather than analytics. On-premises platforms come in at 23%, unchanged from the prior result.

Oracle, Salesforce, and IBM are tied at 21%, marking a slight down-tick for Salesforce, which had an edge in actual deployment a year ago. Databricks, SAP, and Snowflake, not present in last year's data, join private cloud in the over-10% group. All offer data warehouse and data lakehouse solutions to their customers, and these are likely being used for unstructured data.

Not surprisingly, IT industry respondents, who tend to use a broader portfolio of technology, use AWS, Google, IBM, Oracle Cloud, SAP BTP, private cloud, and Cloudera more.

Moving on-premises unstructured data is a hard sell; it represents a sunk cost for storage, movement itself is expensive, and then new storage charges in the cloud will apply. The platform has gravity more than the data itself.

Data Engineers and AI Engineers Lead in Deploying Projects

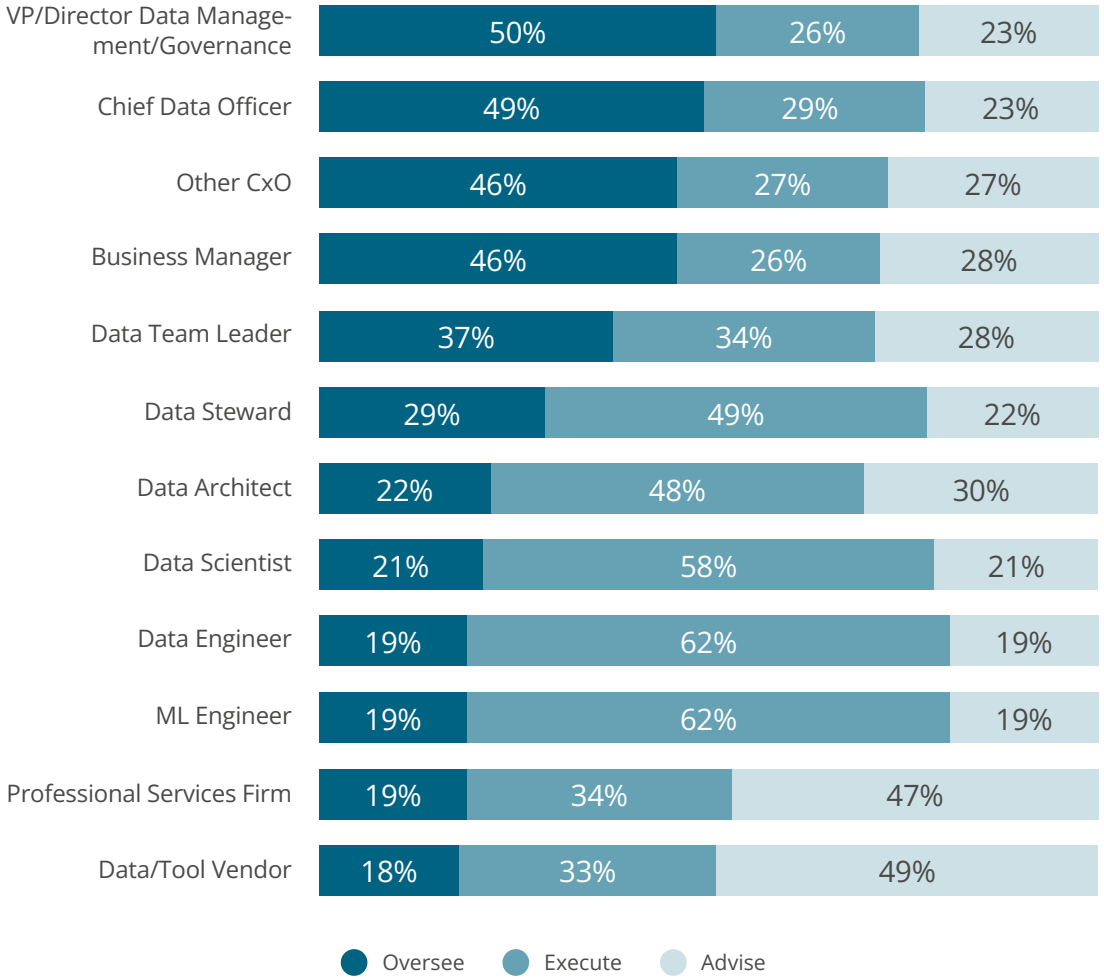


Figure 6: Please describe the main role of the following stakeholders in managing and/or governing unstructured data to support your AI projects. (n=211)

Viewpoint

Execution of unstructured data tasks falls squarely on the shoulders of engineering staff—but in a bit of a surprise, data engineers and ML engineers tied at 62%, outpacing data scientists (at 58%) and data stewards (49%). There are differences between North America and Europe: ML engineers and data scientists are more involved in execution in Europe (this is also true for Leaders compared to the general result), while data engineers in Europe are less so. However, data engineers play a larger role in execution outside the IT industry. Across industries, they need the help of AI/ML experts as roles continue to overlap.

Oversight falls firmly into the management camp, with 50% of VPs, 49% of chief data officers (CDOs), and 46% each of "other CXOs" and business managers, rounding out the top four. CDOs tend to be more hands-on in North America, as do VP/Directors in the IT industry. Overall, managers play a more significant role in execution than engineers in management. Readers are encouraged to draw their own conclusions and compare them with their own experiences. Vendors and professional services (PS) firms are, unsurprisingly, more involved in execution than internal managers. They come close to engineering staff in oversight.

IT and Centers of Excellence Lead Adoption Across the Enterprise

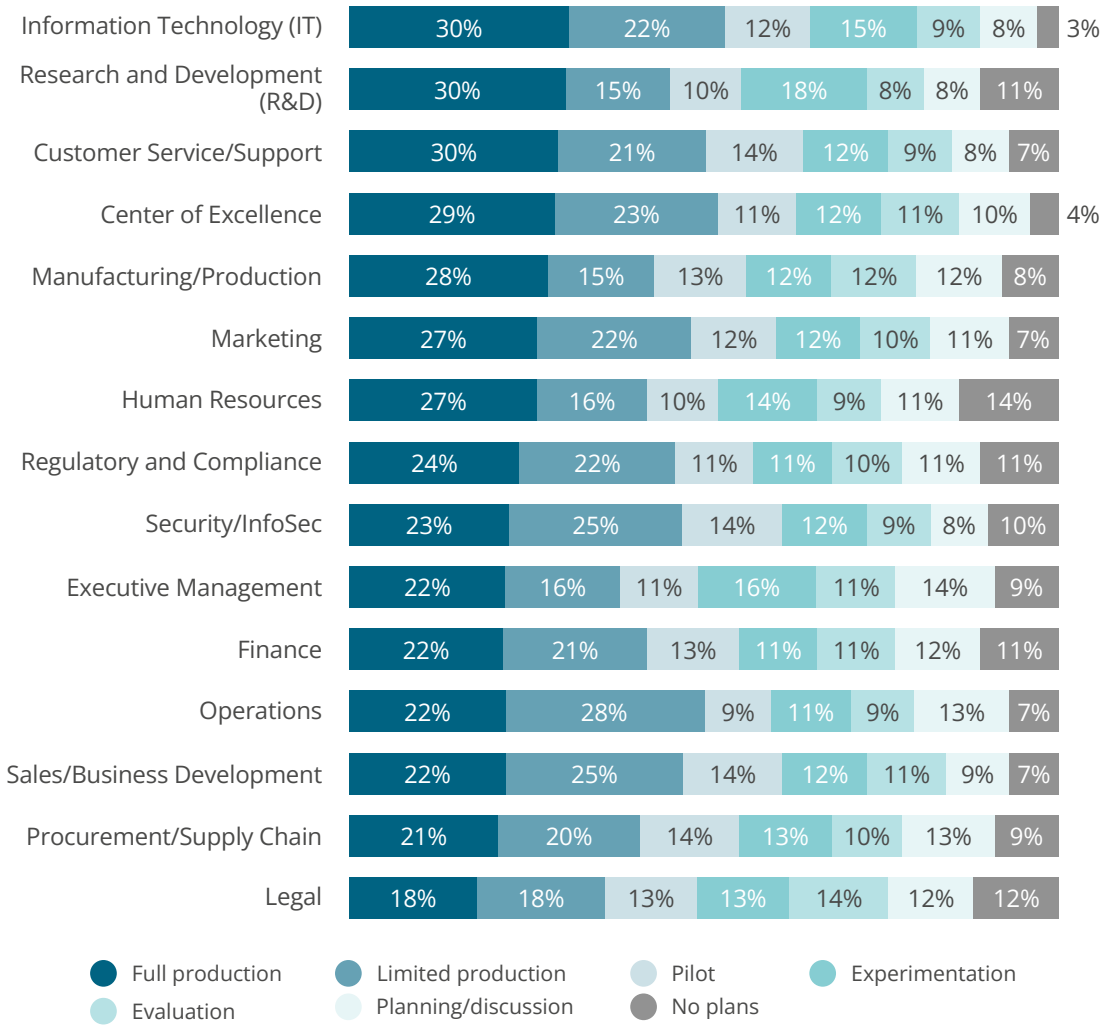


Figure 7: Which of your organization's business functions apply AI to unstructured data, and what is the status of their initiatives? (n=212)

Viewpoint

Production with unstructured data is in full flight across business functions in the respondent firms. IT departments and centers of excellence (COE) lead with full or limited production at 52%, followed closely by customer service at 51%. In earlier surveys where we measured influence and/or strategy, and budget participation, IT led and COEs were close behind. Marketing (49%), operations and security/infosec (48%), sales (47%), regulatory & compliance (46%), and R&D at 45% have all moved well up the rankings since those earlier days. This is not surprising since the outcomes focus more on their needs.

Executive management (38%) reported much lower numbers than in the earlier surveys but engaged in pilots and experiments at rates comparable to the Leaders. These results suggest that "softer" targets with less obvious payback and ROI have lagged slightly but are likely to grow in the near future. But in aggregate, responses to this question demonstrate aggressive consumption of unstructured data across the enterprise as AI goes mainstream.

Organizations Seek Meaning and Trust in Data

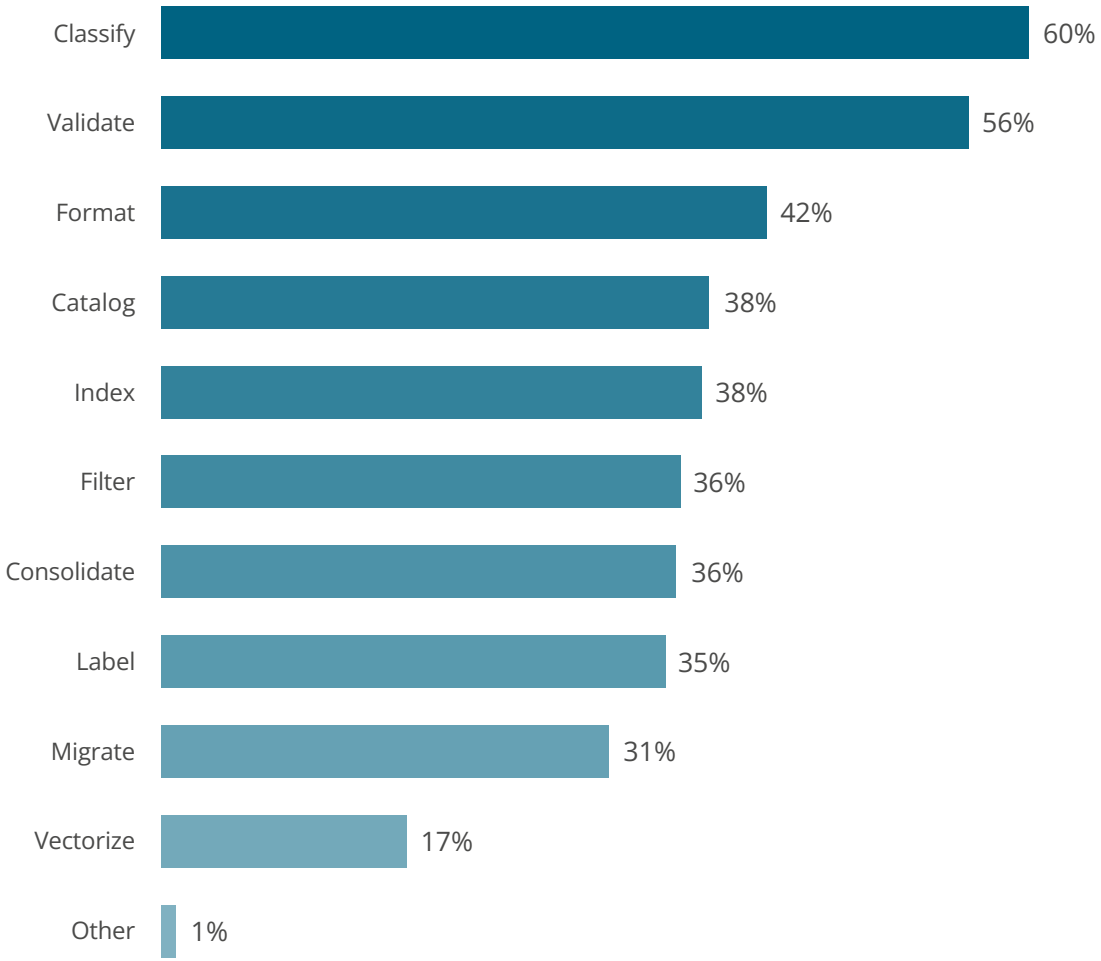


Figure 8: Which of the following tasks does your organization perform to prepare unstructured data for AI? (n=193)

Viewpoint

By far, the most widely used tasks for organizations that seek to use unstructured data are the ones that provide meaning and confidence in the data itself. Classification (60%) and validation (56%) are the leading choices, and there is a drop of 14% before requirements such as formatting (42%), cataloging (38%), and indexing (38%) appear. Aside from labeling (35%), all the remaining tasks relate to usage: filtering, consolidating, migrating, and vectorizing cannot begin until the data is understood. You need your map before you begin the voyage.

It is striking to see just how low vectorization scores on this question. Indexing ranks higher and may seem more approachable as a first step toward improving performance. We expect to see vectorization rise in future surveys, unless it becomes invisible as tools build it in rather than requiring it as an explicit task.

Generative AI Leads the Charge with AI Model Deployment

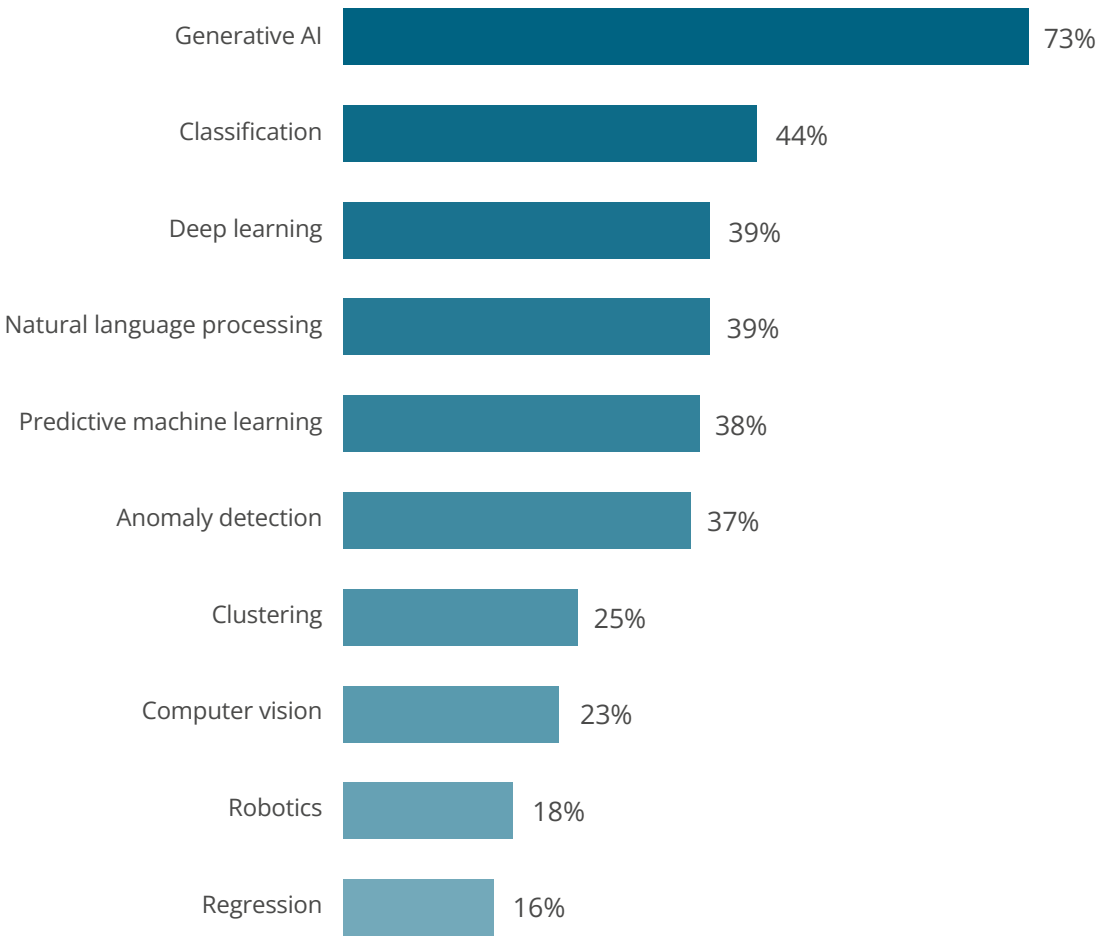


Figure 9: Which of the following model types is consuming unstructured data for training or inference? (n=200)

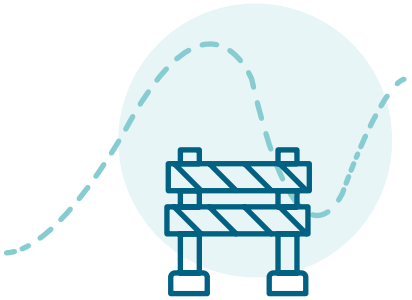
Viewpoint

The significant lead of Generative AI (GenAI) in overall usage in early AI deployments is reflected here in responses that describe what types of models are in use with unstructured data. Coming in at 73%, generative models are used two thirds more often than classification, even though classification is the leader among data management tasks performed (see page 17).

Familiar categories such as deep learning and natural language processing (both at 39%), predictive machine learning (38%), anomaly detection (37%), and clustering (25%) make up the next group. They are largely foundational, establishing the smooth and effective operation and usage of the data with an eye to growing usage. We believe many agentic workflows will involve multiple model types, not just GenAI.

The least used model types are likely to grow substantially after adopters complete this foundational work. Computer vison (23%), robotics (18%), and regression (16%) all lend themselves to practical applications.

Survey Results



Challenges and Priorities

Organizations Prioritize Trust, Training, and Business Results

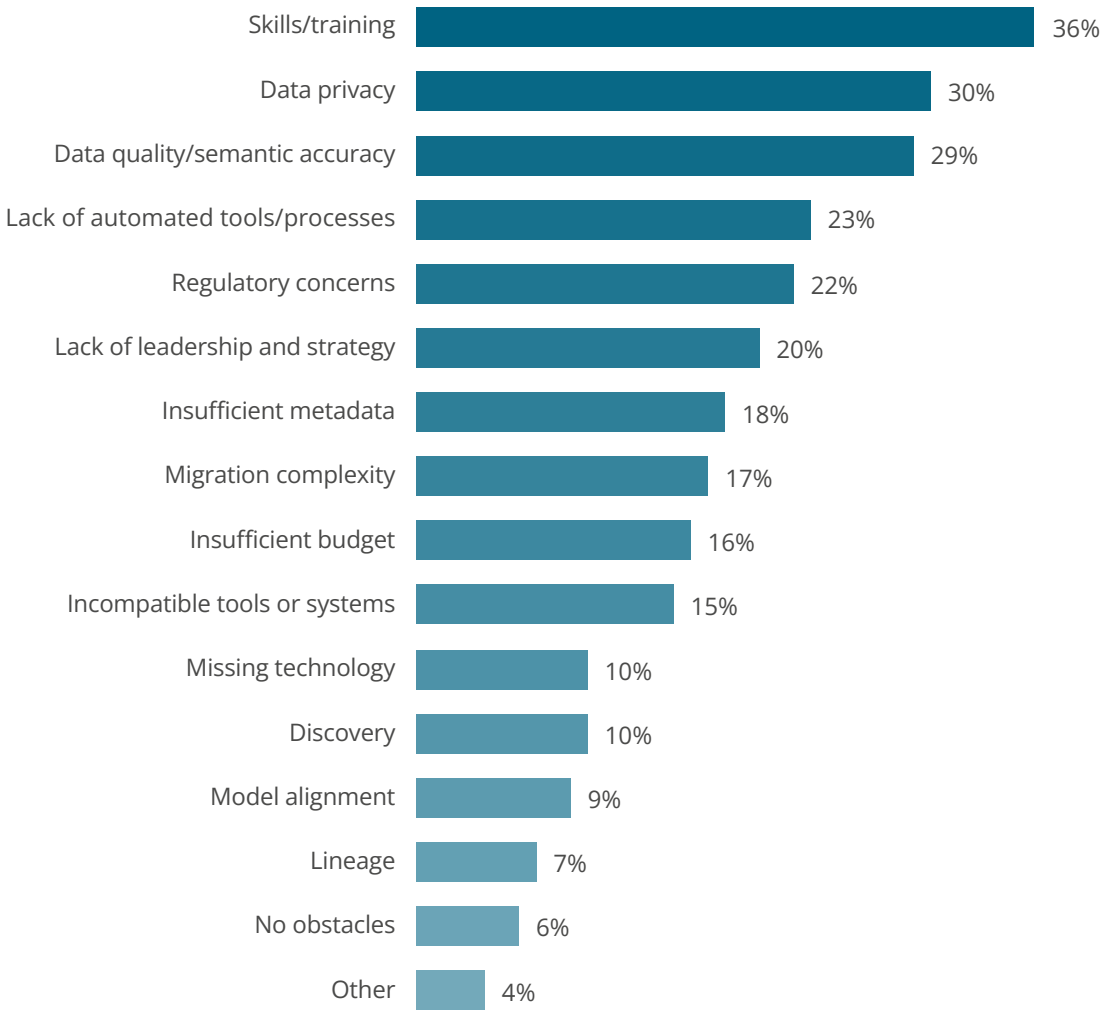


Figure 10: What are the top three obstacles to success with managing unstructured data for AI? (n=216)

Viewpoint

The need for skills and training continues to lead any discussion of obstacles. Even where other steps have been taken to improve, unstructured data is often so underutilized that over a third of respondents (36%) cite this challenge. Skills and training rank first or second in all four surveys.

Data privacy (30%) and quality/semantic accuracy (29%) come next, reflecting the difficulties that often accompany the data itself. Three of four surveys had these in the top three.

The next set of obstacles comes as no surprise: tools/automation (23%), regulatory concerns (22%), and leadership/strategy (20%) have been perennial themes for new technology introductions, and unstructured data for AI is no exception. Both data surveys ranked tools and automation markedly higher than the AI surveys did. As in earlier comments, we believe these are problems that surface in implementation. Leadership and strategy appear in the middle of the pack in all four surveys, with no significant growth or decline.

Some interesting topics appear at the bottom of the list. Lineage (7%) has fallen to last place among obstacles, just behind model alignment (9%). These obstacles featured more prominently in broad AI surveys over the past few years. These numbers, coupled with low adoption of lineage tools (see page 14), refutes the conventional wisdom that lineage is a top priority.

Data Quality Tops the List of Success Measures

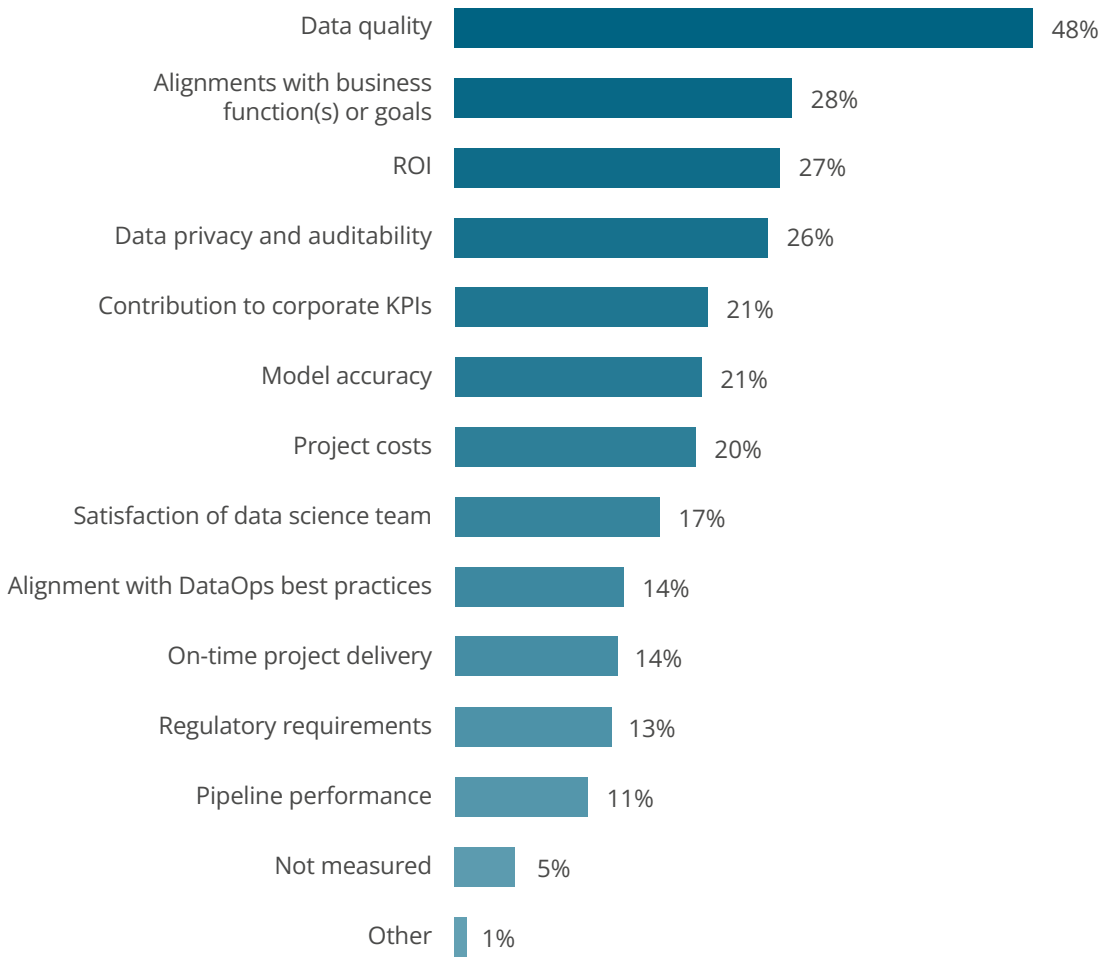


Figure 11: What are the top three ways in which you measure the success of your unstructured data management activities that support AI? (n=210)

Viewpoint

Organizations targeting the use of unstructured data measure success in different ways compared with earlier general AI projects. Data quality (48%), a significant unknown for many organizations tackling unstructured data that has been underutilized, dominates the responses to the question, and led in the *Preparing and Delivering Data for AI* study last year. Data quality has risen to the top of overall AI obstacles over the last two years as organizations learn the risks of faulty model inputs.

Alignment with business functions or goals remains a focus, appearing in the top three in all four surveys. It is second in this study with 28%; in the *Lessons From the Leading Edge* research, it was ranked third with 30%. ROI is third this year at 27%, down from its second place finish last year with 35%, but still well ahead of its 17% result in the *Lessons From the Leading Edge* research. Accuracy also remains high on both lists: tied for fifth at 21% this year and fifth at 24% last year. It's also notable that the earlier research ranked cost second at 30% while this survey had it seventh at 30%. Accuracy and cost are the priorities of production and reflect the mainstream adoption levels of AI.

Data and AI Teams Favor Reliable, Fast, and Simple Data Management Tools

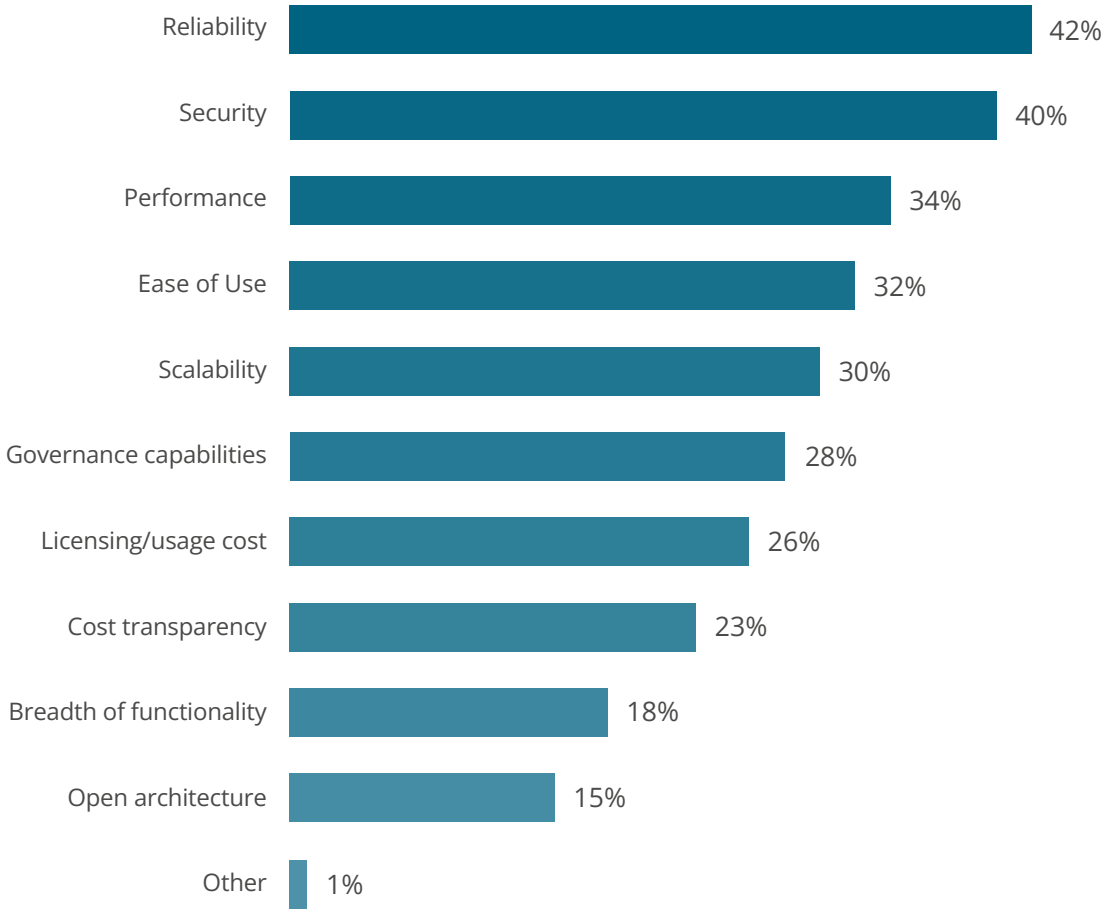


Figure 12: What are your top three criteria for evaluating unstructured data management tools? (n=209)

Viewpoint

Although concerns about costs rank fairly high in other categories, when organizations evaluate the tools to be used with unstructured data, reliability (42%) and security (40%) dominate and cost is far behind in seventh place at 26%. Prior challenges with unstructured data tools (of which there are many in place already) no doubt have made implementers wary. Cost transparency (23%) is right behind—the unpredictable costs of working with unstructured data have made buyers somewhat vigilant.

More typical criteria such as ease of use (32%) and scalability (30%) come behind the top three, joined by governance at 28%.

A surprise here is the lack of interest in open architecture, which trails all other criteria at only 15%. Considering the impact it has had in so many data management software trends over the past few years, it surprisingly lags here. Many vendors are starting to unlock their IT and embracing open-source projects such as Apache Iceberg and open standards such as Model Context Protocol. However, many key links remain closed. Users should advocate for open standards from their vendors and avoid closed tools.

Survey Results

Governance

Responsible AI Means Trusted, Transparent, and Helpful to Humans

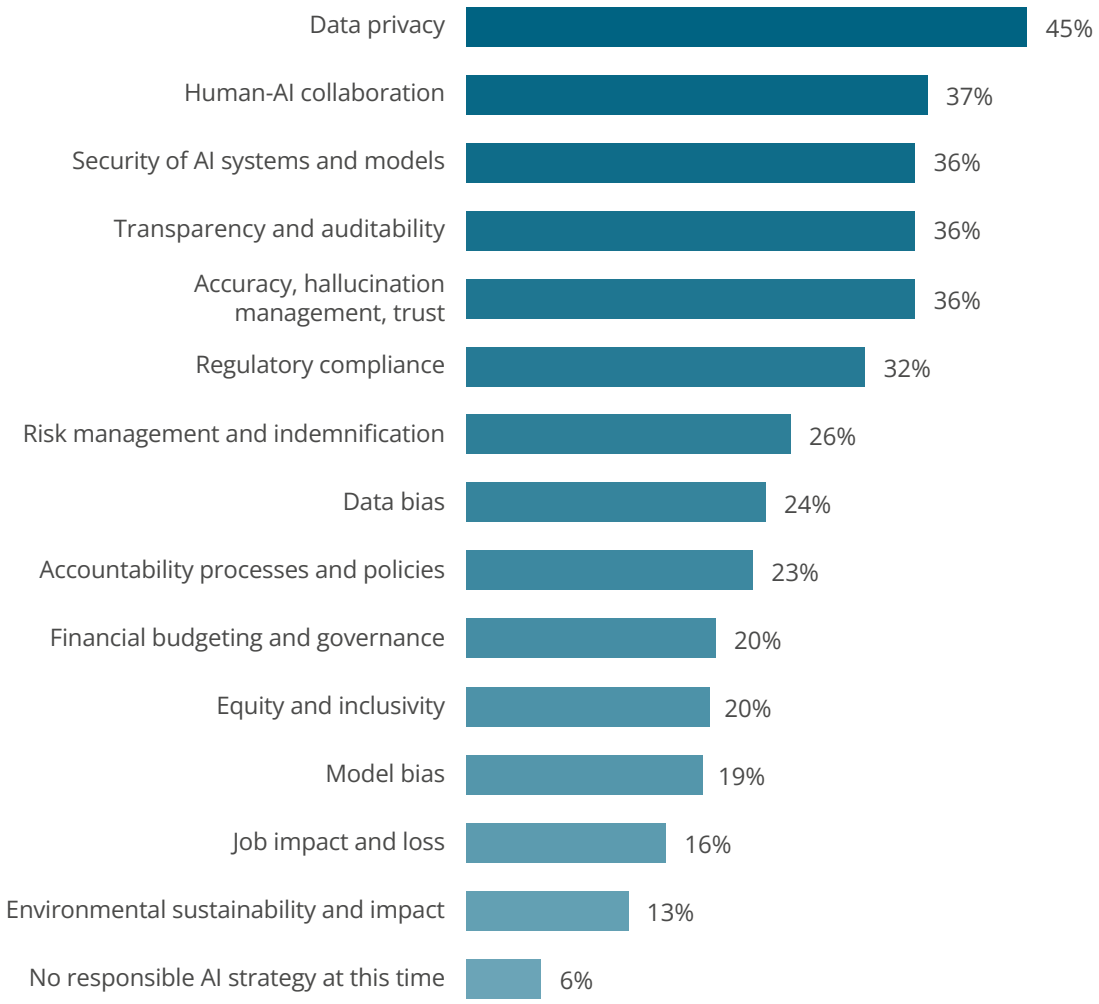


Figure 13: What are your Responsible AI priorities as it relates to unstructured data? (n=205)

Viewpoint

Data privacy, at 45%, continues its top status as a priority, following its move into first place in the *Lessons from the Leading Edge* research last year. It has been in the top three across all four surveys. As elsewhere in this study, security, transparency/auditability, and accuracy (all with 36%) come in near the top. These are the issues of production: as organizations deploy AI with unstructured data, Responsible AI priorities like these become paramount in importance.

Human-AI collaboration drew 37% of this year's respondents for second place, up from sixth place with 21% last year, although it appears in the top three in three of the four surveys. AI adopters clearly see a strategic role for humans in helping models understand the meaning of data. We believe they will do so by building, curating, and enriching semantic layers.

Finally, as we have seen in other studies, environmental sustainability lags all other issues. And equity/inclusivity (20%) and model bias (19%) remain far from their early status as key concerns. Only 6% of respondents have no Responsible AI strategy at this time. It is not zero yet, unfortunately, and reverses the downtick seen in last year's study.

Organizations Show Progress with Governance, but Uneven Maturity Levels

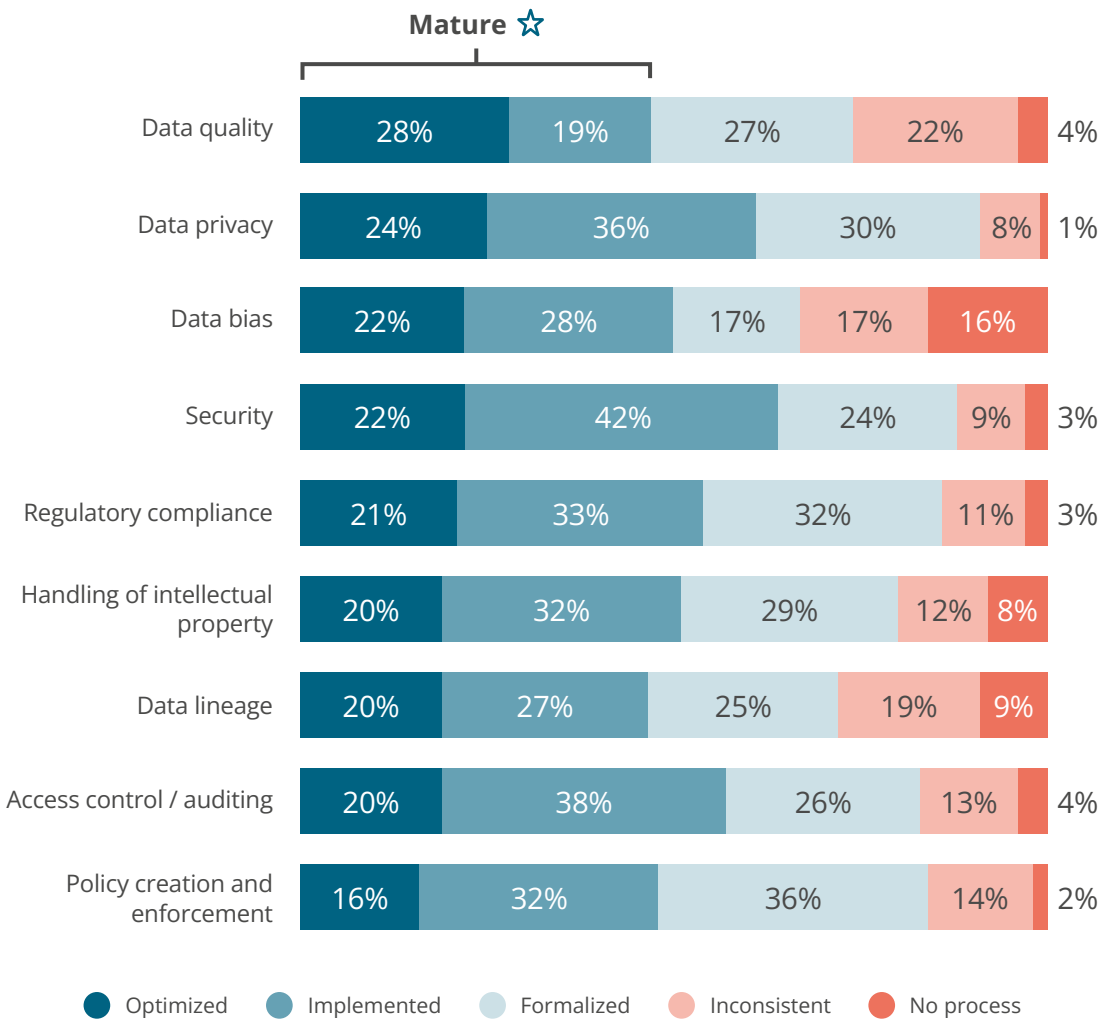


Figure 14: Please rate how well your organization addresses the following aspects of unstructured data governance in support of AI initiatives. (n=208)

Viewpoint

Organizations show surprising progress with governance of unstructured data. Across every category, a higher portion of respondents have implemented or even optimized their governance controls for unstructured data compared to overall data governance for March 2025. They show higher maturity levels with data privacy, security, and access controls in particular. Data quality, the biggest obstacle to AI success, jumped to 47% for optimized/implemented responses from 36% last year. This progress is more stunning when you consider that last year we asked about governance of all data types, including tradition-al tables, while this year we focused on the more challenging frontier of unstructured data.

Leaders outpace other adopters in every category, reflecting the many governance-related aspects of their AI programs. North Americans are more mature than Europeans in every category, even regulatory compliance (58% implemented/optimized vs. 48%). This begs the question of whether General Data Protection Regulation (GDPR) and the EU AI Act have achieved their privacy and safety objectives. It also gives North Americans a stronger springboard for AI innovation. IT industry respondents demonstrate higher maturity than the overall market in many critical categories, including data quality, lineage, bias, and access controls. Like Leaders and North Americans, they set the pace for the rest of the market.

However, many adopters still have plenty of work to do. On average, respondents have not implemented half of these controls yet. Data quality, bias, and lineage need urgent attention at most organizations. In fact, 33% and 28% of respondents have inconsistent or non-existent controls for data bias and data lineage, respectively. Governance teams must close these gaps while extending their policies, rules, and standards to address new risks such as rogue agent behavior.

Recommendations

Recommendations

As organizations deploy AI across multiple business functions, they have a rising need to extract value from unstructured data. Data leaders should take the following steps to start their journey.

1 Audit and catalog your unstructured data before entering production.

Far too few respondents know where their relevant unstructured data resides; in fact, 70% report that less than half of it is discoverable and usable. Without a clear inventory, AI projects will lack the contextual inputs they need to perform reliably at scale.

2 Prioritize classification and validation as the foundation of your data preparation workflow.

These tasks ranked as the two most widely adopted preparation activities, and for good reason. Meaning and confidence in data must precede downstream tasks such as vectorizing or migrating. Data engineers and scientists that focus here early prevent costly rework and reduce the risk that faulty model inputs impact production.

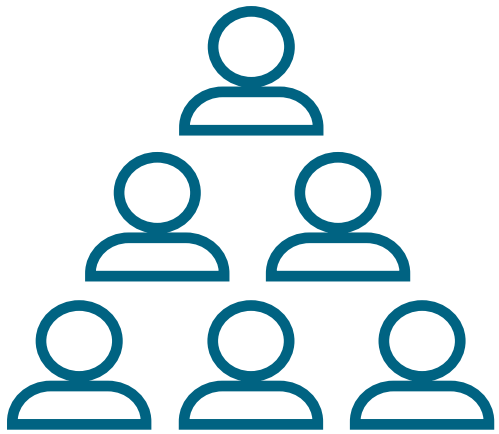
3 Close governance gaps in data bias and lineage… with humans.

Roughly one third of respondents report inconsistent or non-existent controls for bias and lineage, even as some of them deploy agents on unstructured data. The potential for compounding damage from poorly governed inputs grows significantly when human review is no longer part of every decision loop.

4 Build toward a platform-independent semantic layer to unify distributed unstructured data.

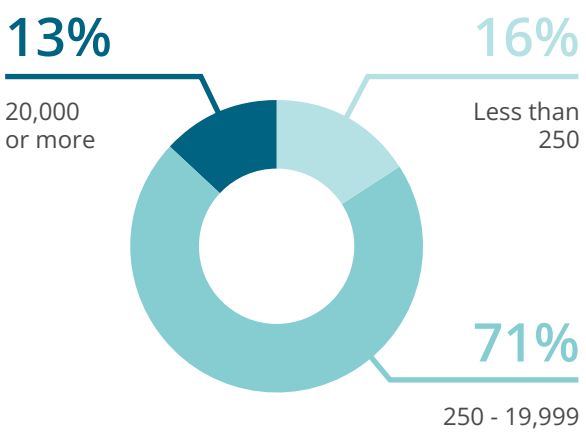
Unstructured data spans on-premises, hybrid, and multi-cloud environments that enterprises are unlikely to consolidate. Define your requirements for an independent semantic layer—covering federated querying, aggregated data views, and integrated metadata—then evaluate and select a commercial product to implement it. This will give your data team the cross-platform access and context it needs to move AI projects into production faster and with greater confidence.

Methodology

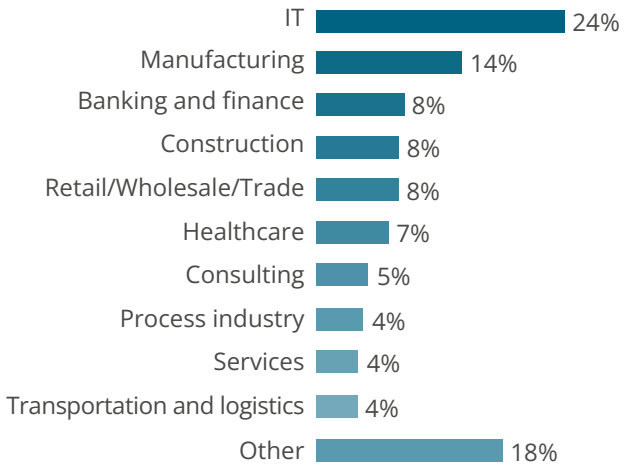


About This Research: A Global Survey of 225 Data and AI Leaders

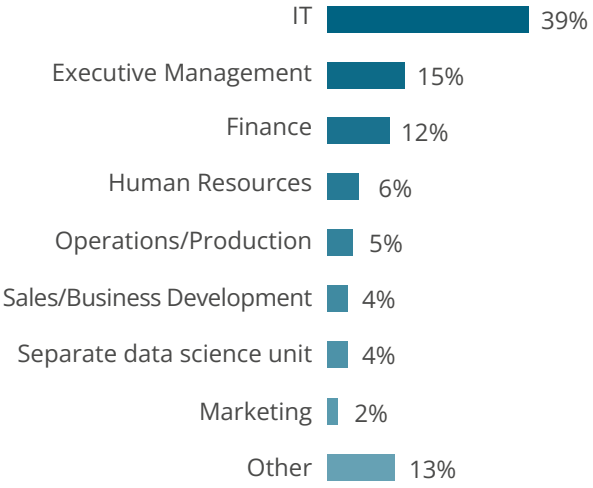
Number of Employees



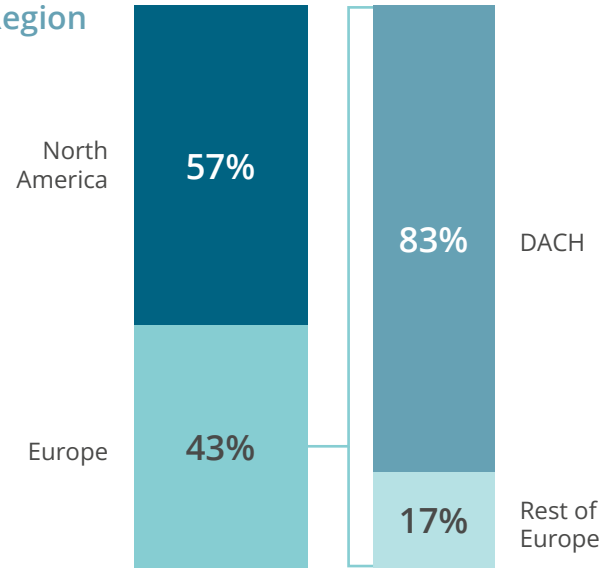
Industry



Department



Region



Information on the Survey

This worldwide study is based on a quantitative online survey conducted in February and March 2026. BARC promoted the survey within the BARC panel, on websites and via newsletter distribution lists. The objective was to understand how organizations capture, govern and operationalize unstructured data for AI innovation, and to identify the practices that distinguish the leading edge from the rest of the market.

After data cleansing, a total of 225 qualified responses remained for analysis. Respondents came from a wide range of industries, company sizes and roles, with a global distribution and a particular focus on Europe and North America. Participants were included based on their direct responsibility for data, analytics or AI strategy; those without project experience in the field were excluded.

BARC used a structured questionnaire specifically designed to capture participants' expertise and practical experience with unstructured data for AI. Due to rounding, percentage totals in the figures may not add up to exactly 100. The "I don't know" response option is not included in the sample size shown below each figure and is not displayed in the charts.

About BARC

BARC

BARC is the leading analyst firm for data & analytics, AI, corporate performance management (CPM), and ESG with a reputation for unbiased and trusted advice. Our expert analysts deliver a wide range of research, events, and consulting services for the data & analytics community. Our innovative research evaluates software, vendors, and service providers rigorously and highlights market trends, delivering insights that enable our customers to innovate with data, analytics, and AI. BARC's 25 years of experience with data strategy & culture, data architecture, organization, and software selection helps clients transform into truly data-driven organizations.

Research

BARC user surveys, software evaluations, and analyst advisory services along with expert driven content such as research notes, trend analysis, and blogs give organizations the confidence to make the right decisions. Our independent research gets to the heart of market developments, evaluates software, vendors, and service providers thoroughly and gives valuable ideas on how to turn data, analytics, and AI into added value and successfully transform businesses.

Consulting

The BARC consulting practice is entirely focused on translating companies' requirements into future-proof decisions. The holistic advice we provide helps companies successfully implement their data & analytics strategy and culture as well as their architecture and technology. BARC's research and experience-founded expert input sets organizations on the road to the successful use of data & analytics, from strategy to optimized data-driven business processes.

Events

At BARC events, leading minds and industry experts come together to share insights and drive innovation. Our conferences, roundtables, and online webinars attract over 10,000 participants annually, offering a unique blend of information, inspiration, and interactivity. These events provide a platform to exchange ideas with peers, explore emerging trends, and gain expert perspectives on market developments. By engaging with thought leaders and industry practitioners, participants discover actionable strategies to enhance their business and stay ahead in the evolving world of data & analytics.

Germany

BARC GmbH
Berliner Platz 7
D-97080 Würzburg

info@barc.de
+49 931 880 6510

Austria

BARC GmbH
Hirschstettner Straße 19 / I / IS314
A-1220 Wien

info@barc.at
+43 660 6366870

Switzerland

BARC Switzerland GmbH
Buchhaldenstrasse 7
CH-5442 Fislisbach

info@barc.ch
+41 76 340 35 16

USA

BARC US
13463 Falls Drive
Broomfield, CO 80020

info@barc.com
+1 720-381-4988

www.barc.com

BARC

Sponsor Profiles

DataHub

DataHub transforms enterprise data into trusted context, enabling intelligent decision making by humans and AI agents. The company was founded by the creators of the popular DataHub open source product that has more than 15,000 contributors and is used by thousands of organizations.

The company's flagship product, DataHub Cloud, is the leading context management platform trusted by the Global 2000 to ensure that context is always relevant, reliable and continuously refreshed across the entire data. estate. DataHub is backed by Bessemer Venture Partners, LinkedIn and 8VC.

For more information, go to: datahub.com.

Contact

DataHub
datahub.com

LinkedIn: linkedin.com/company/datahub-cloud
X: x.com/DataHubCloud
Slack: datahub.com/slack

Ohalo

Eighty percent of enterprise data lives in files: documents, PDFs, emails, contracts, scanned images, and nested archives. Your catalog indexes their metadata. Your DLP watches them move. Your SIEM logs who touched them. But no tool in your stack really knows what's in them. None has ever opened those files, read what is inside, and classified the contents. That is the blind spot where AI initiatives stall, compliance gaps hide, and security exposure compounds.

Ohalo built Data X-Ray because governance stops where files begin. It opens every file across petabyte-scale estates, classifies what is inside using a three-layer approach (NLP/ML annotators for token-level precision, GenAI/LLM for document-level context, and customer-defined rules for business logic), and feeds enriched metadata directly into existing Agents via MCP, data catalogs, DLP tools, and AI pipelines. Every file is read, not inferred from metadata.

Deploys entirely within the customer's environment. Container-based, agentless, operational in hours. On-prem, air-gapped, hybrid, or multi-cloud. 55+ native connectors, 1,000+ upstream data sources. No data migration. No cloud dependency. It clears the regulatory, privacy, and security hurdles that SaaS-dependent tools cannot.

Trusted by financial institutions, defense agencies, and critical infrastructure operators across the US, Europe, and Asia. The result: unstructured data that was previously invisible becomes classified, governed, and ready for AI pipelines, compliance workflows, and security operations, with confidence and without guessing what is in it.

Contact

Ohalo
ohalo.co
hello@ohalo.co

LinkedIn: linkedin.com/company/ohalo-limited

ohalo | Data X-Ray

BARC