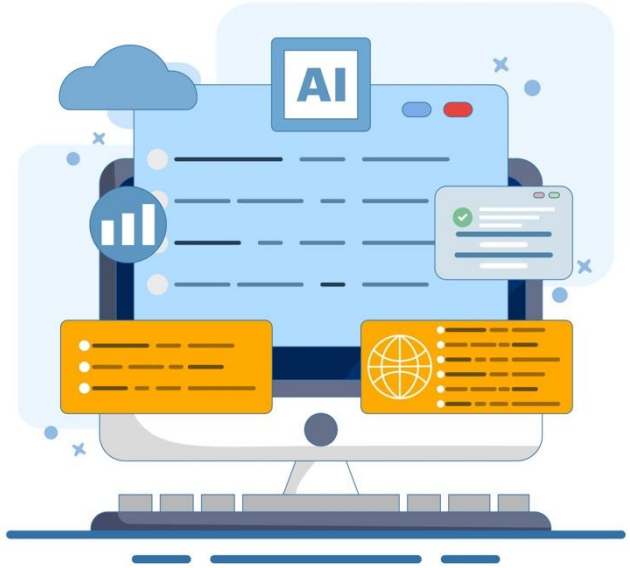
Valora Technologies

— WEBINAR SERIES —

10 Practical Ways
AutoClassification Can Impact AI



Sandy Serkes
Co-Founder & CEO

Jennifer Nelson
VP Strategic Solutions

Webinar Series

www.ValoraTech.com

ROT Remediation:
File Share Clean-up
January 21

How AC Transforms
Enterprise Search

Letting the Robots
Classify: Automating IG
February 18

Managing Data Privacy &
While Managing Records

AutoClassifying the 3 Rs:
ROT, Records & Retention
March 18

From Observational to
Operational InfoGov

Structured Data
Management & Disposition
April 15

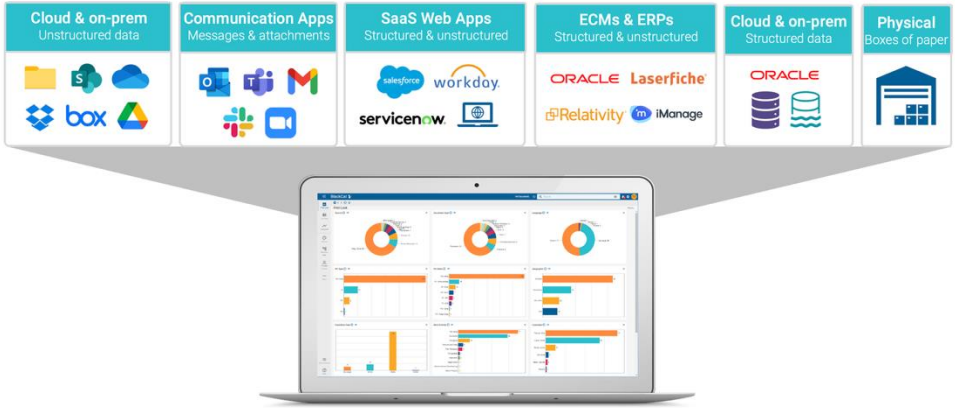
AI: What Happens when
the Model Collapses

10 Ways AutoClassification
Can Impact AI
May 13

AutoClassification & AI
Trends in Enterprise

About Valora

- AutoClassification platform for Information Governance
- Automates discovery, identification, classification & defensible disposition
- Trusted by IG teams, Records & IM teams, Legal, Compliance, IT teams
- Brings multiple enterprise repositories into a single view

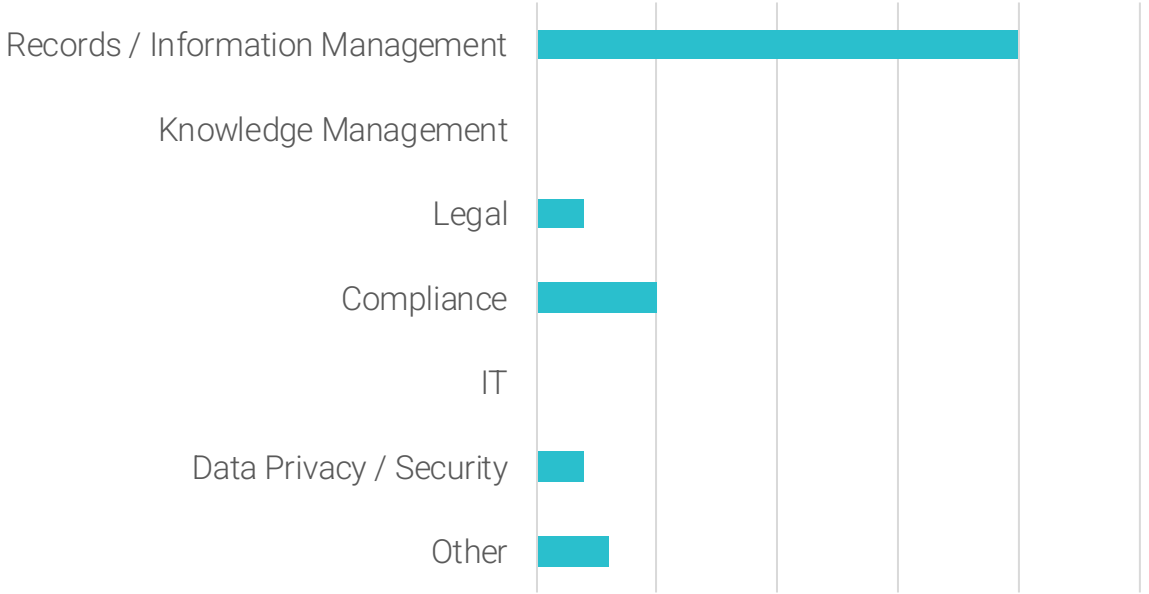


Poll Results

Who’s here today
& what are your interests?

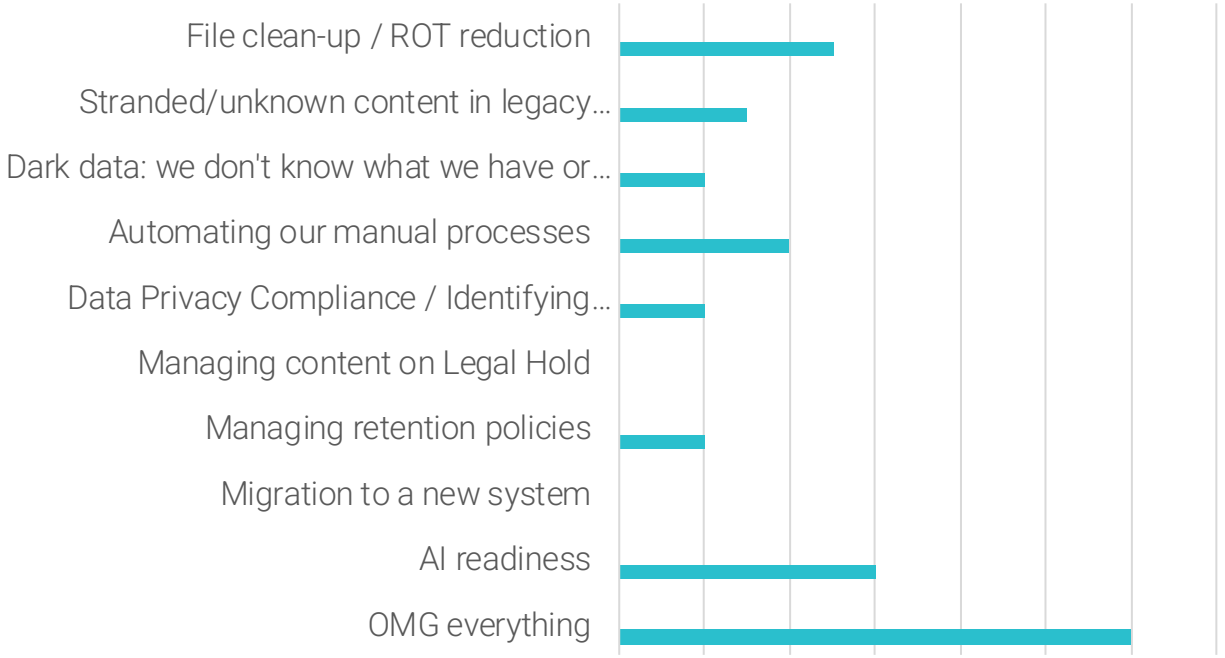
Who you are & what your challenges are

Who You Are



Who you are & what your challenges are

Your RIM & IG Challenges



Housekeeping

Q&A

Recording

Slides

Feedback

Types of AI

Transparent & defensible →

Machine Learning

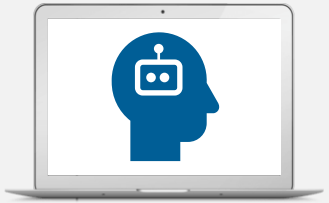
Algorithms that enable computers to learn from and make predictions or decisions based on data.

Extractive AI

Extracts specific information from large datasets or text.

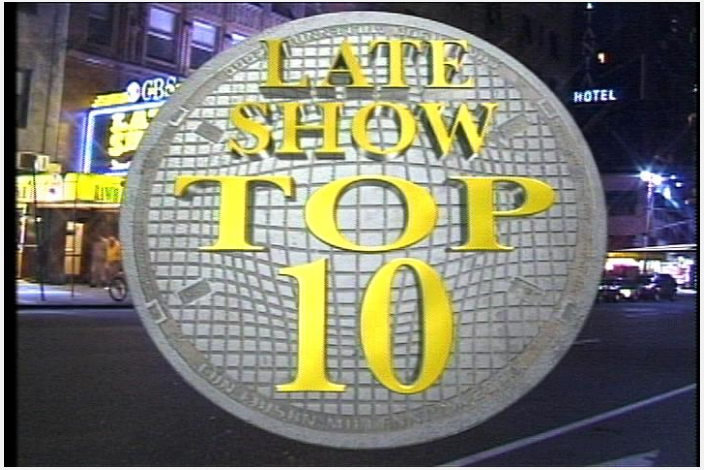
Generative AI

Creates new content on the data it has been trained on.



Agentic AI

Operates independently, making decisions & performing tasks without human intervention.



Agenda

- 10 Practical Ways AutoClassification Impacts AI

1 Prevents AI from Accessing Sensitive Data

Why it Matters

AI cannot distinguish sensitive content from non-sensitive content *unless the organization can*.

How AutoClassification Helps

Identifies PII, PHI, legal, financial & other confidential content

Allows organization to:

- Restrict what AI tools can access
- Apply policy-based guardrails
- Prevent accidental data leakage

Examples

LLM deployment excludes documents labelled:

- “Attorney-Client Privileged”
- “Confidential HR”
- “Patient Data”

✅ **Result :** AI becomes usable without exposing risk

2 Improves Retrieval Accuracy for RAG & Enterprise Search

Why it Matters

Retrieval-Augmented Generation (RAG) is only effective if the *right content is retrieved*.

What is RAG?

RAG = Open Book Exam

Giving AI the "right data" to source answers from, not just what it was trained on.

- Retrieve, then generate
- Grounding responses in data
- Source attribution

Common use cases

- Summarize this document
- What does our policy/manual/SOP say about xyz?
- Find me everything related to xyz
- Helpdesk bots

2 Improves Retrieval Accuracy for RAG & Enterprise Search

Why it Matters

Retrieval-Augmented Generation (RAG) is only effective if the *right content is retrieved.*

How AutoClassification Helps

Enriches docs with metadata:

- Document Type
- Topic
- Record Type
- Business process
- Jurisdiction

Improves:

- Search precision
- Semantic ranking
- Context relevance
- Response quality

Example

Instead of retrieving every “contract”, AI retrieves:

- Active vendor contracts
- Based in North America
- Executed in the last 24 months

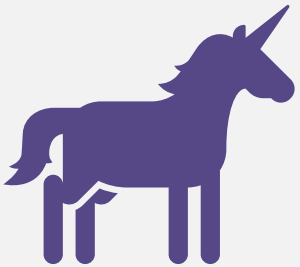
✅ **Result :** More accurate AI answers with less hallucinations.

3 Reduces AI Hallucinations

Why it Matters

LLMs hallucinate when they retrieve *low-quality, outdated, or irrelevant* content.

What is an AI hallucination?



Where an AI (LLM) generates *confident*, *realistic*, and *coherent* content that is factually *incorrect*, *illogical*, or *ungrounded*.

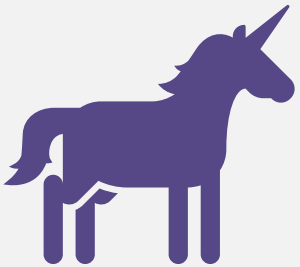
It is not a "bug" but a result of AI predicting plausible patterns rather than verifying facts.

3 Reduces AI Hallucinations

Why it Matters

LLMs hallucinate when they retrieve *low-quality, outdated, or irrelevant* content.

What is an AI hallucination?



Convincing misinformation: looks perfectly credible, makes it hard to distinguish from correct information (fake citations, fabricated legal cases, non-existent events)

Root Cause: LLMs are trained to predict the most statistically likely next word is a sequence, not to maintain a factual database of realty.

Trigger Factors: often occur when the necessary information is *missing, ambiguous* or *outside the training data.*

3 Reduces AI Hallucinations

Why it Matters

LLMs hallucinate when they retrieve *low-quality, outdated, or irrelevant* content.

How AutoClassification Helps

AutoClassification helps AI prioritize:

- Authoritative content
- Final versions (vs drafts)
- Approved policies
- Current procedures
- Trusted repositories

Example

AI references:

- Final policy document instead of Old drafts in shared drives

✅ **Result :** Higher confidence & trust in AI outputs.

4 Enable Policy-Based AI Governance at Scale

Why it Matters

Manual governance *does not scale* across millions of files.

How AutoClassification Helps

AutoClassification operationalizes governance automatically:

- Retention
- Privacy enforcement
- Legal hold
- Geographic & regulatory restrictions

Example

Documents classified as:

- "EU Personal Data" automatically become:
- *Restricted* from non-EU AI processing

✅ **Result :** Governance becomes enforceable instead of aspirational.

5 Reduces Bias in AI Training Data

Why it Matters

Poorly governed data introduces bias into AI models.

What is bias in AI?

Refers to systematic, *unfair outcomes* produced by AI systems that privilege certain groups over others, often *reinforcing existing human prejudices* or *societal inequalities*.

It occurs when AI models are trained on flawed, *unrepresentative data* or designed with skewed assumptions, leading to skewed predictions in areas like hiring and lending.

5 Reduces Bias in AI Training Data

Why it Matters

Poorly governed data introduces bias into AI models.

What causes bias in AI?

Data Bias (unrepresentative data): AI trend on data that does not accurately reflect the target population.

Historical or societal bias: models learning from historical data that contain human prejudice.

Measurement bias: when data collection methods in accurately measure “something” leading to distorted, non-neutral data sets interaction.

5 Reduces Bias in AI Training Data

Why it Matters

Poorly governed data introduces bias into AI models.

What are some examples of bias in AI?

Hiring algorithms *(gender bias):* Preferencing male candidates over female candidates for technical roles based on skewed historical hiring data.

Healthcare algorithms *(socioeconomic/racial bias):* an algorithm used to manage patient care for a major health system prematurely recommended ending coverage for elderly patients based on flawed metrics.

Image generation *(stereotyping):* reproducing gender and racial stereotypes in image generation (ex. associating “nurse” with women, or “entrepreneur” with white men)

5 Reduces Bias in AI Training Data

Why it Matters

Poorly governed data introduces bias into AI models.

What are the consequences of bias in AI?

Discrimination: Perpetuating & amplifying discrimination against marginalized groups.

Reputational/Legal Risk: Organizations may face lawsuits and loss of trust if their AI systems exhibit disparate impacts on protected classes.

Amplification: Unchecked AI can create a feedback loop that makes humans more biased themselves.

5 Reduces Bias in AI Training Data

Why it Matters

Poorly governed data introduces bias into AI models.

How AutoClassification Helps

AutoClassification can identify:

- Redundant data
- Skewed datasets
- Sensitive demographic information
- *Inappropriate* training content (old/now inappropriate data, terms)

Outcomes:

- Curates balanced training sets
- Supports fairness & responsible AI initiatives
- Reduces exposure to discriminatory outcomes

✅ **Result :** Cleaner datasets & reduced algorithmic bias.

6 Makes AI Outputs More Defensible

Why it Matters

Organizations increasingly need to explain *why* AI produced an answer, *what* data it used, & whether it complied with policy.

How AutoClassification Helps

AutoClassification provides:

- Traceability
- Data provenance
- Contextual metadata
- Auditability

Outcomes

- Enables explainability and auditability
- Connects AI outputs to authoritative sources

✅ **Result :** Greater legal defensibility & audit readiness.

7 Reduces AI Costs by Eliminating ROT

Why it Matters

AI processing massive amounts of redundant data is expensive.

Why is AI processing ROT so expensive?

"Tokens" are the currency AI runs on.

Every interaction consumes a token: ex. input / output

- 1 token ≈ 4 characters

Model / Platform	Context Window (Tokens)	Rough Equivalent
OpenAI (ChatGPT – GPT-4o / 4.1 class)	128,000 tokens	~300 pages
Anthropic Claude 3 (Opus/Sonnet/Haiku)	200,000 tokens	~500 pages
Google Gemini (1.5 Pro)	Up to ~1M tokens	~2,500+ pages
Google Gemini (1.5 Flash)	~1M tokens (optimized)	~2,500+ pages
OpenAI (older GPT-4 Turbo)	128,000 tokens	~300 pages

7 Reduces AI Costs by Eliminating ROT

Why it Matters

AI processing massive amounts of redundant data is expensive.

Why is AI processing ROT so expensive?

Hidden Multipliers (ROT):

- Same data, stored differently
- Duplication (Redundancy)
- Repeated efforts
- Re-indexing

Messy data = more complexity = more tokens = *more cost & worse results*

- If ROT is 50%, *AI costs are 50% + inflated*

7 Reduces AI Costs by Eliminating ROT

Why it Matters

AI processing massive amounts of redundant data is expensive.

How AutoClassification Helps

AutoClassification identifies:

- Redundant
- Obsolete
- Trivial data

Outcomes:

- Archive or delete unnecessary content (gasp!)
- At a minimum, not feed it to your AI

✅ **Result :** Lower AI infrastructure costs & faster (correct) retrieval.

8 Enables Context-Aware AI Experiences

Why it Matters

AI becomes more useful when it understands business context.

How AutoClassification Helps

AutoClassification enriches content with business meaning, identifies:

- Project
- Customer – Active / Inactive
- Product
- Jurisdiction / Region / Office
- Process / Lifecycle stage

Outcomes:

- Adds business meaning to enterprise content
- Improves personalization & recommendations
- Enhances AI understanding of business workflows

✅ **Result :** Smarter recommendations & more personalized AI responses.

9 Accelerates AI Readiness & Deployment

Why it Matters

Most AI projects stall because organizations don't trust their data estate.

How AutoClassification Helps

AutoClassification helps organizations:

- Inventory content
- Understand, then mitigate risk
- Identify sensitive data
- Identify corporate data appropriate for training AI model

Outcomes:

- Supports remediation before AI rollout
- Confidence in the data its training on
- Accelerates secure AI adoption

✅ **Result :** Faster AI roll-out with fewer security concerns & objections.

10 Builds Trust in Enterprise AI

Why it Matters

AI adoption ultimately depends on trust.

How AutoClassification Helps

Employees are more likely to use AI when they know:

- (their) sensitive content is protected
- Governance is enforced
- Results come from trusted sources
- Compliance obligations are respected

Example:

Employees confidently use internal AI because:

- HR records are excluded
- Legal data is protected
- Responses cite approved sources

✅ **Result :** Higher adoption & better business outcomes.

11 Enables AI Recovery & Rebuild Readiness

Why it Matters

You have a way back if you need to retrain, rebuild, or replace AI systems after model collapse.

What is an AI model collapse?

Model collapse refers to the *degradation of AI quality* over time when models increasingly train on:

- synthetic AI-generated content
- low-quality data
- duplicated information
- biased outputs
- hallucinated content

11 Enables AI Recovery & Rebuild Readiness

Why it Matters

You have a way back if you need to retrain, rebuild, or replace AI systems after model collapse.

How AutoClassification Helps

AutoClassification preserves:

- Sensitivity labels
- Business context
- Record types
- Retention categories
- Regulatory context
- *Identifies AI-generated content*

Outcome:

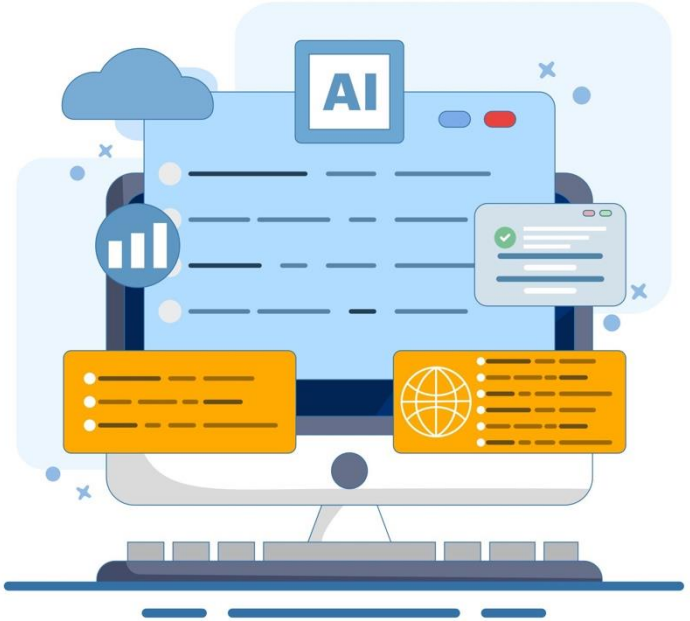
AutoClassification becomes "institutional memory":

- GenAI vendors will change
- Models will fail
- Architectures will evolve

The classified metadata layer remains reusable

✅ **Result :** When AI systems fail, governed data becomes the recovery foundation.

Key Takeaways



- AI is only as good as the data it can access
- AutoClassification is becoming foundational to AI governance
- Classification improves trust, security, accuracy, and defensibility
- Organizations with governed data deploy AI faster, safer, and cheaper

Valora Technologies

www.valoratech.com

Book a Demo

Webinar Series

/valora-technologies

Q&A

Thank you

Sandy Serkes
President & CEO
sserkes@valoratech.com

Jennifer Nelson
VP Strategic Solutions
jnelson@valoratech.com