



International Council on Archives  
Conseil International des Archives

# Artificial Intelligence and Archives

Dr Anthea Seles, Secretary General

31 August 2021

West Nordic Archives Conference



International Council on Archives  
Conseil International des Archives

## Overview

- What is 'Artificial Intelligence'?
- How can it be deployed in archival practice?
- Points to consider when deploying artificial intelligence
- How does the use artificial intelligence improve and/impact dissemination?



International Council on Archives  
Conseil International des Archives

# Artificial Intelligence, Machine Learning and Neural Networks

- *Artificial Intelligence*: It's an all-encompassing definition for any activity where a machine/system takes information (structured and unstructured) to predict an outcome
- *Machine Learning*: Process of training a system to 'learn' how to make a decision using a pre-tagged dataset.
- *Neural Networks*: Just like we use our brains to identify patterns and classify information, neural networks can be trained to accomplish similar tasks.
  - Deep learning: Layering multiple neural networks



International Council on Archives  
Conseil International des Archives

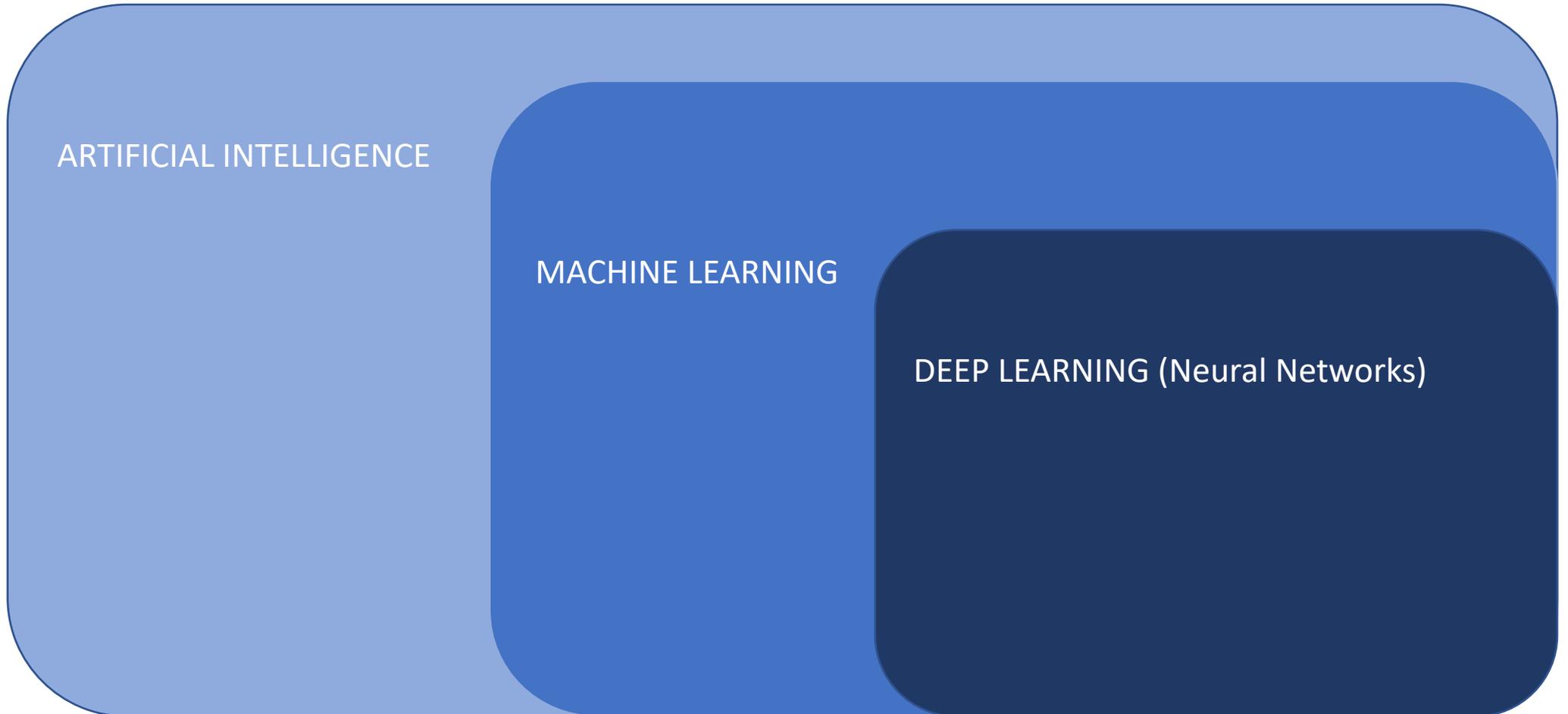
# What is Artificial Intelligence(AI)?

- Artificial intelligence can be defined in many different ways; there is no standard definition
- There are really two categories
  - Supervised
  - Unsupervised
- Supervised: Requires a human to mark up or compile a homogeneous dataset to train an algorithm to recognise patterns or terms in the data. This process requires a lot of up front work and also requires you to have some level of understanding of the dataset.
- Unsupervised: Data is loaded into the system and without any upfront human intervention, analyses the data and provides result.



International Council on Archives  
Conseil International des Archives

# Artificial Intelligence, Machine Learning and Deep Learning



ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

DEEP LEARNING (Neural Networks)



International Council on Archives  
Conseil International des Archives

# How can AI be deployed in archival practice?

## Considerations

- Volume and complexity of digital information complicates all part of the archival work.
- Archivists do not have the capability to carry out large scale evaluations of massive amounts of data (structure or unstructured).
- Researchers will not have the means to review all the digital information that archives hold.
- Legacy and current information management practices impact the deployment of artificial intelligence from the beginning of the archival/information management process to the end.
- Automation is not an option.



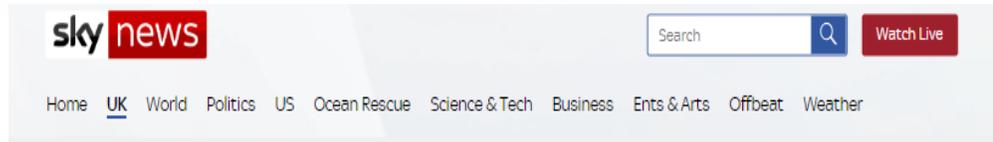
**ICA**

International Council on Archives  
Conseil International des Archives

# How can AI be deployed in archival practice?

- Records management/Information management
  - Classification
  - Categorisation
  - Retention/Disposal
- Appraisal and selection
- Sensitivity review
  - Personal information vs context-based sensitivity
- Digital preservation
  - Identification of at risk format or material
- Access and dissemination

# Government Use of Artificial Intelligence and *machine learning*



sky news Search Watch Live

Home UK World Politics US Ocean Rescue Science & Tech Business Ent's & Arts Offbeat Weather

## Handwriting to help Govt catch gangs behind mass-scale benefit fraud

Artificial intelligence is going to be used to clamp down on cheats claiming bogus benefit payments worth millions of pounds.

19:31, UK  
Sunday 31 December 2017



A record £1.1bn in overpaid benefits was recovered from fraudsters last year



CONCEPTION GRATUITE  
ET SANS ENGAGEMENT !

Les Bons Plans  
sont chez Darty\*

**-15%**

sur les meubles Sorbonne,  
Concorde, Rio et Pérou



The Atlantic Popular Latest Sections Magazine More Subscribe

### TECHNOLOGY

## A Popular Algorithm Is No Better at Predicting Crimes Than Random People

The COMPAS tool is widely used to assess a defendant's risk of committing more crimes, but a new study puts its usefulness into perspective.

ED YONG JAN 17, 2018



International Council on Archives  
Conseil International des Archives

# **Points to consider when deploying artificial intelligence**

# Points to consider when deploying artificial intelligence

## It is a tool

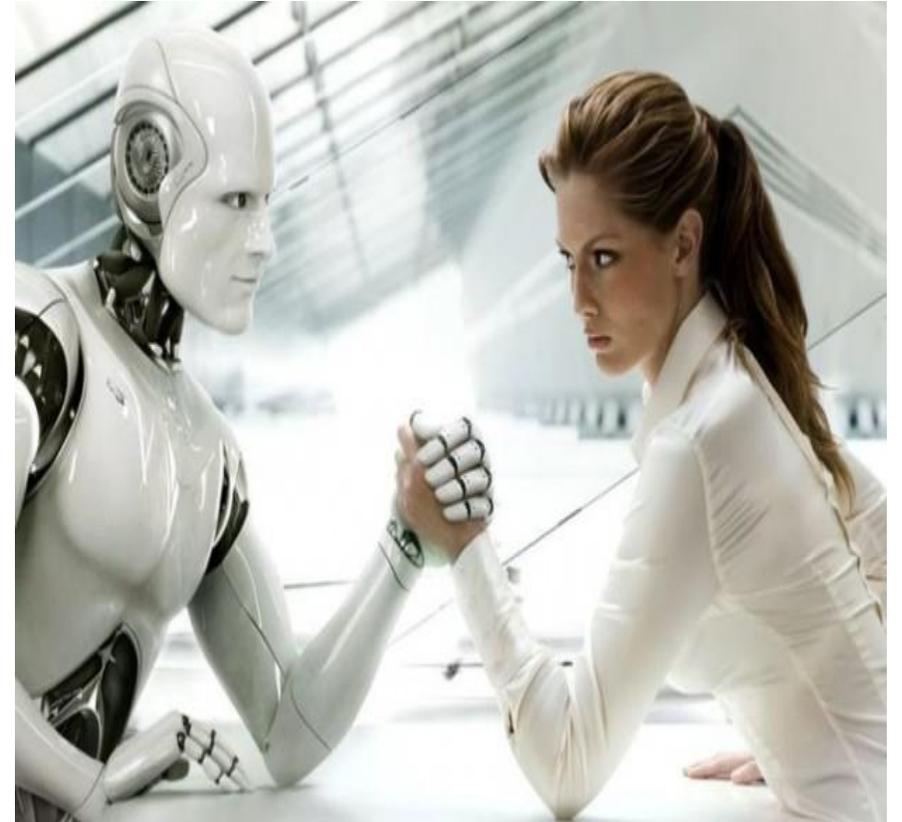
- The machine has strengths and weaknesses, the human has strengths and weaknesses.
- Archivists, information managers and researchers will not be made redundant.
- It cannot think for itself and simply because it gives an answer does not mean the output is correct.

# Humans vs AI: Strengths and weaknesses

- What can machines do well?
  - Boolean and keyword searches ✓
  - Regular expressions ✓
  - Process at scale ✓
  - Understand context and inference ✗
  - Handwriting analysis ✗

---

- What can humans do well?
  - Process at scale ✗
  - Understand and infer context ✓
  - Handwriting analysis ✓



# Artificial Intelligence and *machine learning* in Records Management and Archives

- Problems and limits encountered during testing
  - Lack of understanding regarding the content and the context of creation
  - Corruption or alteration of metadata
  - Difficulty understanding the visualisations generated by the machines
  - Understanding the reliability (precision and recall) of the results and the acceptable level of risk
  - Distrust in technology and the results generated by the systems
    - However in other instances the results are accepted without question with an imprecise understanding of how the results were arrived at.
  - Significant time required to ‘train’ the system, departments wanted something much more automated (i.e. unsupervised)



International Council on Archives  
Conseil International des Archives

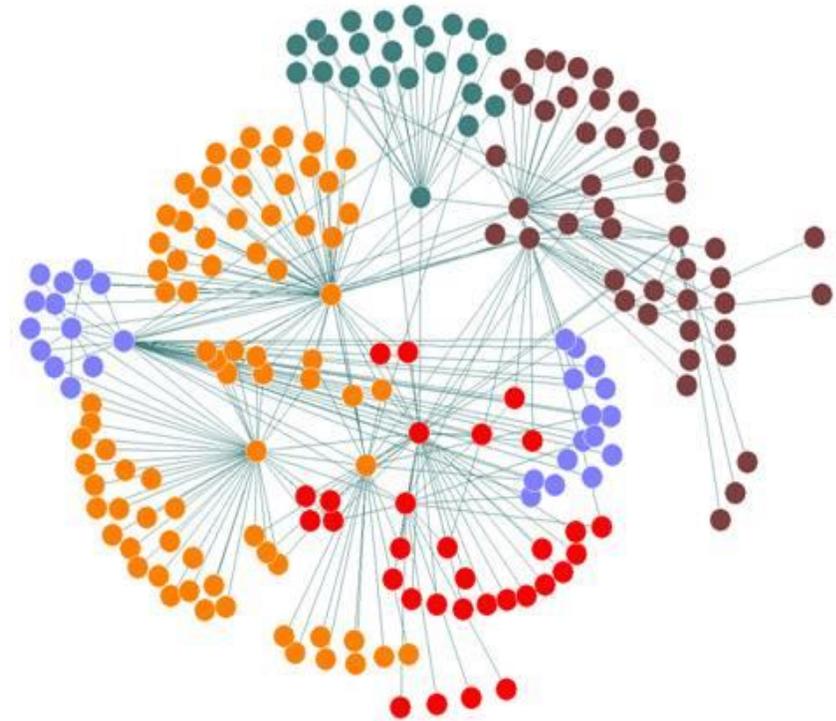
**How does the use Artificial  
Intelligence improve and/or  
impact dissemination?**

# Artificial Intelligence and Dissemination

- Can interconnect previously siloed fonds or collections
- Records in Context- the new ICA descriptive standard allows interconnection of archival description through entities and ontologies.
  - The objective of the standard is to reconcile, integrate, and build on the four existing standards: General International Standard Archival Description (ISAD(G)); International Standard Archival Authority Records—Corporate Bodies, Persons, and Families (ISAAR(CPF)); International Standard Description of Functions (ISDF); and International Standard Description of Institutions with Archival Holdings (ISDIAH).
  - This allows artificially intelligent systems to process and make connections between entities.
  - Breaks down the silos that exist in our descriptive practices.

# Artificial Intelligence and Machine Learning in Research and Dissemination

- There is a question for archivists about how much access we may wish to allow researchers access to public records and data
  - Data mining and machine learning tools breakdown siloes created by archival description (i.e. fonds, series, files)
  - Can reveal unknown connection that become sensitive or problematic by virtue of making that connection
  - Can surface sensitive information that was missed during sensitivity review
  - Also once the data is mined and put into a system outside the archives, what else can it can be combined to?
- Let's not get tunnel vision with AI. There is a danger of focusing too much on the impact on our individual collections, but what about linked data? And the semantic web? What will this mean for archives and opening our collections?



# Artificial Intelligence and Machine Learning in Research

- We also need to consider the impact of future digitisation.
  - The re-purposing and re-use of archival records and data has enormous value and I think we sacrificed much of digitisation and allowing companies to digitize archival records and data, in order that we can get a ‘free’ copy’. We must be savvier.
  - Companies are beginning to realise the value of data held in historical records. Digitising them and applying OCR is a method for gaining access to large volumes of data to train algorithms.
- We need to start asking ourselves:
  - Why is the digitisation free?
  - Will this data be used to train an algorithm?
  - What is the company’s ethical stance?
  - What happens to the data once the digitisation is done?
  - Will there be an impact on people’s lives?
- Scenario: Paper death registrations

## Conclusion

- How will you make your collections more discoverable via machine learning algorithms?
  - Consider the levels of access
  - Consider carefully requests coming from commercial enterprises
- Implementing Records in Context is one way to make information more easily accessible and discoverable.
  - They are developing a converter that converts your descriptions into RiC compliant descriptions
- Will you allow any machine learning algorithm to access your descriptions?

# References and Further Reading

- *The Application of Technology Assisted Review to Born-Digital Records Transfers, Inquiries and Beyond.* (2016) London: The National Archives UK <http://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf>
- Bhaskar, Michael. *Curation: The Power of Selection in a World of Excess.* (2017) London: Piatkus
- Caplan, Robyn, Joan Donovan, Lauren Hanson and Jeanna Matthews. 'Algorithmic Accountability: A Primer' *Data and Society* (2018) [https://datasociety.net/wp-content/uploads/2018/04/Data\\_Society\\_Algorithmic\\_Accountability\\_Primer\\_FINAL-4.pdf](https://datasociety.net/wp-content/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL-4.pdf)
- Chumtong, Jason and David Kaldewey. 'Beyond the Google NGRAM Viewer: Bibliographic Databases and Journal Archives As Tools for Quantitative Analysis of Scientific and Meta-Scientific Concepts. *FIW Working Paper No 8* (2017) <https://www.fiw.uni-bonn.de/publikationen/FIWWorkingPaper/fiw-working-paper-no.-8>
- Delort, Pierre. *Le Big Data* (2015) Paris : PUF
- Domingos, Pedro. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (2015) New York: Basic Books
- Engin, Zeynep and Philip Treleaven. 'Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies' *The British Computer Society* (August 2018) <https://academic.oup.com/comjnl/advance-article/doi/10.1093/comjnl/bxy082/5070384>
- Ertzscheid, Oliver. *L'appétit des géants: pouvoir des algorithmes, ambitions des plateformes* (2017) Paris : C&F
- Information Privacy Commissioner. *Big Data, Artificial Intelligence, Machine Learning and Data Protection.* (2017) London: ICO <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>
- Jerven, Morten. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It.* (2013) Ithaca: Cornell University Press
- LeSueur, Andrew. 'Robot Government: Automated Decision-Making and its Implications for Parliament' [Draft chapter for publication in *Parliament: Legislation and Accountability* (Oxford: Hart Publishing) 2016] [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2668201](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2668201)
- Lorenzi, Jean-Hervé et Mickaël Berrebi. *L'avenir de notre liberté* (2017) Paris : Eyrolle

# References and Further Reading

- Lynch, Clifford. Stewardship in the 'Age of Algorithms' *First Monday* Vol 22 (12) December 2017  
<http://firstmonday.org/article/view/8097/6583>
- Mason, Susan . 'Analysing Stereotypes Across Time Using Google Ngram Viewer' *SAGE Research Methods Cases Part 2* (2018)  
doi:10.4135/9781526436245
- Mason, S. E., C.V. Kuntz, & , C. M. McGill. 'Oldsters and ngrams: Age stereotypes across time'. *Psychological Reports: Sociocultural Issues in Psychology*, (2015),116, 324–329. doi:<http://dx.doi.org/10.2466/17.10.PRO.116k17w6>
- Musser, George. 'Artificial Intelligence: How Machines could learn creativity and common sense, among other human qualities'. *Scientific American* Vol 320, No 5 (May 2019) 47-51
- O'Neill, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (2016) New York: Crown Publishing
- Padilla, Thomas, Laurie Allen, Sarah Potvin, Elizabeth Roke Russey, and Stewart Varner. 'Collections as Data', 7 March 2017.  
<https://doi.org/10.17605/OSF.IO/MX6UK>.
- Rolan, Gregory, Glen Humphries, Lisa Jeffrey, Evanthia Samaras, Tatiana Antsoupova and Katharine Stuart. 'More Human than Human? Artificial intelligence in the archive' *Archives and Manuscripts* Vol 47, No 2 (November 2018) 179-203
- Venkatesan M *Artificial Intelligence vs. Machine Learning vs. Deep Learning* (7 May 2018)  
<https://www.datasciencecentral.com/profiles/blogs/artificial-intelligence-vs-machine-learning-vs-deep-learning>
- Villani, Cédrique. *Donné un sens à l'intelligence artificielle: Pour une stratégie nationale et européenne* (8 mars 2018)  
[https://www.aiforhumanity.fr/pdfs/9782111457089\\_Rapport\\_Villani\\_accessible.pdf](https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf)
- World Wide Web Foundation. 'Algorithmic Accountability: Applying the Concept to Different Country Contexts'. *A Smart Web for a More Equal Future* (2017) [https://webfoundation.org/docs/2017/07/Algorithms\\_Report\\_WF.pdf](https://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf)
- Zambonelli, Franco, Flora Salim, Seng W. Loke, Wolfgang De Meuter and Salil Kanhere. 'The Algorithmic Governance in Smart Cities: The Conundrum and the Potential of Pervasive Computing Solutions' *IEEE Technology and Society Magazine* (June 2018) pp 80-87

# Thank you.

**Dr Anthea Seles**  
Secretary General  
International Council on Archives  
[seles@ica.org](mailto:seles@ica.org)



# ICA

**International Council on Archives**  
Conseil International des Archives