"Through the looking glass: envisioning new library technologies" how artificial intelligence will impact libraries

Peter Fernandez

This is the second in a two-part series on artificial intelligence (AI). The first column summarized some of the basic concepts necessary to understand AI and why recent developments indicate that it is poised to radically transform an array of emerging technologies.

This column will assume some basic knowledge of AI to more fully explore how this technology is likely to impact libraries in the future in the areas of search, educational technology and logistics.

Recap

At its core, AI is a group of technologies that attempt to enable computers to solve problems in more dynamic ways than they previously have been able to do. Often, these efforts are conceptualized as replicating human intelligence in their functionality, even if they often use radically different underlying methods. One key difference is that computers have inherent advantages over humans in absorbing and processing certain types of data. As a result, what can be done once a computer is capable of making more complex inferences can be truly remarkable, particularly when AI is used to empower humans in accomplishing otherwise tedious or difficult tasks.

If this definition seems slightly underwhelming, it may be because AI has already been incorporated into so many of the innovations we use in our daily life. Some form of AI already augments basic online searches, drives the algorithms that power Amazon's recommended products and even helps to augment recommendations given to us by mobile phone personal assistants. The first wave of AI has already arrived.

Even so, such observations should not lead us to underestimate the transformational nature of what AI is on the verge of being able to do. The previous column outlined a number of areas in which various technologies are converging to make AI more impactful than ever before, but the one worth highlighting in this context is the concept of machine learning.

Machine learning allows for AI to be programmed by giving the AI some basic rules and then allowing it to encounter information and run simulations so that it can begin to develop its own conclusions about how to respond to new stimuli. In other words, an AI could master a new task, such as predicting what books a library is likely to need just by having access to relevant data sets without having been given any (or at least minimal) programming directly related to libraries' needs. There is a reason that the term conjures up images of science fiction, where robots and computers have taken over tasks once thought to be the sole domain of humans, and, in many cases, complete those tasks better than humans ever could. As computers begin to be able to absorb information and make sophisticated decisions about what to do with that data, the possibilities for how they can be used grow exponentially. They even begin to have the capacity to surprise the very people who programmed them.

Rate to adoption

Importantly, as we consider how this technology might apply to libraries, it is worth noting that machine learning has a number of implications that need to be teased out. In an effort to be the platform upon which others build their own implementations, many of the most prominent AI developers (Microsoft, Google, Yahoo and Facebook, among others) are making large parts of their code available open source, so that anyone can inspect and improve the code without proprietary restrictions (Shafto, 2016). Ideally, this will mean that other companies will be able to rapidly transform the results of this advanced AI and use it in new applications.

More importantly, however, machine learning holds the promise of extraordinary rapid development once the technology meets certain thresholds. If AI can learn to adapt to new situations, rather than requiring extensive unique programming to take on new tasks, then it will reduce the effort required to adapt AI and use it in relatively niche industries, such as libraries. Many barriers remain to applying this technology fruitfully within the context of a library and it will still likely require considerable effort. But the kind of paradigm shift this technology represents means that it may be relevant to libraries much sooner than we otherwise anticipated.

Library application: search

In recent years, many libraries have begun to move away from library catalogs and toward online "discovery systems" (Sadeh, 2015). The online catalog was an outgrowth of physical card catalogs, where each item in the library was organized, along with relevant metadata, to make the holdings of that particular library findable and accessible to patrons. Discovery systems combine information from multiple sources – the libraries catalog, subscription databases of articles, even free resources and organize them into a single search interface. This allows, in theory, for patrons to "discover" items that would have previously been inaccessible, replicating in-part the

experience they get from searching multi-format interfaces like Google.

Although an improvement over traditional catalogs in most cases, the reality is that many of these systems have struggled in a variety of ways to make meaningful sense of all their different data sources, and there is significant room for improvement. Much of the difficulty occurs because each data source has different kinds of metadata associated with it and these can be challenging to effectively account for (Sadeh, 2015). How to rank an article in an academic journal to which the system has full text access, when compared to a book title to which only the subject headings are available, is a difficult problem to solve. It is even harder when the system has relatively little context about why the patron is searching for that information.

However, a sufficiently advanced AI has a clear and obvious potential to help alleviate many of these concerns. Using the machine learning techniques outlined above, it is easy to imagine AI finding connections and patterns in the existing data that would be nearly impossible for humans to replicate on a large scale. With refinement, AI could analyze data from thousands of previous searches in many different ways and find patterns that emerge based on patron usage. For instance, it might notice that patrons who use certain phrases in their search tend to be most satisfied by items that are written for an academic audience. Further, it may be able to evaluate which items are written for an academic audience, even if that is not in the metadata, using contextual clues about the title and other existing information.

These predictions are not just a matter of wishful thinking, but are efforts that are actively being pursued by a number of library companies. Companies such as IBM, the Allen Institute for Artificial Intelligence and Meta are working to solve these search and discovery problems in a variety of ways (Luther, 2016). Meanwhile, efforts such as open access publishing and alt-metrics are creating robust, copyright-free platforms to draw initial and limited restraint inferences about literature.

Creating metadata

The same technologies that enable better analysis of the information contained in existing resources can, in turn, be used to complete traditional library tasks such as cataloging, and thus create a virtual circle. Already, efforts such as SciELO Suggester attempt to harness the ability of tools to analyze records and enable catalogers to make better, faster decisions about what metadata to give an item (Mitzig *et al.*, 2016).

As these kinds of tools develop, it will be possible for information professionals to make more sophisticated decisions about what kind of metadata is needed about a given item and dynamically generate it with relatively minimal human interaction. When human decision-making is needed, it will be facilitated by AI-powered tools that suggest metadata that, in turn, makes it easier for the AI-powered search to be effective.

Translational research

One of the more intractable problems in connecting patrons to the information they desire is that often it does not exist in a format they want. The salient part of a book that answers a query may not be readily apparent, or even indexed properly. A 10-page feature story could incidentally contain a single chart that is useful to patrons who are uninterested in the rest of the article. Or, the answer might be contained in an academic article written using jargon, that is difficult for non-experts to parse.

Machine learning is already being used to translate texts from one language into another. But it has the potential to do something equally revolutionary: it could help translate information from one format into the learning style the end user most needs. If AI can absorb large amounts of information, determine its relevance and reproduce it in new formats, then it could lead to a revolution in how information is processed by humans. In any instance where copyright is not a barrier, AI will be increasingly good at scanning the entire corpus of information, and then reconfiguring how it is displayed to meet a user's needs.

What is searching?

In this context, the very idea of what it means to search for something will be transformed, and the skills that information professionals need to assist users will change dramatically.

Recently, Google's new CEO released his first founders' letter in which he outlined parts of his vision for the company. Among the many AI-centered revelations was his belief that the very idea of a device will become increasingly irrelevant, as sophisticated AI merges the user experience across devices (Eadicicco, 2016). Distinctions such as mobile and online will become less relevant as the most salient feature of all of these devices becomes that they are helpful personal assistants, able to respond to and anticipate users' needs with the appropriate information.

If Google is correct in its assumption about devices, these advances will also have implications on library spaces. As mobile devices interact dynamically with spaces, users will be able to query their library surroundings for context-specific information, and computer labs will evolve into spaces that allow for a wider array of interactions.

One immediate manifestation of this phenomenon is the chatbot that many companies (Facebook, Microsoft and others) have begun to release (Carey, 2016). These chatbots take natural questions and input from humans, process them and attempt to return meaningful, human-like exchanges that make interacting with a product or service more conversation-like. Such interaction makes interacting with the product easier and ideally more enjoyable.

These are the first, tentative steps in what will be increasingly sophisticated ways of interacting with information. Many libraries already have online expert chat, but in the near future, they may be able to turn over most basic reference questions, including the information contained in FAQs, to such AI services. Indeed, as these various interfaces become more robust, the challenge of searching will transform. Information professionals will be less dependent on concepts like controlled vocabulary and limiters. Instead, what they will have to offer is context that enables patrons to properly use and

navigate the AI systems that can easily process and reconfigure huge data sets of information.

Distorting algorithms

AI is not, however, a magic panacea that will always get everything right. Information processed by AI is filtered and potentially distorted just as much as, if not more than, information processed by humans. For instance, filter bubbles occur when algorithms present only some information to a given user to match that user's perceived preferences, resulting in a "filter" that would minimize users' contact with information that contradicts their already held views. As AI is able to more deeply personalize results based on robust information about whomever they are performing a search for and has ever more control over how information is presented and interpreted, these tendencies could easily become exacerbated. Indeed, if that same AI is helping to inform how other items are cataloged, it could come to distort how we understand whole arenas of knowledge.

It is not just a matter of personalizing information for a given user, but that through machine learning, AI may develop remarkably effective ways of making connections about information that is difficult for humans to fully analyze or account for. Inasmuch as information professionals trust these connections, we are inevitably turning over analysis to systems that we do not fully understand and that our patrons are likely to understand even less. Moreover, even if a particular technique is remarkably effective for 85 per cent of users, that indicates that it is actively problematic for a significant percentage of other users, leaving a number of library patrons without the access they need.

Worth noting is that inadvertently limiting access is not a problem unique to AI, or even to the algorithms increasingly being used to filter information to users online. Library catalogs, and even in-person interactions have inadvertently altered patrons' access to information. But professionals have always done what they can to mitigate these influences, and as AI develops, they will need to continue to do so.

Insights into collections

In addition to transforming what it means to conduct a search, this same kind of assistance will have implications for decision-making of all kinds. In libraries, one of the most obvious uses will be the effective analysis of collections. Libraries tend to be effective at collecting data about the physical items held in their collection, but are often less efficient when it comes to deciding what items to purchase and retain. While information professionals are often well-equipped to make these decisions, it can be time consuming to do it correctly and having access to the appropriate data at the right moment is an ongoing challenge (Williamson *et al.*, 2014).

AI will make it possible to use the same kinds of analysis on existing collections that is done to help patrons during a search. This will allow information professionals to easily draw new conclusions about ways in which their collection is being used and how it should develop. Perhaps, items in certain call number ranges check out more or less during certain time periods, or perhaps books with certain subject headings are popular even when they are out of date. These connections can be found by humans, but AI that can recognize patterns independently will be vastly better at finding them in a timely manner.

Preparing for the future

To understand the implications of AI, it is not enough to see its effects in one arena. It will have cascading implications across a range of services. To extend this example of collections, AI will also affect how easy it is to store and transfer items. This can be seen in companies such as DHL, which has issued a report that envisions AI combined with the Internet of Things remaking the logistics of transportation and supply chains (Kilcarr, 2016). But this concept is equally true for libraries. As a result, it could dramatically aid in attempts by libraries to create multi-institution, shared storage for rarely used items. The AI can identify these rarely used items, which will help reduce storage costs, and then efficiently move these items to the location where they are most likely to be used.

The possible future of storage also points to the way in which libraries can begin to prepare for the implications of AI, even as they acknowledge that the full implications of AI are impossible to predict. One theme that we have seen is how AI will make connections where they were not seen before, making it an inherently collaborative technology. The opportunities are likely to emerge in partnerships of various kinds, both between libraries and with the technology companies that control this powerful technology. Libraries can begin to expend political capital now to create relationships that can harness these abilities.

Working cooperatively with other libraries, information professionals will need to continue to develop and define their areas of expertise. To harness this technology and affect how it is developed, libraries will need to overcome institutional barriers to take advantage of the efficiencies of scale that collaboration can bring. When communicating their needs to technology companies, libraries will need to be able to clearly articulate their ethical and service principles so the technology can assist their patrons in ways that continue the best traditions of librarianship. And finally, given the transformational nature of this technology, libraries will need to prepare for rapid change.

REFERENCES

Carey, S. (2016), "What is a chatbot, how do you build one, who is building them and are they any good?", available at: www.techworld.com/apps/what-is-chatbot-how-do-you-build-one-who-is-building-them-are-they-any-good-3638740/ (accessed 6 May 2016).

Eadicicco, L. (2016), "Read the Google CEO's new letter on the company's future", *Time*, available at: http://time.com/4311233/google-ceo-sundar-pichai-letter/

Kilcarr, S. (2016), "DHL: artificial intelligence will remold logistics world", available at: http://fleetowner.com/fleet-management/dhl-artificial-intelligence-will-remold-logistics-world (accessed 6 May 2016).

Luther, J. (2016), "Discovery in an age of artificial intelligence", *Learned Publishing*, Vol. 29 No. 2, pp. 75-76, available at: http://doi.org/10.1002/leap.1024

Mitzig, N.L., Mitzig, M.S., Martínez, F.A., Piriz, R.A., Ferracutti, V.M., González, M.P. and Maguitman, A.G. (2016), "SciELO suggester: an intelligent support tool for cataloging library resources", *Library & Information Science Research*, Vol. 38 No. 1, pp. 39-51, available at: http://doi.org/10.1016/j.lisr.2016.01.001

Sadeh, T. (2015), "From search to discovery", *Vom Suchen Zum Entdecken*, Vol. 39 No. 2, pp. 212-224, available at: http://doi.org/10.1515/bfp-2015-0028

Shafto, P. (2016), "Why big tech companies are open-sourcing their AI systems", available at: http://phys.org/news/(2016)-02-big-tech-companies-open-sourcing-ai.html (accessed 4 March 2016).

Williamson, J., Fernandez, P. and Dixon, L. (2014), "Factors in science journal cancellation projects: the roles of faculty consultations and data", available at: http://doi.org/10.5062/F4G73BP3

Peter Fernandez (*pfernand@utk.edu*) is based at Research Services Agricultural Sciences and Natural Resources, University of Tennessee, Knoxville, Tennessee, USA.