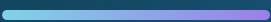


BIGID AI READINESS CHECKLIST

The AI Data Readiness Checklist

A practical, working checklist for security, privacy, and data leaders preparing for AI at scale, organized by the order it actually needs to happen.



WHAT'S INSIDE

- 1 Six ordered checkpoints, from discovery through policy enforcement
- 2 A red flag for each section that signals you're not actually ready
- 3 The non-human identity gap most access reviews still miss
- 4 Three direct questions to ask before you scale AI further

Most "AI readiness" checklists are written for the model side: model selection, evaluation, guardrails. This one isn't. It's built around the question that actually determines whether AI adoption is safe: do you know what's in your data, who and what can reach it, and what happens when an AI system tries to use it.

Work through this in order. Each section builds on the one before it.

1 Know what you have

You cannot govern data you haven't found, and most organizations have more of it, in more places, than they think.

- ☐ Inventory structured data sources: databases, data warehouses, data lakes
- ☐ Inventory unstructured data: file shares, SharePoint, Google Drive, Confluence, email
- ☐ Inventory SaaS application data (Salesforce, Workday, ServiceNow, and similar)
- ☐ Inventory AI-native data stores: vector databases, embedding stores, RAG indexes
- ☐ Identify shadow data: copies, exports, and backups outside primary systems of record
- ☐ Map which of these sources currently feed, or could feed, an AI model, agent, or copilot
- ☐ Confirm discovery covers multi-cloud and on-prem, not just one environment
- ☐ Surface shadow AI: unsanctioned models, unauthorized users running models, and vendors using AI without authorization

Red flag: if your AI inventory and your data inventory are maintained by different teams using different tools, you have a visibility gap, not just a coordination problem.

2 Know what it is

Discovery without classification just tells you where the haystack is.

- ☐ Classify data by sensitivity (PII, PHI, PCI, IP, source code, financials)
- ☐ Classify by regulatory exposure (GDPR, CCPA, HIPAA, sector-specific, and emerging AI regulation)
- ☐ Tag data by business context, not just data type (which function owns it, what it's used for)
- ☐ Identify where sensitive data overlaps with data already in use by AI pipelines
- ☐ Establish classification at the field and document level, not just the dataset level
- ☐ Confirm classification updates continuously, not on a quarterly or annual cycle
- ☐ Maintain a searchable catalog of generative AI data so you can audit what data was used to train or feed which AI system

Red flag: classification that lives in a static report is already out of date by the time someone reads it.

3 Apply privacy controls before AI touches the data

Compliance obligations don't pause for a new use case, and AI adoption tends to outrun privacy review.

- ☐ Identify all AI use cases currently in production, pilot, or planning
- ☐ Map data subject rights obligations (access, deletion, correction) against AI training and retrieval data
- ☐ Confirm consent and lawful basis cover AI use, not just the original collection purpose
- ☐ Apply data minimization before ingestion into training sets or RAG pipelines
- ☐ Build a process for handling deletion requests that reaches data already embedded or fine-tuned into a model
- ☐ Document privacy impact assessments for new AI use cases before launch, not after

Red flag: if legal or privacy is finding out about new AI use cases after they're already live, your AI pipeline is moving faster than your governance.

4 Lock down access, including non-human identities

AI agents don't behave like human users, and most access models were never built for them.

- ☐ Inventory non-human identities: service accounts, AI agents, API integrations, bots
- ☐ Map what data each AI agent or copilot can actually reach, not just what it's supposed to reach
- ☐ Identify over-permissioned access: agents or integrations with broader access than their function requires
- ☐ Apply least-privilege access specifically to AI training data, RAG sources, and agent permissions
- ☐ Confirm access reviews include AI systems on the same cadence as human user access reviews
- ☐ Test what happens when an AI agent attempts to retrieve data outside its intended scope
- ☐ Track agent activity against the data it touches, not just whether access was granted, so you have traceability if something goes wrong

Red flag: if you can produce an access report for every employee but not for every AI agent, the agents are your biggest unknown.

Why this is hard to fix with existing tools: Access platforms typically don't know what data they're protecting, and classification platforms typically don't know who, or what, can reach it. Closing this gap means running both from a shared view of the data, not stitching together exports from two systems that were never designed to talk to each other.

5 Fix retention before AI makes it permanent

Data that should have been deleted doesn't just sit there quietly. It gets pulled into training sets and embeddings before anyone notices it was overdue for deletion.

- ☐ Identify data past its defined retention period across primary and shadow sources
- ☐ Confirm retention policy is enforced at the source, not just documented in a policy PDF
- ☐ Check whether expired or to-be-deleted data has already been pulled into a training set or vector store
- ☐ Build a process for purging data from AI-adjacent stores (embeddings, indexes, caches), not just primary systems
- ☐ Set retention rules for AI-generated outputs and logs, not only the input data
- ☐ Reconcile retention policy against classification, so what's flagged sensitive is also flagged for deletion correctly

Red flag: if your retention policy was written before anyone considered embeddings, it doesn't cover your actual exposure today.

6 Make policy enforceable, not aspirational

A control that isn't enforced at the point of use is a slide in a deck.

- ☐ Translate written data policy into rules enforced automatically at the system level
- ☐ Confirm policy enforcement covers AI pipelines specifically (training, fine-tuning, RAG, agent actions)
- ☐ Test enforcement: attempt an action that should be blocked and confirm it is
- ☐ Establish a single source of truth for policy so classification, access, privacy, and retention rules don't drift apart across tools
- ☐ Build alerting for policy violations involving AI systems, with a clear owner for response
- ☐ Review enforcement logs regularly, not only after an incident

Red flag: a quarterly audit that catches the violation after the fact is detection, not enforcement.

6 checkpoints, read from a single shared view of your data. Run them out of order, or from six disconnected tools, and the gaps between them become the exposure.

Before you scale further, ask three questions

Can you name, today, every AI use case touching sensitive or regulated data, and the data sources feeding it?

If an AI agent tried to access something it shouldn't, would you know, and would it be stopped automatically?

If a regulator or board member asked how you're governing data across AI systems, could you answer with evidence, not assurances?

If any of these gets a soft answer, the data underneath it is where to start, ahead of the model evaluation and ahead of the next pilot.

This checklist is also a map of how BigID works: one platform giving discovery, classification, privacy, access, retention, and policy enforcement a shared view of your data, so AI moves at the pace your governance can actually support.

NEXT STEP

Talk to our AI security experts

Work through this checklist against your own environment with our team. Bring your toughest "soft answer" from the three questions above, and they'll help you turn it into a plan, not just another finding.

CUSTOMER VALIDATION

Proof in Action

"BigID gives me better visibility into sensitive data, helps prioritize security protections, reduces attack surfaces, strengthens compliance, and increases operational efficiency overall — a strategic pillar of our AI-First Cybersecurity Transformation"

- CISO, Global Healthcare Company

"We needed to automate processes and data management across systems for the data of a few million customers across a lot of systems and data. BigID was the one solution that did this in the most efficient and sophisticated way - and had more use cases we could add on moving forward."

- Chief Privacy Officer, Global Telecoms Company

ANALYST RECOGNITION

Validated by Industry Leaders

- 2025 Company of the Year for AI Governance - Frost & Sullivan
- Named a Leader across all DSPM research including Omdia DSPM Universe, CB Insights for DSPM, TAG, GigaOm, Frost & Sullivan
- Named a Leader in all Privacy Management evals including Forrester Privacy Management Wave & IDC Privacy Compliance MarketScape
- BigID named a Leader across 4 GigaOm evals: DSPM, Data Security Platforms, Unstructured Data Management, Data Access Governance
- Named in Gartner's 2025 DLP Market Guide
- Named in Gartner's 2025 AI TRiSM Market Guide
- Represented 30+ times in Gartner's 2025 Hype Cycles for Security, Privacy, AISPM

FREE RISK ASSESSMENT

Discover Your Data & AI Risk

An agentless, cloud-native assessment to enable security, privacy, and compliance: pinpoint where sensitive data and AI risks live - and how to fix them.

What You Get:

- Complementary 2 week engagement on real data to gain visibility on risk across data and AI
- Executive-ready report identifying high-risk data and AI across cloud environments with next-step recommendations

NEXT STEPS

Ready to Connect the Dots in Data & AI?

Schedule your free risk assessment or ask for a deeper demo across DSPM, DLP, AI SPM, Privacy Automation, AI Risk, and Data Lifecycle Management

► www.bigid.com/demo

► www.bigid.com/risk-assessment

► info@bigid.com



BigID Next

"The features they offer are the moat that they have built around the product. This is what makes them the market leader in Data Privacy and Protection."



IT Security & Risk Management Associate - Banking

Gartner
Peer Insights.

Gartner® and Peer Insights™ are trademarks of Gartner, Inc. and/or its affiliates. All rights reserved. Gartner Peer Insights content consists of the opinions of individual end users based on their own experiences, and should not be construed as statements of fact, nor do they represent the views of Gartner or its affiliates. Gartner does not endorse any vendor, product or service depicted in this content nor makes any warranties, expressed or implied, with respect to this content, about its accuracy or completeness, including any warranties of merchantability or fitness for a particular purpose.



BigID Next

"Exceptionally powerful and easy to use tool that integrated smoothly into our environment and delivered value from day one. The BigID team is fabulous to work with."



Senior Vice President & Chief Data Officer - Banking

Gartner
Peer Insights.

Gartner® and Peer Insights™ are trademarks of Gartner, Inc. and/or its affiliates. All rights reserved. Gartner Peer Insights content consists of the opinions of individual end users based on their own experiences, and should not be construed as statements of fact, nor do they represent the views of Gartner or its affiliates. Gartner does not endorse any vendor, product or service depicted in this content nor makes any warranties, expressed or implied, with respect to this content, about its accuracy or completeness, including any warranties of merchantability or fitness for a particular purpose.