



International Council on Archives
Conseil International des Archives

Archival Odyssey: Part One

Dr Anthea Seles, Secretary General

8 April 2019

Norwegian Triennial Archival Conference



International Council on Archives
Conseil International des Archives

Artificial Intelligence= ???

**Don't Gauge Artificial
Intelligence By What You See
in the Movies (or in a Google
image search)**

(October 5, 2017, National Public Radio) <https://www.npr.org/sections/alltechconsidered/2017/10/05/555032943/lawmakers-dont-gauge-artificial-intelligence-by-what-you-see-in-the-movies?t=1554307632807>

Digital Information Overload

- Between 2006 and 2010 the global quantity of digital data will have increased more than six-fold, from 161 exabytes to 988 exabytes (Floridi, 2010)
- That was 2010, almost ten years later we are now talking about zettabytes
- 2025: 163 zettabytes (Forbes, 2017)
- 1 zettabyte = 1 quintillion Word documents

Decimal Value	Metric
1000	Kylobyte (kB)
1000 ²	Megabyte (MB)
1000 ³	Gigabyte (GB)
1000 ⁴	Terabyte (TB)
1000 ⁵	Petabyte (PB)
1000 ⁶	Exabyte (EB)
1000 ⁷	Zettabyte (ZB)
1000 ⁸	Yottabyte (YB)

Curation is now a thing!

- In the archives we have been talking about 'digital curation' for some time, though the ideas being proposed now are slightly different from our understanding
- Digital curation: selection, collection, preservation, maintenance, access, long term archiving, and stewarding digital assets over time
- Curation is now seen as a competitive advantage, along with the ability to exploit data science techniques.
- It is changing the way that retailers and others do their business
- Governments are also 'curating' their data to make decisions
- There is just too much information!

How Retail Changes When Algorithms Curate Everything We Buy

by Bobby Gibbs and Nick Harrison

JANUARY 07, 2019

Summary Save Share Comment ³ Text Size Print \$8.95 Buy Copies



Harvard Business Review, January 07 2019



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 a 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

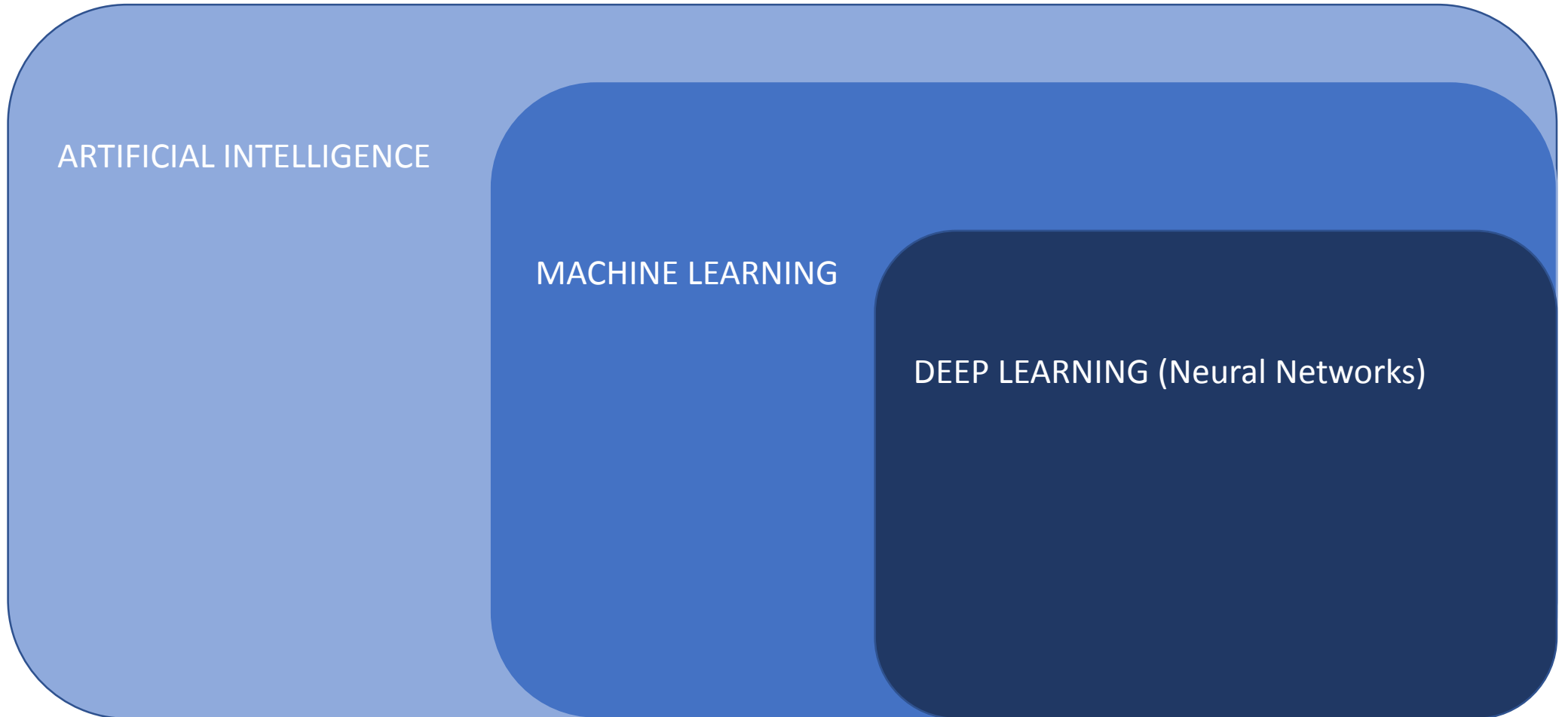
Artificial Intelligence, Machine Learning and Neural Networks

- *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

Artificial Intelligence, Machine Learning and Deep Learning



ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

DEEP LEARNING (Neural Networks)



International Council on Archives
Conseil International des Archives

Artificial Intelligence and Archives

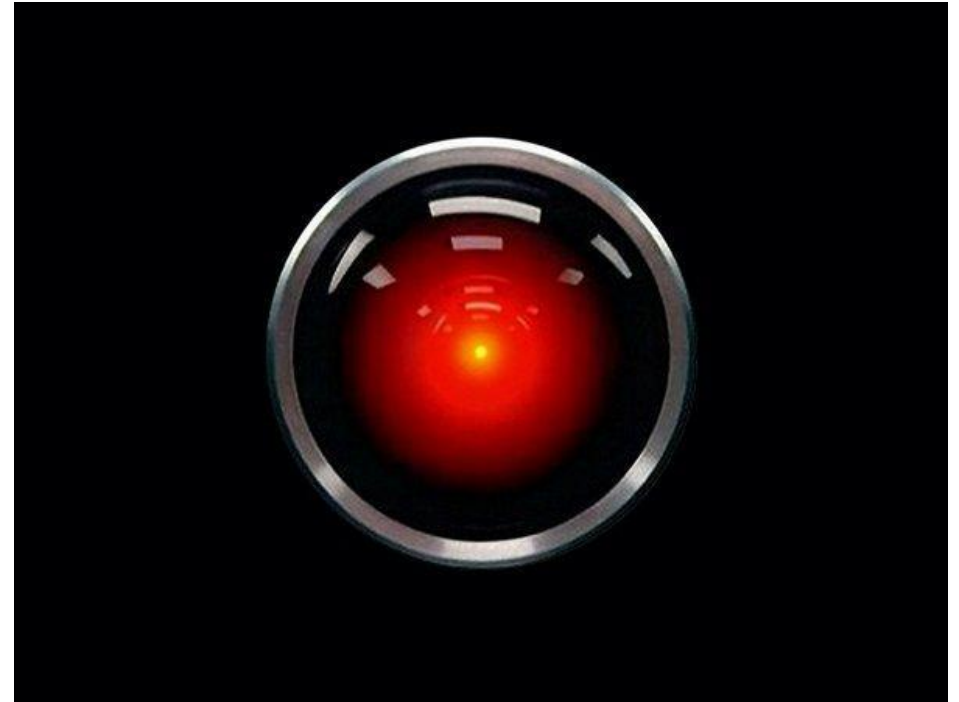
- We have known for 30 years that we need to be involved in the digital records creation process to ensure that the necessary elements are in place to preserve their evidential value
- But how does that work for machine-learning and artificial intelligence? Are the outputs of AI and machine-learning systems records?
 - **YES!**
 - If a government uses the output of a machine-learning or an artificial intelligence process to make a decision that impacts the administration of policy and/or people's lives then it is a record
- This has implications for us in terms of what we select and then how we preserve.
- There are also challenges about how we use these technologies in our practice



International Council on Archives
Conseil International des Archives

Automation and Archives: Reality check

- Automation for archives is not optional, because the volume and variety of material we will need to appraise and select is not getting any smaller. We need to consider the implications of automation in terms of ethics and in terms of practice.
- We are all going to lose our jobs because of automation of archival practice and artificial intelligence- **FALSE!**
- Ex: 'In the 1980s, new technologies can decimate the labour force in the goods producing sectors of the economy...' (Sleepers Wake (1982, revised 1995) by Barry Jones)
- It's a balance. We need to understand the strengths and weaknesses of machines and that of humans



Artificial Intelligence and Archival Processes

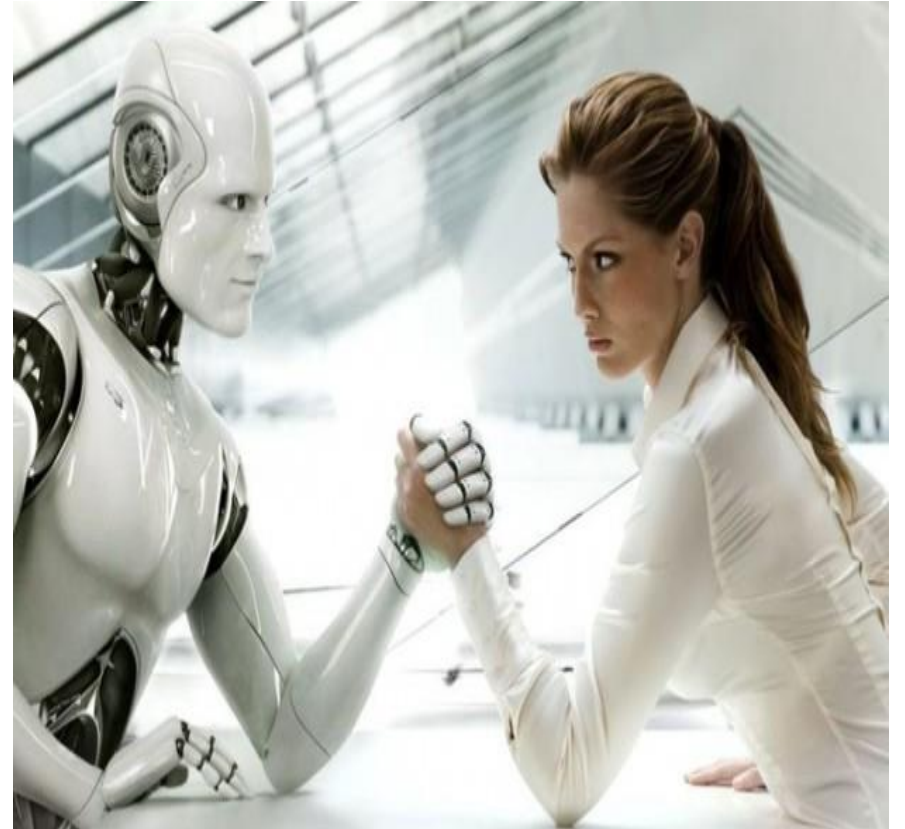
- What can machines do well?

- 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 Implications for Archives

- The challenge of identifying artificial intelligence records:
 - How do we do this? How do we get invited to the table to begin a conversation? How do we get decision-makers to understand that these are records that need to be preserve?
 - What do we preserve? Training data? All the data? Code? Everything? How do we transfer this information to our archives?
 - How do we preserve these records? What infrastructure do we need for this?
 - How do you make AI records accessible to researchers? How do you describe them?
- The challenge with automating appraisal and selection and other parts of the archival process:
 - How do you measure accuracy? What does 'good enough' look like? What are the risks? What is acceptable risk appetite?
 - How can we determine what might be missing?
 - How can be accountable for the decisions we make based on machine outputs? How do we equally hold the machines to account?
 - How do we compensate for the change in the digital record over time? Re-tune the algorithm?
 - What is the risk of breaking down the silos that we built (i.e. archival description)?
- **RISK: Biasing the historical record and by proxy history along with our collective memory**

Conclusion

- We need to get to grips with the fact that the volume of digital information that we will eventually have to deal with is increasing, not decreasing
- Need to recognize that it's only about AI's impact on our practice, but that AI is the record to come and we need to advise on how to create and maintain it
- Automating parts of the archival process is necessary, but we have to engage intelligently with these systems
 - We need to have a basic understanding of how they work
 - We need to know what questions we want to ask and when we need to reformulate the question
- Artificial intelligence (supervised/unsupervised) brings with it many ethical considerations for us as archivists, however we are operating in a legal and legislative vacuum.
- We need to acknowledge the strengths and weaknesses of computers and humans