

ARTICLE



## More human than human? Artificial intelligence in the archive

Gregory Rolan<sup>a</sup>, Glen Humphries<sup>b</sup>, Lisa Jeffrey<sup>c</sup>, Evanthia Samaras<sup>c</sup>,  
Tatiana Antsoupova<sup>d</sup> and Katharine Stuart<sup>e</sup>


<sup>a</sup>Centre for Organisational and Social Informatics, Monash University, Melbourne, Australia; <sup>b</sup>Digital Archives, NSW State Archives and Records, Sydney, Australia; <sup>c</sup>Lotus Notes Email Analysis Project, Public Record Office Victoria, Melbourne, Australia; <sup>d</sup>Information and Technology Branch, National Archives of Australia, Canberra, Australia; <sup>e</sup>Department of Finance, Australian Government, Canberra, Australia

### ABSTRACT

Not a day appears to go by without breaking news of some Artificial Intelligence (AI) advance that seemingly has the potential to transform our lives. As recordkeeping professionals, we can very well ask, 'What about us?' Where is the AI or automation to help us with our classification, appraisal and disposal work? If we are to meet the challenges of managing records in the digital age, such technology – together with appropriate skills and knowledge – will be necessary. How can AI automate our digital recordkeeping and archive work? In this article, the authors provide a snapshot of the practice of AI in Australian recordkeeping. What is the reality versus the hype of such technology, and what is actually being done now? In answering these questions, they first provide a brief introduction into AI techniques and their characteristics in relation to recordkeeping work. They then introduce four case studies from Australian archival and government institutions that have embarked on AI initiatives. In each case, they provide an overview of the project in terms of requirements, activities to date, outcomes and futures. The article concludes with a discussion of the lessons learnt, issues and implications of AI in the archive.

### Introduction

Not a day appears to go by without breaking news of some Artificial Intelligence (AI) advance or, at least, automation product that seemingly has the potential to transform our lives. From face recognition in our phone cameras, to the promise of self-driving cars, to the voice-enabled assistants in our homes, there are AI assistants for all of our everyday needs. Equally pervasive are those inscrutable engines that attempt to predict and control our next purchase or second-guess our information needs; filtering the various streams and feeds that make up our infotainment and advertising networks. Our business lives are also similarly enhanced, for example, with email systems that recognise and divert spam, advise us of possible recipients and prompt for suggested calendar entries based on email conversations. We also hear about breakthroughs in

**CONTACT** Gregory Rolan  [greg.rolan@monash.edu](mailto:greg.rolan@monash.edu)  Centre for Organisational and Social Informatics, Faculty of Information Technology, Monash University, P.O. Box 197, Caulfield East, VIC 3145, Australia

© 2018 Gregory Rolan, Glen Humphries, Lisa Jeffrey, Evanthia Samaras, Tatiana Antsoupova and Katharine Stuart

various types of knowledge work: for example, in medical diagnoses, language translation, image processing, scenario analysis (including game playing) and intelligence gathering. We are surely entering a golden age of AI-assisted automation in all facets of our lives.<sup>1</sup>

As recordkeeping professionals, we can very well ask, ‘What about us?’ Where is the AI or automation that can help us with our classification, appraisal and disposal work? If AI can help us drive our cars, diagnose our diseases and filter our personal information needs, surely we should be able to harness such technologies to address the routine aspects of our work and attack our growing backlogs, if not re-figure our discipline, in line with the realities of digital recordkeeping. Arguably, if we are to meet the challenges of managing records in the digital age,<sup>2</sup> such technological aids – together with the skills and knowledge required to wield them – will be necessary.<sup>3</sup>

Today’s information environments have become a ‘wild frontier’;<sup>4</sup> decentralised and fractured, and subject to pressures that include increasing data volumes, reliance on commercial and proprietary systems, and evolving forms of records and formats.<sup>5</sup> For example, a recent National Archives UK review found that while 30% of agencies held digital records within an Electronic Document and Records Management System (EDRMS), there is generally no active retention/disposal function. The report also highlighted that emails are inconsistently managed. Only 10% of agencies are appraising and capturing their social media content, and shared drives are used to store high volumes of data – ‘for every TB of information held in an EDRMS, there are approximately 10TB of data in shared drives’.<sup>6</sup> Faced with taming this ever-expanding and increasingly ‘wild frontier’, institutions are faced with the prospect of simply storing this overwhelming volume of digital information in the hope that ‘innovative techniques for mining, recovering, and reusing digital materials and their traces’ may eventually be found to separate out the good oil of meaningful records from vast quantities of information sludge.<sup>7</sup>

The ongoing decrease in storage costs is often compared to the cost of manual effort in forming intellectual boundaries around records and determining value in the complex digital realm. Shifting today’s problems to an idealised future is, however a false economy. Instead, there is an imperative to move from passive service provision to proactive systems design and monitoring in order to intervene at ‘risk points of design, migration, and decommissioning’ for digital records.<sup>8</sup> Put simply, without appropriate attention (and suitable technological aids), we will drown in the sludge.<sup>9</sup>

So, what technologies are available now, or in the near future, to automate this digital recordkeeping and archive work? In this article we look beyond digital forensics or even general systems management tools (for example, as described by William P Vinh-Doyle,<sup>10</sup> Anthony Cocciolo,<sup>11</sup> Victoria Sloyan<sup>12</sup> or Ross Spencer<sup>13</sup>) that may form some building blocks of automated pipelines. Instead, we will attempt to provide a point-in-time snapshot of the research, literature and practice of AI in recordkeeping to aid in the evaluation and selection of AI solutions. What is the reality versus the hype of such technology, and what is actually being done now? While we are patently a long way from *Blade Runner*-style ‘Replicant’ recordkeepers, is AI for recordkeeping all talk, or is there substance behind the hype?

The rest of this article is organised as follows. We will first provide a brief introduction to the nature of AI techniques and their characteristics in relation to recordkeeping

work. We will then provide overviews of a small number of contemporary AI-related projects that are taking place in Australian institutions. Finally, we will conclude with some speculation about the short- and longer-term prognoses for the use of AI technologies in recordkeeping work.

But first, what do we mean by the term AI and what affordances does it bring to our mission?

## Forms of AI: from expert systems to deep learning

Artificial intelligence has evolved since the earliest days of computing (and, in fact, from earlier, if one takes a broader view of technologies that help automate human knowledge work)<sup>14</sup> but, due to its ever-shifting nature, it defies easy definition. Part of the problem is that as ‘people become accustomed to this technology, it stops being considered AI, and newer technology emerges’.<sup>15</sup> This constant churn means that, in one sense, AI ‘ends up being viewed as those things which the AI people are doing’.<sup>16</sup> For the purposes of this article, we understand AI as involving digital systems that automate or assist in ‘activities that we associate with human thinking, activities such as decision-making, problem solving, learning, [and] creating’.<sup>17</sup> In fact, we take the broad view that AI manifests ‘on a multi-dimensional spectrum [comprising] scale, speed, degree of autonomy, and generality’ – thus encompassing a range of automation techniques and technologies.<sup>18</sup>

Additionally, like many technology domains, the AI space has its fair share of jargon: expert systems, rule engines, machine learning, deep learning, neural networks – the list goes on. Moreover, confusion about the status or efficacy of these technologies is exacerbated by the succession of hype cycles,<sup>19</sup> with peaks and troughs of interest (and therefore research funding, development and application) as various techniques and technologies fall in and out of favour and are rediscovered.<sup>20</sup> In this section, we will try to put these technologies and techniques in perspective for knowledge work such as recordkeeping. One way of doing this is through consideration of domain expertise, and the degree and manner in which this knowledge is embodied in an automated system. Such a taxonomy encompasses rule-based systems, statistical model, and deep learning systems. While this is a useful taxonomy for understanding various AI techniques, it should be noted that real world implementations might comprise one or more of these techniques, in various hybrid combinations. The following is therefore a brief and simplified account.

### Rule-based systems

At its core, a *rule-based* system involves the externalisation of domain expert knowledge into a working system that can apply this knowledge to future cases.<sup>21</sup> Rule-based systems are sometimes called *expert systems* as they rely on the elicitation and articulation of domain expertise. For example, an early illustration of automated electronic communications appraisal employed an expert system that embodied the appraisal expertise of a number of archival domain experts.<sup>22</sup>

Rules can be simple ‘If this, then do that’ procedural statements that may be run algorithmically when presented with new input data. In fact, a set of document search terms (where the implicit ‘do that’ is ‘apply to search results’) may be considered the trivial case of this, particularly if such a search is saved and routinely used to filter

results. More complex systems can comprise sets of declarative statements comprising facts and formal rules termed a *knowledge base*. Such a knowledge base can then be applied to the additional (if provisional) facts of new input cases via an *inference engine* to arrive at decisions or outputs based on the new facts.<sup>23</sup>

One characteristic of rule-based systems is that inspection of their internal processing is straightforward. If a decision or output is questioned or determined to be incorrect, the application of rules and facts can be traced for a given input. If found to be erroneous, the knowledge base can then be amended if necessary. This capability aids in the development and debugging of systems as well as the post-hoc justification of processing outcomes.

However, rule-based systems can be brittle, tending to ‘fail badly for problems even slightly outside their area of expertise and in unforeseen situations’.<sup>24</sup> If circumstances change, then rules and/or facts in a knowledge base need to be revised or rewritten. As this is done by domain experts together with those skilled in the expert system, adapting a rules engine developed in one context to another can be a time-consuming and expensive exercise. Part of the reason for this brittleness is that rule-based systems do not typically *learn* per se – even if the additional facts and inference outputs are added to the knowledge base. For this reason, the focus in AI has swung from mechanisms for formalising knowledge to various techniques of *machine learning*,<sup>25</sup> including statistical models and deep learning networks. Nonetheless, rule-based expert systems are employed today, albeit in well-defined procedural and/or slowly changing environments.

### **Statistical models**

Rather than hard-coding rules that express domain expertise, an alternative approach is to identify patterns in input data and their mapping to desired outputs.<sup>26</sup> Given a sufficient number of high-quality samples, a statistical relationship between inputs and outputs can then be developed; and this statistical model may then be applied to new inputs. The *representation* of inputs in a manner suitable for statistical analysis underpins machine-learning design; each piece of information identified from an input is known as a *feature* and a set of features is known as a *feature vector*. The choice of representation and selection of features is a key task in developing machine-learning algorithms. For example, in image processing, features are likely to be the characteristics of individual or groupings of pixels (colour, brightness and so on, or, at a higher level of granularity, edges and shapes and so on). In document analysis, features may comprise the existence and relative locations of particular words, phrases or other grammatical elements as well as metadata clues. In medical diagnosis, features may comprise the existence or not of symptoms or pathologies.

Thus, it should be emphasised that such statistical models do not ‘understand’ the images, documents or medical conditions per se. Rather, the data is pre-processed to transform the input image, document or medical case (for example) into a suitable representation as a set of (numerical) features amenable to statistical interpretation.<sup>27</sup> Note too, that the term *machine-learning algorithm* concerns the manner in which the features are interpreted by the statistical model rather than any semantic sequence by which outputs are derived, as is the case with rule-based systems. Investigating the reasons for a particular output involves understanding how features contribute to the

model rather than following the model's computational logic, which largely concerns numeric processing.

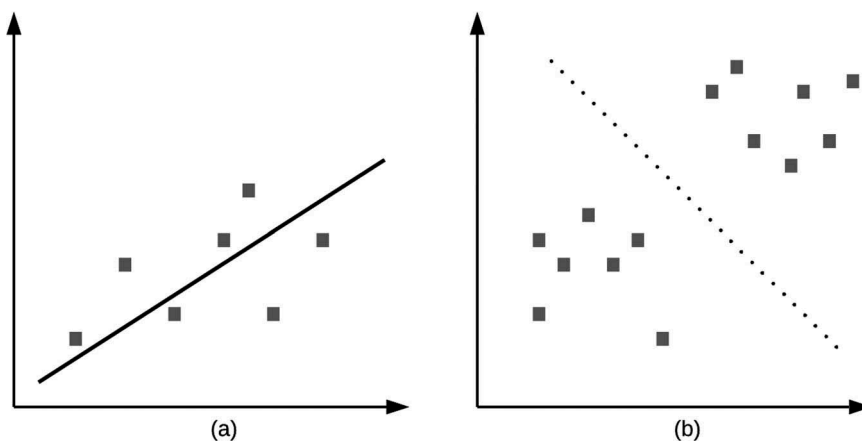
Machine learning can be used for a number of problem types including classification, transcription (for example, speech to text, optical character recognition), translation, prediction, anomaly detection, noise/outlier reduction and the synthesis of new cases.<sup>28</sup>

*Classification* – the task, perhaps, of most interest to recordkeeping knowledge work today – involves determining which of an a priori set of classifications apply to an input case. Such classification can be binary (for example, is/is-not SPAM, perform/not-perform medical procedure, preserve/destroy record), 1-of-N classes or multi-valued (for example, work is on Subject X, record is covered by schedule M, image contains elements A, B & C).<sup>29</sup>

The actual statistical mechanisms employed in such models are beyond the scope of this article, but conceptually they are similar to the exercise of fitting a line to data points as shown in [Figure 1\(a\)](#), whereby mathematical techniques are used to determine a formula that best fits the data. The presentation of sample data to the machine-learning algorithm for analysis and building of the model is termed *training*. In this example, each training data point relates two features (say  $x$  &  $y$ ) and the resulting formula is a straight-line relationship that can provide a 'best'  $y$ -value for a new  $x$ -input. In reality, there will typically be more than two dimensions: the feature vector will be much larger, and the mathematics more complex.

If the sample data includes the target output so that the algorithm can determine the 'formula of best fit' – for example, the feature vector derived from a set of documents and associated classification terms – then the training is termed *supervised*.<sup>30</sup> The aim of the machine-learning algorithm in this case is to learn how to match the supplied inputs to the supplied target results. Once trained, the system – when presented with a new case represented as a feature vector – runs the 'formula' developed during training against this new numeric data and then translates the result back into the matching target result, for example, matching a new document to classifications used in the training data.

A different kind of problem concerns modelling the structural properties of input data – particularly if each input comprises many features. In this case, the sample data



**Figure 1.** Examples of fitting mathematical formulae to data points.

is not labelled with the target results and such learning is termed *unsupervised*.<sup>31</sup> An example of unsupervised learning is *topic-modelling*, whereby features of the input documents are analysed to discern *clusters* of similar text that are then interpreted as topics against a standard vocabulary of terms. In the simple example of Figure 1(b), the mathematical techniques are used to distinguish clusters of two-dimensional training data points. Again, in real life, the feature vectors will have many more than two dimensions.

An extreme version of this sort of system is IBM Watson,<sup>32</sup> which was famously trained on 200 million pages from 600,000 sources including encyclopaedias, dictionaries, thesauri, Newswire articles, literary works and so on, to play (and win) the quiz game Jeopardy.<sup>33</sup>

With statistical models, then, the domain expertise is embodied firstly through the identification of a suitable representation and feature set, and secondly by preparation of suitable training data. The preparation of quality training data (whether supervised or unsupervised) that is germane to the specific task and problem domain, together with the iterative processing and adjustment of feature weightings and other parameters required during training to build the model, is a significant part of the effort and cost of developing a machine-learning application. On the other hand, even though statistical models may be more difficult to develop, debug and interrogate, they do provide a built-in capability for reporting confidence in a given output. For example, outputs may comprise ranked decisions or classifications, together with an estimate of correctness leading to greater understanding of the meaning or relevance of results. This is discussed in more detail below.

### **Deep learning models**

One of the challenges in machine learning is the identification of features that directly influence the output results. Often, such features are not found directly in raw data but are higher-level abstractions. For example, facial recognition does not depend on the nature of individual pixels in an image, but on features such as eye colour and separation, jaw-line angle and so on. Manually identifying, extracting and weighting such features can be very difficult,<sup>34</sup> even for domain experts – particularly when dealing with highly variable input data such as images or documents.<sup>35</sup> *Deep learning* systems, designed to classify patterns based on sample data,<sup>36</sup> address this problem by being able to extract features out of raw data; for example, from pixels in an image, or words in a document.

Deep learning systems do this through multiple, highly interconnected layers of large numbers of processors that abstract higher-order features from simpler ones at the layer below – in effect, learning the abstract or complex features that may be discerned from the data. Deep learning systems are most often implemented using *neural networks*, so-named due to their implementation similarity to the brain's structure of highly interconnected neurons. For example, a simplistic deep learning neural network model may have an input as an array of pixels, a first layer that detects edges given the pixels, a second layer that detects corners and contours given edges, a third layer that recognises shapes given corners and contours, and an output layer that labels objects.<sup>37</sup>

Deep learning has been remarkably successful in solving problems in a wide variety of domains.<sup>38</sup> For example, Google has recently introduced deep learning algorithms into its Internet search service alongside existing rule- and statistical-based models, and has promoted the previous head of AI to the search portfolio.<sup>39</sup> Nonetheless, deep learning approaches have their drawbacks. As with statistical models, the application of domain expertise in deep learning is relegated to the preparation of quality training data and working with deep learning specialists to tune the models. The training of deep learning models can be supervised or unsupervised as with statistical models, however deep learning requires more training data than statistical models. For example, in 2016 it was reported that ‘a rough rule of thumb is that a supervised deep learning algorithm will generally achieve acceptable performance with around 5,000 labelled examples per category and will match or exceed human performance when trained with a dataset containing at least 10 million labelled examples’;<sup>40</sup> Google Street View’s recognition of street numbers was initially trained on a set of 200,000 images;<sup>41</sup> deep learning model accuracy increases when training sets are increased from 1 million to 300 million samples.<sup>42</sup>

Significantly though, deep learning systems may be even harder to develop, tune and debug than other types of automation due to their opacity. The ever-increasing number of layers (hundreds) and nodes (millions) of deep learning systems,<sup>43</sup> together with their complex interconnectedness and parameters, result in ‘black box’ systems with workings that are difficult to interpret.<sup>44</sup> Despite the success and ubiquity of deep learning systems, the extent of this significant issue is being recognised and its effects slowly being addressed from social, methodological and technological perspectives.<sup>45</sup> An even greater problem is that contemporary deep learning systems can be very brittle, particularly when dealing with textual data.<sup>46</sup> However, research suggests that, given enough training data, this second issue may be solved over time.<sup>47</sup>

## Applying AI to recordkeeping

Given the above taxonomy of AI techniques, how may they be applied to recordkeeping knowledge work? AI systems have the potential to automate processes such as classification and disposal, and some academic research has been performed in this area.<sup>48</sup> Simple rule-based triggers may be used to automate records closure scheduling within an EDRMS.<sup>49</sup> More complex AI is also being used for textual classification and analysis in the digital humanities, for example the work done to explore the Irish Ryan Report of the Commission to Inquire into Sexual Abuse.<sup>50</sup> There are also commercial records management products available that use machine learning, such as auto-classification, aimed at improving recordkeeping knowledge work.<sup>51</sup> A notable initiative has been reported by the National Archives, UK regarding the exploration of technology that could assist with born-digital records collections in appraisal, selection and sensitivity review processes when transferring records. This trial focused on eDiscovery tools (software designed for the discovery or disclosure of digital information for the purposes of lawsuits).<sup>52</sup> The report concludes that eDiscovery software can assist and support government departments during appraisal, selection and sensitivity review as part of digital transfers. Beyond these examples, however, the literature is rather thin on the ground, and there are few clear success stories being trumpeted.



Perhaps barriers to the uptake of AI technology for recordkeeping knowledge work include:

- A lack of compelling case studies – maybe this is the hype-cycle effect, but while commentary abounds,<sup>53</sup> there are not many real-world examples within the academic or professional literature.
- The cost and time needed to configure machine-learning solutions – in particular, developing large, clean and labelled training and testing sets of data (once trained, systems require testing on different data before being applied in production), especially as changes occur over time.<sup>54</sup> The need for large datasets and sufficient computing resources could also be barriers, especially with smaller organisations.
- The difficulty of integrating the technology with complex human-oriented tools such as retention and disposal authorities. When applying these authorities to sentence records for retention or disposal, familiarity with the content of records and knowledge of relevant contexts is essential.<sup>55</sup> Some classes require deciphering what is a ‘significant’ or ‘major’ record from a class of records. Common sense also plays a part in sentencing records. Retention and disposal authorities are often written to be format neutral, but are they really?<sup>56</sup>
- The heterogeneity of recordkeeping data (and metadata) across files, series, collections and institutions that presents difficulties in generalising automation initiatives (see below).

If ‘a proper machine learning application is built combining access to data, access to domain expertise and access to the data scientists who developed the algorithms’,<sup>57</sup> it may be no wonder that there has not been much progress in our field. While recordkeeping institutions certainly have the data – even if not cleaned and labelled in a format for AI training – perhaps it is direct connection to the data science sector that is lacking.

### **Asking the right questions**

Another problem may be in the difficulty of posing the right questions; framing system requirements in a manner that is amenable to solution by AI. Consider for the moment, a ‘simple’ binary classification for the appraisal/sentencing process – keep or destroy? Due to the complexity of automating the understanding of context,<sup>58</sup> perhaps the negative question is easiest: what is definitely *not* a record of value? The consideration of context to answer the affirmative version of the question is much harder to frame as a representation – take for example, the two-word ‘well done’ email from the Iran/Contra scandal of the 1980s.<sup>59</sup>

Further, there is the brittleness mentioned above. The *no free lunch theorem* articulates limits to the portability of current AI models,<sup>60</sup> even to ostensibly similar datasets. Egregious examples of the impact of biased or off-the-mark image training data include embarrassingly offensive labels for images of African Americans applied by Google, or sexist labels assigned by Microsoft and Facebook.<sup>61</sup> Perhaps, even a binary classifier AI model developed for one context may not be applicable to another domain (even one that is seemingly close), and we need to be vigilant for introduced assumptions, bias or errors if we suffer scope creep or subtle changes in context.<sup>62</sup>



Classification of material into 1-of-N classes (for example, schedule class IDs) is more complex still.<sup>63</sup> Some *natural language processing* (NLP) work has been done as part of the Bitcurator-NLP project, which aims to ‘develop software for collecting institutions to extract, analyze, and produce reports on features of interest in text’.<sup>64</sup> Similarly, some work has been performed on the extraction of simple bibliographic metadata from highly constrained information sources.<sup>65</sup> However, the generation of recordkeeping metadata for responsive, dynamic documentation, sentencing and access may require complex, interdependent, multi-stage, deep learning algorithms that remain still within the realm of fundamental AI research.

### **Cloud services**

While there are freely available tools, engines and systems for AI development and execution – from rules engines to deep learning neural networks – the hardware requirements for AI solutions could be off-putting for institutional use. Depending on the application, the run-time hardware requirements can be quite modest (for example, the facial recognition feature in camera phones) but the training of the system can take considerable resources, particularly in the case of deep learning systems. For example, at the high end are systems such as IBM’s Watson, comprising 90 servers in 10 racks totalling 2880 cores and 16 TB of RAM,<sup>66</sup> not to mention the engines that run Facebook, and Google’s services. Of course, workloads that are more modest would require less computing power. Interestingly, though, it turns out that the type of graphical processing unit (GPU) hardware that has evolved to support intensive graphics processing on desktop and even portable computers (for example, for gaming, animation or video effects) is equally applicable to AI workloads due to its ability to perform parallel mathematical operations.<sup>67</sup> The wide availability of such hardware at consumer pricing has meant that AI techniques can start to be cost-effectively achievable at the scale necessary for acceptable performance and results.

Particularly in the case of deep learning, however, institutions may elect to avail themselves of cloud-based platforms for AI systems development and deployment – though at some risk for data governance (see the NSW State Archives case study below). Most of the larger vendors of cloud services such as Google, Microsoft, Amazon, IBM, Oracle and so on provide AI development and deployment platforms as part of their cloud-computing service offerings with a range of pricing and performance options. There are also a number of contender start-ups that provide more tailored platforms for vertical markets.

### **Measuring results**

Finally, it is worth discussing the ways in which success (or otherwise) of an AI initiative may be measured. Perhaps the simplest measure is an *accuracy* metric, for example: ‘determinations are 85% correct when compared to human results’. Such a metric should be attainable from straightforward measurement of results. It is worth pointing out that human accuracy rates are rarely 100%,<sup>68</sup> and so a realistic accuracy threshold informed by real-world human performance should form the basis for AI system design and/or evaluation. Accuracy (and indeed the other metrics described

below) should be determined for data in the training set (how well the model fits its training data) as well as for data in a testing set (how well it deals with data that it has not seen before).<sup>69</sup>

It is also useful to understand how such an accuracy figure is determined – particularly the incidence of Type-I (false positive) and Type-II (false negative) errors.<sup>70</sup> Perhaps, in recordkeeping, when determining what to preserve, a Type-I error (that is, a record retained when it need not be) is less egregious than a Type-II error (that is, a required record is destroyed)<sup>71</sup> and so, perhaps, accuracy for both cases should be reported. The metrics from information retrieval can be applied here, with measurement of *precision* (the proportion of returned results that are ‘correct’) and *recall* (the proportion of ‘correct’ results that are returned)<sup>72</sup> providing insight as to how well the system is performing.<sup>73</sup> Sometimes this is expressed as an *F<sub>1</sub>-Score*, this value being the average (harmonic mean) of the precision and recall measurements.

Additionally, many statistical (and some deep learning) algorithms can provide confidence levels of their outputs. Rather than absolute numbers, an indication of *coverage* can give additional insight to the meaning of measures related to recall and precision figures. Coverage relates to the proportion of decisions (or outputs) that are obtained at an acceptable level of confidence.<sup>74</sup> For example, the Google Street View project described above achieved 95.64% coverage at 98% accuracy (with 98% being the performance of human operators).<sup>75</sup> The ranking of candidate outputs may also help in understanding the confidence level of outputs and their alternatives.<sup>76</sup>

## Australian initiatives

Having provided a brief background to the field of artificial intelligence and some considerations for its application to recordkeeping knowledge work, we now present four case studies from state and national institutions that highlight some current initiatives.

### ***Public Record Office Victoria – machine-assisted email appraisal, Proof of Concept***

The first case study concerns how Public Record Office Victoria (PROV) is tackling the problem of email appraisal. Email plays a crucial role in how organisations exchange ideas, and enact and document decisions. The volume and unstructured nature of email makes its management, disposal and sensitivity review difficult. IBM’s Lotus Notes (LN) product has been used across Victorian Government to manage email since the mid-1990s. Over 20 years of routine backups have resulted in an unwieldy backlog amounting to 67,000 tapes and 28 petabytes of content. This backlog can no longer be efficiently queried to respond to requests for information – compromising the Government’s reputation for transparency and accountability.

#### ***The approach – eDiscovery***

The purpose of the Proof of Concept (PoC) project was to explore the use of an eDiscovery tool to appraise large volumes of email. eDiscovery tools are used to perform Technology-Assisted Review of documents for legal purposes, using a variety of rule-based, statistical models, and deep learning techniques in order to ‘find as nearly all

of the relevant documents in a collection as possible, with reasonable effort'.<sup>77</sup> This is an example of binary classification described earlier.

While eDiscovery approaches have been used elsewhere in recordkeeping contexts, this PoC was unique in its combination of the Nuix tool,<sup>78</sup> volume of data (1.5 TB) and focus on the email file format. To deliver the PoC, PROV worked with a partner agency in order to obtain both a meaningful sample and valuable input into the development of the appraisal criteria. Collaboration with a technical consultant was also vital to gain the required skills to apply and configure the eDiscovery tool's functionality. The approach was iterative, with this first stage run over six months with minimal staffing and computing outlay.

Out of the box, Nuix was used to perform:

- technical appraisal, to better understand the composition of the dataset (custodians, format, volume, creation date and so on); and
- de-duplication, applying an MD5 hash to tag identical emails (roughly 40% of the 4.6 million emails in the dataset were found to be duplicates).<sup>79</sup>

To find emails worth retaining, three approaches applied with Nuix were tested on the dataset:

(1) *Positive* – to identify valuable emails. Search terms comprised:

- (a) partner agencies' role definitions;
- (b) action verbs/objects; and
- (c) function/activity terms.

(2) *Negative* – to identify low-value emails:

- (a) emails from non-relevant addresses and domains were excluded.

(3) *Macro* – to assess and apply additional metadata to the emails according to their functional contexts, utilising organisational charts to understand the roles of email creators.

The positive approach identified 93% of the de-duplicated emails as business-related records (albeit with many false positives). The negative approach identified 7% of the emails as non-records – possibly indicating high precision but lower recall of this negative case.

## Results

The project team concluded that the lowest-risk/highest-benefit approach was a multi-layered approach comprising:

- de-duplication to reduce the volume of email;
- identification of non-valuable emails with the Negative approach; and
- identification and application of additional metadata as part of the Macro approach.

The PoC demonstrated that the Nuix eDiscovery tool could effectively be used to reduce the volume of email needed to be analysed by PROV for appraisal. It also showed that a technology-assisted workflow could enhance the manageability and discoverability of email that had previously been difficult to process from within historical LN file formats and backups.

The tool can also be used to perform appraisal tasks and detect the presence of sensitive and protected information. However, it is not a completely automated solution; as with traditional appraisal, a knowledge of the business and its functions was critical in constructing meaningful searches, as was manual sampling to assess accuracy. Looking forward, the tool could be run periodically to prevent further backlogs occurring; applied over a broader range of documents, that is, shared drives and other repositories; and used to explore more automated, AI approaches to appraisal.

### **NSW State Archives and records**

The second case study documents a NSW State Archives (NSWSAR) internal pilot that the Digital Archives team conducted in November and December 2017.<sup>80</sup> The goal of this pilot was to apply off-the-shelf machine-learning software to the problem of classifying a corpus of unstructured data against a retention and disposal authority. The main aim was to test machine-learning algorithms on a corpus of records that had previously been manually sentenced against a disposal authority. With what level of accuracy could multi-valued classification be performed to automatically match the corpus against the same disposal classes?

#### **Preliminary set-up**

The internal pilot was constrained by limited resources and there was no specific budget for the project. However (and very fortunately), an ICT graduate placement was available who had recent university experience in machine learning. To identify suitable technologies for the pilot, the project team investigated low-cost, off-the-shelf solutions.

The first product explored was Microsoft Azure's cloud-based AI and Cognitive Services. This platform has pre-built algorithms and classifiers and a great interface: the 'Machine Learning Workbench'. Although this option looked very promising, it ultimately had to be ruled out for this project because of uncertainties about the retention and management of data submitted to the platform. While many Microsoft Azure services are available locally in New South Wales, AI and Cognitive Services were only available on offshore servers.<sup>81</sup> Interestingly, since the pilot was conducted Microsoft Azure Cognitive Services' terms and conditions of agreement have been updated with changes that are now more aligned with other Azure storage services and would now be acceptable.

However, at the time of the pilot, these conditions meant that a risk assessment was required to meet the requirements of *Transferring Records out of NSW for Storage with and Maintenance by Service Providers Based Outside of the State* (GA35).<sup>82</sup> Key steps in this assessment included ensuring that records stored on the service would be managed in accordance with the *State Records Act 1998* (NSW) and standards and also vetting contractual arrangements to ensure that ownership of the records was retained by the

State and that, once the project concluded, all records would be returned – see *Storage of State Records with Service Providers Outside of NSW*.<sup>83</sup>

Because the internal pilot involved a corpus of closed records that had been transferred as State Archives, and because the project time-frame precluded a more detailed risk assessment (that may have involved following up with vendors for more detailed information and clarification and/or seeking legal advice), the project decided against any cloud-based solution and looked instead for off-the-shelf software that could be run locally.

The project quickly settled on scikit-learn,<sup>84</sup> a free and open-source machine-learning library for the Python programming language. This is a simple and accessible set of tools that provide a wide variety of rule-based, statistical models, and deep learning algorithms. Like Microsoft's Cognitive Services, scikit-learn includes pre-built classifiers. The decision to avoid cloud-based services raised the question of finding sufficient computing power to train and run the various AI test cases. Fortunately, the project had access to a high-powered machine for this purpose.<sup>85</sup>

### *The corpus*

The records that were chosen for the internal pilot had been transferred to the Digital State Archive in 2016 by a central government department. This corpus was unusual in that it contained a complete corporate folder structure extracted from an Objective EDRMS. The full corpus comprised 30 GB of data, in 7561 folders, containing 42,653 files; no disposal rules had been applied to the files. In a joint effort with the department the corpus was manually sentenced (at a folder level) against the General Retention and Disposal Authority Administrative Records (GA28),<sup>86</sup> resulting in a total of 12,369 files Required as State Archives.

The following options were considered for the internal pilot:

- Apply all Required as State Archives classes from GA28 (75 in total). Folders that didn't fit these classes would remain unclassified.
- Apply the subset of Required as State Archives classes that had been manually identified in the corpus (23 in total). Folders that did not fit these classes would be excluded from the corpus.
- Apply all of the GA28 classes (686 in total). This would require a complete test of all folders.
- Pre-treat the corpus by removing all folders which would be covered by Normal Administrative Practice (NAP), for example, duplicates or non-official/private records.

The decision was made to pre-treat the corpus and remove all folders which would be covered by NAP and to take the subset of 12,369 files that were identified as being Required as State Archives, which used only 23 classes of GA28. Further manual preparation of the subset involved assigning the classification from the folder level at the level of the individual files. Refer to [Table 1](#) for details.

### *Text-extraction steps*

**Text extraction.** To be usable, the documents chosen for analysis need to be easily text-extractable. This was to ensure performance and ease of conducting further text manipulation later in the project. Only 8784 of the 12,369 files which were classified

**Table 1.** Breakdown of the NSW corpus.

Dataset	No. of files
Complete corpus	42,653
NAP (Normal Administrative Procedures)	25,643
Corporate file plan	17,307
Required as State Archives	12,369
Usable sample set: Required as State Archives and suitable format	8784

As Required as State Archives were selected for use because their file types allowed simple text extraction. After sorting the sample set, a Python program using various libraries was developed to extract text from PDF, DOCX and DOC file types. The text that was extracted from documents was then placed within a single Comma Separated Value (CSV) file. The CSV file was divided into three columns: the file name (unique identifier), classification (GA28 class) and lastly the text extract.

**Data cleaning.** A very basic approach to data cleaning was taken. The following filters were applied: removal of document formatting; removal of stop words; removal of documents that are not required; and conversion of all letters to lower case.

**Text vectorisation and feature extraction.** Text vectorisation is the process of turning text into numerical feature vectors. For this pilot, the Bag-of-Words approach was used for text vectorisation. Bag-of-Words is a simple model that disregards the location of words within documents but instead focuses on presence and frequency of individual words and considers each unique word as a feature. This approach can represent any document as a fixed-length vector corresponding to the set of unique words known as the *vocabulary* of features. Each position for the unique word in the feature vector is filled by the frequency of the particular word appearing in that document, thus creating a document-term matrix that describes the frequency of terms that occur in a collection of documents.

For example, suppose a vocabulary includes the following words:

- brown, dog, fox, jumped, lazy, over, quick, the, zebra

Then, given an input document:

- the quick brown fox jumped over the lazy dog

The resultant feature vector is shown in Table 2.

One issue with using simple word frequencies in feature vectors is that some often-recurring words with large frequency values may not be meaningful to the vector representations of documents. An alternate metric, Term Frequency Inverse Document Frequency, works by calculating the term frequency (frequency of a particular word within a document) and then multiplying it by the Inverse document frequency.<sup>87</sup> This helps decrease the weighting of words that appear too frequently in the document set in favour of unique or unusual words.

**Table 2.** Example feature vector.

[illegible]

Classification

The project elected to compare two widely used machine-learning classification algorithms against the test data: Multinomial Naïve Bayes and the Multi-Layer Perceptron.

- (1) Multinomial Naïve Bayes (MNB) is a statistical model algorithm that is part of a family of simplistic probability-based classifiers. This classifier is based on Bayes’ theorem, which strongly assumes independence between features.
- (2) The Multi-Layer Perceptron (MLP) is a form of deep learning network that may be used for classification or regression.

Both cases employed supervised learning for training their models.

Two versions of the corpus were used with each of the algorithms: a cleaned version and an original version. The corpus was split 75%/25% for training and testing data. To begin with, 75% of the pre-classified Required as State Archive content was used to train each algorithm. Once trained, the same algorithm and model was used to process the 25% test set and obtain classifications that were then compared with the original manual sentencing determinations.

Results

The results from the four cases (clean/original corpus, MNB/MLP algorithm) are shown in Table 3, which includes the feature vector size, the training time, the accuracy measure and the F<sub>1</sub>-Score for each case.

The resulting statistics indicate that the Multi-Layer Perceptron algorithm with cleaned data was the most successful, with a maximum of 84% success rate in the test portion of the corpus. These results suggest that this technology is capable of assisting with the classification and disposal of unclassified, unstructured data. While 84% is probably not yet human-level accuracy (though the actual human accuracy rate in this case is not known), this was a good result given that it was obtained from a comparatively brief pilot, and simple models comprising approximately 100 lines of code. In particular, the corpus used was manually sentenced at folder level with only a sampling of individual documents, whereas the model was able to sentence directly at document level more quickly. However, the dependency of the result on the accuracy of representation and training should be emphasised, as any error or bias introduced into the training data during sentencing will only increase in the model over time.

Table 3. NSW results.

Multinomial Naïve Bayes				Multi-Layer Perceptron			
No data cleaning		Cleaned data		No data cleaning		Cleaned data	
<b>Features:</b>	<b>5000</b>	<b>Features:</b>	<b>5000</b>	<b>Features:</b>	<b>5000</b>	<b>Features:</b>	<b>5000</b>
Accuracy:	65.4%	Accuracy:	69%	Accuracy:	77%	Accuracy:	82.7%
F1 score:	0.624	F1 score:	0.648	F1 score:	0.767	F1 score:	0.812
Train time:	109ms	Train time:	108ms	Train time:	2m 23s	Train time:	2m 43s
<b>Features:</b>	<b>10,000</b>	<b>Features:</b>	<b>10,000</b>	<b>Features:</b>	<b>10,000</b>	<b>Features:</b>	<b>10,000</b>
Accuracy:	64%	Accuracy:	68%	Accuracy:	78%	Accuracy:	84%
F1 score:	0.622	F1 score:	0.638	F1 score:	0.777	F1 score:	0.835
Train time:	111ms	Train time:	109ms	Train time:	3m 28s	Train time:	4m 02s



### ***National Archives of Australia***

The third case study is a research project at the National Archives of Australia (NAA) to investigate how to create and issue disposal and retention authorisations in a format that supports digital business in the Australian Government. Currently the National Archives, as with many other archives around Australia and throughout the world, issues general and agency-specific records authorities that document retention and disposal decisions. These traditional archival tools were designed within and for a paper-based environment and with the intent that humans execute any decisions. Truly digital disposal authorisations should enable records creators to automate such decisions and allow digital systems that manage information to execute them with the minimal involvement of humans.

Since 2015, a team of four has been working part-time on this project, learning throughout the process. Several consultants and vendors were engaged at different times to run workshops, which helped the team to become familiar with available technology and to test theoretical and methodological approaches to appraisal. The Archives team looked into entity and automated metadata extraction, semantic analysis, taxonomy and ontology building, and linked data approaches. Work done internationally, by W3C on the PROV Data Model and by the International Council on Archives on 'Records in Contexts – Conceptual Model',<sup>88</sup> has also stimulated the thinking of the team.

The first stage of the project resulted in a conceptual model and semantic analysis of the disposal authorisations. The schema is represented in xml and contains many elements similar to the existing retention and disposal schedules in that it comprises:

- (1) A header with contextual information about the disposal authorisation and government agency or agencies to which they are issued.
- (2) Functions at the highest level of business, establishing the context for the records created to document and support associated activities.
- (3) Disposal classes and, within those, criteria which would, theoretically, allow automated linking of a record with the appropriate disposal class and action.

These criteria would be composed of the so-called test groups and some work has been done to analyse the language used in formulating these in current disposal classes.

The next step will be to test these theoretical assumptions in practice on the National Archives' own core business records authority, which is more than 10 years old and is due for revision. With advice from specialists in machine learning and artificial intelligence, the project will test and either demonstrate or disprove the validity of our assumptions. The project expects to evaluate the creation and implementation of various types of machine learning, including auto-classification, clustering, and indexing tools. It is expected that this stage will be a pilot for a future digital approach to all agency records authorities.

### ***The Australian Government Department of Finance***

The final case study is an ongoing PoC project led by the Australian Government Department of Finance (DoF). This is a 12-week research initiative that began in

February 2018 to test the application of microservices architecture and linked data technologies for automating records management. In this instance, it will attempt to capture data from an agency's email server.

We intent to create prediction models that will automatically identify the ongoing business value of the captured email data and to interlink these records using features from the Australian Government Records Interoperability Framework.<sup>89</sup> The captured data will be classified against retention schedules. Testing of the classification results will be performed by Agency records managers, who will confirm that the correct disposal classes have been applied.

It should be emphasised that this is a research initiative leading to a demonstration of the PoC system. We expect that the market will provide any future records-management solutions for government.

### **Implementation**

The pilot agency's business activities and existing records will be analysed and user stories that describe the processes and information artefacts associated with the business will be identified.

These user stories will drive the modelling of the pilot agency's business, resulting in reusable machine-and-human-readable artefacts that can be used by the demonstration services to categorise and support discovery. We expect that this modelling will be able to be extended as needed, with the products created as features of the demonstration.

The microservices will be deployed on Amazon Web Services, and will have a 'light touch' interaction with pilot agency's Exchange server and Office instances. This light touch will include a web interface for performing tasks such as discovery and access to records for use by general users, as well as the ability to support management activities for agency records managers. The use of microservices means that it is possible to compose new services from microservice components should there be a need for a service that was not in the original demonstration scope.

The demonstration system will be based on the following technologies:

- Netflix open source technology stack for microservices;
- Big data technologies such as Apache Spark and Scala for machine learning;
- Facebook open-source technology REACT.js for user interface implantation; and
- Amazon Web Services for cloud service provision.

### **Conclusion – AI in the archive**

Anecdotal evidence, together with these case studies, suggests that there is significant interest in the area of AI technologies and applications that will impact on and produce benefits for recordkeeping knowledge work. While AI promises efficiencies in the support of digital appraisal, documentation and disposal, it appears to be an emerging capability and certainly not a production-ready 'silver bullet'. However, AI *has* arrived in our field and it will produce profound changes in our working environments in the years to come. Perhaps this is a symptom of the 'hype-cycle' effect that is also known as Amara's law: 'we tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run'.<sup>90</sup>

What the case studies highlight is that the introduction of AI technologies requires effort and, depending on the goals, may require significant groundwork. Both PROV and NSW SAR found that data preparation is necessary, for example data cleaning, deduplication,<sup>91</sup> developing toolchains and workflows. Toolkits such as those assembled by the Bitcurator Consortium or those described by Spencer may play an important role here – firstly, in the identification and preparation of suitable training and test data and, secondly, as a pre-processing step for data in the automated workflow.<sup>92</sup> Moreover, the NAA determined that archival processes themselves might need to change in order for technologies to enhance description, discovery and interoperability of archival information. Similarly, the DoF initiative has determined that fundamental analysis and redesign of a flexible microservices architecture is necessary for delivery of their services.

The NSW SAR case study, in particular, highlighted the necessity to adequately resource such initiatives: not only in terms of large training sets of classified data to achieve results over the test data, but also having sufficient computational power on local machines to process the model. That project also highlighted the potential risks in using cloud services, especially around issues of personal privacy of individuals and legal ownership of the data being processed – as well as the rapidly evolving nature of such services. Perhaps a first step when contemplating such an initiative is to prepare for the introduction of AI as described by Carlton Sapp.<sup>93</sup>

Importantly, it is apparent that there is a need to develop an understanding of AI concepts within the profession to ensure that recordkeeping needs are met into the future. Archivists need to take advantage of the opportunity presented by the emergence of this suite of new technologies to explore the potential of AI and to expose achievements and learnings to the broader professional community. As with the impact of earlier technological waves (data modelling, Web, Semantic Web and so on) on our profession, such retraining will be necessary to mitigate the knowledge and skills gap that is beginning to emerge between those information managers and archivists who are familiar with such concepts, and those who are not.<sup>94</sup>

To digress slightly, such familiarisation is also important from another standpoint. The opacity of AI algorithms directly impacts the kind of recordkeeping that may be performed in relation to transactions driven by such technologies. Certainly, the outcomes of decisions may be recorded but, crucially, not the detailed rationale for such decisions that form part of business documentation. The design of infrastructure using AI technologies needs to incorporate adequate recordkeeping or the current crisis will be exacerbated.<sup>95</sup> If we are to contribute to discussion about AI and recordkeeping we need to be informed and preferably experienced in AI techniques, to counteract technological arguments for non-compliance. Significantly, this also applies to our own housekeeping. In order to meet the documentation requirements for our own records – such as those described in ISO 23081 – we will need to understand how to deploy AI that contributes to documentation of recordkeeping business.<sup>96</sup>

Our goal should be good recordkeeping and to achieve it we must move with and learn new technologies. We must use them to achieve efficiencies where they are available and ensure that the accountability of our systems and processes is not compromised. We may be human, but the AI discussion has begun, and we all need to join in.

## Notes

- 1 Stanford University, 'Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence; Report of the 2015 Study Panel', 2016, available at <[https://ai100.stanford.edu/sites/default/files/ai\\_100\\_report\\_0916fml\\_single.pdf](https://ai100.stanford.edu/sites/default/files/ai_100_report_0916fml_single.pdf)>, accessed 14 February 2018.
- 2 National Archives of Australia and Council of Australasian Archives and Records Authorities, 'Digital Archiving in the 21st Century: Archives Domain Discussion Paper', Collections Council of Australia, 2006; Frank Upward, Barbara Reed, Gillian Oliver and Joanne Evans, 'Recordkeeping Informatics: Re-Figuring a Discipline in Crisis with a Single Minded Approach', *Records Management Journal*, vol. 23, no. 1, 2013, pp. 37–50, doi:10.1108/09565691311325013.
- 3 Ross Harvey and Dave Thompson, 'Automating the Appraisal of Digital Materials', *Library Hi Tech*, vol. 28, no. 2, 2010, pp. 313–22, doi:10.1108/07378831011047703.
- 4 John McDonald, 'Managing Records in the Modern Office: Taming the Wild Frontier', *Archivaria*, no. 39, Spring 1995, pp. 70–9.
- 5 Kate Cumming and Anne Picot, 'Reinventing Appraisal', *Archives and Manuscripts*, vol. 42, no. 2, 2014, pp. 133–45, doi:10.1080/01576895.2014.926824.
- 6 The National Archives UK, 'Digital Strategy', available at <<https://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf>>, accessed 14 February 2018.
- 7 Anne Gilliland, 'Archival Appraisal: Practising on Shifting Sands', in Caroline Brown (ed.), *Archives and Recordkeeping: Theory into Practice*, Facet, London, 2014, p. 50. Attributed to Clive Humby and, perhaps, first quoted in Michael Palmer, 'Data is the New Oil', ANA Marketing Maestros, 2006, available at <[http://ana.blogs.com/maestros/2006/11/data\\_is\\_the\\_new.html](http://ana.blogs.com/maestros/2006/11/data_is_the_new.html)>, accessed 14 February 2018; Upward et al., 'Recordkeeping Informatics'.
- 8 Cassie Findlay, 'Appraisal: An Essential Tool for Digital Recordkeeping', August 2015, available at <<http://www.informationstrategy.tas.gov.au/Publications/Documents/Cassie-Findlay-Appraisal-Essential-tool-for-digital-recordkeeping-Aug-2015.pptx>>, accessed 14 February 2018.
- 9 Frank Upward, Barbara Reed, Gillian Oliver and Joanne Evans, *Recordkeeping Informatics for a Networked Age*, Social Informatics, Monash University Publishing, Clayton, Vic., 2018, pp. xix–xx.
- 10 William Vinh-Doyle, 'Appraising Email (Using Digital Forensics): Techniques and Challenges', *Archives and Manuscripts*, vol. 45, no. 1, 2017, pp. 18–30, doi:10.1080/01576895.2016.1270838.
- 11 Anthony Cocciolo, 'Finding Inactive Records on Institutional Networks: An Evaluation of Tools', *Practical Technology for Archives*, June 2016, available at <[https://practicaltechnologyforarchives.org/issue6\\_cocciolo/](https://practicaltechnologyforarchives.org/issue6_cocciolo/)>, accessed 14 February 2018.
- 12 Victoria Sloyan, 'Born-Digital Archives at the Wellcome Library: Appraisal and Sensitivity Review of Two Hard Drives', *Archives and Records*, vol. 37, no. 1, 2016, pp. 20–36, doi:10.1080/23257962.2016.1144504.
- 13 Ross Spencer, 'Binary Trees? Automatically Identifying the Links Between Born-Digital Records', *Archives and Manuscripts*, vol. 45, no. 2, 2017, pp. 77–99, doi:10.1080/01576895.2017.1330158.
- 14 Pamela McCorduck, *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, AK Peters, Natick, MA, 2004, p. 523.
- 15 Stanford University, p. 12.
- 16 Edward Feigenbaum as quoted in McCorduck, p. 326.
- 17 Richard Bellman, *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser, San Francisco, 1978, p. 3.
- 18 Stanford University.
- 19 Gartner, 'Hype Cycle Research Methodology Gartner Inc.', 2018, available at <<https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>>, accessed 14 February 2018.

- 20 McCorduck, *Machines Who Think*, 423.
- 21 Ikujiro Nonaka and Hirotaka Takeuchi, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, New York, 1995, pp. 86, 154.
- 22 Anne J Gilliland. 'Designing Expert Systems for Archival Evaluation and Processing of Computer-Mediated Communications', in Anne J Gilliland, Sue McKemmish and Andrew J Lau (eds), *Research in the Archival Multiverse*, Monash University Publishing, Clayton, Vic., 2016, pp. 686–722. For more detail, seek access to Anne Jervois Gilliland-Swetland, 'Development of an Expert Assistant for Archival Appraisal of Electronic Communications: An Exploratory Study', PhD dissertation, University of Michigan, 1995.
- 23 Daniel G Bobrow, Sanjay Mittal and Mark J Stefik, 'Expert Systems: Perils and Promise', *Communications of the ACM*, vol. 29, no. 9, 1986, pp. 880–94.
- 24 Craig Stanfill and David Waltz, 'Toward Memory-Based Reasoning', *Communications of the ACM*, vol. 29, no. 12, 1986, pp. 1213–28.
- 25 Herbert Gelerntner, 'Realization of a Geometry-Theorem Proving Machine', in Edward A Feigenbaum and Julian Feldman (eds), *Computers and Thought: A Collection of Articles*, McGraw-Hill, New York, 1963, pp. 134–52.
- 26 Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning*, MIT Press, 2017, p. 2, available at <<http://www.deeplearningbook.org/>>, accessed 14 February 2018.
- 27 *ibid.*, p. 3.
- 28 *ibid.*, p. 98.
- 29 Steven Bird, Ewan Klein and Edward Loper, *Natural Language Processing with Python*, O'Reilly Media, Sebastopol, CA, 2016, p. 221.
- 30 Goodfellow, Bengio and Courville, p. 103.
- 31 *ibid.*
- 32 The following names mentioned in this article are registered trademarks<sup>(TM)</sup>: IBM Watson, Jeopardy, Google Street View, Facebook, IBM Lotus Notes, Nuix, Microsoft Azure, Microsoft Azure Cognitive Services, Python, Objective, Amazon Web Services, Microsoft Exchange, Microsoft Office, Netflix, Apache Spark and Scala
- 33 David Ferrucci et al., 'Building Watson: An Overview of the DeepQA Project', *AI Magazine*, vol. 31, no. 3, 2010, pp. 59–79; Jo Best, 'IBM Watson: The Inside Story of How the Jeopardy-Winning Supercomputer Was Born, and What It Wants to Do Next', *Tech Republic*, September 2013, available at <<https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/>>, accessed 14 February 2018.
- 34 Goodfellow, Bengio and Courville, p. 3.
- 35 Ian Goodfellow et al., 'Multi-Digit Number Recognition from Street View Imagery Using Deep Convolutional Neural Networks', arXiv Preprint arXiv:1312.6082, 2013; Li Deng and Dong Yu, 'Deep Learning: Methods and Applications', *Foundations and Trends in Signal Processing*, vol. 7, nos. 3–4, 2013, p. 219, doi:10.1561/20000000039.
- 36 Gary Marcus, 'Deep Learning: A Critical Appraisal', arXiv Preprint arXiv:1801.00631, January 2018.
- 37 Goodfellow, Bengio and Courville, *Deep Learning*, 6.
- 38 Stanford University; Goodfellow, Bengio and Courville, pp. 22–5.
- 39 Cade Metz, 'AI is Transforming Google Search. The Rest of the Web Is Next', *WIRED*, February 2016, available at <<https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/>>, accessed 14 February 2018.
- 40 Goodfellow, Bengio and Courville, p. 20.
- 41 Goodfellow et al.
- 42 Chen Sun et al., 'Revisiting Unreasonable Effectiveness of Data in Deep Learning Era', paper presented at IEEE, Venice, 22–29 October 2017.
- 43 Goodfellow, Bengio and Courville, pp. 21–3.
- 44 Davide Castelvecchi, 'Can We Open the Black Box of AI?' *Nature News*, vol. 538, no. 7623, October 2016, pp. 20–3, doi:10.1038/538020a.

- 45 For a discussion of the effects of these issues, see Joanna Redden and Jessica Brand, 'Data Harm Record', December 2017, available at <<https://datajusticelab.org/data-harm-record/>>, accessed 14 February 2018; and Cathy O'Neil, *Weapons of Math Destruction*, Penguin Books, New York, 2017. For initiatives that are looking at ways of addressing these problems, see FAT/ML, 'Home:: FAT ML', Fairness, Accountability, and Transparency in Machine Learning, 2017, available at <<https://www.fatml.org/>>, accessed 14 February 2018; Paul VoosenJul, 'How AI Detectives Are Cracking Open the Black Box of Deep Learning', *Science AAAS*, July 2017, available at <<http://www.sciencemag.org/news/2017/07/how-ai-detectives-are-cracking-open-black-box-deep-learning/>>, accessed 14 February 2018.
- 46 Robin Jia and Percy Liang, 'Adversarial Examples for Evaluating Reading Comprehension Systems', arXiv Preprint arXiv:1707.07328, 2017.
- 47 Marcus.
- 48 André Vellino et al., 'Assisting the Appraisal of E-Mail Records with Automatic Classification', *Records Management Journal*, vol. 26, no. 3, 2016, pp. 293–313, doi:10.1108/RMJ-02-2016-0006; Floriana Esposito et al., 'Machine Learning Methods for Automatically Processing Historical Documents: From Paper Acquisition to XML Transformation', paper presented at IEEE, Palo Alto, CA, 23–24 January 2004, doi:10.1109/DIAL.2004.1263262.
- 49 Kye O'Donnell, 'Taming Digital Records with the Warrior Princess: Developing a Xena Preservation Interface for TRIM', *Archives and Manuscripts*, vol. 38, no. 2, 2010, pp. 37–60.
- 50 Susan Leavy, Emilie Pine and Mark Keane, 'Mining the Cultural Memory of Irish Industrial Schools Using Word Embedding and Text Classification', paper presented at DH2017, Montreal, Canada, 8–11 August 2017, available at <<https://dh2017.adho.org/abstracts/098/098.pdf/>>, accessed 14 February 2018.
- 51 IBM, 'Auto-Classification Models', IBM, 2014, available at <[https://www.ibm.com/support/knowledgecenter/SSDUBN\\_7.5.1/Administrator/cpt/cpt\\_autoclassificationmodel.html](https://www.ibm.com/support/knowledgecenter/SSDUBN_7.5.1/Administrator/cpt/cpt_autoclassificationmodel.html)>, accessed 14 February 2018; Open Text, 'Auto-Classification for Records Management', OpenText, 2018, available at <<https://www.opentext.com.au/what-we-do/products/discovery/auto-classification>>, accessed 14 February 2018; Integro, 'Auto-Classification – Integro: Experts in Information Governance and Enterprise Content Management', 2017, available at <<https://www.integro.com/ecm-solutions/auto-classification>>, accessed 14 February 2018; Concept Searching, 'Auto-Classification, Taxonomy Management, Metadata Generation', Concept Searching, 2017, available at <<https://www.conceptsearching.com/>>, accessed 14 February 2018.
- 52 Jason R Baron, 'Toward a Federal Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery', *Sedona Conference Journal*, vol. 6, Fall 2005, pp. 237–9.
- 53 Nicholas Fripp, 'Is Machine Learning the Future of Records Management?' *IQ: The RIM Quarterly*, vol. 33, no. 1, 2017, pp. 22–3; George Parapadakis, 'A Clouded View of Records and Auto-Classification', For What It's Worth..., June 2013, available at <<https://4most.wordpress.com/2013/06/26/clouded-view-of-records-and-classification/>>, accessed 14 February 2018; Ronald Layel, 'Auto-Classification for RM – Beginning to See It as Possible – Association for Information and Image Management International', February 2012, available at <<http://community.aiim.org/blogs/ron-layel/2012/02/16/auto-classification-for-rm---beginning-to-see-it-as-possible>>, accessed 14 February 2018.
- 54 Tim Shinkle, 'Automated Electronic Records Management? Are We There yet? IDM Magazine', *Image and Data Manager*, December 2016, available at <<http://idm.net.au/article/0011369-automated-electronic-records-management-are-we-there-yet>>, accessed 14 February 2018.
- 55 Parapadakis.
- 56 For example, see Department of Corporate and Information Services, 'Guidelines for the Development of a Functional Records Disposal Schedule – Records Policy and Standards – NTG IT and Communications – Department of Corporate and Information Services,'



- Northern Territory Government, 2014, available at <[http://www.nt.gov.au/dcis/info\\_tech/records\\_policy\\_standards/records\\_disposal/index.shtml](http://www.nt.gov.au/dcis/info_tech/records_policy_standards/records_disposal/index.shtml)>, accessed 16 April 2018; Public Record Office Victoria, 'Retention and Disposal Authorities (RDAs)', PROV, 2018, available at <<https://www.prov.vic.gov.au/recordkeeping-government/how-long-should-records-be-kept/retention-and-disposal-authorities-rdas>>, accessed 16 April 2018; Queensland State Archives (Department of Science, Information Technology and Innovation), 'Use a Retention and Disposal Schedule', Queensland Government, 2017, available at <<https://www.forgov.qld.gov.au/use-retention-and-disposal-schedule>>, accessed 16 April 2018; State Archives and Records, 'Disposal Authorisation Procedures', State Archives and Records NSW, 2015, available at <<https://www.records.nsw.gov.au/recordkeeping/rules/procedures/disposal-authorisation>>, accessed 16 April 2018.
- 57 Erika Morphy, 'How to Differentiate Machine Learning from Dressed-up BI', CMSWire.com, January 2018, available at <<https://www.cmswire.com/digital-experience/how-to-differentiate-machine-learning-from-dressed-up-bi/>>, accessed 14 February 2018.
  - 58 Parapadakis.
  - 59 Harold Greene, 'United States V. Poindexter, 725 F. Supp. 13 (D.D.C. 1989)', Justia Law, October 1989, available at <<https://law.justia.com/cases/federal/district-courts/FSupp/725/13/1406938/>>, accessed 14 February 2018.
  - 60 Goodfellow, Bengio and Courville, p. 141.
  - 61 Tom Simonite, 'Machines Learn a Biased View of Women', *WIRED*, August 2017, available at <<https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>>, accessed 14 February 2018.
  - 62 Brian d'Alessandro, Cathy O'Neil and Tom LaGatta, 'Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification', *Big Data*, vol. 5, no. 2, 2017, pp. 120–34, doi:10.1089/big.2016.0048.
  - 63 Shinkle; Parapadakis.
  - 64 Bitcurator Consortium, 'BitCurator NLP', Bitcurator, 2018, available at <<https://bitcurator.net/bitcurator-nlp/>>, accessed 14 February 2018.
  - 65 Hui Han et al., 'Automatic Document Metadata Extraction Using Support Vector Machines', paper presented at IEEE, Houston, TX, 27–31 May 2003.
  - 66 Tami Deedrick, 'It's Technical, Dear Watson', *IBM Systems Magazine*, February 2011, available at <<http://ibmsystemsmag.com/ibmi/trends/whatsnew/it-s-technical,-dear-watson/>>, accessed 14 February 2018.
  - 67 And not just AI. GPUs are also being employed in cryptographic applications, most notoriously in the mining of virtual currencies.
  - 68 Layel; Vellino et al.; Goodfellow et al.
  - 69 Goodfellow, Bengio and Courville, p. 108.
  - 70 Judy Sheard, 'Quantitative Data Analysis', in Kirsty Williamson and Graeme Johanson (eds), *Research Methods: Information, Systems and Contexts*, Tilde University Press, Prahran, Vic., 2013, p. 408.
  - 71 Baron; Richard J Cox, 'The Documentation Strategy and Archival Appraisal Principles: A Different Perspective', *Archivaria*, vol. 38, Fall 1994, pp. 11–36.
  - 72 Philip Hider and Ross Harvey, *Organising Knowledge in a Global Society: Principles and Practice in Libraries and Information Centres*, Topics in Australasian Library and Information Studies 29, Centre for Information Studies, Charles Sturt University, Wagga Wagga, 2008, p. 191.
  - 73 Goodfellow, Bengio and Courville, p. 418.
  - 74 *ibid.*, p. 419.
  - 75 Goodfellow et al.
  - 76 Eugene Yang, David Grossman, Ophir Frieder and Roman Yurchak, 'Effectiveness Results for Popular E-Discovery Algorithms', paper presented at the 16th International Conference on Artificial Intelligence and Law, London, 12–16 June 2017.



- 77 Gordon V Cormack and Maura R Grossman, 'Evaluation of Machine-Learning Protocols for Technology-Assisted Review in Electronic Discovery', *ACM*, 2014, pp. 153–62, doi:10.1145/2600428.2609601. Also see Baron; Yang et al.
- 78 See Nuix, 'Electronic Discovery', Investigation, Cybersecurity, Information Governance and eDiscovery Software, 2017, available at <<https://www.nuix.com/problems-we-solve/electronic-discovery>>, accessed 14 February 2018.
- 79 For discussion of this method, see Spencer.
- 80 Glen Humphries, 'Machine Learning and Records Management', 14 September 2014, 2017, available at <<http://futureproof.records.nsw.gov.au/machine-learning-and-records-management/>>, accessed 2 March 2018.
- 81 At the time of the pilot, Microsoft's Trust Center included the following advice about use of its Cognitive services: 'Data that is sent to Cognitive Services is treated differently than other customer data. Microsoft may use Cognitive Services data to improve Microsoft products and services. For example, we may use content that you provide to the Cognitive Services to improve our underlying algorithms and models over time. To do that, we may retain Cognitive Services data after you are no longer using the services.' An additional clause, under the Privacy tab, stated: 'Cognitive Services collect and use many types of data, such as images, audio files, video files, or text, all of which may be retained by Microsoft indefinitely to improve Microsoft products and services, without a means for you to access or delete that retained data. Unless otherwise stated in documentation for a particular service, these services provide no means for you to store, access, extract, or delete customer data.'
- 82 State Records Authority of New South Wales, 'Transferring Records Out of NSW (GA35)', State Archives and Records NSW, November 2015, available at <<https://www.records.nsw.gov.au/node/649>>, accessed 14 February 2018.
- 83 State Records Authority of New South Wales, 'Storage of State Records with Service Providers Outside of NSW', State Archives and Records NSW, November 2015, available at <<https://www.records.nsw.gov.au/recordkeeping/advice/storage-and-preservation/service-providers-outside-nsw>>, accessed 14 February 2018.
- 84 Scikit-learn developers, 'Scikit-Learn: Machine Learning in Python', 2017, available at <<http://scikit-learn.org/stable/>>, accessed 14 February 2018. Also see Fabian Pedregosa et al., 'Scikit-Learn: Machine Learning in Python', *Journal of Machine Learning Research*, vol. 12, October 2011, pp. 2825–30.
- 85 The actual machine specifications were: HP Z440 Workstation, CPU: 2x Intel® Xeon® E5-1650 v3 (12 cores total), RAM: 64 GB DDR4, Storage: 2x Micron M600 1 TB SSD.
- 86 State Records Authority of New South Wales, 'Administrative Records (GA28)', State Archives and Records NSW, November 2015, available at <<https://www.records.nsw.gov.au/recordkeeping/rules/gdas/ga28>>, accessed 14 February 2018.
- 87 Formally, the Inverse Document Frequency =  $\text{Log}(\text{Total number of documents} / \text{Number of documents having the particular word})$ .
- 88 W3C, 'PROV-DM: The PROV Data Model', 2013, available at <<http://www.w3.org/TR/2013/REC-prov-dm-2013043>>, accessed 2 March 2018. International Council on Archives, 'Records in Contexts: A Conceptual Model for Archival Description', 2016, available at <<https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf>>, accessed 2 March 2018.
- 89 Available at <<https://www.finance.gov.au/archive/policy-guides-procurement/interoperability-frameworks/information-interoperability-framework>>, accessed 15 July 2018.
- 90 Matt Ridley, 'Amara's Law', November 2017, available at <<http://www.rationaloptimist.com/blog/amaras-law/>>, accessed 14 February 2018.
- 91 The identification and classification of 'duplicates' is not without its tensions. See Geoffrey Yeo, 'Nothing Is the Same as Something Else: Significant Properties and Notions of Identity and Originality', *Archival Science*, vol. 10, no. 2, 2010, pp. 85–116, doi:10.1007/s10502-010-9119-9.
- 92 BitCurator Consortium; Spencer.

- 93 'Preparing and Architecting for Machine Learning', Gartner, January 2017, available at <[https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/preparing\\_and\\_architecting\\_for\\_machine\\_learning.pdf](https://www.gartner.com/binaries/content/assets/events/keywords/catalyst/catus8/preparing_and_architecting_for_machine_learning.pdf)>, accessed 14 February 2018.
- 94 Richard Marciano et al., 'Archival Records and Training in the Age of Big Data', in Johnna Percell, Lindsay Sarin, Paul Jaeger and John Bertot (eds), *Re-envisioning the MLS: Perspectives on the Future of Library and Information Science Education*, Advances in Librarianship 44, Emerald Group Publishing, Bingley, UK, 2017, p. 210.
- 95 Upward et al., 'Recordkeeping Informatics'.
- 96 International Organization for Standardization, ISO 23081 Information and Documentation – Records Management Processes – Metadata for Records – Principles, International Organization for Standardization, 2006.

## Acknowledgments

The authors would like to acknowledge their project collaborators, without whom these initiatives would not have been possible: Richard Lehane and Malay Sharma – NSW State Archives and Records; David Header, Marian Kearney and Sean Wright – National Archives of Australia; and John Machin and Wicka Simet – Commonwealth Government, Department of Finance.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

Dr **Gregory Rolan** is currently a research fellow at the Centre for Organisational and Social Informatics at Monash University. Following a 30-year career in IT, Dr Rolan returned to study, obtaining a PhD in recordkeeping informatics, investigating participatory recordkeeping systems. His research comprises the design-science study of systems interoperability; conceptual modelling in recordkeeping informatics; metadata standards setting; and organisational/social factors in information systems design and implementation.

**Glen Humphries** is currently a project officer with the Digital State Archive at NSW State Archives and Records. Glen has previously worked for Archives New Zealand from 2006 to 2014 where he gained a wide knowledge of archival practices before moving to Australia in 2014. Glen joined the Digital State Archive in August 2015 and has been working on a number of digital transfers of various sizes and ages. Glen also currently has been leading pilot projects that look at the capabilities of machine-learning technologies and records management specifically at disposal of structured and unstructured data.

**Lisa Jeffrey** is a Melbourne-based information professional recently engaged at Public Record Office Victoria on a machine-assisted appraisal Proof of Concept. She has over 10 years experience as a records manager, archivist and information strategist across the private and public sectors (both Federal and State) and has completed graduate and postgraduate study at Monash University. Lisa is interested in how and why individuals self-document (or do not) in different contexts and how technology can support organisational and community memory.

**Evanthia Samaras** has worked in the Australian archive sector since 2013 and is currently the Victorian Electronic Records Strategy, Senior Officer at Public Record Office Victoria. She is presently undertaking a PhD at the University of Technology Sydney to explore how computer-generated imagery projects produced by the film visual effects industry can be archived and preserved for future use.

*Tatiana Antsoupova* is the acting Chief Information Governance Officer at the National Archives of Australia in Canberra. She has been working at the National Archives since 2005 and was Archives Officer at the Noel Butlin Archives Centre of the Australian National University from 1996 to 2005. Before that, she was a government archivist in Russia. She has a degree in archives and history from the Moscow State Institute of Archival and Historical Studies (now Moscow State University of Humanities) and spent one year at the University of Pittsburgh studying archives and records management with Professor Richard Cox.

*Katharine Stuart* works for the Australian Department of Finance as the project lead for the Australian Government Records Interoperability Framework. This framework adopts linked data standards to create a semantically interoperable environment for government records. Katharine has previously worked at the National Archives of Australia where she contributed to the development of records management standards, policy and strategy. At the National Archives Katharine led the project team which delivered the Digital Continuity 2020 Policy. Prior to the National Archives, Katharine worked for the State Records Authority of New South Wales on the NSW digital strategy Future Proof. Katharine is a PhD candidate at the University of Canberra, undertaking research into digital government and the effects on records management. Katharine has previous degrees from the University of Canberra and Macquarie University including a Master of Knowledge Management (Information Studies) and Master of Museum Management.

## ORCID

Gregory Rolan  <http://orcid.org/0000-0001-5891-3732>