

Conclusion

Jason Griffey

The incredible pace of Moore's Law has brought artificial intelligence (AI) systems down into the range where technologists at even small organizations can afford to have the computing power necessary to run machine learning systems locally.¹ From running open-source systems like TensorFlow, Keras, or Theano on local hardware like high-end GPUs, all the way down to \$100 neural net engines like Intel's Movidius Neural Compute Stick, which allows for pre-trained neural nets to run almost anywhere, there is an enormous wealth of options for programmers who are interested in experimenting with AI. It's even easier if you're running something that doesn't require local processing power, since every major provider of cloud services has some option for running machine learning systems in the cloud. Amazon has Machine Learning on AWS, Microsoft has Azure Machine Learning Studio, Google has Cloud AI, and IBM has Watson Machine Learning. Even your phone has chips in it dedicated just to AI processing; the newest iPhones have a dedicated Apple-designed 8-core neural chip in them just for doing AI work for apps and iOS.

TensorFlow
<https://www.tensorflow.org>

Keras
<https://keras.io>

Theano
www.deeplearning.net/software/theano

Intel: Movidius Neural Compute Stick
<https://developer.movidius.com>

Amazon: Machine Learning on AWS
<https://aws.amazon.com/machine-learning>

Microsoft: Azure Machine Learning Studio
<https://azure.microsoft.com/en-us/services/machine-learning-studio>

Google: Cloud AI
<https://cloud.google.com/products/ai>

IBM: Watson Machine Learning
<https://www.ibm.com/cloud/machine-learning>

It's never been easier to experiment with machine learning and AI systems. This situation is giving rise to an explosion of different services, systems, and apps that use AI as their primary processing function. The next five to ten years will be full of these same services and systems finding customers either directly or through business-to-business arrangements, such as being sold to libraries. Any provider of electronic books or journals, really anyone with a large corpus of digitized text, will be the first to begin experimenting with new indexing and finding services that have AI and machine learning at their base. It's low-hanging fruit for them and an easy upsell to libraries to have access to new discovery tools for their journals. The downside is that, because data is the lifeblood of machine learning systems, they are only as good as the amount of text (or photos, or videos) you can feed them. This gives existing vendors enormous leverage and little incentive to cooperate to allow for consolidation of systems in the same way that libraries could with federation of metadata in the past. The immediate effect will likely be highly siloed and limited to

being viable for only the largest players because they will provide the most value for payment for a library's money.

There are a number of other possible AI implementations that could impact libraries, which I'll discuss here very briefly. This is not meant to be a complete list by any means, but rather to consider the strengths of AI and machine learning as they relate to the work of libraries and see where the likely overlaps are.

The potential for machine learning systems to be trained to create metadata from any number of media types is very high. Throwing text, photos, and video at a machine learning system for subject heading assignments is not an incredibly difficult challenge for AI. Current incarnations wouldn't be perfect, and some secondary analysis may be needed, but given appropriate training set data, it wouldn't surprise me to see more automated cataloging over the next five years in libraries. I do think that given the speed of development, this AI cataloging system would be a brief and ultimately unnecessary part of the development of AI in libraries. Chris Bourq, Director of Libraries at MIT, wrote a prescient essay in 2017 titled "What Happens to Libraries and Librarians When Machines Can Read All the Books?" which I think gets at the longer-term issues relating especially to text, but also to video and photographs.² That is, as AI systems are increasingly better at understanding media, classical techniques in library and information science will become less effective and ultimately unable to keep pace with the increasingly capable automated systems.

Libraries and librarians have enormous sunk-costs in cataloging, in the assignment of category and subjects, ranging from call numbers to more modern descriptive technologies like RDA and Linked Data descriptions. When AI systems start bypassing these previously necessary stages in discoverability by directly parsing the texts themselves for semantic connections between them (à la HAMLET), a lot of traditional library science is at risk of being rendered at best irrelevant and at worst actively wasteful. This isn't to say there's no role for humans in this new world of AI-enhanced discoverability, but their role is much changed and more focused on preparation of training data and evaluation of outputs rather than direct creation of the descriptions. There are also roles that would be far more technical, involved in working with the algorithms that make up the various machine learning systems.

As we move forward through the development of increasingly more complex AI systems, even without getting all the way to general AI, we will quickly move into AI systems that are highly tied to individual users and learn from their activities in order to automate needed outcomes. We are starting to see this type of system in things like Google's Assistant and Apple's Siri virtual assistant. In both cases, the

systems "learn" from use and are supposed to suggest things to the user and pre-analyze some expected behaviors: for example, when Google's Assistant on Android will preemptively warn you about upcoming appointments that require driving or other transit and will take into account current driving conditions when it does so (e.g., I have an appointment across town that would normally take me thirty-five minutes to get to, but traffic is a little busy so right now, so travel time is more like forty-five minutes. Google will warn me forty-five to fifty minutes before the appointment and give me the updated directions on how to get there on time).

Another more recent example is in iOS 12 (the most recent version, as of this writing, of Apple's mobile operating system), where Siri watches all your activities on the phone and collects your most commonly performed tasks in a dedicated app called Shortcuts. Shortcuts then suggests new automation and triggers for your most common activities. For example, it might suggest after a week or so of seeing your behavior that it should automate your morning routine and automatically build a routine that would turn on the lights in your house, unlock your door, start playing the news, and pull up the weather and traffic report. All of this could be triggered by telling your phone, "Good morning." This is all backed by the local AI system described in the introduction to this report and is driven by local decisions. Each person's system will be very distinct and will continue to diverge over time as the system trains itself from the user's behavior.

One can easily imagine systems built to do this sort of automation work for researchers and students. As AI systems continue to be easier to implement, having a system local to your device that learns your preferences, your interests, and your needs will be commonplace. Researchers and students will have AI systems that find sources for them, summarize them, help them build bibliographies, and more. Over time, these systems will become irreplaceable archives of the learning and thinking history of individuals, a sort of universal diary of their activities. Now, imagine for a moment that this sort of system exists and is used by most learners. Who would you prefer be the developer of such a system: a large corporation like Facebook, or a collaborative effort by educational institutions and libraries?

Farther Future Issues

The far future of these AI systems will be far stranger than we can imagine. This report has focused mostly on the analysis and use of media as input and the resulting user outputs, but the future includes AI as a creator of media as well. *WIPO* and others have discussed the intellectual property implications of

creative works that emerge from AI systems.³ How these systems are treated in regard to intellectual property will have long-lasting effects on how libraries can use, collect, share, and archive media in the future. It's worth libraries and librarians paying close attention to these efforts and systems.

Academic libraries and higher education are going to have to deal with a whole different set of issues. AI that is smart enough to read, understand, and summarize a text will soon be smart enough to read several texts and show connections between them in an analytical way, and it's only a short step from there to automating the research paper process. How will education change when robots are capable of writing a paper that's indistinguishable from one that a human would write? And while I know you're already thinking "But it will be obvious that a machine wrote it," remember that these new systems will be learning from the individual that they are writing for and will absolutely be "smart" enough to tailor the language to sound like the person they are representing. AIs are already producing original works of visual art,⁴ and we have examples of AI-driven systems writing stories as well. How will the expectations of education change to accommodate this new digital capacity? I'm not yet sure, but I do know that libraries and librarians will be in the center of the discussions.

I'm Sorry, Dave . . .

The risks associated with AI shouldn't be understated. The risks of bias and error are present in ways that are not directly predictable, and the black box nature of machine learning systems provides an extra barrier to understanding and preventing negative outcomes from the use of systems trained on biased or incorrect data sets. It is possible that if AI systems are fully integrated into individuals' lives, it might increase the problem of filter bubbles and confirmation bias that exists in modern media discourse. Since your personal AI will be trained on the data that you yourself provide to it through your habits and information-seeking behavior, it is entirely possible that said systems will simply become a reality filter in horribly negative ways.

There are also the usual concerns about user and patron privacy in regard to the information-seeking process. If the near future of information searching entails siloed AI search driven from publisher's digital libraries, we should be very concerned about the possible leakage of patron information to the third-party systems (in the same way we should be concerned about any mediated access to resources). That a given system is driven by machine learning isn't necessarily worse than a non-AI system vis-à-vis privacy, but

since these systems will be new to the library world, it may be more difficult for us to determine how they are acting and what they are collecting. It is worth proceeding carefully anywhere that patron privacy is concerned.

The opportunities associated with new machine learning systems to reform large portions of library activities will be rich and varied. While it will be some time before general AI will be having full conversations and conducting reference interviews with students and patrons à la HAL from *2001*, the use of AI as increasingly powerful levers inside other systems will progress very quickly over the next three to five years. As with much of the modern world, automating the interaction between humans is often the most difficult challenge, while the interactions between humans and systems are less difficult and are the first to be automated away. In areas where human judgment is needed, we will instead be moving into a world where machine learning systems will abstract human judgment from a training set of many such judgments and learn how to apply a generalized rubric across any new decision point. This change will not require new systems short term, but in the longer term a move to entirely new types of search and discovery that have yet to be invented is very likely.

I'm very excited about the possibilities, and very concerned about the risks. Let's hope that libraries watch these systems as they develop, work with vendors, and create their own services and systems so that library values and ethics are baked into the technology at the outset. These systems will serve our patrons far better if we are concerned and focused early in their development, rather than waiting until after they are commonplace.

Notes

1. Wikipedia, s.v. "Moore's law," last updated October 6, 2018, 05:25, https://en.wikipedia.org/wiki/Moore%27s_law.
2. Chris Bourg, "What Happens to Libraries and Librarians When Machines Can Read All the Books?" *Feral Librarian* (blog), March 16, 2017, <https://chrisbourg.wordpress.com/2017/03/16/what-happens-to-libraries-and-librarians-when-machines-can-read-all-the-books>.
3. Andres Guadamuz, "Artificial Intelligence and Copyright," *WIPO Magazine*, no. 5/2017 (October 2017), www.wipo.int/wipo_magazine/en/2017/05/article_0003.html.
4. Naomi Rea, "Why One Collector Bought a Work of Art Made by Artificial Intelligence—and Is Open to Acquiring More," *Artnet News*, April 3, 2018, <https://news.artnet.com/art-world/art-made-by-artificial-intelligence-1258745>.