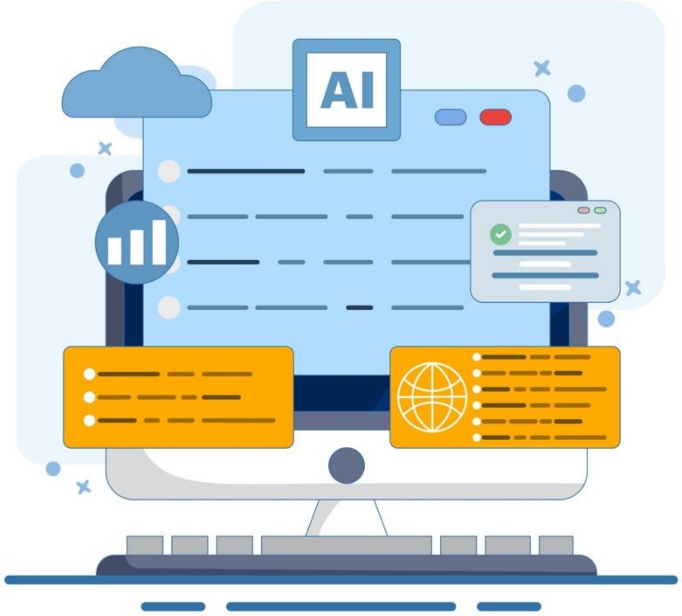
Valora Technologies

WEBINAR SERIES

The Role of Information Governance in AI Readiness

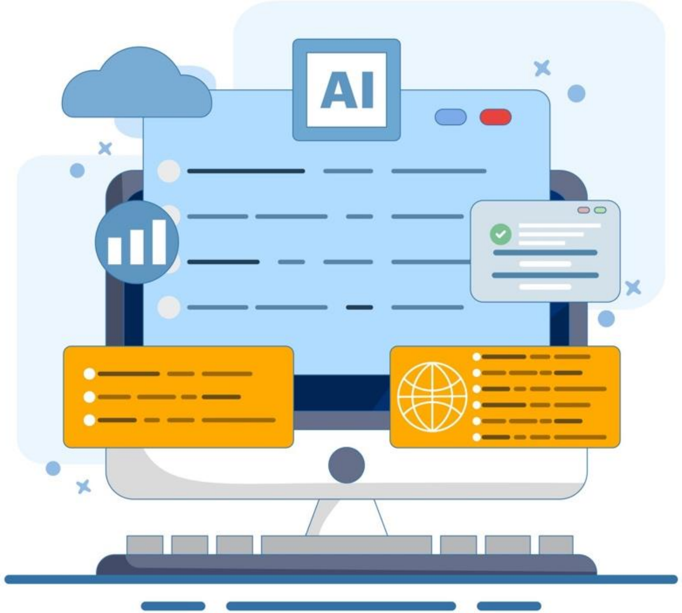


Valora Technologies

— WEBINAR SERIES —

Sandra Serkes
Co-Founder & CEO

Jennifer Nelson
VP Strategic Solutions



Webinar Series

ROT Remediation:
file share clean-up
February 12

Managing Data Privacy
while Managing Records
March 19

The Role of InfoGov
in AI Readiness
April 16

AI & AC for Structured
Data Management
May 14

AutoClassifying the 3 Rs:
ROT, Records & Retention
September 17

Getting Ready for
AutoClassification
October 15

Demo Day
November 12

www.ValoraTech.com

About Valora

- Information Governance platform
- 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

About Valora

Information Governance / Records Management

Classification

ROT

Retention

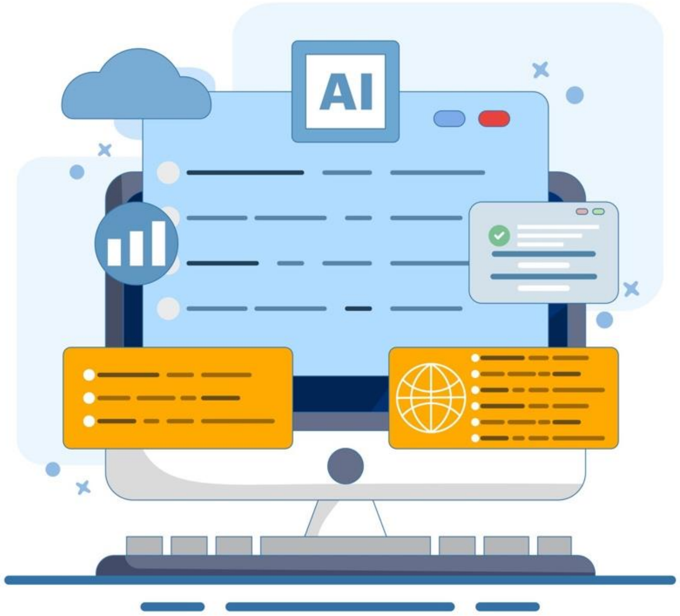
Disposition

Archive

About Valora

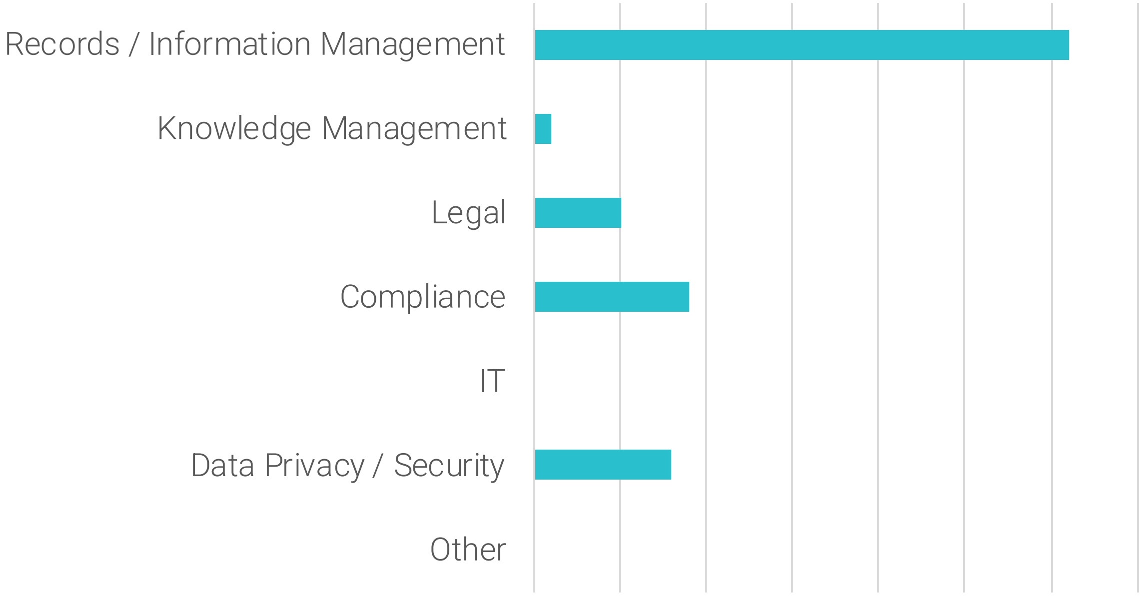
Poll Results

Who's here today
& what are your interests?



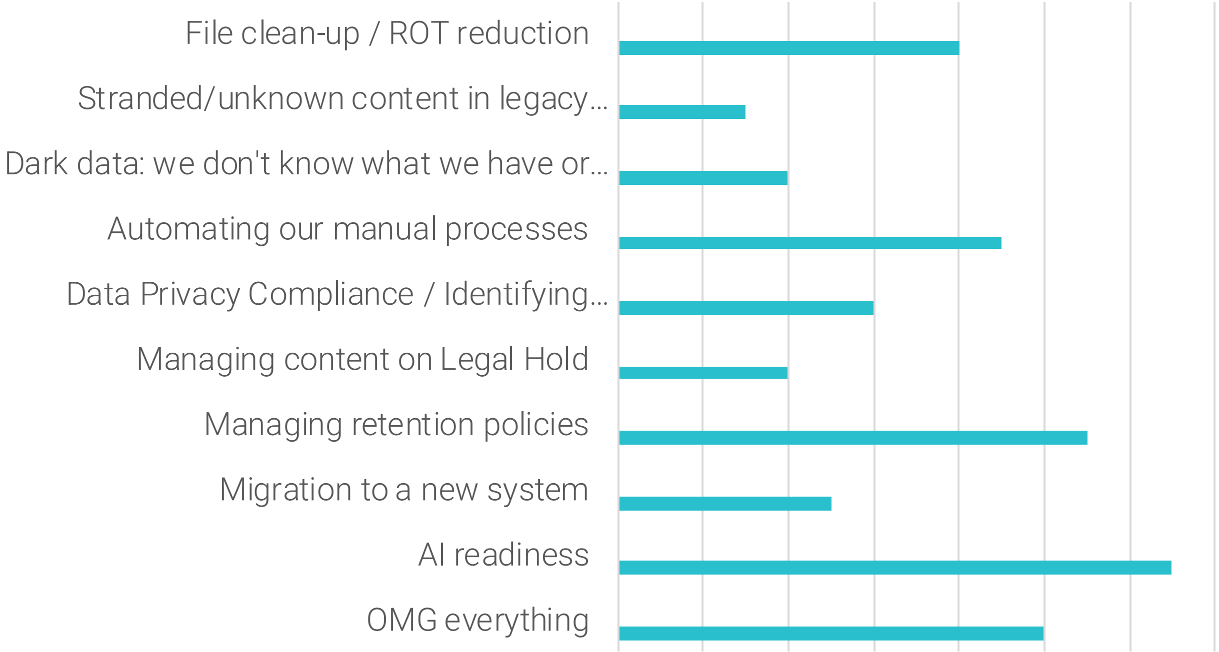
Who you are & what your challenges are

Who You Are



Who you are & what your challenges are

Your RIM & IG Challenges



Housekeeping

Q&A

Recording

Slides

Feedback

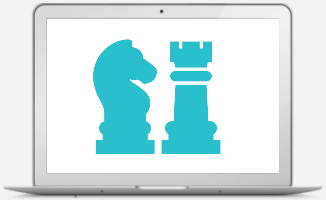
Agenda

- What is AI?
- What is AI Readiness?
- The connection between IG and AI readiness
- Challenges & Opportunities
- Things the IG team can do to promote good GenAI outcomes
- How to get started

What is AI?

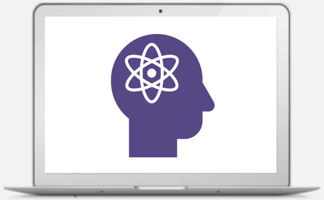
Definition:

Artificial intelligence refers to the simulation of human intelligence by machines, including learning, reasoning, and self-correction.



Early Use Cases

OCR, chess playing,
speech recognition, robotics



Neural Networks

Interactive
search & retrieval

1956 — 2025

Types of AI

Artificial Intelligence

Development of systems to perform tasks that typically require human intelligence.

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.

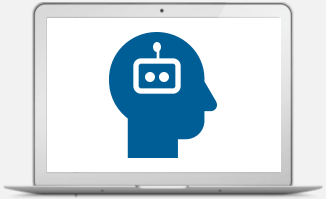
Types of AI

Artificial Inte

Development of s
perform tasks th
require human in

Learning

nable computers
make predictions
ased on data.

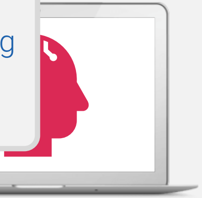


Agentic AI

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

Extractive AI

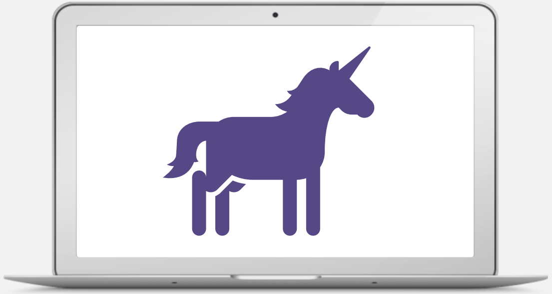
Extracts specific information from large datasets or text.



Generative AI

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

Gen AI
reality vs. fantasy



Trained on datasets

Learns the data to
generate new content

Not magical search engine

Creates new content based on
learned patterns

Hallucinations

Takes a best guess based on
training data, does not fact check

Not up to date

Only as accurate as its
last run of source data

Machine Learning / Extractive AI

- ✓ Auditable AI/ML algorithms
- ✓ Reuse & repurpose existing content
- ✓ Predictive analytics
- ✓ Consistent output

Generative AI

- ✓ Generates net new content
- ✓ Blackbox: can describe it but can't explain it
- ✓ Statistical amplifier
- ✓ Can't untrain/undo it
- ✓ Hard to govern

AI Governance Challenges

Real: is pulling from authentic content & you can prove it?

Reliable: is it producing reliable results?

Reversable: can you undo what it has done?

Recordable: can you maintain traceability and records?

What is
AI readiness?

The ability to integrate AI technologies into your organization effectively and responsibly.

Data quality
Clean, correct version, most recent, usable (formatted) data

Governance policies
Execution, corporate sensitive data, acceptable usage

Privacy compliance
Respecting data privacy regulations/redacting from LLMs

Business goals
Does this align with and support business goals?

The Enterprise Data Estate

4 key areas
are interconnected

A weakness in one
impacts the others

What GenAI needs to Operate Properly

Large bodies of data

Content to train on

- Unstructured & structured data
- Respecting content access/permissions
- A way to get at the data in multiple systems
- Content organized by authenticity & relevance
- Clean data (data hygiene)

Content that is unique

Specialized/proprietary

- Unique content has the highest value
- Creates the most competitive advantage
- Example: Google Maps

Deeper Dive on Access & Permissions

Understanding and respect of content access/permissions:

- ✓ Who is allowed to query (or use) this content?
- ✓ Under what conditions?
- ✓ How is its use being tracked and/or audited?
- ✓ Who or what is the gatekeeper?
- ✓ Restrictions based on the content itself vs. on the user identity/role.

GenAI can be easy to trick without restrictions.

Data Authenticity

Makes the data trustworthy

- Proper/appropriate disclosure (access) builds authenticity
- What should be included? (doc types)
- What should be excluded? (trade secrets)
- Identify redundant data (old versions)
- Identify personal data (patient/employee)
- Identifying unblinding data

Identity & Permission

Who is asking? What can they see?

- What is the sensitivity of the data?
- Should sensitive docs be tagged?
- Event-based trigger access (ex. DSAR = employee records)

Deeper Dive on Content Analysis

For GenAI to work,
the starter/training content must be:

- ✓ Indexed (tagged)
- ✓ Scored/sorted (relevance)
- ✓ Served (in a manner the GenAI expects), and
- ✓ Gated (permissioned)

Repeat for every repository

Repeat for every file

- Every file in every data repository needs to undergo these steps, or analysis
- AIIM estimated most large organization have 15-20 currently unconnected systems
- Baseline indexing vs. delta indexing
- Consistency is key

Deeper Dive on Content Analysis

AI and data go hand in hand.

AI *needs* data to learn and produce useful outputs.

The better the training data, the better the AI results.

Do you know what data you have?

- ✓ ROT? Duplicates? Past-retention?
- ✓ Obsolete? Dated/old? Incorrect? Drafts and interim versions?
- ✓ Sensitive? Inappropriate? Personal data?
- ✓ Protected? Under legal hold?
- ✓ Inaccessible? Acquired? Archived? Offsite?

AutoClassification
is your best friend here

- ✓ It will tell you what you have.
- ✓ AutoClassification sets up the AI LLMs for optimum performance

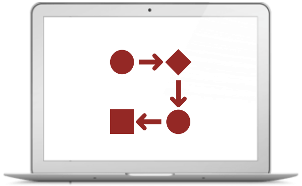
Content Analysis = AutoClassification

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

Language
Personal Data
Word & page count
Tone/Sentiment
Attachment Range
Security Level

Author
Custodian
Content Owner
Recipient
Copyees (CC & BCC)
Assignee
Signatories
Patient Name

Brand
Legal Hold
Employee
Audience
Document Type

Creation Date
Signature Date
Effective Date
Last Activity Date
Expiration Date
Destruction Date

File type/extension
Created date
Created by
Last modified date
File path

DocType as defined by RRS

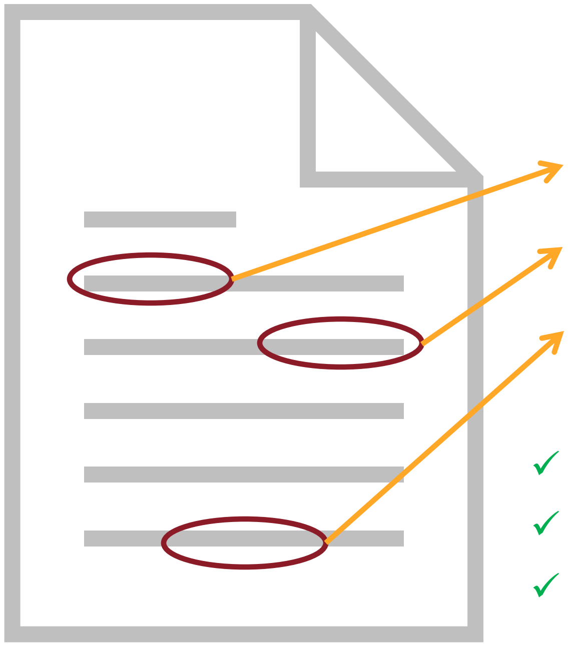
File & Rich Metadata
Determine Attributes

+

Recognition Algorithms
Determine DocType

+

Disposition Algorithms
Rules & Actions



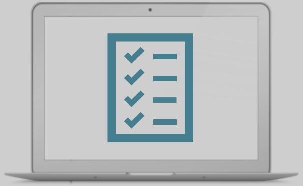
Facet 1: Service
Facet 2: Effective Date
Facet 3: T&Cs

- ✓ Service Agreement
- ✓ Termination Letter
- ✓ Financial Report

Defensible Disposition

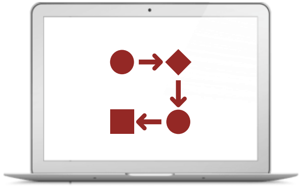
File & Rich Metadata
Attributes

+



Recognition Algorithms
Determine DocType

+



Disposition Algorithms
Rules & Actions

- 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

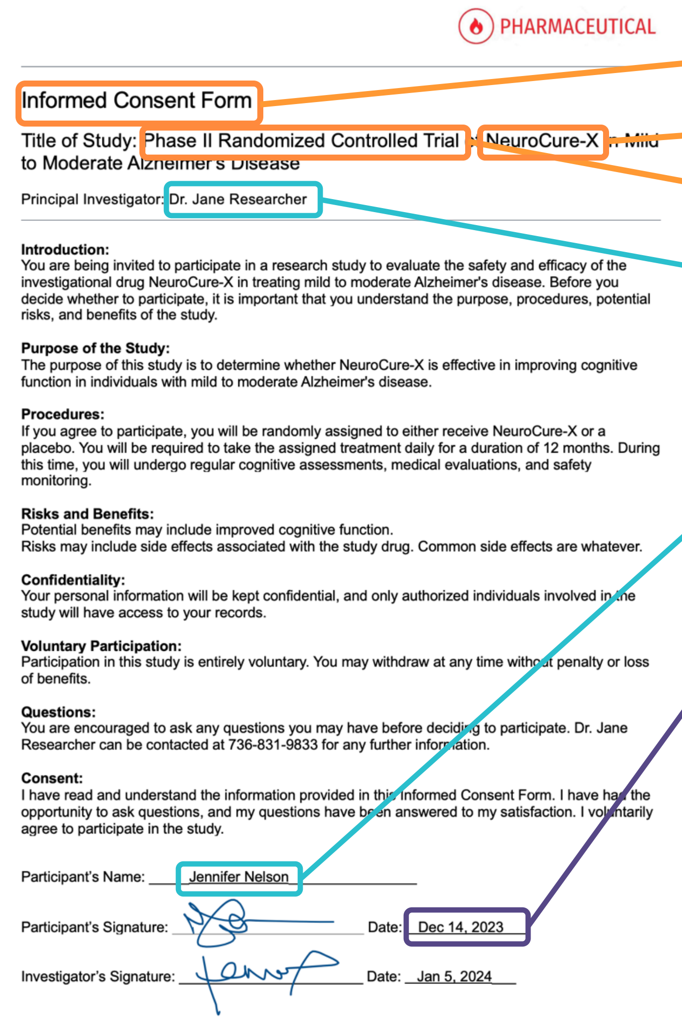
Automated, Manual and/or Hybrid Disposition

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 NeuroCure-X 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

Retain.
Delete on Dec 14, 2031

Reducing Risk & Exposure

- ✓ Uncover dark data
- ✓ Inform & action content
- ✓ Ensure compliance

Data Governance Risk

Where is it? What is it?
Is it being handled properly?

Privacy Risk

Does it contain personal
or sensitive data?

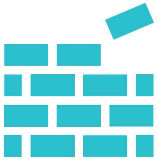
Compliance Risk

Onside with your own policies &
regulatory obligations?

AI Risk

Source data clean & trusted for
Generative & Extractive AI?

CIO: 12 famous AI disasters



Grok AI falsely accuses NBA star of vandalism

Ingested articles about Thompson "throwing bricks" (missed shot)

Chatbot makes up fare policies

Air Canada held liable, ordered to pay to what the chatbot said.

ChatGPT hallucinates court cases

Lawyer uses ChatGPT to reference 6 prior cases that did not exist.

https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html

Key Things the IG Team Can Do to Promote Good GenAI Outcomes

What is the impact of training the algorithms on wrong or old data?

Stress the need for good data hygiene

Remove ROT = remove low-value content that will screw up your LLM training

Make sure the GenAI algorithm matches your RIM policies

First instance of a dupe, last/most recent version, most trusted source?

Establish protocols for how data should be tagged & served to the GenAI engines

Taxonomies, field tables, RRS's are all good

Establish what GenAI results auditing must cover

What will you, as RIM/IG professionals, need to see to be convinced the "right things are happening"?

Key Things the IG Team Can Do to Promote Good GenAI Outcomes

Establish Robust Data Governance Frameworks

Clear data ownership, classification, and lifecycle management policies in place: access, retention, storage

Develop Clear Policies for Ethical AI Use

Create AI governance frameworks that address bias mitigation, transparency, and accountability

Address Data Privacy & Security Concerns

Update policies, ensure data used for AI training is anonymized, properly consented, track data usage.

Create Guidelines for Model Transparency and Explainability

Establish guidelines for explainability: how decisions or outputs are generated, data lineage, human-validated

Key Things the IG Team Can Do to Promote Good GenAI Outcomes

Prepare for Intellectual Property and Legal Challenges

Policies for clarifying ownership and use of outputs created by GenAI, licensing & copyright compliance

Train Employees on AI Literacy and Governance

Training on AI literacy, focusing GenAI use within organization, data governance policies, impact on roles

Monitor and Audit AI Outputs for Compliance and Quality

Update policies, ensure data used for AI training is anonymized, properly consented, track data usage.

Where you will likely get pushback

Balancing Innovation with Regulation

The fast pace of AI development can outstrip existing regulatory frameworks and governance structures, making it difficult for policymakers and end-users to keep up with innovations and potential risks. Beware of "move fast and break (or violate) things."

Ensuring Data Privacy and Security

AI systems rely on enormous datasets that could – purposely or accidentally – access sensitive or private data. As AI tools continue to evolve, so does the need for privacy and security systems, notably sensitive data tagging, that can keep up with that level of advancement.

Eliminating Bias

There are inherent biases in AI systems, but detecting and removing these is equally complex and time consuming. This work is required, however, as addressing bias in AI systems is the only way to work towards fairness and equity, which ultimately drives trust and acceptance.

Steps to establish AI readiness through IG

1. **Audit your data estate** – Identify gaps in the data quality & accessibility
2. **Align governance policies with AI goals** – Define processes for collecting, labeling, and securing data.
3. **Address compliance** – Ensure alignment with GDPR, CCPA, and other privacy regulations.
4. **Automate repetitive tasks** – Use tools to manage data lifecycle processes at scale
5. **Monitor & adapt** – Regularly review data policies to stay ahead of changing technology and regulations

Start with cleaning up your data

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

Cool. How long is that going to take?

Complexity

Of your data environment: cloud, on-prem, structured, unstructured data environments

Goals & Requirements

Input from stakeholders for enterprise-wide approach to data governance.

Amount of Data

Smaller organizations with less data will be faster to process than large enterprise.

Weeks

Months

Phased Implementation

Planning
Prioritize data environments, align stakeholders, set KPIs

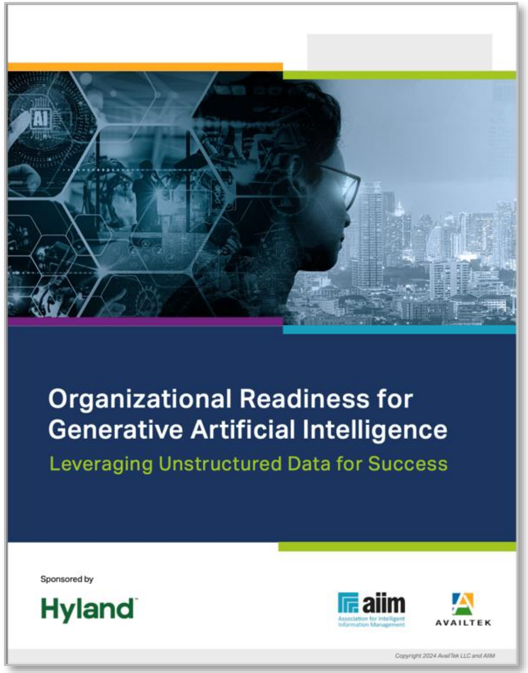
Proof of Concept
Limited data set to benchmark and forecast scope/scale

Implementation
Implement scale to process all enterprise content

Delta Monitoring
Continuously look for and process new & edited content

Weeks → Months

Additional Resources



AIIM White Paper:

Organizational Readiness for Generative Artificial Intelligence, 2024

Click to Download

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