

Chapter 5

Cultures of Innovation: Machine Learning as a Library Service

Sue Wiegand
Saint Mary's College

Introduction

Libraries and librarians have always been concerned with the preservation of knowledge. To this traditional role, librarians in the 20th century added a new function—discovery—teaching people to find and use the library’s collected scholarship. Information Literacy, now considered the signature pedagogy in library instruction, evolved from the previous Bibliographic Instruction. As Digital Literacy, the next stage, develops, students can come to the library to learn how to leverage the greatest strengths of Machine Learning. Machines excel at recognizing patterns; researchers at all levels can experiment with innovative digital tools and strategies, and build 21st century skill sets. Librarian expertise in preservation, metadata, and sustainability through standards can be leveraged as a value-added service. Leading-edge librarians now invite all the curious to benefit from the knowledge contained in the scholarly canon, accessible through libraries as curated living collections in multiple formats at distributed locations, transformed into new knowledge using new ways to visualize and analyze scholarship.

Library collections themselves, including digitized, unique local collections, can provide the data for new insights and ways of knowing produced by Machine Learning. The library could also be viewed as a technology sandbox, a place to create knowledge, connect researchers, and bring together people, ideas, and new technologies. Many libraries are already rising to this challenge, working with other cultural institutions in creating a culture of innovation as a new learning paradigm, exemplified by Machine Learning instruction and technology tool exploration.

Library Practice

The role of the library in preserving, discovering, and creating knowledge continues to evolve. Originally, libraries came into being as collections to be preserved, managed, and disseminated, a central repository of knowledge, possibly for political reasons (Ryholt and Barjamovic 2019, 1–2). Libraries founded by scholars and devoted to learning came later, during the Middle Ages (Casson 2001, 145). In more recent times, librarians began “[c]ollecting, organizing, and making information accessible to scholars and to citizens of a democratic republic” based on values developed during the Enlightenment (Bivens-Tatum 2012, 186).

Bibliographic Instruction in libraries, and later Information Literacy, embodied the idea of learning in the library as the next step beyond collecting, with librarians instructing on information infrastructure with the goal of empowering library users to find, evaluate, and use scholarly information in print and digital formats, with an emphasis on privacy and intellectual freedom as core library values. Now, librarians are also contributing to and participating in the learning enterprise by partnering with the disciplines to produce new knowledge. This final step of knowledge creation in the library completes the scholarly communications cycle of building on previous scholarship—“standing on the shoulders of giants.”

One way to cultivate innovation in libraries is to include Machine Learning in the library’s array of tools, resources, and services, both behind-the-scenes and public-facing. Librarians are expert at developing standards, preserving the scholarly record, and refining metadata to enhance interdisciplinary discovery of research, scholarship, and creative works. Librarian expertise could go far beyond local library collections to a global perspective and normative practice of participation at scale in innovative emerging technologies such as Machine Learning.

For instance, citations analysis of prospective collections for the library to collect and of the institutions’ research outputs would provide valuable information for both further collection development and for developing researchers’ toolkits. Machine Learning with its predilection for finding patterns, would reveal gaps in the literature and open up new questions to be answered, solving problems and leading to innovation. As one example, Yewno, a multi-disciplinary platform that uses Machine Learning to help combat “Information Overload,” advertises that it “helps researchers, students, and educators to deeply explore knowledge across interdisciplinary fields, sparking new ideas along the way...” and “makes [government] information accessible by breaking open silos and comprehending the complicated interconnections across agencies and organizations,” among other applications to improve discovery (Yewno n.d.). Also, in 2019, the Library of Congress hosted a Summit as “part of a larger effort to learn about machine learning and the role it could play in helping the Library of Congress reach its strategic goals, such as enhancing discoverability of the Library’s collections, building connections between users and the Library’s digital holdings, and leveraging technology to serve creative communities and the general public” (Jakeway 2020). Integration of Machine Learning technologies is already starting at high levels in the library world.

New Services

A focus on Machine Learning can inspire new library services to enhance teaching and learning. Connecting people with ideas and with technology enables library virtual spaces to be used as a learning service by networking researchers at all levels in the enterprise of knowledge creation. Finding gaps in the literature would be a helpful first step in new library discovery tools. A way

this could be done is through a “Researchers’ Workstation,” an end-to-end toolkit that might start by using Machine Learning tools to automate alerts of new content in a narrow area of interest and help researchers at all levels find and focus on problem-solving. A Researchers’ Workstation could contain a collection of analytic tools and learning modules to guide users through the phases of discovery. Then, managing citations would be an important step in the process—storing, annotating, and sorting out the most relevant. Starting research reports, keeping lab notebooks, finding datasets, and preserving the researcher’s own data are all relevant to the final results. A collaboration tool would enable researchers to find others with similar interests and share data or work collaboratively from anywhere, asynchronously. Having all these tools in one serendipitous virtual place is an extension of the concept of the library as the physical place to start research and scholarship. It is merely the containers of knowledge that are different.

Some of this functionality exists already, both in Open Source software such as Zotero for citation management, and in proprietary tools that combine multiple functions, such as Mendeley from Elsevier.¹ Other commercial publishers are developing tools to enable researchers to work within their proprietary platforms, from the point of searching for ideas and finding research gaps through the process of writing and submitting finished papers for publication. The Coalition of Open Access Repositories (COAR) is similarly developing “next generation repositories” software integrating end-to-end tools for the Open Access literature archived in repositories, to “facilitate the development of new services on top of the collective network, including social networking, peer review, notifications, and usage assessment.” (Rodrigues et al, 2017, 5).

What else might a researcher want to do that the library could include in a Researchers’ Workstation? Finding, writing, and keeping track of grants could be incorporated at some level. Generating a timeline might be helpful, and infographics and data visualizations could improve research communication and even help make the case for the importance of the study with others, especially the public and funders. Project management tools might be welcomed by some researchers, too.

Finally, when it’s time to submit the idea (whether at the preliminary or preprint stage) to something like an ArXiv-like repository or an institutional repository, as well as to journals of interest (also identified through Machine Learning tools), the process of submission, peer-review, revision, and re-submitting could be done seamlessly. The tools and functions in the Workstation would ideally be modular, interoperable, and easy to learn and use, as well as continuously updated. The Workstation would be a complete ecosystem in the research cycle—saving time in the Scholarly Communications process and providing one place to go to for discovery, literature review, data management, collaboration, preprint posting, peer review, publication, and post-print commenting.²

Collections as Data, Collections as Resources

Exemplified by the literature search that now includes a myriad of Open content on a global basis, collections is an area that provides the greatest scope for library Machine Learning innovations to date, both applied and basic/theoretical. Especially if the pathway to using the expanded collections is clear and coherent, and the library provides instruction on why and how to use the various tools to save time and increase impact of research, researchers at all levels will benefit from

¹See <https://www.zotero.org> and <https://www.mendeley.com>

²In 2013, I wrote a blog that mentions the idea (Wiegand).

partnering with librarians for a more comprehensive view of current knowledge in an area. The *Always Already Computational: Collections as Data* final report and project deliverables and *Collections as Data: Part to Whole* Project were designed to “develop models that support collections as data implementation and holistic reconceptualization of services and roles that support scholarly use....” The Project specifically seeks “to create a framework and set of resources that guide libraries and other cultural heritage organizations in the development, description, and dissemination of collections that are readily amenable to computational analysis.” (Padilla et al 2019).

As a more holistic approach to data-driven scholarship, these resources aim to provide access to large collections to enable computational use on the national level. Some current library databases have already built this kind of functionality. JSTOR, for example, will provide up to 25,000 documents (or more at special request) in a dataset for analysis.³ Clarivate’s Content as a Service provides Web of Science data to accommodate multiple purposes.⁴ Besides the many freely available bibliodata sources, researchers can sign up for developer accounts in databases such as Scopus to work with datasets for text mining and computational analysis.⁵ Using library-licensed collections as data could allow researchers to save time in reading a large corpus, stay updated on a topic of interest, analyze the most important topics at a given time period, confirm gaps in the research literature for investigation, and increase the efficiency of sifting through massive amounts of research in, for instance, the race to develop a COVID-19 vaccine (Ong 2020; Vamathevan 2019).

Learning Spaces

Machine Learning is a concept that calls out for educating library users through all avenues, including library spaces. Taking a cue from other GLAM (Galleries, Libraries, Archives, and Museums) cultural institutions, especially galleries and museums, libraries and archives could mount exhibits and incorporate learning into library spaces as a form of outreach to teach how and why using innovative tools will save time and improve efficiency. Inspirational, continuously-updating dashboards and exhibits could show progress and possibilities, while physical and virtual tutorials might provide a game-like interface to spark creativity. Showcasing scholarship and incorporating events and speakers help create a new culture of ideas and exploration. Events bring people together in library spaces to network for collaborative endeavors. As an example, the Cleveland Museum of Art is analyzing visitor experiences using an ArtLens app to promote its collections.⁶ The Library of Congress, as mentioned, hosted a summit that explored such topics as building Machine Learning literacy, attracting interest in GLAM datasets, operationalizing Machine Learning, crowdsourcing, and copyright implications for the use of content. As another example, in 2017 the United Kingdom’s National Archives attempted to demystify Machine Learning and explore ethics and applications such as topic modeling, which

was used to find key phrases in Discovery record descriptions and enable innovative exploration of the catalogue; and it was also deployed to identify the subjects being discussed across Cabinet Papers. Other projects included the development

³See <https://www.jstor.org/dfr/about/dataset-services>.

⁴See <https://clarivate.com/search/?search=computational%20datasets>

⁵See <https://dev.elsevier.com/> and <https://guides.lib.berkeley.edu/text-mining>

⁶See <https://www.clevelandart.org/art-museums-and-technology-developing-new-metrics-measure-visitor-engagement> and <https://www.clevelandart.org/artlens-gallery/artlens-a>

of a system that found the most important sentence in a news article to generate automated tweeting, while another team built a system to recognise computer code written in different programming languages — this is a major challenge for digital preservation. (Bell 2018)

Finally, the HG Contemporary Gallery in Chelsea, in 2019, mounted an exhibit that utilized a “machine-learning algorithm that did most of the work” (Bogost 2019).

Sustainable Innovation

Diversity, equity, and inclusion (DEI) concerns with the scholarly record and increasingly with recognized biases implicit in algorithms can be addressed by a very intentional focus on the value of differing perspectives in solving problems. Kat Holmes, an inclusive design expert previously at Microsoft and now a leading user experience designer at Google, urges a framework for inclusivity that counteracts bias with different points of view by recognizing exclusion, learning from human diversity, and bringing in new perspectives (Bedrossian 2018). Making more data available, and more diverse data, will significantly improve the imbalance perpetuated by a traditional-only corpus. In sustainability terms, Machine Learning tools must be designed to continuously seek to incorporate diverse perspectives that go beyond the traditional definitions of the scholarly canon if they are to be useful in combating bias. Collections used as data in Machine Learning might undergo analysis by researchers, including librarian researchers, to determine the balance of content. Library subject headings should be improved to better reflect the diversity of human thought, cultures, and global perspectives.

Streamlining procedures is to everyone’s benefit, and saving time is universally desired. Efficiency won’t fix the time crunch everyone faces, but with too much to do and too much to read, information overload is a very real threat to advancing the research agenda and confronting a multitude of escalating global problems. Machine Learning techniques, applied at scale to large corpora of textual data, could help researchers pinpoint areas where the human researcher should delve more deeply to eliminate irrelevant sources and hone in on possible solutions to problems. One instance—a new service, Scite.ai “can automatically tell readers whether papers have been supported or contradicted by later academic work” (Khamsi 2020). WHO (World Health Organization) is providing a Global Research Database that can be searched or downloaded.⁷ In research on self-driving vehicles, a systematic literature review found more than 10,000 articles, an estimated year’s worth of reading for an individual. A tool called Iris.ai allowed groupings of this archive by topic and is one of several “targeted navigation” tools in development (Extance 2020). Working together as efficiently as possible is the only way to move ahead, and Machine Learning concepts, tools, and techniques, along with training, can be applied to increasingly large textual datasets to accelerate discovery.

Machine Learning, like any other technology, augments human capacities, it does not replace them. If 10% of library resources (measured in whatever way works for each particular library), including both time resources of expert librarians and staff and financial resources, were utilized for innovation, libraries would develop a virtuous self-sustaining cycle. Technologies that are not as useful can be assessed and dropped in an agile library, the useful can be incorporated into the 90% of existing services, and the resources (people and money) repurposed. In the same way, that

⁷See <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov>

10% of library resources invested into innovations such as Machine Learning, whether in library practice or instruction and other services, will keep the program and the library fresh.

Creativity is key and will be the hallmark of successful libraries in the future. Stewardship of resources such as people's skills and expertise, and strategic use of the collections budget, are already library strengths. By building out new services and tools, and instructing at all levels, libraries can reinvent themselves continuously by investing in creative and sustainable innovation, from digital and data literacy to assembling modules for a library-based customized Researchers' Workstation that uses Machine Learning to enhance the efficiency of the scholars' research cycle.

Results and more questions

A library that adapted Machine Learning as an innovation technology would improve its practices; add new services; choose, use, and license collections differently; utilize all spaces for learning; and role model innovative leadership. What is a library in rapidly changing times? How can librarians reconcile past identity, add value, and leverage hard-won expertise in a new environment? Change management is a topic that all institutions will have to confront as the digital age continues, as we reinvent ourselves and our institutions in a fast paced technological world.

Value-added, distinctive, unique—these are all words that will be part of the conversation. Not only does the library add value, but librarians will have to demonstrate and quantify that value while preparing to pivot at any time in response to crises and innovative opportunities. Distinctive library resources and services that speak to the institutions' academic mission and purpose will be a key feature. What does the library do that no other entity on campus can do? At each particular inflection point, how best to communicate with stakeholders about the value of the distinctive library mission? Can the library work with other cultural heritage institutions to highlight the unique contributions of all?

One possible approach—develop a library science/library studies pedagogy as well as outreach that encompasses the Scholarship of Teaching and Learning (SoTL) and pervades everything the library does in providing resources, services, and spaces. Emphasize that library resources help people solve multi-dimensional, complex problems, and then work on new ideas to save the time of researchers, improve discovery systems, advocate and facilitate Open Access and Open Source alternatives while enabling, empowering, and yes, inspiring all users to participate in and contribute to the record of human knowledge. Librarians, as the traditional keepers of the scholarly canon in written form, have standing to do this as part of our legacy and as part of our envisioned future.

From the library users' point of view, librarians should think like the audience we are trying to reach to answer the question—why come into the library or use the library website instead of more familiar alternatives? In an era of increasing surveillance, library tools could be better known for an emphasis on privacy and confidentiality, for instance. This may require thinking more deeply about how we use our metrics and finding other ways to show how use of the library contributes to student success. It is also important to gather quantitative and qualitative evidence from library users themselves, and apply the feedback in an agile improvement loop.

In the case of Open Access vs. proprietary information, librarians should make the case for Open Access (OA) by advocating, explaining, and instructing library users from the first time they do literature searches to the time they are graduate students, post-docs, and faculty. Librarians should produce Open Educational Resources (OER) as well as encourage classroom faculty to adopt these tools of affordable education. Libraries also need to facilitate Open Access content

from discovery to preservation by developing search tools that privilege OA, using Open Source software whenever possible. Librarians could lead the way to changing the Scholarly Communications system by emphasizing change at the citations level—encourage researchers to insist on being able to obtain author-archived citations in a seamless way, and facilitate that through development of new discovery tools using Machine Learning. Improving discovery of Open Access, as well as embarking on expanded library publishing programs and advancing academic research, might be the most important endeavors that librarians could undertake at this point in time, to prevent a repeat of the “serials crisis” that commoditized scholarly information and to build a more diverse, equitable, and inclusive scholarly record. Well-funded commercial publishers are already engaging scholars and researchers in new proprietary platforms that could lock in academia more thoroughly than “Big Deals” did, even as the paradigm shifts away from large, expensive publishers’ platforms and library subscription cancellations mount due to budget cuts and the desire to optimize value for money.

The concept of the “inside-out library” (Dempsey 2016) provides a way of thinking about opening local collections to discovery and use in order to create new knowledge through digitization and semantic linking, with cross-disciplinary technologies to augment traditional research and scholarship. Because these ideas are so new but fast-moving, librarians need to spread the word on possibilities in library publishing. Making local collections accessible for computational research helps to diversify findings and focuses attention on larger patterns and new ideas. In 2019, for instance, the Library of Congress sought to “Maximize the Use of its Digital Collection” by launching a program “to understand the technical capabilities and tools that are required to support the discovery and use of digital collections material,” developing ethical and technological standards to automate in supporting emerging research techniques and “to preprocess text material in a way that would make that content more discoverable” (Price 2019). Scholarly Communication, dissemination, and discovery of research results will continue to be an important function of the library if trusted research results are to be available to all, not just the privileged. The so-called Digital Divide isolates and marginalizes some groups and regions; libraries can be a unifying force.

An important librarian role might be to identify gaps, in research or in dissemination, and work to overcome barriers to improving highly distributed access to knowledge. Libraries specialize in connecting disparate groups. Here is what libraries can do now: instruct new researchers (including undergraduate researchers and up) in theories, skills, and techniques to find, use, populate, preserve, and cite datasets; provide server space and/or Data Management services; introduce Machine Learning and text analysis tools and techniques; provide Machine Learning and text analysis tools and/or services to researchers at all levels. Researchers are now expected or even required to provide public scholarship, i.e., to bring their research into the public realm beyond obscure research journals, and to explain and illuminate their work, connecting it to the public good, especially in the case of publicly-funded research. Librarians can and should partner in the public dissemination of research findings through explaining, promoting, and providing innovative new tools across siloed departments to catalyze cross-disciplinary research. Scholarly Communications began with books and journals shared by scholars over time, then libraries were assembled and built to contain the written record; librarians should ensure that the Scholarly Communications and information landscape continues into the future with widely-shared, available resources in all formats, now including interactive, web-based software, embedded data analysis tools, and technical support of emerging Open Source platforms.

In addition, the flow of research should be smooth and seamless to the researcher, whether

in a Researchers' Workstation or other library tools. The research cycle should be both clearly explained and embedded in systems and tools. The library, as a central place that cuts across narrowly-defined research areas, could provide a systemic place of collaboration. Librarians, seeing the bigger picture, could facilitate research as well as disseminate and preserve the resulting data in journals and datasets. Further investigations on how researchers work, how students learn, best practices in pedagogy, and life-long learning in the library could mark a new era in librarianship, one that involves teaching, learning, *and* research as a self-reinforcing cycle. Beyond being a purchaser of journals and books, libraries can expand their role in the learning process itself into a cycle of continuous change and exploration, augmented by Machine Learning.

Library Science, Research, and Pedagogy

In Library and Information Science (LIS), graduate library schools should teach about Machine Learning as a way of innovating and emphasize pervasive innovation as the new normal. Creating a culture of innovation and creativity in LIS classes and in libraries will pay off for society as a whole, if librarians promote the advantages of a culture of innovation in themselves and in library users. Subverting the stereotypes of tradition-bound libraries and librarians will revitalize the profession and our workplaces, replacing fear of change and an existential identity crisis with a spirit of creative, agile reinvention that will rise to challenges rather than seek solace in denial, whether the seemingly impossible problem is preparedness in dealing with a pandemic or creatively addressing climate change.

Academic libraries must transition from a space of transactional (one-time) actions into a transformational learning-centered user space, both physical and virtual, that offers an enhanced experience with teaching, learning, and research—a way to re-center the library as the place to get answers that go beyond the Internet. Libraries add value: do faculty, students, and other patrons know, for instance, that when they find the perfect book on a library shelf through browsing (or on the library website with virtual browsing), it is because a librarian somewhere assigned it a call number to group similar books together? The next step in that process is to use Machine Learning to generate subject headings, and also show the librarians accomplishing that. This process is being investigated in different types of works from fiction to scientific literature (Golub 2006, Joorabchi 2011, Wang 2009, Short 2019). Cataloging, metadata, and enabling access through shared standards and Knowledge Bases are all things librarians do that add value for library users overwhelmed with Google hits, and are worthy of further development, including in an Open environment.

Preservation is another traditional library function, and now includes born-digital items and digitization of special collections/archives, increasing the library role. Discovery will be enhanced by Artificial/Augmented Intelligence and Machine Learning techniques. All of this should be taught in library schools, to build a new library culture of innovation and problem-solving beyond just providing collections and information literacy instruction. The new learning paradigm is immersive in all senses, and the future, as reflected in library transformation and partnerships with researchers, galleries, archives, museums, citizen scientists, hobbyists, and life-long learners re-tooling their careers and life, is bright. LIS programs need to reflect that.

To promote learning in libraries, librarians could design a "You belong in the Library" campaign to highlight our diverse resources and new ways of working with technology, inviting participation in innovative technologies such as Machine Learning in an increasingly rare public, non-commercial space—telling why, showing how. In many ways, libraries could model ways to

achieve academic success and life success, updating a traditional role in educating, instructing, preparing for the future, explaining, promoting understanding, and inspiring.

Discussion

The larger questions now are, who is heard and who contributes? How are gaps, identified in needs analysis, reduced? What are sources of funding for libraries to develop this important work and not leave it to commercial services? Library leadership and innovative thinking must converge to devise ways for libraries to bring people together, producing more diverse, ethical, innovative, inclusive, practical, transformative, and novel library services and physical and virtual spaces for the public good.

Libraries could start with analyses of needs—what problems could be solved with more effective literature searches? What research could fill gaps and inform solutions to those needs? What kind of teaching could help build citizens and critical thinkers, rather than simply encouraging consumption of content? Another need is to diversify collections used in Machine Learning, gathering cultural perspectives that reflect true diversity of thought through inclusion. All voices should be heard and empowered. Librarians can help with that.

A Researchers' Workstation could bring together an array of tools and content to allow not only the organization, discovery, and preservation of knowledge, but also facilitate the creation of new knowledge through the sustainable library, beyond the literature search.

The world is converging toward networking and collaborative research all in one place. I would like the library to be the free platform that brings all the others together.

Coming full circle, my vision is that when researchers want to work on their research, they will log on to the library and find all they need.... The library is the one place ... to get your scholarly work done. (Wiegand 2013)

The library as a platform should be a shared resource—the truest library value.

Here is a scenario. Suppose, for example, scholars wish to analyze the timeline of the beginning of the Coronavirus crisis. Logging on to the library's Researchers' Workstation, they start with the Discovery module to generate a corpus of research papers from, say, December 2019 to June 2020. Using the Machine Learning function, they search for articles and books, looking for gaps and ideas that have not yet been examined in the literature. They access and download full-text, save citations, annotate and take notes, and prepare a draft outline of their research using a word processing function, writing and citing seamlessly. A Methods (protocols) section could help determine the most effective path of the prospective research.

Then, they might search for the authors of the preprints and articles they find interesting, check the authors' profiles, and contact some of them through the platform to discern interest in collaborating. The profile system would list areas of interest, current projects, availability for new projects, etc. Using the Project Management function, scholars might open a new workspace where preliminary thoughts could be shared, with attribution and acknowledgement as appropriate, and a peer review timeline chosen to invite comments while authors can still claim the idea as their own.

If the preprint is successful, and the investigation shows promise after the results are in, the scholars could search for an appropriate journal for publication, the version of record. The au-

thor, with researcher ID (also contained in his/her profile), has the article added to the final published section of the profile, with a DOI. The journal showcases the article, sends out tables of content alerts and press releases where it can be picked up by news services and authors invited to comment publicly. Each institution would celebrate its authors' accomplishments, use the Scholars' Workstation to determine impact and metrics, and promote the institutions' research progress.

Finally, the article would be preserved through the library repository and also initiatives such as LOCKSS. Future scholars would find it still available and continue to discover and build on the findings presented. All of this and more would be done through the library.

Conclusion

Machine Learning as a library service can inspire new stages of innovation, energizing and providing a blueprint for the library future—teaching, learning, and scholarship for all. The teaching part of the equation invokes the faculty audience perspective: how can librarians help classroom faculty to integrate both library instruction and library research resources (collections, expertise, spaces) into the educational enterprise (Wiegand and Kominkiewicz 2016)? How can librarians best teach skills, foster engagement, and create knowledge to make a distinctive contribution to the institution? Our answers will determine the library's future at each academic institution. Machine Learning skills, engagement, and knowledge should fit well with the library's array of services.

Learning is another traditional aspect of library services, this time from the student point of view. The library provides collections—multimedia or print on paper, digital and digitized, proprietary and open, local, redundant, rare, unique. The use of collections is taught by both librarians and disciplinary faculty in the service of learning, including life-long learning for non-academic, everyday knowledge. Students need to know more about Machine Learning, from data literacy to digital competencies, including concerns about privacy, security, and fake news across the curriculum, while learning skills associated with Machine Learning. In addition, through Open Access, library “collections” now encompass the world beyond the library’s physical and virtual spaces.

Then, as libraries, like all digitally-inflected institutions, develop “change management” strategies, they need to double-down on these unique affordances and communicate them to stakeholders. The most critical strategy is embedding the Scholarship of Teaching and Learning (SoTL) in all aspects of the library workflow. Instead of simply advertising new electronic resources or describing Open Access versus proprietary resources, libraries should broadly embed the lessons of copyright, surveillance, and reproducibility into patron interactions, from the first undergraduate literature search to the faculty research consultation. Then, reinforce those lessons by emphasizing open access and data mining permissions in their discovery tools. These are aspects of the scholarly research cycle over which libraries have some control. By exerting that control, libraries will promote a culture that positions Machine Learning and other creative digital uses of library data as normal, achievable parts of the scholarly process.

To complete the Scholarly Communications lifecycle, support for research, scholarship, and creative works is increasingly provided by libraries as a springboard to creation of knowledge, the library’s newest role. This is where Machine Learning as a new paradigm fits in most compellingly as an innovative practice. Libraries can provide not only associated services such as Data Management of the datasets resulting from analyzing huge textual corpora, but also databases of propri-

etary and locally-produced content from inter-connected, cooperating libraries on a global scale. Researchers—faculty, students, and citizens (including alumni)—will benefit from crowdsourcing and citizen science while gaining knowledge and contributing to scholarship. But perhaps the largest benefit will be learning by doing, escaping the “black box” of blind consumerism to see how algorithms work and thus develop a more nuanced view of reality in the Machine Age.

References

- Bedrossian, Rebecca. 2018. “Recognizing Exclusion is the Key to Inclusive Design: In Conversation with Kat Holmes.” *Campaign* (blog). July 25, 2018. <https://www.campaignlive.com/article/recognizing-exclusion-key-inclusive-design-conversation-kat-holmes/1488872>.
- Bell, Mark. 2018. “Machine Learning in the Archives.” *National Archives* (blog). November 8, 2020. <https://blog.nationalarchives.gov.uk/machine-learning-archives/>
- Bivens-Tatum, Wayne. 2012. *Libraries and the Enlightenment*. Los Angeles: Library Juice Press. Accessed January 6, 2020. ProQuest Ebook Central.
- Bogost, Ian. 2019. “The AI-Art Gold Rush is Here.” *The Atlantic*. March 6, 2019. <https://www.theatlantic.com/technology/archive/2019/03/ai-created-art-invades-chelsea-galler>.
- Casson, Lionel. 2001. *Libraries in the Ancient World*. New Haven: Yale University Press. Accessed January 6, 2020. ProQuest Ebook Central.
- Dempsey, Lorcan. 2016. “Library Collections in the Life of the User: Two Directions.” *LIBER Quarterly* 26: 338–359. <https://doi.org/10.18352/lq.10170>
- Extance, Andy. 2018. “How AI Technology Can Tame the Scientific Literature.” *Nature* 561: 273–274. <https://doi.org/10.1038/d41586-018-06617-5>
- Golub, K. 2006. “Automated Subject Classification of Textual Web Documents.” *Journal of Documentation* 62: 350–371. <https://doi.org/10.1108/00220410610666501>
- Jakeway, Eileen. 2020. “Machine Learning + Libraries Summit: Event Summary now live!” *The Signal* (blog), *Library of Congress*. February 12, 2020. <https://blogs.loc.gov/the-signal/2020/02/machine-learning-libraries-summit-event-summary-now-live/>
- Joorabchi, Arash and Abdulhussin E. Mahdi. 2011. “An Unsupervised Approach to Automatic Classification of Scientific Literature Utilising Bibliographic Metadata.” *Journal of Information Science*. <https://doi.org/10.1177/0165551500000000>
- Khamsi, Rozanne. 2020. “Coronavirus in context: Scite.ai Tracks Positive and Negative Citations for COVID-19 Literature.” *Nature*. <https://doi.org/10.1038/d41586-020-01324-6>
- Padilla, Thomas, Laurie Allen, Hannah Frost, et al. 2019. “Final Report — Always Already Computational: Collections as Data.” *Zenodo*. May 22, 2019. <https://doi.org/10.5281/zenodo.3152935>
- Price, Gary. 2019. “The Library of Congress Posts Solicitation For a Machine Learning/Deep Learning Pilot Program to ‘Maximize the Use of its Digital Collection.’” *Library Journal*. June 13, 2019.
- Rodrigues, Eloy et al. 2017. “Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group.” *Zenodo*.

- November 28, 2017. <https://doi.org/10.5281/zenodo.1215014>.
- Ryholt, K. S. B., and Gojko Barjamovic, eds. 2019. *Libraries Before Alexandria: Ancient near Eastern Traditions*. Oxford: Oxford University Press.
- Vamathevan, Jessica, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Lee, Anant Madabhushi, Parantu Shah, Michaela Spitzer, and Shanrong Zhao. 2019. “Applications of Machine Learning in Drug Discovery and Development.” *Nat Rev Drug Discov* 18: 463–477. <https://doi.org/10.1038/s41573-019-0024-5>
- Wang, Jun. 2009. “An Extensive Study on Automated Dewey Decimal Classification.” *Journal of the American Society for Information Science & Technology* 60: 2269–86. <https://doi.org/10.1002/asi.21147>
- Wiegand, Sue. 2013. “ACS Solutions: The Sturm und Drang.” *ACRLog* (blog), Association of College and Research Libraries. November 8, 2020. <https://acrlog.org/2013/04/06/acs-solutions-the-sturm-und-drang/>
- Wiegand, Sue and Frances Kominkiewisz. 2016. Unpublished manuscript. “Integration of Student Learning through Library and Classroom Instruction.”
- Yewno. n.d. “Yewno — Transforming Information into Knowledge.” Accessed January 6, 2020. <https://www.yewno.com/>

Further Reading

- Abbattista, Fabio, Luciana Bordoni, and Giovanni Semeraro. 2003. “Artificial Intelligence for Cultural Heritage and Digital Libraries.” *Applied Artificial Intelligence* 17, no. 8/9: 681. <https://doi.org/10.1080/713827258>
- Ard, Constance. 2017. “Advanced Analytics Meets Information Services.” *Online Searcher* 41, no. 6: 21–24.
- “Artificial Intelligence and Machine Learning in Libraries.” 2019. *Library Technology Reports* 55, no. 1: 1–29.
- Badke, William. 2015. “Infolit Land. The Effect of Artificial Intelligence on the Future of Information Literacy.” *Online Searcher* 39, no. 4: 71–73.
- Boman, Craig. 2019. “Chapter 4: An Exploration of Machine Learning in Libraries.” *Library Technology Reports* 55: 21–25.
- Breeding, Marshall. 2018. “Chapter 6: Possible Future Trends.” *Library Technology Reports* 54, no. 8: 31–32.
- Dempsey, Lorcan, Constance Malpas, and Brian Lavoie. 2014. “Collection Directions: The Evolution of Library Collections and Collecting” *portal: Libraries and the Academy* 14, no. 3 (July): 393–423. <https://doi.org/10.1353/pla.2014.0013>
- Enis, Matt. 2019. “Labs in the Library.” *Library Journal* 144, no. 3: 18–21.
- Finley, Thomas. 2019. “The Democratization of Artificial Intelligence: One Library’s Approach.” *Information Technology & Libraries* 38, no. 1: 8–13. <https://doi.org/10.6017/ital.v38i1.10974>
- Frank, Eibe and Gordon W. Paynter. 2004. “Predicting Library of Congress Classifications From Library of Congress Subject Headings.” *Journal of The American Society for Information Science and Technology* 55, no. 3. <https://doi.org/10.1002/asi.10360>
- Geary, Daniel. 2019. “How to Bring AI into Your Library.” *Computers in Libraries* 39, no. 7: 32–35.

- Griffey, Jason. 2019. "Chapter 5: Conclusion." *Library Technology Reports* 55, no. 1: 26–28.
- Inayatullah, Sohail. 2014. "Library Futures: From Knowledge Keepers to Creators." *Futurist* 48, no. 6: 24–28.
- Johnson, Ben. 2018. "Libraries in the Age of Artificial Intelligence." *Computers in Libraries* 38, no. 1: 14–16.
- Kuhlman, C., L. Jackson, and R. Chunara. 2020. "No Computation without Representation: Avoiding Data and Algorithm Biases through Diversity." ArXiv:2002.11836v1 [cs.CY], February. <http://arxiv.org/abs/2002.11836>
- Lane, David C. and Claire Goode. 2019. "OERu's Delivery Model for Changing Times: An Open Source NGDLE." Paper presented at the 28th ICDE World Conference on Online Learning, Dublin, Ireland, November 2019. <https://oeru.org/assets/Marcoms/OERu-NGDLE-paper-FINAL-PDF-version.pdf>
- Liu, Xiaozhong, Chun Guo, and Lin Zhang. 2014. "Scholar Metadata and Knowledge Generation with Human and Artificial Intelligence." *Journal of the Association for Information Science & Technology* 65, no. 6: 1187–1201. <https://doi.org/10.1002/asi.23013>
- Mitchell, Steve. 2006. "Machine Assistance in Collection Building: New Tools, Research, Issues, and Reflections." *Information Technology & Libraries* 25, no. 4: 190–216. <https://doi.org/10.6017/ital.v25i4.3353>
- Ojala, Marydee. 2019. "ProQuest's New Approach to Streamlining Selection and Acquisitions." *Information Today* 36, no. 1: 16–17.
- Ong, Edison, Mei U. Wong, Anthony Huffman, and Yongqun He. 2020. "COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning." *Frontiers in Immunology* 11. <https://doi.org/10.3389/fimmu.2020.01581>
- Orlowitz, Jake. 2017. "You're a Researcher Without a Library: What Do You Do?" *A Wikipedia Librarian* (blog), Medium. November 15, 2017. <https://medium.com/a-wikipedialibrarian/youre-a-researcher-without-a-library-what-do-you-do-6811a30373cd>
- Padilla, Thomas. 2019. *Responsible Operations: Data Science, Machine Learning, and AI in Libraries*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/xk7z-9g97>
- Plosker, George. 2018. "Artificial Intelligence Tools for Information Discovery." *Online Searcher* 42, no. 3: 31–35. <https://www.infotoday.com/OnlineSearcher/Articles/Features/Artificial-Intelligence-Tools-for-Information-Discovery-124721.shtml>
- Rak, Rafal, Andrew Rowley, William Black, and Sophie Ananiadou. 2012. "Argo: an Integrative, Interactive, Text Mining-based Workbench Supporting Curation." *Database: the Journal of Biological Databases and Curation*. <https://doi.org/10.1093/database/bas010>
- Schmidt, Lena, Babatunde Kazeem Olorisade, Julian Higgins, and Luke A. McGuinness. 2020. "Data Extraction Methods for Systematic Review (Semi)automation: A Living Review Protocol." *F1000Research* 9: 210. <https://doi.org/10.12688/f1000research.22781.2>
- Schonfeld, Roger C. 2018. "Big Deal: Should Universities Outsource More Core Research Infrastructure?" *Ithaka S+R*. <https://doi.org/10.18665/sr.306032>
- Schockley, Nick. 2013. "How Open Access Empowered a 16-year-old to Make Cancer Breakthrough." June 12, 2013. <http://www.openaccessweek.org/video/video/show?id=5385115%3AVideo%3A90442>

- Short, Matthew. 2019. “Text Mining and Subject Analysis for Fiction; or, Using Machine Learning and Information Extraction to Assign Subject Headings to Dime Novels.” *Cataloging & Classification Quarterly* 57, no. 5: 315–336. <https://doi.org/10.1080/01639374.2019.1653413>
- Thompson, Paul, Riza Theresa Batista-Navarro, and Georgio Kontonatsios. 2016. “Text Mining the History of Medicine.” *PLoS One* 11, no. 1:e0144717. <https://doi.org/10.1371/journal.pone.0144717>
- White, Philip. 2019. “Using Data Mining for Citation Analysis.” *College & Research Libraries* 80, no. 1. https://scholar.colorado.edu/concern/parent/cr56n1673/file_sets/9019s3164
- Witbrock, Michael J. and Alexander G. Hauptmann. 1998. “Speech Recognition for a Digital Video Library.” *Journal of the American Society for Information Science* 49, no. 7: 619–32. [https://doi.org/10.1002/\(SICI\)1097-4571\(19980515\)49:7<619::AID-ASIT4>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-4571(19980515)49:7<619::AID-ASIT4>3.0.CO;2-A)
- Zuccala, Alesia, Maarten Someren, and Maurits Bellen. 2014. “A Machine-Learning Approach to Coding Book Reviews as Quality Indicators: Toward a Theory of Megacitation.” *Journal of the Association for Information Science & Technology* 65, no. 11: 2248–60. <https://doi.org/10.1002/asi.23104>