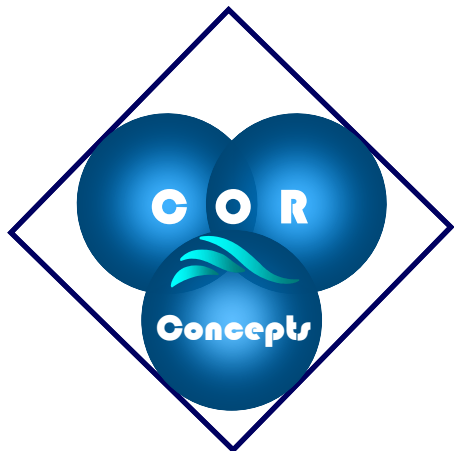




Rethinking Records Management in the Multiverse: Leveraging Artificial Intelligence and Emerging Technologies in South Africa

Artificial Intelligence (AI) in Intelligent Document Processing (IDP), Records Management, and Digital Preservation

National Archives and Records Service of
South Africa/UNISA

Annual Conference 2025

Dr. Paul Mullan

paulm@corconcepts.co.za

+27 (0) 832736087



Agenda

- What is Artificial Intelligence?
- What is Intelligent Document Processing?
- IDP, Records management and AI
 - New “Records” that need to be managed
 - How AI can assist Records Managers and Archivists
- AI and Digital Preservation
- Paradigm shifts



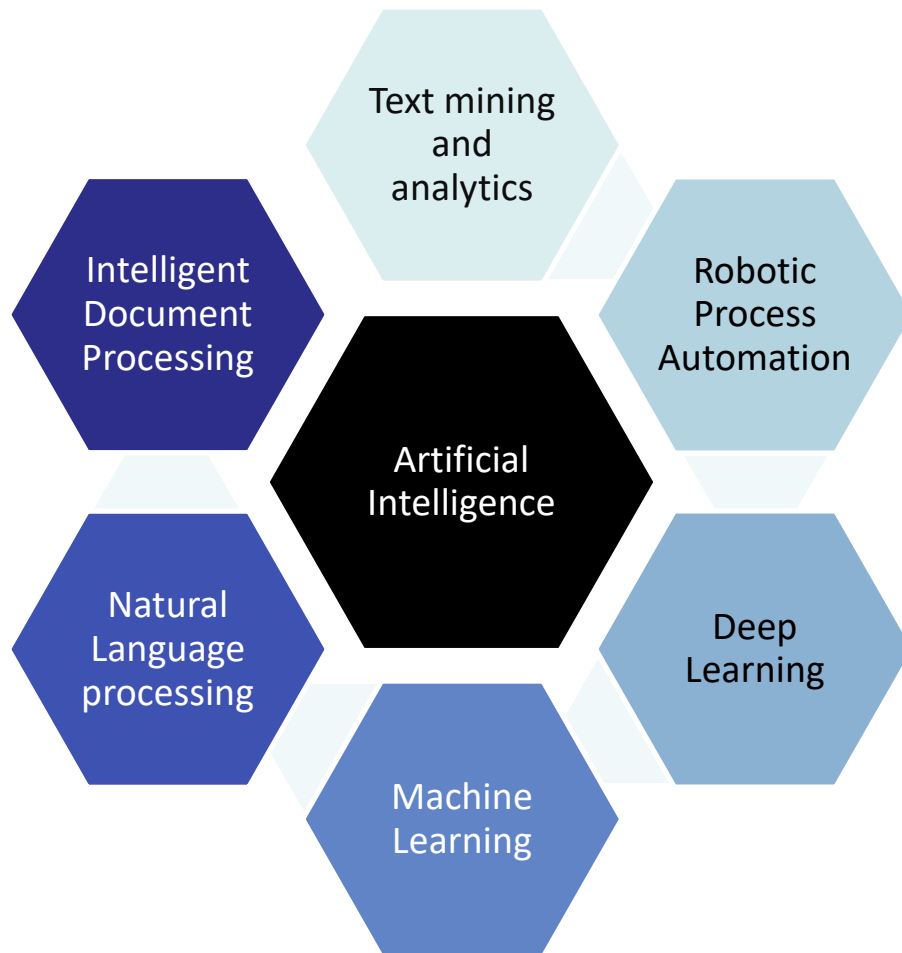
What is AI?

Source: ChatGPT

- A branch of computer science that focuses on creating intelligent machines that can perform tasks that typically require human intelligence....enable computers to perceive, reason, learn, and solve problems.
- Mimics cognitive functions such as speech recognition, visual perception, decision-making, problem-solving, and language translation.
- They can process large amounts of data, identify patterns, and make predictions or recommendations.



What is AI? Different types.



AI Can only work with data,
so all paper needs to be
converted to digital if AI is to
assist.

Scanned images need to have
the data extracted.

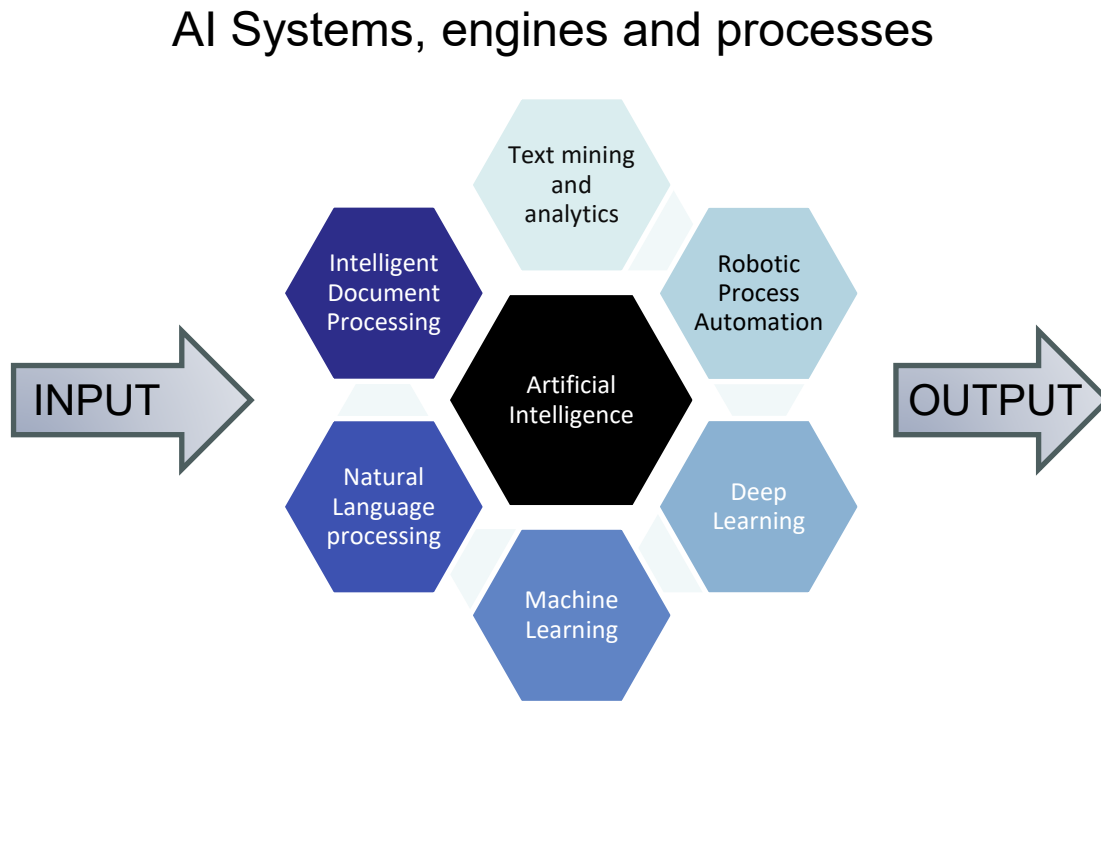
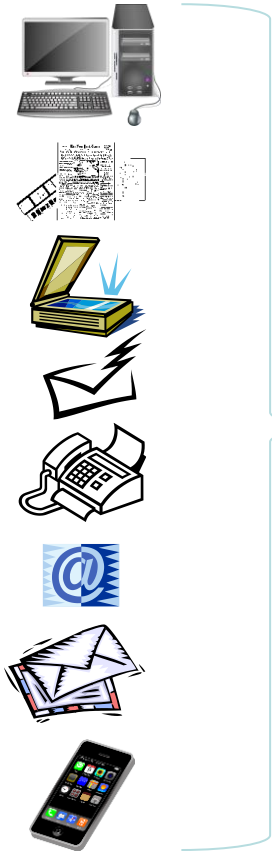
AI plus IDP is going to become
very important for records managers.

Lots of overlaps and interrelationships



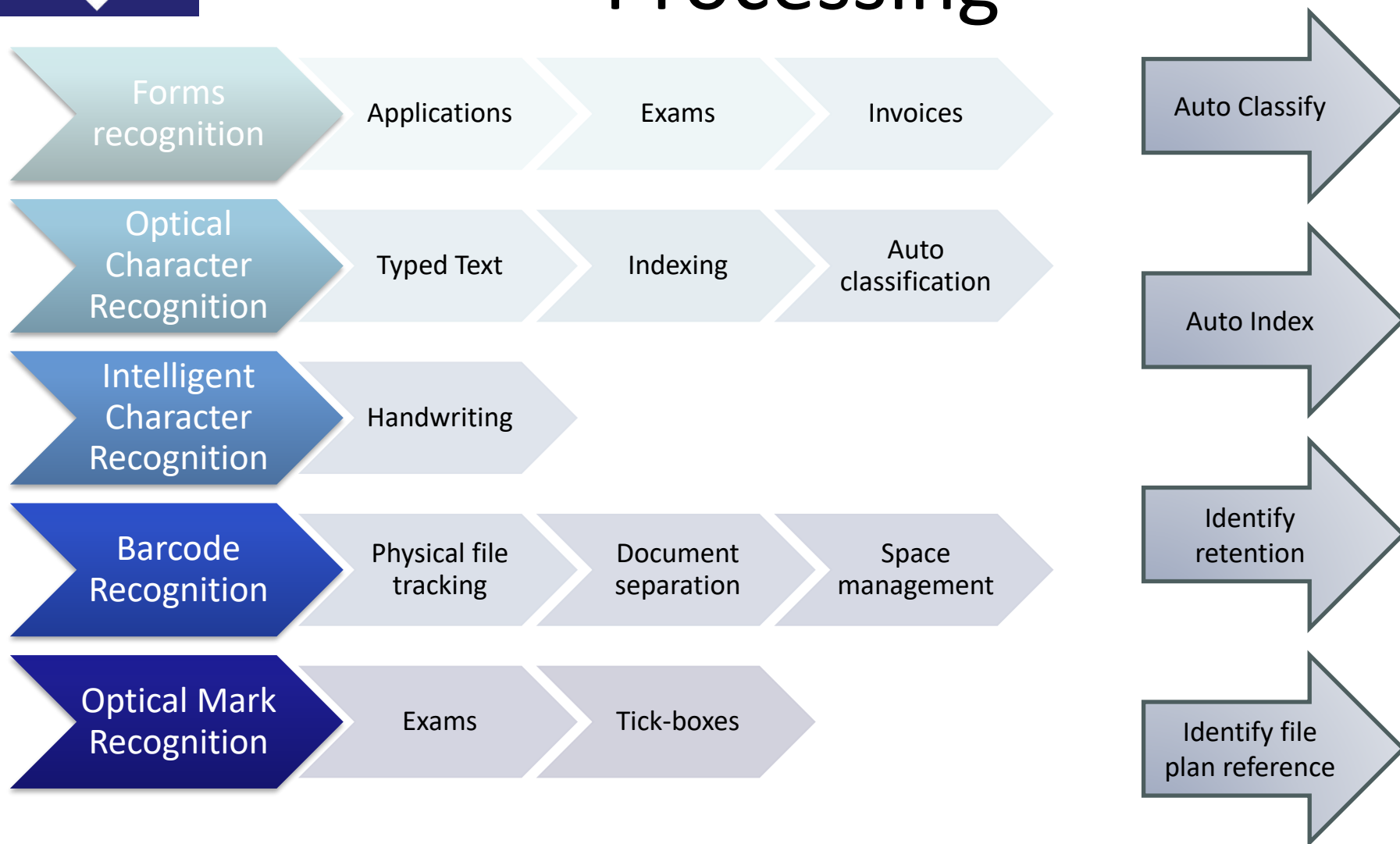
Simplified AI process

DATA SOURCES





Intelligent document Processing





Modern data sources

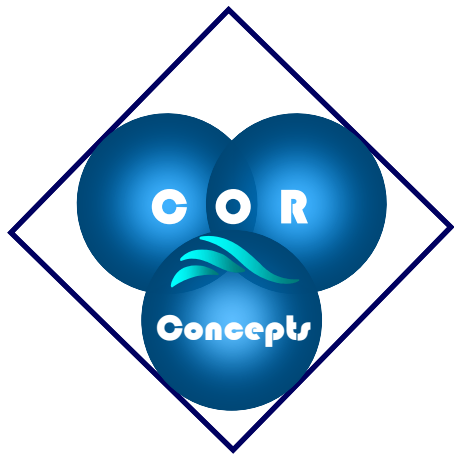
- Email messages and attachments
- Social media and blog posts (consumer and business)
- Text messages and instant messages
- Photos, videos and audio (multi-media, photos, videos, conference calls, voice mails, memos, etc.)
- Machine-generated data (log files, call detail records, etc.)
- Technology data – from cars, robotics, Process automation, factory automation
- Structured data (database/transactional)
- Scanned images
- Other Electronic content (Born digital documents and records)
- Paper records

Few of these are under Records and Archives control!
Think of a Schedule of Records Other than Correspondence!



What AI information will need to be managed?

- Datasets used as AI Inputs.
- Output from AI driven or created processes
- AI Generated documents and records
- Fewer “records” which look like “documents”
- Less human intervention at records creation and capture time
- Some critical records:
 - System generated reports
 - Audit reports – what the system did
 - Logic and coding as determinants of AI Actions, - why the system did it
 - Analytics – how the system derived the answer



Applications in records, archives & digital preservation



Text mining and analytics

Despite Rumours, Not Everything That Towers Is Eiffel's ...

www.nytimes.com/2014/10/29/world/americas/despite-rumors-not-everything-that-towers-is-eiffels.html

Key sentences

- Despite Rumours, Not Everything That Towers Is Eiffel's
 - The San Camilo market, with tall iron columns supporting a corrugated metal roof, was actually built in the 20th century by a local company, years after Eiffel had left his engineering firm.
 - Professor Gutierrez said that the bridge closely matched others built by Phoenix, and that there was no evidence of any Eiffel connection with either the company or the bridge.
 - [Cogito Intelligence API metadata](#) Taxonomies (16) Relations (11) Places (22)
- Related documents: Real time analysis with Cogito Intelligence API

People, organizations and places mentioned



Other key entities



Auto Classify

Auto Index

Identify retention

Identify file plan reference

Source: Expert Systems (From AIIM CIP)



AI and Records/Archives Management

- E-Discovery
 - Searching across systems and collections and finding related information
 - Identification of exceptions (Missing information)
 - Linking information across systems ERP, EDRMS, File servers, e-mail, collaboration (Chat, Teams)
 - Automating triggers for discovery and audit processes (e-discovery in anticipation)
 - Automated gathering of related data and imposing records holds



AI Records, Archives and Digital preservation

- Determination of types of formats and media in use
- Determination of records activity patterns and developing new retention rules for active/semi-active, archival records (Appraisal within appraisal)
- Internet searches (Digital Preservation strategy)
 - Technology/industry scans (internal and external)
 - New emerging technology
 - Threats to existing technologies and formats
 - Creation of lists of “Records at Risk”



Aspects of digital preservation

Creation	Classification
Capturing data	Transfer lists
Storage and preserving	Disposition lists
Analysis	Querying
Search	Updating
Sharing	Information privacy
Archiving	Security



Robotic Process Automation

- Handles dull, repetitive, non-value added tasks
- Identification of and creation of lists of records due for appraisal, transfer and destruction.
- Identification of duplicates
- Extraction and or copying of records from source systems and movement to electronic records centres/archives
- Database lookups for accuracy of indexing and classification
- Identification of invalid or missing metadata
- Application of additional metadata on transfer
- Conversion/migration of software formats
 - i.e. MS Word to PDF/A
 - Older versions of PDF to PDF/A
 - .dwg or other proprietary formats to PDF/E
 - Automated sampling and quality checks



AI and Records/Archives Management

- Automatic classification
 - Sensitivity (secret, top secret, confidential)
 - Personal Information (Privacy/Data Protection)
 - Classification scheme (file plan) and taxonomies
 - Retention periods
 - Determining disposal dates
 - Machine learning (better classification over time)



AI and Records/Archives Management

- Using e-discovery to “know your data” in archival collections and repositories
- Earlier and more automated transfer from Units, Departments, other organisations
- Deep learning
 - Analysis of massive quantities of data
 - Automatically analysing and extracting context and focus areas/insights
- Machine learning
 - Automated appraisal
 - Improvement of learning engines all the time
- Robotic process Automation
 - Automation of non-value adding mundane, repetitive tasks



Using AI to assist

- Generative AI
 - Create drafts of reports, policies procedures
 - Improve presentations with graphics
- Quality control
 - Check quality of policies, procedures and processes for completeness and compliance
 - Ask AI to check your documents
- Speed
 - Reduce time to find required information



Possibly the biggest opportunity for Government?

BAS, Persal,
SAGE, Oracle, SAP...

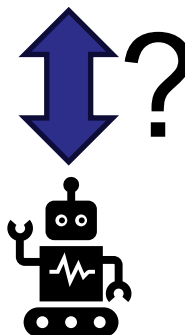
Expenditure



Daily transactions

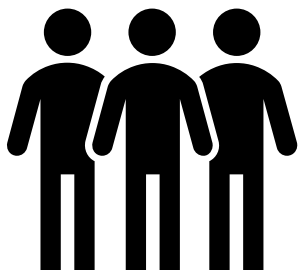


Daily logs



- How many transactions?
- What type?
- What documents should have been created?
- Where are the approvals?

Daily



Records Team.

Departmental Records Liaisons

Where are the supporting documents?

Robotic
Process
Automation



Opportunity or Threat (Paradigm shifts)

- Information is at the heart of business disruption
- AI has become mainstream
- Treat AI as your assistant – not your replacement
- The choice to disrupt or be disrupted lies with you
- There will be pain. How much, depends on whether you force the changes, or are forced to change.
- Your organisation will change. Your organisation is changing.
- Understand, manage, optimise, leverage information assets.
- You cannot do it alone. Partner with IT and Business Units
- Add value. Be relevant.



Thank you

Dr. Paul Mullan

paulm@corconcepts.co.za

+27 (0) 832736087