

Context Engineering for AI



EXECUTIVE SUMMARY

AI systems are only as reliable as the data that trains and informs them. BigID provides the foundation for context engineering — discovering, classifying, correlating, and governing enterprise data across environments to build trusted, explainable, and secure AI systems.

The AI Context Problem

Organizations are rapidly deploying AI across analytics, automation, and generative applications. However, most AI initiatives face a common challenge: AI models lack context about the data they consume.

Training pipelines often pull data from diverse environments without understanding:

- Whether the data contains sensitive information
- Which datasets are authoritative
- Who owns the data
- How recently the data was updated
- How frequently it is accessed or used
- Whether better training sources exist
- Which users or applications should be allowed to access the data

Without context, AI systems introduce serious challenges:

Poor Model Quality

Models trained on duplicated, inconsistent, or outdated data produce unreliable outputs.

Sensitive Data Exposure

Personal data, secrets, or regulated information may unintentionally enter training datasets or prompts.

Lack of Governance and Accountability

Organizations cannot explain which datasets were used to train models or support AI responses.

Inefficient AI Development

Data scientists spend significant time identifying and preparing appropriate training datasets.

Context engineering solves these challenges by enriching data with the metadata and controls required to make AI systems trustworthy and governable.

What is Context Engineering?

Context engineering is the practice of enriching data used in AI systems with meaningful metadata, governance controls, and preparation pipelines that help AI models interpret and safely use enterprise data.

Instead of feeding AI models raw datasets alone, context engineering adds multiple layers of intelligence that help organizations:

- Curate trusted training datasets
- Identify the most relevant information for prompts and agents
- Prevent sensitive data exposure
- Align AI responses with business context
- Ensure AI systems access only authorized information

Effective context engineering requires multiple layers of metadata and data preparation, including:

- Technical metadata describing data structure and location
- Operational metadata describing usage and access patterns
- Inferred metadata describing business meaning and ownership
- Risk metadata identifying sensitive or regulated data
- Entity correlation connecting related records and datasets to specific entities
- Freshness metadata identifying stale or outdated data

Together, these layers provide the data intelligence required to safely power AI systems.

The Role of Metadata in AI Context Engineering

AI systems require signals that help them interpret data and determine which information is relevant and trustworthy. Metadata provides these signals. By enriching

datasets with contextual metadata, organizations can guide AI systems toward the most relevant, authorized, and trusted information.

Technical Metadata

Describes the structure and location of enterprise data — including schema definitions, file types, storage location, dataset relationships, and lineage across pipelines.

Operational Metadata

Provides insight into how data is used — including access patterns, user activity, query frequency, application usage, and data movement. Identifies high-value datasets actively used by the business.

Inferred Business Metadata

Connects technical datasets to business meaning through glossary terms, ownership, data domains, and functional categorization — aligning AI systems with business context.

Risk Metadata

Identifies sensitive and regulated data requiring protection or exclusion — including personal data, financial records, regulated information, confidential IP, and secrets and credentials.

Entity Correlation

Correlates data across systems to identify records belonging to the same entity, enabling AI systems to receive complete and accurate contextual views — improving analytics and decision-making.

Data Freshness & Staleness

Tracks dataset age and activity via last updated/accessed timestamps and lifecycle stage, ensuring AI systems rely on current and relevant information.

Preparing AI Context with BigID

Beyond metadata discovery, context engineering requires preparing data so that AI systems receive clean, normalized, and policy-compliant inputs. BigID helps prepare enterprise data for AI by standardizing formats, filtering sensitive information, and enriching context before it reaches AI systems.

Normalize Data Across Formats

Enterprise data exists across structured, semi-structured, and unstructured formats. BigID connects to diverse data environments and normalizes content through consistent processing pipelines, allowing organizations to deliver clean and consistent context to AI models and agents — regardless of source format.

Mask Sensitive Data Before AI Use

Sensitive data should not always be exposed to AI models. BigID identifies sensitive information and can apply masking or transformation policies before data enters AI pipelines or prompt contexts, helping organizations safely leverage enterprise data without exposing regulated or confidential information.

Deliver Context Based on Authorization

Not every user or application should receive the same information. BigID enriches data with identity and access context to ensure AI systems receive information aligned with user permissions and policies. Role-based access controls help ensure AI agents and applications only access information they are authorized to view.

Access Metadata Context Through the BigID MCP Server

AI systems often need structured metadata about enterprise data in order to interpret and safely use it. This includes information such as ownership, sensitivity classification, business context, and related datasets.

BigID provides an MCP (Model Context Protocol) server that allows AI applications and agents to retrieve metadata context directly from the BigID platform. Through MCP tools and APIs, AI systems can request metadata about specific data objects and receive enriched context describing the data.

For example, AI systems can retrieve:

- Technical metadata describing the dataset or file
- Sensitivity classifications and risk indicators
- Ownership and stewardship information
- Business domain and glossary mappings

- Related or correlated data objects

This allows AI systems to dynamically access the metadata context required to interpret enterprise data more accurately.

By exposing trusted metadata through MCP, BigID enables organizations to integrate data intelligence directly into AI workflows while maintaining governance and visibility.

Governing AI Context with Data Intelligence

AI governance extends beyond preparing datasets for training. Organizations must also control what data AI systems are allowed to access and use as context. Without clear boundaries, AI applications and agents may retrieve information outside their intended scope, increasing the risk of exposing sensitive or irrelevant data.

BigID provides the data intelligence required to govern the datasets used by AI systems. By discovering enterprise data, identifying sensitive content, and enriching data with business and domain context, BigID helps organizations determine which data is appropriate for different AI use cases.

Define Domain-Appropriate AI Context

AI applications are typically designed to operate within specific business domains such as finance, customer support, or operations. BigID helps organizations organize data by business domain, ownership, and sensitivity — identifying which datasets are appropriate for specific AI use cases and ensuring AI systems retrieve information aligned with their intended purpose.

For example:

- A financial AI application should access finance and accounting datasets, not HR records.
- A support assistant may retrieve customer documentation but not confidential engineering designs.

Benefits of Context Engineering with BigID

Higher Quality AI Systems

Models train on relevant, accurate, and trusted datasets, improving output quality and reliability.

Reduced Data Risk in AI Pipelines

Sensitive data can be detected, masked, or excluded before reaching AI systems.

Faster AI Development

Data scientists spend less time preparing and validating training datasets.

Stronger AI Governance

Organizations gain visibility into how AI initiatives use enterprise data.

Safer AI Agent Deployment

Context-aware controls help ensure AI systems retrieve only authorized information.

The Future of AI Requires Context

AI adoption is accelerating across every industry. However, organizations cannot rely on raw data alone to power intelligent systems. AI requires context-rich data environments that describe what data represents, how it should be used, and whether it can be trusted.

Context engineering transforms enterprise data into governed AI inputs. By providing deep metadata intelligence, data correlation, policy governance, and context preparation, BigID enables organizations to build AI systems that are trusted, explainable, and secure.

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.