



— WEBINAR SERIES —

Rich Lauwers
IG Program Manager

Jennifer Nelson
VP Strategic Solutions

Letting the Robots Classify: Automating IG with Confidence



Webinar Series



www.ValoraTech.com



ROT Remediation:
File Share Clean-up
January 21



How AC Transforms
Enterprise Search
June 17



Letting the Robots
Classify: Automating IG
February 18



Managing Data Privacy &
While Managing Records
September 16



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



Demo Day
October 14



Structured Data
Management & Disposition
April 15



TBD
November 12



10 Ways AutoClassification
Can Impact AI
May 20



AutoClassification & AI
Trends in Enterprise
December 9

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

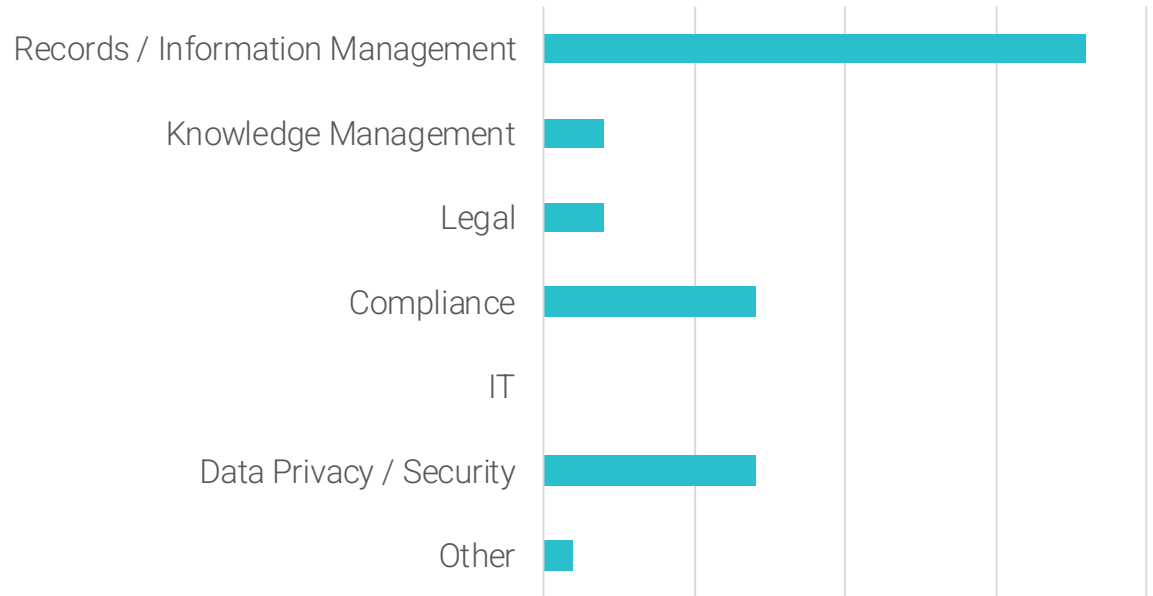
Who's here today
& what are your interests?



Who you are & what your challenges are



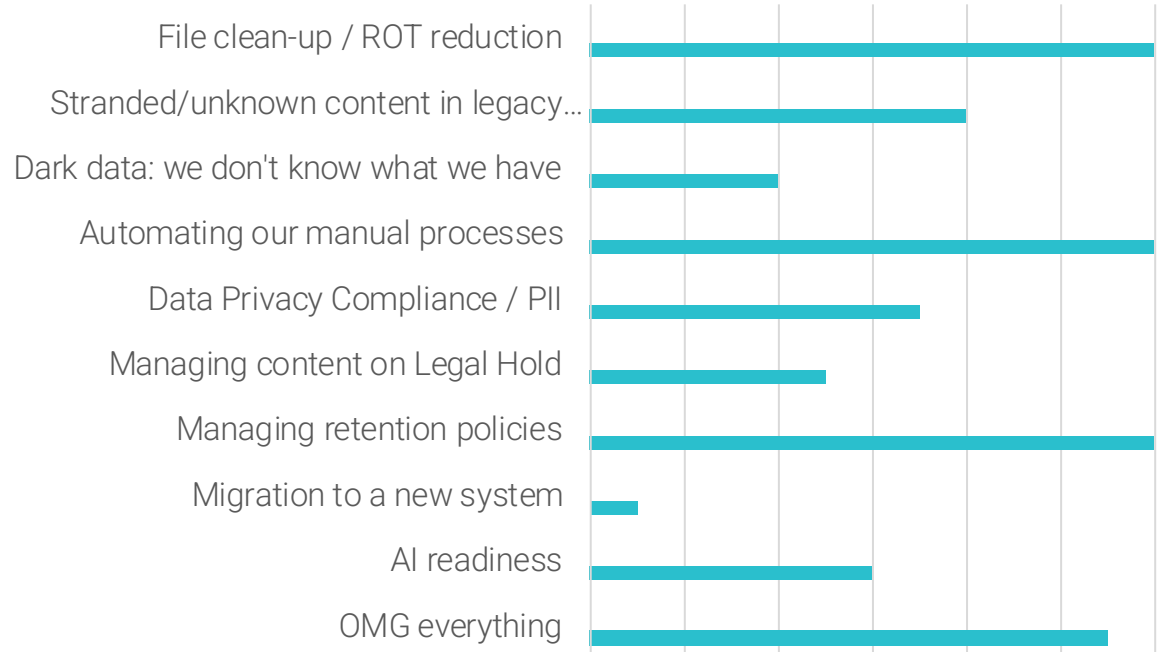
Who You Are



Who you are & what your challenges are



Your RIM & IG Challenges



Housekeeping



Q&A



Recording

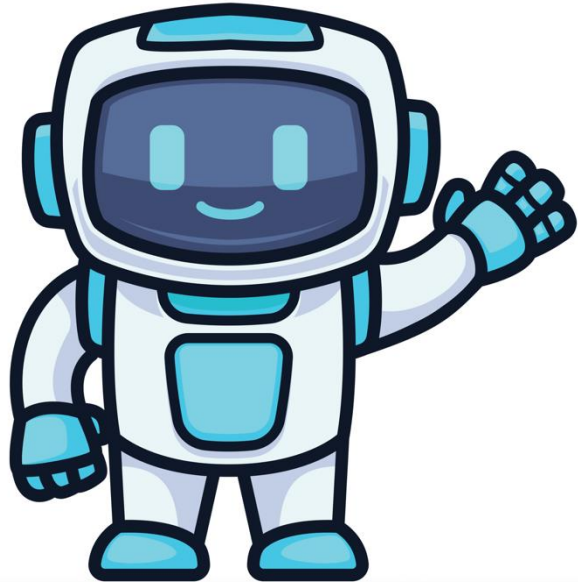


Slides



Feedback

Agenda



- Data Sprawl & Risk
- Why Now for AutoClassification?
- AI & AutoClassification
- AutoClassification Use Cases
- Case Study: real-world example
- How to get started with AutoClassification

Evolution: From Records Management

Information Governance / Records Management



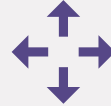
Classification



ROT



Retention

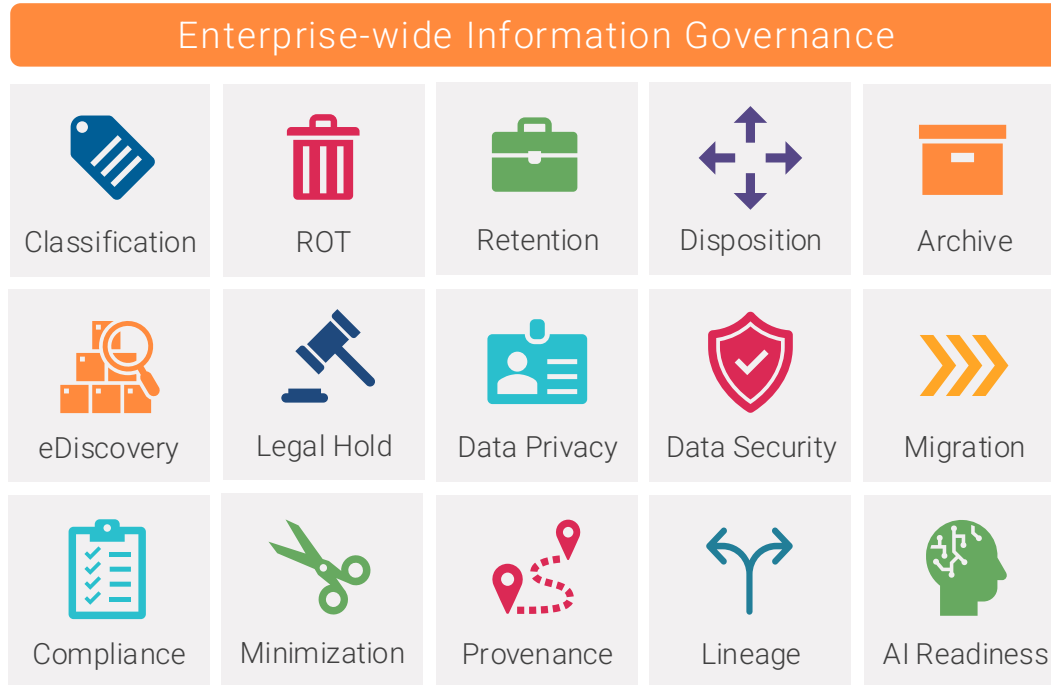


Disposition

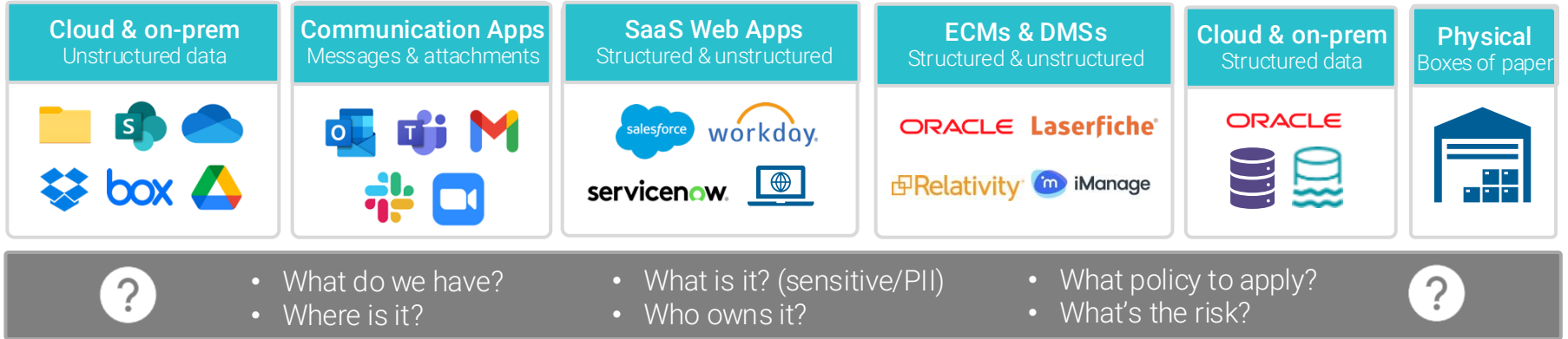


Archive

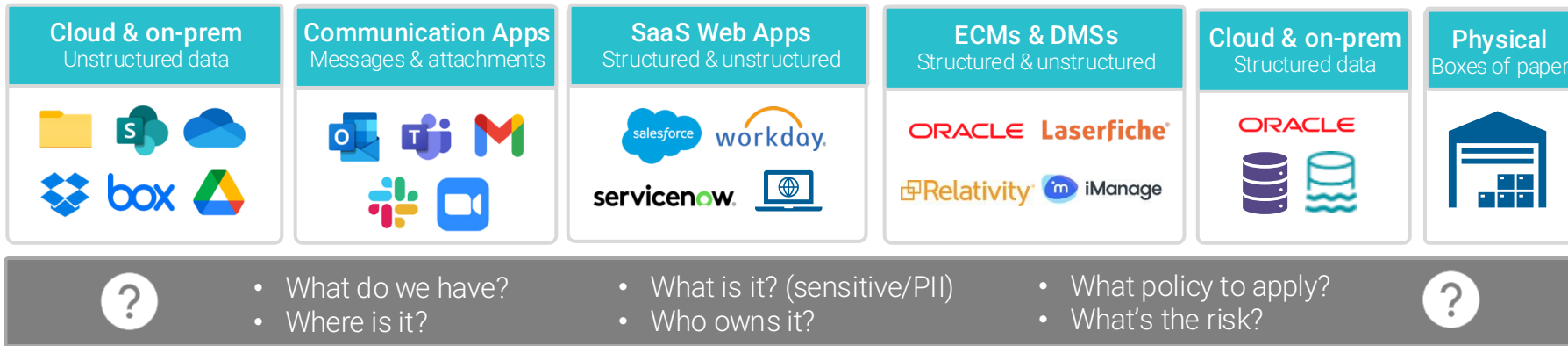
Evolution: to Enterprise-wide Information Governance



Challenge: Data Sprawl

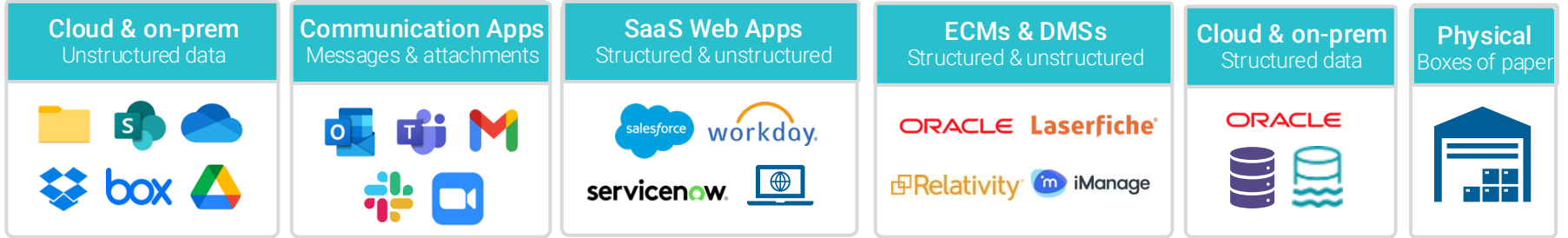


Challenge: Data Sprawl



- ~50%+ of content is ROT
- Regulatory fines
- Litigation exposure
- Operational inefficiency
- AI model training

Why Not Humans?



- | | | | | |
|--|---|--|--|--|
| | <ul style="list-style-type: none">• What do we have?• Where is it? | <ul style="list-style-type: none">• What is it? (sensitive/PII)• Who owns it? | <ul style="list-style-type: none">• Where are the duplicates?• What's the risk? | |
|--|---|--|--|--|



Human Error

Inconsistency, interpretations/
reluctance, "not my job"

Scalability Issues

Impossible to keep up with
volume of data

Compliance Risk

Regulatory non-compliance,
retention & deletion errors

Limited Auditability


Lack of transparency, why
things were categorized

Time Consuming & Costly

Really costly....

Why Not Humans?

Cloud & on-prem
Unstructured data




Communication Apps
Messages & attachments



SaaS Web Apps
Structured & unstructured



ECMs
Structured & unstructured



- What do we have?
- Where is it?

- What is it? (sensitive/PII)
- Who owns it?



Human Error

Inconsistency, interpretations/
reluctance, "not my job"

Scalability Issues

Impossible to keep up with
volume of data

Compliance Risk

Regulatory non-compliance,
retention & deletion errors

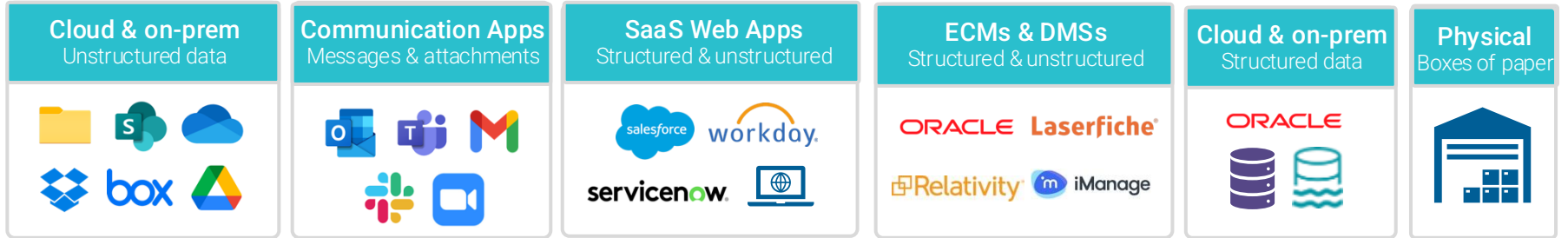
Limited Auditability

Lack of transparency, why
things were categorized

Time Consuming & Costly

- 30 seconds = employee to manually classify 1 doc
- 1M docs
 - = 30M seconds / 8000 hrs ÷ 2,080 work hrs/year
 - = 3.8 employee-years
 - = 1 employee manually classify 264k docs/year
- Average employee in a business:
 - = sends/receives 100 emails, works on 10 docs/day
- 1000-person organization
 - = 110,000 doc/day
 - = 1.32M docs/year
- Been in business 20 years
 - = backlog of 438M docs
 - = 1,664 employee-years to get through the backlog
- Get it done over 5 years = 87M docs/year
 - = 264k docs/year/person
 - = require a team of 329 humans
- US @ \$65k/year average salary
 - = \$21.3M in wages per year
 - = \$106.9M in wages over 5 years
- Overseas @ \$5k/year average salary
 - = \$1.65M in wages per year
 - = \$8.25M over 5 years
- That's just to get through the backlog

Technology really is the only way



- What do we have?
- Where is it?

- What is it? (sensitive/PII)
- Who owns it?

- Where are the duplicates?
- What's the risk?



Reduces Human Error

eliminates the possibility of human error or oversight

Scalability Issues

process & classifies 100,000s docs per day

Reduces Costs / Reduces Time

- Reduce ROT / reduces hosting costs
storage costs
back-up costs
- Reduces manual resources time
- Reduces manual searches: data requests
DSAR
eDiscovery

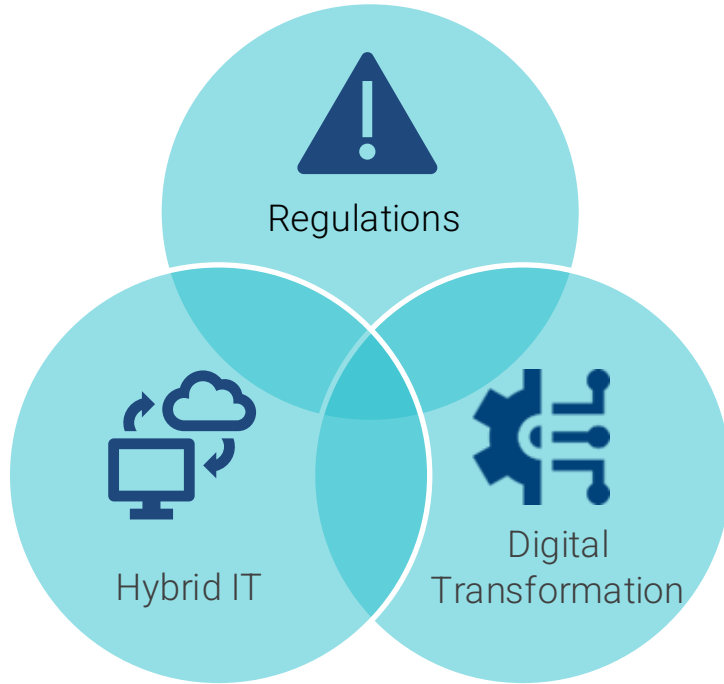
Reduces Risk

Confidence in your data & processes, ready for audits

Audited Decision Making

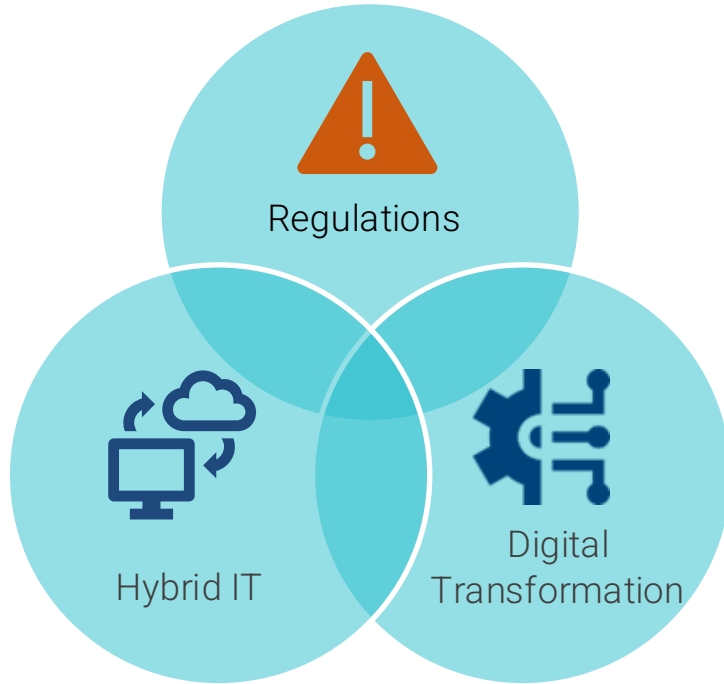
Transparent confidence scoring and decisioning

Why Now?



- New compliance mandates (privacy, DOJ DSP, industry regulations)
- Shift to hybrid cloud environments
- Need for operational efficiency

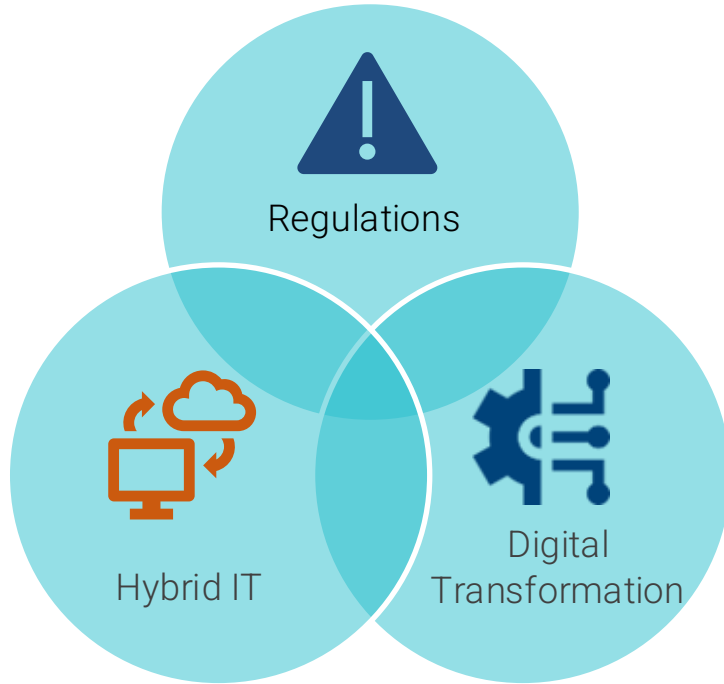
Why Now?



Regulations & Compliance Pressure

- **Growing complexity:**
Evolving regulations — privacy laws (GDPR, CCPA, CPRA), industry mandates (HIPAA, DOJ’s DSP), and sector-specific rules
- **Audit readiness:**
Regulators increasingly demand demonstrable, defensible controls over data. Manual methods no longer withstand scrutiny.
- **Risk of non-compliance:**
Financial penalties, reputational damage, and litigation exposure make compliance not optional but mission-critical.

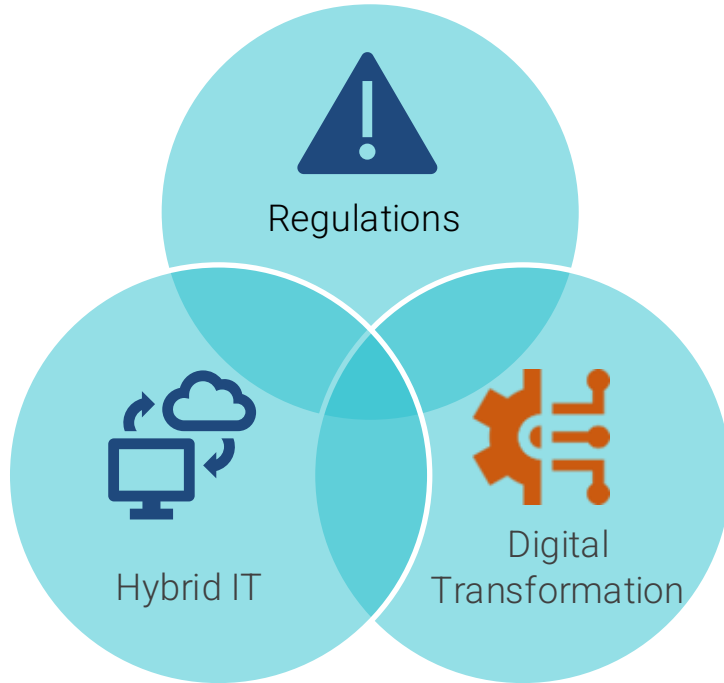
Why Now?



Hybrid IT Environments

- **Multiple data silos:** Enterprise content lives across on-prem systems, cloud apps, collaboration tools, and legacy repositories.
- **In-place management challenge:** Moving or consolidating data is costly and risky, but leaving it unmanaged creates compliance blind spots.
- **Need for platform-agnostic governance:** IG teams require solutions that enforce policies across all systems without rip-and-replace.

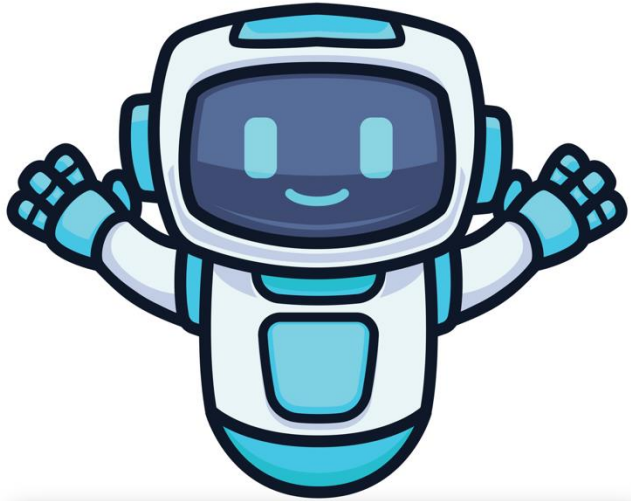
Why Now?



Digital Transformation & Operational Efficiency

- **Data as an asset/ business enabler:**
Beyond risk management, data is an asset for analytics, AI, and automation — but only if it's clean, classified, and governed.
- **Workforce productivity:**
Automating classification and retention frees up employees from manual, error-prone tasks.
- **Cost pressures:**
Organizations can't afford ballooning storage and eDiscovery costs caused by unmanaged ROT data. HR: "do more with less"

Is AutoClassification AI?



Machine Learning

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



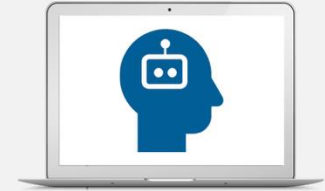
Extractive AI

Extracts specific information from large datasets or text.



Generative AI

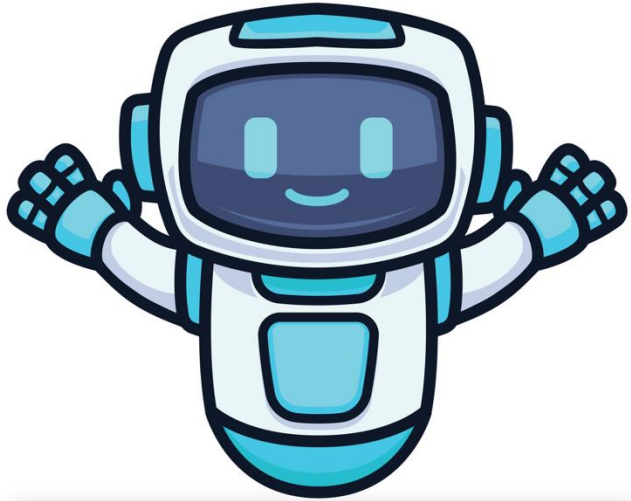
Creates new content on the data it has been trained on.



Agentic AI

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

Is AutoClassification AI?



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

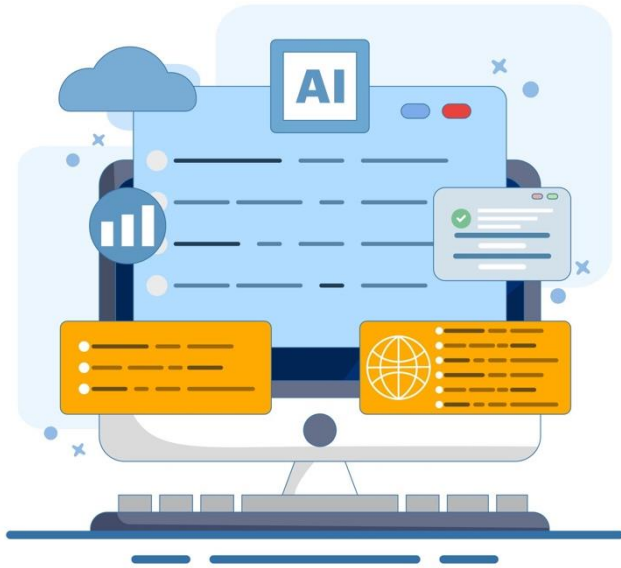
Creates new content on the data it has been trained on.



Agentic AI

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

The Role of AutoClassification & AI



- Use AI to **automatically analyze and categorize** data based on content, context, and metadata
- Create a "single pane of glass," allowing users to search, view, and **manage structured and unstructured data in one place.**
- AI **enhances searchability** and ensures *appropriate, permissioned & relevant* data is surfaced, regardless of its original format or source.

AutoClassification: How it Works



File & Rich Metadata
Determine Attributes

+



Recognition Algorithms
Determine DocType

+



Disposition Algorithms
Rules & Actions

AutoClassification = Rich Metadata + Rules

Software that performs automated analysis & disposition of file/document content



Answers the question:

What is this thing & what do I do with it?

...and then does it.

File attributes become rich metadata



File & Rich Metadata
Determine Attributes

+

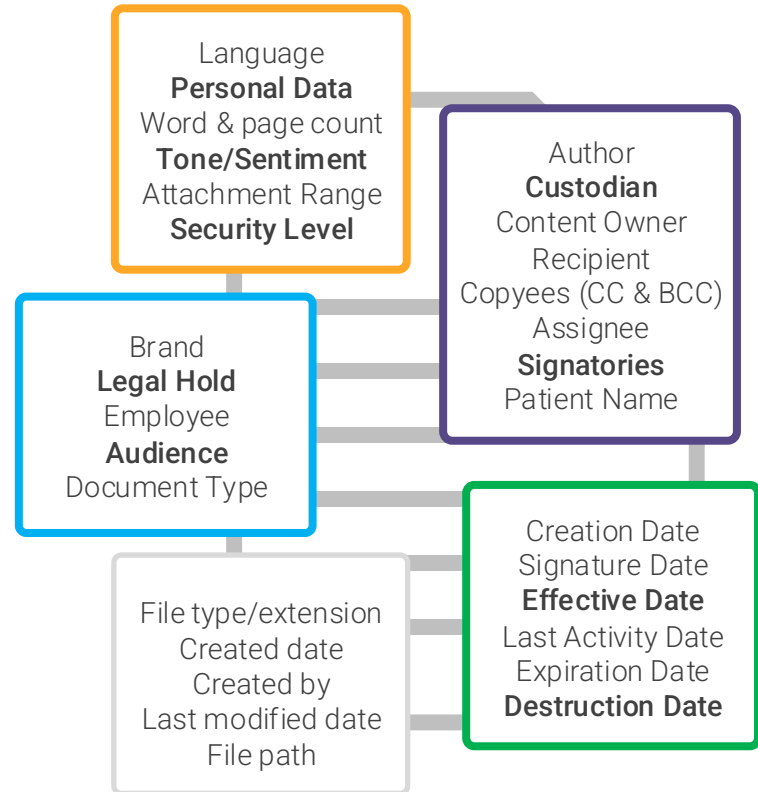


Recognition Algorithms
Determine DocType

+



Disposition Algorithms
Rules & Actions



Determines DocType as defined by RRS



File & Rich Metadata
Determine Attributes

+

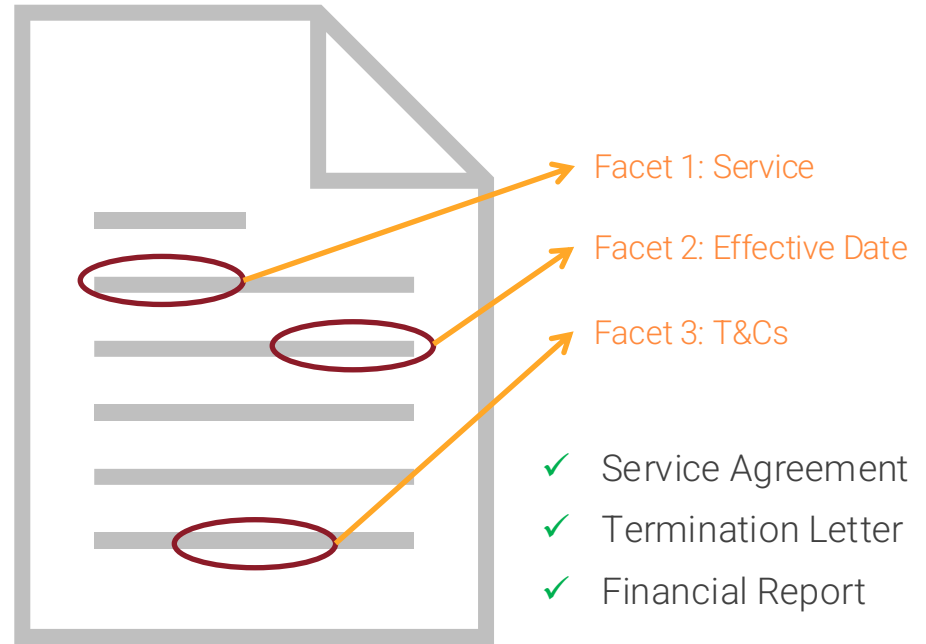


Recognition Algorithms
Determine DocType

+



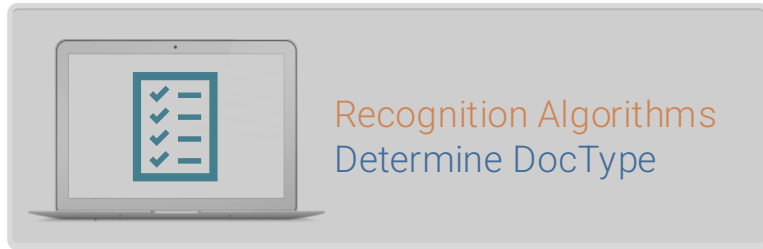
Disposition Algorithms
Rules & Actions



Defensible Disposition



+



+



- Determine the disposition & handling of content
- Typically follow an IF-THEN format, often nested
- Ex: Employment Contract:
 - ✓ Mark as Confidential
 - ✓ Move it to SharePoint
 - ✓ Retain for 8 years after termination
 - ✓ HR approval required before deletion

Automated, Manual and/or Hybrid Disposition



Tag



Delete



Report



Move



Archive



Hold




Migrate



Anonymize

AutoClassifying Documents

Name ^	Date modified	Type	Size
 LGPIF Final.pdf	8/13/2016 9:26 AM	Adobe Acrobat Document	171 KB

AutoClassifying a document

🔥 PHARMACEUTICAL

Informed Consent Form

Title of Study: Phase II Randomized Controlled Trial of NeuroCure-X in Mild to Moderate Alzheimer's Disease

Principal Investigator: Dr. Jane Researcher

Introduction:

You are being invited to participate in a research study to evaluate the safety and efficacy of the investigational drug NeuroCure-X in treating mild to moderate Alzheimer's disease. Before you decide whether to participate, it is important that you understand the purpose, procedures, potential risks, and benefits of the study.

Purpose of the Study:

The purpose of this study is to determine whether NeuroCure-X is effective in improving cognitive function in individuals with mild to moderate Alzheimer's disease.

Procedures:

If you agree to participate, you will be randomly assigned to either receive NeuroCure-X or a placebo. You will be required to take the assigned treatment daily for a duration of 12 months. During this time, you will undergo regular cognitive assessments, medical evaluations, and safety monitoring.

Risks and Benefits:

Potential benefits may include improved cognitive function. Risks may include side effects associated with the study drug. Common side effects are whatever.

Confidentiality:

Your personal information will be kept confidential, and only authorized individuals involved in the study will have access to your records.

Voluntary Participation:

Participation in this study is entirely voluntary. You may withdraw at any time without penalty or loss of benefits.

Questions:

You are encouraged to ask any questions you may have before deciding to participate. Dr. Jane Researcher can be contacted at 736-831-9833 for any further information.

Consent:

I have read and understand the information provided in this Informed Consent Form. I have had the opportunity to ask questions, and my questions have been answered to my satisfaction. I voluntarily agree to participate in the study.

Participant's Name: Jennifer Nelson

Participant's Signature:  Date: Dec 14, 2023

Investigator's Signature:  Date: Jan 5, 2024

DocType = Informed Consent Form

Drug Name = NeuroCure-X

Trial = Phase 2 Randomized
Controlled Trial

Investigator = Dr. Jane Researcher

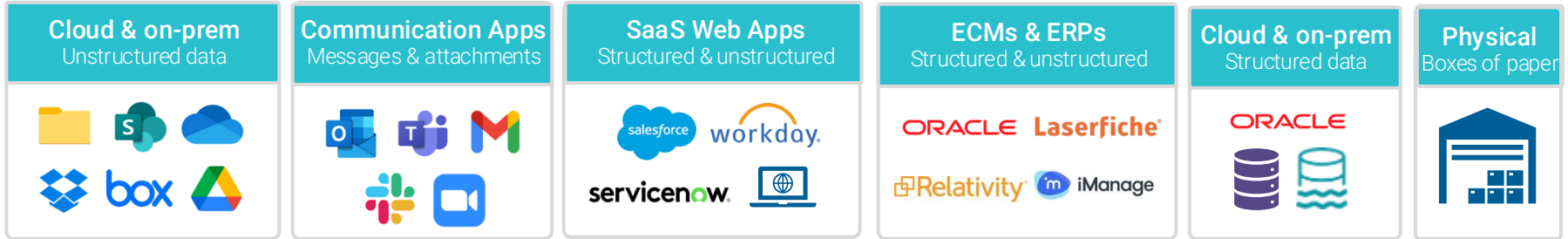
Participant = Jennifer Nelson

Date Executed = Dec 14, 2023
[Date format → 12/14/2023]

RecordType = Consent Form
RecordClass = Clinical Trial
Retention = Executed + 8Y
Sensitivity = Confidential
Handling = Delete after
authorization
AI Training = No

Retain.
Delete on Dec 14, 2031

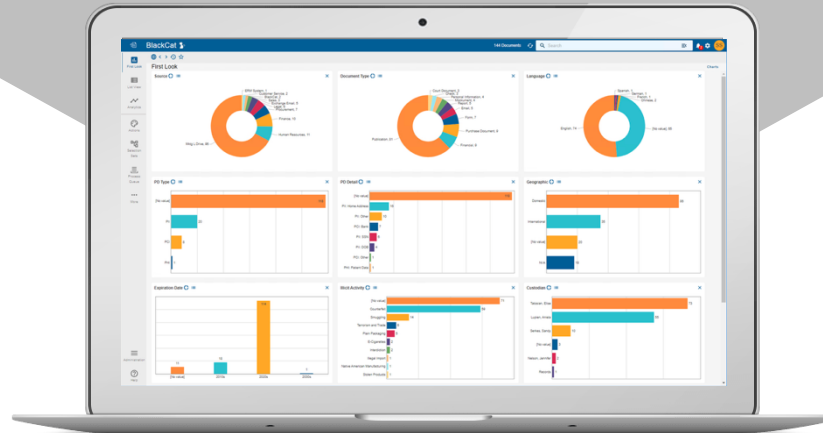
AutoClassifying at the Enterprise Level



- ✓ What do we have?
- ✓ Where is it?

- ✓ What is it? (sensitive/PII)
- ✓ Who owns it?

- ✓ What policy to apply?
- ✓ What's the risk?



How Organizations are using AutoClassification



Data Discovery & Classification

- Identify content: *where* it is and *what* it is to determine *how* and *when* it should be managed.



ROT & File Clean-up

- Identify, disposition, and defensibly delete Redundant, Obsolete and Trivial data across systems.



Records Management

- Automate and manage routine and specialized Records handling: expiration, destruction, migration.



Data Privacy & GRC

- Locate, identify, redact, manage Personal Data and other corporate sensitive content; DSAR mgmt.



Legal Hold/eDiscovery

- Reliable data collection for efficient and defensible Legal Hold & eDiscovery processes.



AI Readiness

- Reduce risk & maximize GenAI initiatives with clean, curated, and trustworthy AI-ready content.

Case Study: Challenges

ORACLE

- 25+ years
- 150 acquisitions
- Millions of inherited files
- 140,000+ employee records
- 10TB+ data



Data relevancy: what to keep, what to delete



People who know the data are no longer around



Out of compliance with their own retention policies

How to:

Execute on formal records retention policies?

Integrate with new systems of record?

Maintain & respect Legal Holds?

State of Affairs at the time



The HR M&A Universe

- Over 25 years of stored files (10 TB and growing)
- From hundreds of deals, many long-past
- Stored in separate folders in on-prem repository
- Files migrated "as is" from acquired companies



Its Impact

- No centralized content controls
- Inconsistent Attribute Tagging
- Searches dependent on dedicated staff and tribal knowledge
- Challenges enforcing retention policies
- High volumes of duplicate and expired records



Why Now?

- Mass migration of content to cloud
- Loss of tribal knowledge
- Compliance requirements
- Demand for faster & easier search results
- Set the stage for Content Lifecycle Management

Plan of Attack: Disparate Data



Digital Files

HR M&A Digital Personnel Documents + IPM and Legal Hold Personnel Records



Paper Files

Personnel files in Offsite Storage or Former HR Record Centers



Legacy Data

Legacy Oracle Personnel Data on Old PeopleSoft or Other HR ERP Servers

Goals

Liabilities & Disadvantages

- ✗ Dark data
(it's unknown)
- ✗ Disparate data
(it's everywhere)
- ✗ Duplicative data
(copies for convenience)

Control over distributed content

- ✓ Consolidate resources
- ✓ Consolidate investment
- ✓ Consolidate storage/SOR
- ✓ Single source of truth
- ✓ Compliance

Building the on-ramp



Planned the route



The right tools



Built the workflows



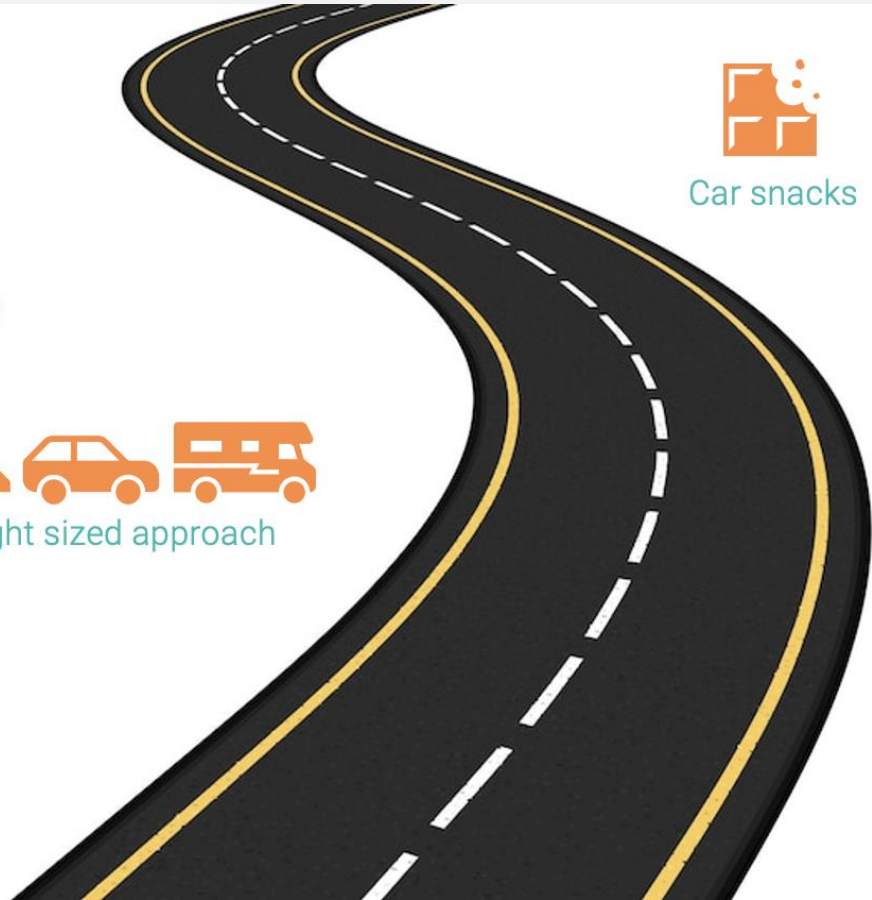
The right sized approach



Ready for hazards



Car snacks



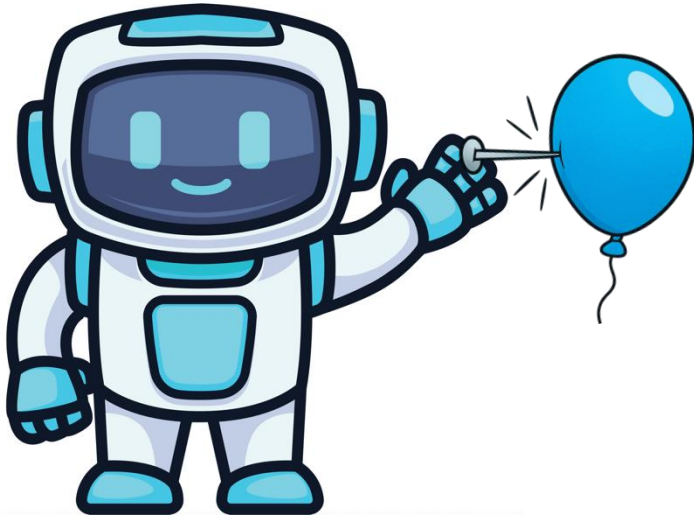
Results



Properly managed content.

- ✓ Reduced manual effort by **95%**
- ✓ Searches in seconds, not days = rapid response to stakeholders
- ✓ Enabled centralized compliance
- ✓ Build once, measure and promote the benefits, refine & repeat process with other content in the organization
- ✓ Team became company role model for global RIM compliance across the board

Reality Check



like other forms of AI
AutoClassification is not magic.

- Requires ongoing optimization, fine tuning
- Human in the loop
- Evolution of your organization
- Evolution of requirements
- Evolution of your data

Getting Started with AutoClassification

Pilot Use Case

Start with a compelling use case

- Quick and easy that will prove ROI quickly
- Highest risk?
- Prove ROI



Look Forward

Always have an eye on future use cases

- What do you need for this application or use case?
- What will you need for future use cases?



Measure Results

Benchmark results and accuracy of pilot

- Accuracy of results?
- Increase guidance or seek SME?
- What reporting was useful? What else is needed?



Build on Success

Build business case on pilot alone then expand

- Prove ROI functioning system
- Where to next?
- Additional use cases running on same platform



Spread the Word

Sharing is caring, expand internally

- Host internal showcases for mgmt & other departments
- Design pilots that highlight important content, results or exposure points



Final Thoughts



Efficient data governance & lifecycle management



Increased productivity & operational efficiency



Scalability & adaptability



Strike while the iron is hot



GenAI needs clean data



IG's time to shine



— WEBINAR SERIES —

Q&A / Thank you!



Sandra Serkes
Co-Founder & CEO

sserkes@valoratech.com



Jennifer Nelson
VP Strategic Solutions

jnelson@valoratech.com



www.ValoraTech.com



Book a Demo



Webinar Series



[/valora-technologies](https://www.linkedin.com/company/valora-technologies)