An overview of the NFAIS Conference: Artificial intelligence: Finding its place in research, discovery, and scholarly publishing

Bonnie Lawlor[*]
Guest Editor, NFAIS Honorary Fellow, 276 Upper Gulph Road, Radnor, PA 19087, USA

Abstract. This paper offers an overview of the highlights of the NFAIS Conference, *Artificial Intelligence: Finding Its Place in Research, Discovery, and Scholarly Publishing,* that was held in Alexandria, VA from May 15–16, 2019. The goal of this conference was to explore the application and implication of Artificial Intelligence (AI) across all sectors of scholarship. Topics covered were, among others: the amount of data, computing power, and the technical infrastructure required for AI and Machine Learning (ML) processes; the challenges to building effective AI and ML models; how publishers are using AI and ML in order to improve discovery and the overall user search experience; what libraries and universities are doing to foster an awareness of AI in higher education; and an actual case study of using AI and ML in the development of a recommendation engine. There was something for everyone.

Keywords: Scholarly publishing, artificial intelligence, machine Learning, AI models, high dimensional data, wikipedia, overfitting, TrendMD, recommendation engines, AI labs, AI hubs, Amazon SageMaker

1. Introduction

> "Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI (Artificial Intelligence) will transform in the next several years [1]".

The scientific publishing industry is no exception to the above quote. This conference provided a glimpse as to how the use of Artificial Intelligence (AI) and Machine Learning (ML) are allowing innovative publishers to mine their data and provide information seekers with knowledge rather than a list of answers to queries. Fortunately for me, this was not a conference for the AI computer-savvy, rather it was for non-techies who wanted to learn how ML and AI are being used within the Information Community - by publishers, librarians, and vendors - and as you read this article I hope that you will agree with me that the goal was met.

But listening to the speakers and subsequently writing this article motivated me to learn more about AI and ML simply because I needed to understand the jargon and how it was being used. How old are the technologies? Do they differ and if so, how? I found the answers and will briefly share them with you before you move on to the meat of this paper.

The concept of AI was captured in a 1950's paper by Alan Turing, "Computing Machinery and Intelligence [2]", in which he considered the question "can machines think". It is well-written, understandable

---

[*]E-mail: chescot@aol.com.

by a non-techie like me, and worth a read. Five years later a proof-of-concept program, Logic Theorist [3], was demonstrated at the Dartmouth Summer Research Project on Artificial Intelligence [4] where the term "Artificial Intelligence" was coined and the study of the discipline was formally launched (for more on the history and future of AI, see, "The History of Artificial Intelligence [5]").

A concise one-sentence definition of AI is as follows: "Artificial Intelligence is the science and engineering of making computers behave in ways that, until recently, we thought required human intelligence [6]". The article from which that definition was extracted went on to define "Machine Learning" as "the study of computer algorithms that allow computer programs to automatically improve through experience [7]". ML is one of the tools with which AI can be achieved and the article goes on to explain why the two terms are so often incorrectly interchanged (ML and AI are *not* the same. ML, as well as Deep Learning, are subsets under the overarching concept of AI). That same article defines a lot of the ML jargon and is worth a read as a complement to some of the more technical papers that appear elsewhere in this issue of *Information Service and Use*.

AI and ML are critical to scholarly communication in general. On January 16, 2019 the Forbes Technology Council posted the top thirteen industries soon to be revolutionized by AI, one of which is business intelligence;

> "Enterprises are overwhelmed by the volume of data generated by their customers, tools and processes. They are finding traditional business intelligence tools are failing. Spreadsheets and dashboards will be replaced by AI-powered tools that explore data, find insights and make recommendations automatically. These tools will change the way companies use data and make decisions [8]".

The term "enterprises" can easily be replaced by "publishers", "funding agencies" research labs", "universities", etc. One of the conference speakers, Rosina Weber, stated that the richest source of data is the proprietary information held within scientific publications. With access to this information AI can generate the most accurate information with which to make informed decisions - including research decisions for the advancement of science. So read on learn about some of the more recent AI-related initiatives in publishing. Enjoy!

2. Opening keynote

The conference opened with a presentation by **Christopher Barbosky**, Senior Director, Chemical Abstracts Service (CAS; see: https://www.cas.org/). The point of his presentation was that Machine Learning (ML) has always required high quality data and lots of it. However, data alone is not enough. Successful implementation of AI technologies also requires significant investments in talent, technology, and the underlying processes essential to maximizing the value of all that data.

He began by giving some examples of the amount of data required. It takes four million facial images to obtain 97.35% accuracy in facial recognition; eight hundred thousand grasps are required to train a robotic arm; fifty thousand hours of voice data is required for speech recognition; and it takes one hundred and fifty thousand hours to train an ML model to recognize a hot dog. To get the data, clean it, and maintain it requires a great deal of computing power.

And why is so much data required? He said that in the past the process was rules (human devised algorithms) plus data gave answers. Today it is answers plus data creates the rules - computers can now learn functions by discovering complex rules through training. He suggested that those interested read the book, *Learning from Data - a Short Course* [9]. I took a quick look at the book and it is technical, but it

also has excellent examples of how to build ML models and provides understandable definitions of the jargon used in the field (and by some of the speakers at this conference; e.g., supervised modelling vs. unsupervised modelling, overfitting, etc.). Worth a look!

Barbosky said that ML, AI, and computing power are necessary to keep up with the amount of information being generated today. The volume of discovery and intellectual property is increasing dramatically. He added that 90% of the world's data was created in the past two years at a rate of 2.5 quintillion bytes per day [10]. In addition 73% of the world's patent applications are not written in the English language and they are increasing both in their number and in their complexity. He said that today publishers not only must invest in building a robust data collection, but they must also take the time to structure and curate that data. In addition, they must allot appropriate resources to support ongoing data maintenance. Once the "factual" data collection has been structured they must create "knowledge" from those facts to generate understanding, and the third step is to apply AI and ML to create search and analytic services that ultimately supply the user with "reasoning." The goal is to minimize the time and effort required by the user to discover, absorb, and make "sense" of the information that is delivered to them.

Using CAS as an example he noted that they gather journals, patents, dissertations, conference materials, and technical reports. They then perform concept indexing, chemical substance indexing, and reaction indexing to gather facts. And then, using machines and humans, perform analytics in order to provide the users with insights. The machines handle the low-hanging fruit and the humans handle the more complex material. Using AI and ML, CAS builds knowledge graphs and is able to discern increased connections across scientific disciplines.

Barbosky said that one of the biggest problems today is finding people with the right skills as there is a shortage of data scientists. Indeed, 27% of companies cite a skills gap as the major impediment to data-related initiatives. He compared the data scientist to a chef who needs a pool of talented people to support him/her. The expertise of the team must cover: data acquisition and curation; data governance stewards; data modelers; data engineers; data visualizers (reports/descriptive analytics); and those who develop and maintain the requisite infrastructure. But in addition to the skilled team, a technology "stack" is required to make it all happen. He compared the stack to a pyramid: the foundation is the hardware, the next layer Hadoop [11], the third layer NoSQL [12], the next layer data analytics software; with the resulting Peak being the value provided to users. The peak is where the business opportunities appear: in product data features, in better search results, and in value from unstructured data.

Barbosky summarized as follows: (1) Big data technologies offer vast potential for innovation and efficiency gains across a wide array of industries; (2) AI/ML projects require high quality and varied data; (3) Scientific information will continue to grow in size and complexity; and (4) Invest in a data platform with support from technical and domain experts.

He closed with the following quote:

"The computer is incredibly fast, accurate and stupid. Man is unbelievably slow, inaccurate, and brilliant. The marriage of the two is a challenge and opportunity beyond imagination".

Barbosky used the quote when showing a photo of a painting entitled *Edmond de Belamy* [14]. It is believed to be the first piece of artwork created by Artificial Intelligence to be auctioned at Christie's. It sold for a whopping $432,500!!!

Barbosky's slides are not available on the NFAIS web site.

3. Building AI and machine learning models

The second speaker was **Dr. Daniel Vasicek**, Data Scientist and Engineer, from Access Innovations, Inc. (see: https://www.accessinn.com/) who discussed how the application of artificial intelligence (AI) to the indexing process can make discovery services more valuable (although he primarily focused on the challenges faced when building ML models). He noted that Machine Learning (ML) and AI have great potential in every field, including publishing, as text is simply one type of data. Potential applications include optical character recognition (OCR), sentiment analysis, trend analysis, concept extraction, entity extraction, topic assignment, quality analysis, and more.

Vasicek talked about neural networks - a set of algorithms modeled loosely after the human brain - that are designed to recognize patterns. They interpret sensory data through a kind of machine perception, labeling or clustering raw input. He uses them for analysis to fit complex models to the high dimensional data that often occurs in the publishing industry. High dimensionality arises in publishing because we are usually interested in a large number of concepts; for example, a moderate thesaurus will contain thousands of concepts. Useful text will have a large number of words. Therefore, the utilization of AI and ML for the discovery of ideas, sentiments, tendencies, and context requires that the algorithms be aware of many different features such as the words themselves, the length of sentences (and paragraphs), word frequency counts, phrases, punctuation, number of references, and links.

He noted that all data has noise - even publishing data - and that large numbers of adjustable parameters can result in the "overfitting" of analysis models such as those used in ML and AI. Overfitting occurs when a proposed predictive model becomes so complex that it fits the noise in the data and this is a very common problem encountered in ML. Two critical factors lead to overfitting - data uncertainty and model uncertainty. To reinforce these factors he added that that real data has measurement errors and that only the use of quantum mechanics guarantees uncertainty. He noted that even if a room was built to be a perfect square it probably is not "perfect", just as the orbits of the planets are not perfect ellipses because they are perturbed by Jupiter, Saturn, etc. The point is, models are *never* perfect!

He said that AI goes back at least two hundred years. In support of this fact he talked about the asteroid Ceres (originally thought to be a planet) that was discovered January 1, 1801 by Giuseppe Piazzi. Piazzi followed it every day as it moved, but had to stop due to an illness. When he returned to his efforts in February of that same year he could not find it (Ceres was "lost" due to the glare of the sun). One of his students, Carl Friedrich Gauss, using mathematics to predict its location, rediscovered Ceres on December 31, 1801 [15]. In his attempt to find it, Piazzi used a circular orbit as a model while Gauss used ellipses as the orbit model. This is an example of how "regularization" methods offer techniques that will reduce the tendency of the model to overfit the data. As a result of regularization, models may do a much better job on new data that is not part of the original training data set. It can force the model to be smoother and reduce the impact of noise.

Vasicek then went on to provide detailed examples that clarified the concepts of "overfitting" and "regularization". He gave examples of how the same data can be matched perfectly by different models and asked which is correct? He answered that technically all of them are correct, but that the simpler model is most likely to be the better one.

He said that models are always imperfect and that data is always noisy and that we must be aware that both of these problems can affect analysis and reduce the usefulness of our predictions. You need to really "know" the data that you are using and its level of uncertainty. It is important to remember that models that are complex enough to fit noise are not useful and can prove to be useless. When we have an imperfect model and erroneous data, we need to balance the cost associated with the imperfect model

against the cost of the erroneous data. He gave the following example. A publisher has five million articles in its databases and wants to develop author profiles for peer review. They apply ML to develop author profiles. Unfortunately, overfitting of the model used leads the AI to assign *specific* concepts to authors; e.g., "Quantum entanglement" rather than the general term "Physics". The generation of very specific topics can lead the machine to *not* suggest an author for peer review because they are not an expert in broader topics such as "Physics".

In closing, Vasicek said that overfitting is a constant challenge with any ML task as most ML applications may generate very complex models which will actually fit the data "better", but are actually fitting the noise. He added that the model can learn the noise as well as it can learn the real information and that using concept identification rules for text helps mitigate overfitting. He suggested that those interested take a look at Google TensorFlow. This is an open source ML tool with tutorials and the software is ready to use out of the box. It can be used to build multi-layer neural networks and is good tool with which to experiment with deep learning [16].

Vasicek's slides are available on the NFAIS website and an article based upon his presentation appears elsewhere in this issue of *Information Services and Use.*

4. AI trends around the globe

The next speaker was **Dr. Bamini Jayabalasingham**, Senior Analytical Product Manager at Elsevier (see: https://www.elsevier.com/), who spoke on the use of A rtificial Intelligence to create, transfer and use knowledge in addition to highlighting the current trends for the use of AI around the globe. She noted that AI is often used as an umbrella term to describe computers applying judgment just as a human being would. Yet, there is no universally agreed-upon definition of AI as the term is frequently interchanged (incorrectly, I must add) with Machine Learning (ML). She believes that such a definition is required to ensure that policy objectives are correctly translated into research priorities, that student education matches job market needs, and that the media can compare knowledge being developed. She said that Elsevier is the first to characterize the field of AI in a comprehensive, structured manner using extensive datasets from their own and public sources. These datasets were examined by Elsevier's data scientists through the application of ML principles and the results were validated by domain experts from around the world.

The body of data was extracted from textbooks, funding information, patents, and media reports and resulted in eight hundred thousand keywords. Eight hundred query terms were mapped against six million articles in Scopus and the output was filtered based on metadata, resulting in six hundred thousand documents that were then analyzed to identify topics.

The results showed that globally, research in the field of AI clusters around seven topics: search and optimization; fuzzy systems; planning and decision making; natural language processing and knowledge representation; computer vision; neural networks; and ML and probabilistic reasoning. The data that they analyzed showed that there are more than sixteen million active authors in the field who are affiliated with more than seventy-thousand organizations, and that there are more than seventy thousand articles, conference papers, and book records related to AI. The number of AI research papers is growing at 12% per year globally. China is the most prolific, followed by the USA, India, Germany, Japan, Spain, Iran, France, and Italy. The share of world publications on AI is growing. The data from 2013 to 2017 show that China has 24%, Europe 30%, the USA 17%, and all other 29%.

She noted that China is very focused on the applied side of AI, with publications targeted towards the topics of computer vision, neural networks, planning and decision making, fuzzy systems, natural language processing, and knowledge representation. Europe is the most diverse, and the USA is very strong in the corporate sector, with a lot of research coming out of IBM and Microsoft. Also, the USA is attracting a lot of talent from overseas to both the corporate and academic sectors.

She said that Elsevier plans to continue its research in the hopes of answering the following questions:

- How can we further improve the AI ontology and field definition?
- Is there a relationship between research performance in AI and research performance in the more traditional fields that support AI (such as computer science, linguistics, mathematics, etc.)?
- How does AI research translate into real-life applications, societal impact, and economic growth? How sustainable is the recent growth in publications, and how will countries and sectors continue to compete and collaborate?

The results of the first stage of research has been published in a report, *Artificial Intelligence: How Knowledge is Created, Transferred, and Used: Trends in China, Europe, and the United States* and is freely-available to download as a full report [17] and as an Executive Summary [18].

Jayabalasingham's slides are available on the NFAIS website and her article that discusses global AI trends appears elsewhere in this issue of *Information Services and Use*. Note that her article has links to many AI reports and studies that you may find to be of interest. Also, another Elsevier speaker at the conference, Ann Gabriel, has a paper in this issue on a similar topic.

5. Reinforcing the need for publishers to evolve

The next speaker was **Sam Herbert**, Co-founder, 67 Bricks (see: https://www.67bricks.com), who spoke on how publishers need to change in order to remain relevant in an era of Artificial Intelligence. He opened with a quote from Sergey Brin, Co-Founder of Google:

"AI is the most significant development in computing in my lifetime. Every month, there are stunning new applications and transformative new techniques [19]".

Herbert noted that the time it takes to do an AI analysis has gone from weeks to hours. He mentioned an organization, Open AI (see: https://openai.com), whose charter is to see that Artificial Intelligence benefits all of humanity and the resource section of their website offers free software for training, benchmarking, and experimenting with AI - definitely worth taking a look. He gave the example of how medical researchers today are able to generate data from brain scans (*Note:* the importance of this data was discussed at the 2019 NFAIS Annual Conference by Dr. Daniel Barron, a resident psychiatrist at Yale University. An overview of the conference will be published in *Information Services and Use* this fall [20]. I was fascinated by the use of AI to decode brain data and found an article that discusses software developed in 2017 that allows for 'decoding digital brain data' to reveal how neural activity gives rise to learning, memory, and other cognitive functions [21].)

Herbert went on to say that if publishers want to remain relevant, they have to provide a better user experience. He gave Uber, Google, and Amazon as examples of major innovators who have revised traditional industries by doing just that. The following is an example that he provided of how the transportation industry has been transformed:

Transportation Industry - 1999	Transportation Industry - 2019
Go outside (in the cold) and hail a taxi	Take out your phone (from anywhere) and request an Uber
Don't know how long it will take	Know how long it will take
Don't know your vehicle or what passengers think of your driver…	Know your vehicle and passenger feedback ratings about your driver
Enter a (mostly) dirty taxi	Enter a (mostly) clean car
Must converse with your driver about where you are going, etc.	Not necessary to speak with your driver, it's up to you
Upon arrival, pay driver via cash	Upon arrival, payment including pre-determined tip
Ask for and wait for hand written receipt	Receipt automatically sent via email

In comparison, he said that in his opinion, the user experience offered by scholarly publishers has not really changed much since 1999, and offered the following example:

Scholarly Publishing - 1999	Scholarly publishing - 2019
Research written up as human readable 'paper	Research written up as human readable 'paper' (data and video in early stages)
Painful non-user friendly submission process	Painful non-user friendly submission process (mostly)
Subjective untrained human peer review process	Subjective untrained human peer review process
Research packaged in periodical journals (this structure dates from the 17th century)	Research packaged in periodical journals
Scientific research behind paywalls	Most scientific content still behind paywalls
Researchers value mostly determined by journal impact factor	Researchers value still mostly determined by journal impact factor
Powerful incentives to withhold results for months or years until research is published…	Powerful incentives to withhold results for months or years until research is published (not all fields)

He said that the importance of ease, speed, and access to content is as important as the quality of the information, and noted that academic publishers can deliver better user experiences by developing a product mindset underpinned by data "maturity" and new software paradigms. He said that data maturity is defined by the publisher's ability to store, manage, create, and use data to deliver value to their users. He said that "low" maturity (e.g., document-based storage and access), requires much effort on the part of the user to access, digest, and use the data. On the other hand "high" maturity (/e.g., information in context, insights (both predictive and prescriptive) offers an experience that requires far less effort on the part of the user. He added that users now expect customized knowledge and tailored recommendations delivered digitally. Data is at the heart of everything and by increasing their data maturity publishers will be able to develop the innovative products and that users demand. (This echoes the comments made during the opening keynote!)

Herbert gave an example of how his organization was able to help SAGE Publishing improve their customer relations. Their sales team worked to create tailored value reports containing detailed information about how SAGE products fit each customer's information needs. In academic institutions this involved examining course curricula together with research profiles of each institution and analyzing them against SAGE content - a highly-manual, time-consuming effort. SAGE asked 67 Bricks to undertake an R&D project to create a working prototype tool that would take a description of an educational course and return a list of SAGE content relevant to the course. Using Neuro-linguistic Programming Techniques, a working prototype was developed and has since been rolled-out to members of the sales, marketing, and editorial teams [22]. This has allowed SAGE to save time and money and has removed some tedious manual tasks from the editorial staff.

In closing, Herbert noted that the publishing industry is one of many that is rapidly changing as a result of AI applications. Scholarly publishers need to focus on delivering new and improved user experiences and he believes that by becoming a data-driven product company publishers will be well- positioned in this changing environment. He concluded by saying that the future of data-driven products is as follows:

- Predicting high impact research
- Improving marketing communications
- Delivering customized/personalized experiences
- Helping researchers discover content
- Delivering information services rather than documents
- Selling data to machine learning companies
- Creating automated or semi-automated content

- Understanding the links between research objects
- Augmenting/automating peer review
- Predicting emerging subject areas
- Identifying peer reviewer
- Adaptive learning products Cross selling across content domains
- Unlocking value in legacy content
- Automated creation of marketing materials Improve internal content discovery and research

Herbert's slides are available on the NFAIS website and a brief article based upon his presentation appears elsewhere in this issue of *Information Services and Use.*

6. Managing manuscript submissions and Peer review

The next speaker was **Greg Kloiber,** Senior User Experience Designer, ScholarOne Manuscripts, at Clarivate Analytics (see: https://clarivate.com/), who spoke about a pilot project that they implemented

that uses Artificial Intelligence (AI) to help authors submit higher-quality papers and also speeds up the editorial evaluation process by providing insightful statistics. ScholarOne Manuscript is an online tool that allows hundreds of publishers to manage their submission and peer-review workflows and currently has ten million active users. He noted that the volume of papers going through the publishing workflow continues to rise and that there has been a 6.1% annual growth of submissions since 2013. In the 2015–2017 time period 5.5million new papers have been submitted to ScholarOne journals. 33% were accepted after review; 27% were rejected after review; and 40% were rejected without review. While screening is important, it takes a lot of time and effort for editors [23]. He noted that 2.3 reviewers are needed for each submission and that it takes five hours for each review [24]. In addition, it takes a little over sixteen days (16.4) to turnaround a review [25]. He referred the audience to another in which similar statistics were shown [26].

What ScholarOne wanted to do was to reduce the time required to check and screen every paper and also reduce the number of papers that are rejected after peer review. They partnered with UNISILO (see: https://unsilo.ai), a company in Denmark that provides AI tools for publishers, and entered into a pilot project using UNISILO's Evaluate, a suite of APIs (Application Programming Interfaces) that help publishers screen and evaluate incoming manuscript submissions, assisted by Artificial Intelligence (AI) and Neuro-Linguistic Programming tools.

The six-month pilot consisted of seven major publishers with forty-three participating editors from around the world. More than twenty qualitative interviews were conducted before, during, and after the pilot. The tool had six screening widgets: key statements, key words (topics), technical checks, related papers, journal matching, and reviewer finder (the latter was not used as ScholarOne already had a similar tool). More widgets are planned and include industry trends and a language qualifier. The initial response from editors was the following statement: "Usually people have a suspicion that something might need to be checked, not the other way around. This (tool) changes that!"

Kloiber said that as a result of the pilot they learned that the title "editor" does not always translate into the same job function - those participating had multiple roles: Editor, Associate Editor, Senior Editor, Managing Editor, Executive Editor, Editor-in-Chief, Clinical Editor, Development Editor, Section Editor, Publisher, Publishing Editor, Senior Publishing Editor, Head of Education, International Audience Editor, and Journal Manager/Manager Online Submission Systems. They found the widgets to be valuable and believe that authors also might benefit by the tool to help with abstracts, keywords, and the identification of journals to which they should submit their manuscript. He said that ScholarOne will begin with the key statement and keyword widgets as reviewers said that the combination of those two were sufficient to make a decision for accepting/rejecting a manuscript. They also used "key statements" in preparing rejection letters.

He said that an interesting side fact learned was that 25% of the editors fear that AI will eliminate their jobs and he had to reinforce the fact that this was to assist them and improve productivity, not take away positions. Some AI lessons learned were that:

- Can a keyword have two or more words? The answer was "yes" and AI had to determine if it was a sentence or what.
- The AI needs to address row/page numbering within a PDF proof.
- The AI needs to consider when a dot (.) is a decimal point or a period.
- The AI needs to be able to handle the various taxonomies within a research field.

Kloiber said that the next steps include using authors and editors to "train" the AI and to roll out the keyword and key statement widgets. He recommended the following additional reading material for those interested in learning more:

- Using AI to solve business problems in scholarly publishing Author: Michael Upshall, UKSG, https://insights.uksg.org/articles/10.1629/uksg.460/, accessed September 12, 2019.
- Mythbusting AI: What is all the fuss about? Author: Michael Upshall, Wiley Online Library, https://www.onlinelibrary.wiley.com/doi/full/10.1002/leap.1211, accessed September 12, 2019.
- Video: The Silverchair Universe Presents: UNSILO, available at: https://www.youtube.com/watch?v=u4U5TM7APLs, accessed September 12, 2019.

Kloiber's slides are available on the NFAIS website.

The final speaker of the day was **Donald Samulack**, President, U.S. Operations, Editage/Cactus Communications (see: https://www.editage.com), who also discussed a decision-making tool for manuscript evaluation, entitled Ada (see: https://www.editage.com/Ada-by-editage.html).

Samulack opened with a description of the evolution of the modern information era as follows:

- *The Internet:* If only I had access to more information
- *Computing Power:* If only I could process more information
- *Semantic Search:* If only I could extract the right information easier/faster
- *Machine Learning/Artificial Intelligence:* If only I was presented with information that was relevant, digested, and tailored to my needs
- *Wearable/Implanted Technologies:* If only I was integrated with the knowledge and world around me

Editage serves both authors and publishers - the former by providing diverse editing services and the latter by providing a manuscript evaluation tool that can be integrated into their workflow systems. They also have a Data Science lab, a Natural Language Processing lab, and an Artificial Intelligence lab.

He noted that the challenge that they address for publishers is that it can take up to four months from the time that an author submits their manuscript to the time when they are informed if the manuscript has been accepted or rejected. Also, during that process the screening by editors is largely manual and *subjective*. Their goal was to provide editors with an *objective* tool with which to make their decisions and, as mentioned by Kloiber earlier, to reduce the processing time. He reinforced the prior speaker's comments by saying that the scientific research output doubles every six years; that editorial screening is increasingly critical, but time-consuming; that one out of five manuscripts are rejected during "desk review"; and that there is a downward pressure on publishing time and costs. As a result of these trends the mission of Editage's Ada is to be "an industry-first, automated, customizable, manuscript screening solution that saves publication costs and time downstream for academic publishers". "Ada" stands for Automated Document Assessment", but the name is also in honor of Augusta Ada King, Countess of Lovelace, who was an English mathematician and writer, chiefly known for her work on Charles Babbage's proposed mechanical general-purpose computer, the Analytical Engine [27].

One of their key focuses was being able to perform a readability assessment of manuscripts. They defined readability as the ease with which a reader can understand a written text. Samulack said that according to Peter Thrower, PhD, Editor-in-Chief of *Carbon*, one of the top reasons of manuscript rejection is language comprehension [28]. Editage started with public algorithms, but soon realized that none of them provided the fine-tuning required for academic content. In addition, the accuracy was very low

and inconsistent with human assessment. So they went back to the drawing board with their own in-house linguists and studied hundreds of manuscript - a effort that resulted in a proprietary formula that looks at more than thirty-plus different metrics of language structure, composition, spelling, sentence and paragraph lengths, and writing complexity (think back to Vasicek's comments about the ML and AI problems that are related to modelling large numbers of variables). Their algorithm has four components: manuscript assessment, publication ethics, peer review assistance, and post-acceptance assistance.

The output is a very clear, visual "report card" that ranks a document's overall readability and compliance with the various elements checked; e.g. spelling, reference structure, potential plagiarism, etc.

He noted that Ada has a secure RESTful API (Application Programming Interface) that can be integrated into any publisher manuscript management system, and that it offers a secure web-based portal that can be used by a publisher's editorial desk to upload/bulk upload and get assessment reports of the manuscript(s). They have now partnered with Frontiers, an open access publisher, who currently uses the system as part of their own workflow [29]. It has been integrated into their manuscript management workflow and intercepts manuscript pre-editorial desk workflow via an API. More than twenty thousand manuscripts were processed in the first three months.

Samulack said that other enhancements are coming down the road. These include:

Automated discoverability using concept tagging and meta tagging; a custom report to verify specific publisher checks (computer assisted - expert led; an expert verifies automated checks, performs some manual ones, and creates a custom report that allows journal editors to reach quicker decision); the provision of badges for readability, publication ethics, reproducibility, statistical, image integrity, presentation quality, research quality, and data availability; and an outsourced, expert-led technical check/peer review. He noted that some alternate use cases for the system are: (1) Instant feedback to authors on submission quality (an editing service recommender); (2) Automatic identification of required copyedit level; and (3) Automated copyedit quality assessment to improve efficiency and reduce manual error through a pre/post quality comparison.

He added that a demonstration of the system can be requested at: https://www.editage.com/Ada-by-editage.html.

Samulack's slides are available on the NFAIS website.

7. AI and the science & technology cycle

The second day of the conference began with a plenary session given by **Dr. Rosina O. Weber**, Associate Professor, Department of Information Science, College of Computing and Informatics, Drexel University, who spoke about the challenges and potential benefits of applying Artificial Intelligence in scholarship. She opened her presentation with a definition of Artificial intelligence as follows:

> "A field of study dedicated to the design and development of software agents that exhibit rational behavior when making decisions."

She added that AI agents make decisions and exhibit rational behavior as they execute complex tasks such as classification, planning, design, natural language, prediction, image recognition, etc. She added that there are three waves of AI: describe, categorize, and explain. The first wave describes represented knowledge (intelligent help desks, ontologies, content-based recommenders, etc.). But she added that as a Society we have accumulated a great deal of data and it changes quickly, so that the original methods

of converting data into information and letting humans make decisions no longer work effectively. The second wave, "categorize", is based on statistical learning and, according to Weber, is where AI and Machine Learning (ML) are today (image and speech recognition, sentiment analysis, neural networks, etc.). The data deluge can only be managed with automated decision-making agents. But again, this wave has limitations, with the major problem being that humans do not want a machine to make decisions for them [30]. The third wave, "explain", is where software agents and humans become partners, with humans serving as "managers" of the work performed by multiple decision-making agents. It is in this wave that the AI systems are capable of interacting with humans, explaining themselves, and adapting to difference contexts.

Weber then went on to describe the Science and Technology Cycle (S&T Cycle) in which legislators identify societal demands and then provide funding agencies with the resources with which to support the requisite research. Researchers then use the funds for the efforts required to create novel scientific output that ultimately will be summarized in scientific publications. Researchers teach new scientists, disseminate their research results, and transfer the copyright of their manuscripts in order to be published in peer reviewed journals. The new knowledge that is an outcome of the funded research is then expected to be converted into wealth as it provides the predicted solutions to societal demands.

Weber noted that the S&T Cycle has problems:

- Neither the U.S. Congress nor the Funding Agencies really know exactly how much money is required to obtain the outcomes needed to meet specific societal demands.
- A great deal of money supports the administration of funding agencies and their processes.
- Education itself is expensive. Teachers continually design, deliver, and revise existing and new courses, design new assignments, grade them, and provide repetitive feedback.
- Textbooks are continually being revised and updated and new ones created - often being out of date by the time they make it through the publishing infrastructure.
- The production of scientific knowledge is inefficient - much research is actually redundant.
- Scientific publishing, like other industries, are at risk to losing their niche markets to the big Tech Giants such as Microsoft, Google, Amazon, etc. (e.g. Microsoft has 176 million papers; Elsevier has 75 million papers!)
- No one knows how many or which specific scientific contributions actually fulfill the desired needs.

Weber said that the use of Artificial Intelligence can solve these problems. It can be used to automate literature reviews [31]. Doing this would have the added benefit of being able to answer the following questions: What are the open research questions? Which research contributions are redundant? And which research contributions are contradictory? AI can be used to determine the effectiveness and efficiency of funding allocations by associating scientific publications to funding projects and analyzing the outcomes. What is the impact of funded vs. non-funded research? An automated manipulation of the contents of scientific publications will ultimately enable studies to determine how much wealth is driven by federally-funded research. AI can help to populate expert-locator systems by analyzing the most recent projects of researchers. AI can help identify the true cost of research. The data will reveal patterns in expenditures combined with scientific contributions. AI can provide the basis for textbook content, curricula, etc. based upon an analysis of the most recent scientific contributions and job market predictions.

Weber said that leveraging the combination of scientific data and AI will have many benefits. It will allow humans to maximize their expertise while managing AI agents to do the work. New automated services will emerge to help all those working in the S&T Cycle. Aligning education content to the job

markets will increase the effectiveness and efficiency of education in populating requisite workforces. Indeed, applying AI to the S&T cycle will;

- Increase the understanding of the Cycle itself
- Make the Cycle truly transparent
- Identify research gaps
- Identify redundant work
- Identify conflicting results
- Provide easy access to scientific facts
- Facilitate better planning and budgeting of scientific research
- Identify which scientific disciplines generate wealth and which do not

According to Weber the future of AI is the third wave, explainable AI (XAI). Humans will supervise their partners, the AI agents, who can explain their decisions, adapt to specific contexts, learn from experience, adopt ethical principles, and comply with regulations. The challenges to the realization of the third wave are: the change in the nature of work - are humans ready to be the "boss" of AI agents; the need for humans to be able to trust the agents; and collaboration among publishers.

In closing Weber said that the S&T Cycle is a cycle of data and of users in need of services. The richest source of data is the proprietary information held in scientific publications - *all* of the information, not only the abstracts. Without this data AI cannot generate the best information that will inform the best decisions and she made a plea to the publishers in the audience to create alliances to make the third wave of AI a reality. She also suggested the following reading materials for those interested in learning more:

- Qazvinian, V., Radev, D. R., Mohammad, S. M., Dorr, B., Zajic, D., Whidby, M., Moon, T., "Generating Extractive Summaries of Scientific Paradigms," *Journal of Artificial Intelligence Research,* Vol. 46, pp. 165–201, February 20, 2013, available at: https://www.jair.org/index.php/jair/article/view/10800/25780, accessed September 13, 2019.
- Weber, R. O., Johs, A. J., Li, J., Huang, K., "Investigating Textual Case-Based XAI," the 26th International Conference on Case-based Reasoning, Stockholm, Sweden, July 9–12, 2018, available at: https://www.researchgate.net/publication/327273653_Investigating_Textual_Case-Based_XAI, accessed September 13, 2019.
- Weber, R. O., Gunawardena, S., "Designing Multifunctional Knowledge Management Systems," *Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (HICSS-41), Jan. 2008 Page(s): 368–368.
- Weber, R. O., Gunawardena, S., "Representing Scientific Knowledge. Cognition and Exploratory Learning in the Digital Age," Rio de Janeiro, Brazil. 6–8 November, 2011, available at: https://idea.library.drexel.edu/islandora/object/idea:4146, accessed September 13, 2019.

Weber's slides are available on the NFAIS website and an article based upon her presentation appears elsewhere in this issue of *Information Services and Use*.

8. Adding value to the research process

The next speaker was **Ann Gabriel**, Senior Vice President Global Strategic Networks, Elsevier who discussed how Elsevier is using Artificial Intelligence (AI) and Machine Learning (ML) to bring new value to research. She said that Elsevier is attempting to evolve from a traditional publisher into a

technology information company with AI included in its arsenal of technologies. (*Note:* her presentation complemented the earlier talk by Dr. Bamini Jayabalasingham).

She said that their goal is to provide their customers with the right insights based on Elsevier's high quality content combined with deep data science and domain expertise. By 2021 they plan to deliver AI capabilities that are scalable, re-usable, and at the best quality possible through: enabling digital data enrichments and structuring at the earliest point and in the most automated ways; by moving from computer-assisted to human-assisted enrichment; by incorporating usage data for the best recommendations and personalizations; by identifying the best features and algorithms to be able to do state-of-the-art predictive analytics; and by offering AI as a service. They are using new capabilities (ML, AI and Natural Language Processing (NLP) to increase the utility of their content via text mining and data analytics and have built a data science team of approximately one hundred full time equivalent employees. Elsevier has about seven thousand employees around the globe, and one thousand of them are technologists. Why? Users want knowledge that is tailored to their exact needs of the moment in their workflow and their wants and expectations are technology-driven having been shaped by the emergence of AI, Knowledge Graphs, and new user experiences.

She noted that when users come across an unknown term in an article, they stop reading, open up Wikipedia, and look up the unknown term to get definitions and background information about the concept. What Elsevier has done via a combination of data, AI, and product development, is to make the top pages of ScienceDirect, Elsevier's website that provides subscription-based access to a large database of scientific and medical research [32], far more user-friendly so users get exactly what they need when they need it - summarized article content, additional relevant, enriched content with links, and good definitions. She admitted that they faced challenges in doing this; e.g., the need to automatically identify good definitions from text; the fact that they are dealing with a large amount of data; that most sentences are not definitions and that sentences that look like definitions often are not; being able to rank definitional sentences; and the fact that many concepts are ambiguous. Elsevier relies on automation, strong predictive ML models; human feedback; and disambiguation tools. The automated processes make Topic Page creation and content enrichment scalable; about one hundred thousand Topic Pages are generated from ScienceDirect content and more than a million articles are being enriched with links to Topic Pages. They have plans to further expand the coverage to more domains. As of the conference there had been more than three million page views and it is freely-accessible to ScienceDirect customers. An article describing the technology and examples of output can be accessed at: https://www.elsevier.com/solutions/sciencedirect/topics/technology-content-a-better-research-process.

In addition, Elsevier is using technology to identify scientific topics that are gaining momentum and therefore are likely to be funded. Their ground-breaking, new technology takes into consideration 95% of the articles available in Elsevier's abstract, indexing and citation database, Scopus [33] and clusters them into nearly ninety-six thousand global, unique research topics based on citation patterns. Aimed at portfolio analysis, Elsevier has identified these "Topics of Prominence" in science using direct citation analysis on the citation linkages in the full Scopus database. As a result, they have created an accurate model that is suitable for portfolio analysis (the methodology can be easily reproduced, but requires a full database). Gabriel added that they have created a topic-level indicator - Prominence - that is strongly correlated with future funding and noted that funding-per-author increases with increasing topic prominence. This tool [34] (Topics and the Prominence Indicator) enables stakeholders in the science system to have the knowledge necessary to make portfolio decisions.

Gabriel closed by saying that Elsevier has additional AI initiatives planned, and recommended that the audience take the time to read the Elsevier report on AI that was noted earlier by Dr. Bamini Jayabalasingham.

Gabriel's slides are not on the NFAIS website, but a paper based upon her presentation appears elsewhere in this issue of *Information Services and Use*.

9. Using AI to connect content, concepts and people

The next speaker in this session was **Bert Carelli**, Director of Partnerships, TrendMD (see: https://www.trendmd.com), a company that uses Artificial Intelligence (AI) and Machine Learning (ML) to help publishers and authors expand their readership to very targeted audiences.

Carelli opened his presentation with a brief overview of TrendMD. The company was founded in 2014 by professionals from academic research, scholarly publishing, and digital technology. It was nurtured by Y Combinator - the startup accelerator that incubated Reddit, Dropbox, and Airbnb (see: https://www.ycombinator.com/). As of the conference the company had sixteen employees, with management in Toronto and California, and was being used by more than three hundred scholarly publishers on nearly five thousand websites.

He noted that more than 2.5 million scholarly articles are published each year - more than 8,000+ each day. Fifty percent of the articles are never read [36] and a much higher percentage are never cited [37]. He said that there are two sides to the problem: (1) researchers want to find papers that are most useful to their work and (2) publishers want to ensure that their journal articles are found and read. He said that a recent report [38] indicates that browsing is core to discovery. Whether it's the Table of Contents of a journal or a search results page, browsing is something that we all do - information seekers are open to serendipity. That same report indicates that users find that the most useful feature on a publisher's website is an indication of content that is related to the articles that they retrieve from a search (the "recommender" feature). TrendMD does just that: it allows users of a publisher website to discover new content, based on what they are reading; what other users like them have read; and on what they have read in the past. He added, however, that simply displaying "Related Content" is not particularly a new or unique idea; e.g. PubMed related content and other "More Like This" links have been around for many years.

TrendMD generates recommendations via collaborative filtering - similar to how Amazon product recommendations are generated. The *Journal of Medical Internet Research* (*JMIR*) performed a six-week A/B test comparing recommendations generated by the TrendMD service, incorporating collaborative filtering, with recommendations generated using the basic PubMed similar article algorithm, as described on the NCBI website [39]. The test showed that the quality of recommended articles generated by TrendMD outperforms PubMed related citations by 272% [40]. But it is not just a recommender system. What is truly unique (and what impressed me most) is that their system has created a cross-publisher network that helps to increase article readership significantly [41], and a system that is supported by the TrendMD credit system. Publishers earn credits when they recommend (sponsor) an article on another publisher's site and "spend" credits when the reverse happens.

Carelli noted that recommendations are broken into two sections when displayed on the screen: internal links (non-sponsored articles on the publisher's own site) on the left of the screen (or on top in a vertical widget) and sponsored links (recommended articles on another publisher's site) on the right of the screen. Publishers earn one-half a traffic credit for each reader that they send to another website within the TrendMD network and spending one traffic credit gets them one new reader to their website from other

websites within the TrendMD network (*Note*: is this perhaps the beginning of the potential collaboration across publishers for which Rosina Weber gave a plea during her presentation??)

Why do so many publishers use TrendMD? Carelli said that it boils down to one thing – the company helps them maximize their growth. The statistics that he quoted were impressive: five million monthly clicks by one hundred twenty-five million unique users around the world; eighty-five percent of U.S. academics are reached as well as ninety percent of U.S. doctors (*note:* usage is forty percent via mobile devices and sixty percent via desktops). But users are not limited to the USA. While almost forty-nine percent of users are U.S.-based, he noted that access is from twenty-three countries, including Malaysia, China, Indonesia, etc. He noted that the **TrendMD** traffic increase is correlated with a seventy-seven percent increase in Mendeley saves [42], and that TrendMD readers view more articles per visit than all other traffic sources. He also noted that TrendMD readers have the lowest bounce rate of all traffic sources.

Carelli added that TrendMD has increased the number of authors submitting to journals published by participants in the TrendMD network. JMIR Publications used TrendMD to drive author submissions and the TrendMD campaign resulted in twenty-six author submissions at a conversion rate of 0.26%. Their return on investment with TrendMD was nearly ten times and Trend MD outperformed Google Ads by more than two times and Facebook by more than five times.

In closing, Carelli added that while TrendMD identifies those readers most likely to be interested in a publisher's content, finer targeting is available under three Enterprise Plans: (1) Country or region (Global); (2) Institutional (Global) - target hospital networks, universities, colleges, organizations; and (3) User-targeting (Global) - target specific types of researchers and/or Health Care Professionals in the network.

Carelli's slides are not available on the NFAIS website, but an article based upon his presentation appears elsewhere in the issue of *Information Services and Use.*

10. AI and academic libraries

The next presentation was done jointly by **Amanda Wheatley**, Liaison Librarian for Management, Business, and Entrepreneurship, and **Heather Hervieux**, Liaison Librarian for Political Science, Religious Studies and Philosophy, both from McGill University (see: https://www.mcgill.ca/). The primary focus of their talk was to discuss the results of a survey that they initiated in order to determine what role the librarian will play in an Artificial Intelligence (AI)-dominated future, as well as to reinforce the importance of autonomous research skills.

They opened with a discussion of the role of reference librarians saying that information seekers are looking to find, identify, select, obtain, and explore resources. Libraries clearly enable those efforts - specifically the reference librarians who recommend, interpret, evaluate, and/or use information resources to help those information seekers. Wheatley and Hervieux put forth six literary concepts: (1) information has value; (2) information creation as a process; (3) authority as constructed and contextual; (4) research as inquiry; (5) searching as strategic exploration; and 6) scholarship as conversation [43]. Their discussion focused on "research as inquiry" and "searching as strategic exploration" within the context of the role of AI and the role of librarians.

With regards to "research as inquiry", the practices of researchers are to: formulate questions based on information gaps; determine appropriate research; use various research methods; organize information in meaningful ways; deal with complex research by breaking complex questions into simple ones.

Researchers consider research as an open-ended exploration; seek multiple perspectives during information gathering; value intellectual curiosity in the development of their questions; and they demonstrate intellectual humility.

Similarly, with regards to "searching as strategic exploration", the practices of researchers are to: identify interested parties who might produce information; utilize divergent and convergent thinking; match information needs and search strategies; understand how information systems are organized; and use different types of searching language. Researchers exhibit mental flexibility and creativity; understand that first attempts at searching do not always produce the desired results; realize that information sources vary greatly; seek guidance from experts (such as librarians); and recognize the value in browsing and other serendipitous methods.

Wheatley and Hervieux then asked the following questions: Is AI prepared to allow researchers to continue their information literacy process? And is AI capable of being information-literate?

AI is used to a certain extent in libraries; e.g., agent technology is used to streamline digital searching and suggest articles; "conversational agents" or chatbots using natural language processing have been implemented; AI supports digital libraries and information retrieval techniques; and libraries' use of radio frequency identification (RFID) tags in circulation. AI is also used to a certain extent in higher education; e.g. digital tutors and online immersive-learning environments; programs and majors dedicated to the study across disciplines; and student researchers utilizing AI hubs.

Wheatley and Hervieux noted that smart speaker ownership (e.g. Alexa, Siri, etc.) in the United States increased dramatically between 2017 and 2018, growing from 66.7 million to 118.5 million units, an increase of almost 78 percent [44]. Since humans are basically creatures of habit, they expressed concern that as younger people become more and more used to simply asking a question to these smart devices rather than doing their own research, not only will perhaps the quality of research inquiries and strategic searching diminish, but also the role of librarians may need to shift to ensure that the research quality is maintained. It is inevitable that as these devices become a common part of people's everyday lives, their use will extend from personal space to the professional and academic environments. Asking Google for the news could soon become asking for the latest research on a given subject. The potential for Virtual Assistants to become pseudo-research assistants is a reality of which all information professionals should be aware. Their goal became to see if this type of AI has the potential to accurately provide research support at the level of an educated librarian with a Master's degree.

By determining whether or not these devices are capable of providing high quality answers to reference questions, they hope to begin to understand how users might utilize them within the research cycle. Their research plan consists of six steps: (1) An environmental scan; (2) the identification of librarians' perceptions of AI; (3) device testing (phase 1); (4) identification of student perceptions; (5) device testing (phase 2); and (6) an evaluation of the AI experience. Only the first step had been completed as of the conference.

The methodology used for the first step was to evaluate the university and university library websites of twenty-five research-intensive institutions in the USA and Canada. They searched for keywords such as Artificial Intelligence, Machine Learning, Deep Learning, AI hub, etc. And they looked at library websites to see if they could find mention of AI in strategic plans/mission/vision; in topic/research/subject guides; in programming; and in partnerships. Similarly, they looked at the University websites for mention of AI hubs, courses, and mention of major AI researchers.

Of the twenty-five universities reviewed not one mentioned Artificial Intelligence in their strategic plans. However, all did have some sort of AI presence (e.g., AI hubs or course offerings). Only one of the academic libraries has a subject guide on AI (Calgary University). Three libraries offer programming and

activities related to AI. Sixty-eight percent of the universities have significant researchers in the field of AI. And although some libraries have digital scholarship hubs, these hubs are not involved with AI. Of the twenty-five academic libraries sampled, only two are collaborating with AI hubs.

Wheatley and Hervieux went on to discuss three case studies that were undertaken by three of the top universities actively looking at AI: the Massachusetts Institute of Technology, Stanford University, and Waterloo University.

MIT started a project in 2006 entitle SIMILE (see: http://simile.mit.edu/) with the goal to create "next-generation search technology using Semantic Web standards". The site is no longer updated or supported.

Stanford has implemented a library initiative to "'identify and enact applications of Artificial Intelligence [...] that will help us (the university) make our rich collections of maps, photographs, manuscripts, data sets and other assets more easily discoverable, accessible, and analyzable for scholars". The site also sponsors talks and discussions about AI (see: https://library.stanford.edu/projects/artificial-intelligence). They also have an AI Studio (see: https://library.stanford.edu/projects/artificial-intelligence/sul-ai-studio) that applies AI in projects to make their collections more discoverable and analyzable. The studio is staffed by volunteer librarians and the library provides access to Yewno, discovery tool that provides a graphical display of the interrelationships between concepts (*Note:* Yewno was discussed by its founders at both the 2017 [45] and 2018 [46] NFAIS Annual conferences. It is a very interesting concept that uses AI to augment the information discovery process).

Waterloo University has established the Waterloo Artificial Intelligence Institute: Centre for Pattern Analysis and Machine Learning. While not affiliated with any library projects, current initiatives have great potential for impact on the research process (see: https://uwaterloo.ca/centre-pattern-analysis-machine-intelligence/research-areas/research-projects).

Wheatley and Hervieux said that their conclusions at this stage of their research are the following: (1) AI is already operating behind the scenes in libraries and the larger research process; (2) Universities and libraries are not doing enough to work together in this field; and (3) Research habits *are* indicative of personal habits and that the personal use of virtual assistants is growing exponentially.

They also raised the following questions:

- As AI builds within the research process do these two concepts become one and the same?
- How do we adjust our standards and frameworks for teaching to account for this change?
- Are citation counts and impact factors still the preferred metric or does AI impact the way we perceive search results and scholarship?
- What does this mean for reference librarians and job security going forward?

They hope to gain insights to these questions when they ultimately conclude their study. Their next step is to focus on the second phase of their plan - to identify the perceptions of librarians with regards to AI.

Wheatley and Hervieux's slides are available on the NFAIS website and an article based upon their presentation appears elsewhere in this issue of Information Services and Use.

11. Increasing the awareness of AI in higher education

The next speaker was **James J. Vileta**, Business Librarian, University of Minnesota Duluth, who spoke passionately and articulately about the need to include the topic of Artificial Intelligence in diverse curricula. He was in total agreement with the prior speakers, Wheatley and Hervieux, who

expressed concern that of the twenty-five universities that they surveyed, only one of the libraries (Calgary University) have a subject guide on AI. He then went on to mention three books:

- *Rise of the Robots: Technology and the Threat of a Jobless Future* by Martin Ford
- *The End of Jobs: Money, Meaning and Freedom Without the 9-5*, by Taylor Pearson
- *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies* by Erik Brynjolfsson and Andrew McAfee

He said that reading them raised a lot of questions for him regarding the impact of AI, such as: Is it just math, computer science, and engineering or does it touch all disciplines? How is it applied to professions, commerce, and industry? How is it applied to careers and personal life planning? How will it impact our social, political and economic world? Should young people, in particular, be more aware of its potential impact on their lives? He noted that the business school at his university does not address these issues at all. Their focus is on globalization since so much of U.S. business is conducted out of the United States. They give no attention at all to Artificial Intelligence. He believes that universities and libraries need to do more - that they should "AI & Robots" all feasible parts of the curriculum. Institutions of higher education have a duty - they need to prepare students for *real* life - a life that is even now being impacted by AI.

Vileta noted that Northeastern University's President, Joseph E. Aoun, has written a book on how colleges need to reform their entire approach to education, *Robot-Proof: Higher Education in the Age of Artificial Intelligence* published by MIT Press. The book was highlighted in *The Washington Post* within the context of today's job market and the article stated that the nearly nine in ten jobs that have disappeared since the year 2000 were lost to automation [47]. He said that Aoun lays out the framework for a new discipline, "humanics", that builds on our innate strengths and prepares students to compete in a labor market in which smart machines work alongside human professionals (does this sound like Weber's Third Wave of AI?).

Vileta agrees with Aoun, but said that we need to do more. We need to consider the following questions: What kind of world do we want to live in? How do we live without incomes from work? How do we achieve a balanced society? How do we find meaning without work? What do we do with our abundant leisure? What new laws and regulations do we want? He said that he himself wants to do more and, as a librarian, he wants to use his expertise to create an awareness of AI on his campus. So he first studied the landscape of the AI topic and then went out and bought some books!

He bought books on all sorts of subjects that touched on AI: universal basic income [48], AI in video games, autonomous cars, AI and robots in the military, AI and robots in healthcare, AI ethics, AI in management, AI in banking, the Internet of things, etc. What he found was that there are a number of recent books and articles related to Artificial Intelligence that touch on all aspects of work and play and that most people are unaware of the variety and quality of these books. Also, he noted that large libraries get most of these books automatically, while smaller libraries have to be mindful of budgets and be more selective in their purchases. He warned, however, that when you are developing new programs in the curriculum, you must be proactive and direct. You cannot be passive.

Taking his own advice he created a web page using Dreamweaver. The page currently has eighteen sections, over two hundred and eighty books and five hundred and fifty clickable links. The page was launched to positive reviews around the time of the conference and will ultimately be transformed into its own website. He has been advertising the page via selective dissemination of information, through Social Media Marketing, strategic viral marketing, and traditional marketing. In addition he has formed alliances with key faculty members at the university, set up special lectures and presentations, and testing in the university's College Writing Courses.

Vileta added that there are many good AI-related books on diverse topics that can be brought into library collections, along with countless scholarly articles on all levels. There are many opportunities for libraries and librarians to bring awareness of AI into higher education.

Libraries can provide focused finding aids, and also initiate innovative projects, such as the AI Lab at the University of Rhode Island Library (subject of the following presentation). He said that librarians should help everyone prepare for the future and that this applies to all types of libraries - academic and public. Librarians need to take the initiative and do more - kudos to Wheatley and Hervieux for their call to action as well!! He added that organizations and vendors who work with libraries should help them in this work by providing more content, better indexing, and more AI-focused online products.

In closing, Vileta suggested that in addition to Auon's book mentioned earlier, those in the audience read *The Future of the Professions: How Technology Will Transform the Work of Human Experts* [49].

Vileta's slides are not available on the NFAIS website.

12. University launches a new AI laboratory

The final speaker of the morning was **Karim Boughida**, Dean of University Libraries at the University of Rhode Island, who spoke about the University's Artificial Intelligence Laboratory. He opened by saying that a library's brand is a book, but libraries are so much more than books! (*Note:* library identification with books was the subject of a presentation by Scott Livingston of OCLC at the 2019 NFAIS Annual Conference [50]).

The lab opened in the fall of 2018 and, unlike AI labs on other campuses that are usually housed in computer or engineering departments, this is housed in the library and it is believed to be the first one placed in such an environment. The location is strategic as the organizers hope that students majoring in diverse fields will visit the lab and use it to brainstorm about important social and ethical issues today and create cutting-edge projects. Also, when surveyed recently about topics they wished to see in their curriculum, AI was among the top requests from the university's students. This was what motivated the university to create the lab and to put it where the majority of students would have access - the library.

The lab has two complementary goals: "On the one hand, it will enable students to explore projects on robotics, natural language processing, smart cities, smart homes, the Internet of Things, and big data, with tutorials at beginner through advanced levels. It will also serve as a hub for ideas - a place for faculty, students, and the community to explore the social, ethical, economic and even artistic implications of these emerging technologies [51]".

The six-hundred square foot lab has three zones. The first zone has AI workstations where a student or a team of students can learn about AI and relevant subject areas. The second zone has a hands-on project bench where, after students receive basic training on AI and data science, they can move on to advanced tools to design hands-on projects in which they can apply AI algorithms to various applications such as deep learning robots, the Internet of things for smart cities, and big data analytics. Located in the center of the lab, the third zone is an AI hub where groups can get together for collaborative thinking.

The lab was funded by a $143,065 grant from the Champlin Foundation, one of Rhode Island's oldest philanthropic organizations, and unlike AI labs focused on research, this lab is focused on providing students and faculty with the chance to learn new computing skills. Also, through a series of talks and workshops, it encourages them to deepen their understanding of AI and how it might impact their lives.

In closing, like Vileta, Wheatley, and Hervieux before him, Boughida stressed the importance of embedding AI in curricula and he referred to the fact that Finland is being aggressive in training its citizens

in AI. In May of 2018, Finland launched *Elements of AI* (see: https://course.elementsofai.com), a first-of-its-kind online course that forms part of an ambitious plan to turn Finland into an AI powerhouse [52]. It's free and anyone can sign up (I just did!). It is definitely work a look.

Boghida did not use slides for his presentation.

13. A call to publishers to act now

The lunch speaker was **Michael Puscar**, Founder and CEO of Oiga Technologies. He has partnered with Mike Hooey, CEO and Founder of Source Meridian, to provide their customers with early access to state-of-the-art technology (AI, Blockchain, Big Data analytics etc.), helping them to create and maintain their competitive advantage in the market. Puscar focuses on the publishing industry and Hooey focuses on the healthcare industry. Puscar spoke last year at the NFAIS conference on the use of Blockchain technology in scientific publishing [53].

Puscar opened by saying that Artificial Intelligence is no longer the future - it is the *now*! He created a sense of urgency by reminding the audience that the information industry is very dynamic. For publishers, content is the crown jewel, and they have spent the last decade refining how it is stored, processed, and transformed. He added that use of AI can provide competitive advantages and that publishers can either be a part of shaping the future or simply adjust and adapt after their competitors have already done so.

He noted that his organization already has had success in healthcare where much of the content is now in digital format. By applying AI to that content, their customers are working to find unexpected drug interactions; detect and eliminate insurance fraud; identify the best doctors, hospitals, and treatment plans; predict the outcome of clinical trials and reduce costs; conduct derivative clinical trials; more efficiently address R&D budgets in healthcare; and to create objective, quantitative predictive performance measures.

He added that most of us are unaware of what we don't know and AI can bring that knowledge to the surface. Technology is ahead of industry applications, and the software already exists that publishers can implement now. It is not as expensive as one would think, but an understanding of how AI works is necessary to set realistic expectations (perhaps the online course from Finland will help?).

He pointed out that when a publisher's content has defined data structures, they need to know the questions for which AI will be used to find the answers (echoing Vasicek and Ciufetti). He added that content does not necessarily need to be unified or normalized, and that AI algorithms often find trends where least expected, noting that human reinforcement is often, but not always, necessary.

Puscar used the example of the process of learning. Often we do not know exactly what we will learn, but we do so through trial and error, or reinforcement. When we achieve a goal, it is positively reinforced. With neural networks, machines can learn in the same way, building unique neural pathways. Machines learn to achieve a stated goal through trial and error. Each success and failure builds neural pathways that improve subsequent tries. Actions in an environment maximize a cumulative reward and optimality criterion governs success or failure.

In closing he said that there is no need to wait to get started. Open Source software is available, but the bad news is that computing time and expertise will require a substantial investment. He added that cloud-based services such as those available from Google, Microsoft, and Amazon all have Machine Learning platform services. For example take a look at Google's offerings - AutoML Natural Language and Natural language API - at https://cloud.google.com/natural-language/. He added that whether you choose Google, Amazon, IBM or Microsoft Azure - they all have what is needed: horizontal scalabilities, API-based

access, Machine Learning tools, out-of-the-box algorithms, examples, and supportive communities. (*Note: the following speaker provides a perfect case study!*)

Puscar's slides are not available on the NFAIS website, but a brief article based upon his presentation appears elsewhere in this issue of *Information Services and Use*.

14. Case study: Using machine learning to build a recommendation engine

The next speaker was **Peter Ciufetti,** Director of Product Development at ProQuest, who discussed how ProQuest used Amazon's SageMaker, a fully-managed Machine Learning system [54], to build a video recommendation system. He opened by saying that he is not giving a commercial for Amazon Web Services, although it may seem so at points, and that he will try not to be overly technical, though it's a bit hard to avoid this given the topic, especially since his goal was to make Machine Learning concepts accessible and inspire those in the audience to try it on their own. He noted that he did receive a degree in computer science from Harvard a long time ago, but that he is not a data scientist, which is an important take-away for this presentation.

He then asked the question - why use Machine Learning? He noted that 75% of Netflix views come from recommendations (identified via ML techniques) presented to users by the system - the user does not have to do much work. Indeed, the big players are setting the user's expectations for how a repository of content should work: Netflix and YouTube in terms of recommending videos; Google in terms of understanding a user's search; and iTunes and Spotify in what to listen to. Ciufetti said that it became clear to ProQuest that if they did not implement recommendations into their video user experience their search page might soon become a 'bounce' page. He used JW Player as an example. He said that when their recommendation service was activated, JW Player customers who hosted videos with them saw a 35% average increase in the number of video views and a 25% increase in the time spent watching. In addition to "recommendations", there are other reasons for publishers to used ML: topic analysis; entity extraction; sentiment analysis; image recognition; speech-to-text; and improving optical character recognition (OCR) – all of the reasons mentioned by mentioned earlier by Vasicek.

Ciufetti then went on to talk about his own ML project. He said that a ML model is going to make a prediction on some that question you ask it. This process is called inference. He added that a ML model is designed to answer a single question; e.g., What objects do you see in this picture? (regression); Is this article a book review or not a book review? (classification). So it really helps (*no - it is essential*) to fine-tune the question that you want the model to answer.

He went on to say that a "supervised" ML model needs to be trained with data, for example, it needs positive examples (this is a book review), negative examples (this is not a book review), and inferences (is this article a book review?) On the other hand an "unsupervised" ML model does not need to be trained, but its application is limited to clustering; e.g. inference: which articles are like each other? He said that for a supervised ML model you need to know (or learn) if you actually have the data that will support the model. If you do have such data, the typical workflow to build the model can take a couple of months - ten days for data analysis, twenty days of ML modeling, twenty days of production work, and another ten days of production validation.

Ciufetti then turned his discussion to Amazon's SageMaker. He said that it is a ML learning platform launched by Amazon in November of 2017. With it Amazon is trying to put decades of ML capability into the hands of non-data scientist programmers. There are three ways to use SageMaker: (1) develop an ML model using a provided algorithm; (2) develop an ML model using your own algorithm; and (3)

develop a model elsewhere and use SageMaker to run it in production. He then went on to briefly discuss its technical underpinnings and moved to ProQuest's s specific use case.

When he became aware of SageMaker in December 2017 he decided to use it to build ProQuest's recommendation system between Christmas and New Year's. His objective was to be able to accurately recommend a video to a user based upon the user's historical interest. Although ProQuest already had a system, it was not personal - everyone was given the same recommendation. Challenges to moving forward were that video usage by individuals is not that high and even when aggregated to the customer level, there is a long tail. Also, the size of ProQuest's video corpus is about three times that of Netflix and ProQuest had to be able to select recommendations from among thousands of unwatched videos.

The steps that he used to create the system were to:

• Prepare the model interaction data: users, customers, videos
• Train the Factorization Machine Model [55] as a Regressor
• Deploy it
• Test and measure its prediction capability
• Repeat the above until the desired level of utility was reached
• Run predictions to make recommendations
• Measure user interactions with it closely (their impressions, click-through rates, total time watching)
• Feed new interaction details back into the model

Once the model was developed a query was created for each customer and user with regards to every title to which they had full access. These queries were converted to a JSON string and run against the deployed model to generate raw predictions - the predictions are the model's estimation of how likely the customer is to view the title. The raw predictions were stored in a database table. Then an API command was used to extract the top one hundred predictions for each customer across a number of category groups (i.e. format, field of study, unit access code) which were then copied to a live Amazon Relational Database Service (RDS) recommendations mysql database. This was done to restrict the total number of recommendations to a manageable number of more relevant titles. The recommendations database was versioned so that new recommendations could be added after a new prediction was run and old ones removed without interrupting live recommendation usage.

Ciufetti closed by summarizing what ProQuest learned along the way: They had two desired use cases, but decided that they did not have the data to support the second one; it is difficult not to introduce bias; getting the question right is important; algorithms have "tuning parameters" neither well-defined nor understood by the non-data scientist; they ended up getting some help from real data scientists and it made a lot of difference; and if the data does not support the decisions, you won't like the answers.

He said that after "going under the hood" with regards to ML, he is not so sure that he would get into a self-driving car.

Ciufetti's slides are not on the NFAIS web site, but a paper summarizing his presentation appears elsewhere in this issue of *Information Services and Use.*

15. Wikipedia and AI

The closing keynote speaker was **Dr. Aaron Halfaker,** American Computer Scientist and a Principal Research Scientist, Wikimedia Foundation, who provided an overview of the most pressing research questions of strategic importance for Wikipedia and the larger Wikimedia movement.

He noted that Wikipedia was launched January 15, 2001 and is the world's largest encyclopedia, with about five million articles in English (*Note*: as of July 2019, there were three hundred and four languages into which Wikipedia has been translated, with two hundred and ninety-four active and ten closed [56].).

He said that a Wiki is a website that allows collaborative editing of its content and structure by users. It is a "flipped" publication model in that content is published first and reviewed (maybe) later. And then he attempted to provide a description of Wikipedia's "structure".

Wikipedia represents an online community with about one hundred thousand active volunteers and is a socio-technical system that merges technology and social production. He said that fishing villages are often shaped by "Dunbar's Number [57]", a suggested cognitive limit to the number of people with whom one can maintain stable social relationships - relationships in which an individual knows who each person is and how each person relates to every other person. When a village grows to one hundred and fifty inhabitants, it then splits. When each section grows again to one hundred and fifty, they each split again. The composite Wikis in Wikipedia are similar.

He also compared Wikipedia to a paramecium - a system with specialized subsystems and said that it is governed by Linus' Law [58] which is "given enough eyeballs, all bugs are shallow"; or more formally: "Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone". The corollary for Wikipedia is given that if enough people see an incomplete article, all potential contributions to that article will be easy for someone to undertake.

Halfaker said that work allocation for a specific Wiki is controlled by those who are creating the entry - they are volunteers who identify and assign tasks within their community. They create the article/essay and request feedback. However, behavior is governed by overall Wikipedia guidelines and policies to enforce consistent behavior across all Wikis and to ensure quality. These policies and guidelines are "developed by the overall community to describe best practices, clarify principles, resolve conflicts, and otherwise further their goal of creating a free, reliable encyclopedia. The policies are standards that all users should normally follow, and guidelines are generally meant to be best practices for following those standards in specific contexts. Policies and guidelines should always be applied using reason and common sense [59]".

Wikipedia's quality control has two components. The first is fully-automated. It is based upon Machine Learning techniques and can identify "vandalism" in about five seconds [60]. The second component is a semi-automated computation, but it is still pretty fast - about thirty seconds, and it helps to minimize human effort. He noted that humans catch most vandalism at a glance [61].

Halfaker said that there are about six thousand newcomers to the global Wikipedia community per day. The automated quality control system can separate vandals from "good faith" contributors and helps to manage the community. Overall, the community itself socializes and trains newcomers and also mediates disputes.

Halfaker then turned his discussion to the future - Where is Wikipedia going? Where does it want to go? And what should they being doing to get there? As of January 10, 2019, planet earth had 7.6 billion inhabitants and Wikipedia had one billion monthly visitors with its content managed by 113,304 editors. The staff of the Wikimedia Foundation, the nonprofit that hosts Wikipedia and their other free knowledge projects, totaled about 225 (see: https://wikimediafoundation.org/about/). He noted that their core research team totals six, while Google has nineteen hundred and fifty-three; Facebook has five hundred and eighty; and Yahoo has eighty-four. But he added that Wikimedia is a volunteer organization and it is interesting because:

- It is huge and read by one-tenth of the planet.
- It is a trusted source of information.

- It is weird. It has a flipped publication model and a decentralized governance. It should not have worked at all.
- It is Open. Datasets, proposals, and initiatives are all made freely-available.

He then compared Wikipedia to Google+ and asked why no one looked at the latter [62]. His answer was that while Wikipedia is open with its datasets, APIs, and benefits humanity, Google+ was a walled garden with publishing embargoes and primarily benefited Google.

However, Wikipedia has faced its fair share of problems. It needs to maintain a pool of volunteer contributors to remain relevant. It was created and fueled initially through a tremendous number of contributions by millions of contributors. However, by 2009 research had shown that the number of active contributors in Wikipedia had been declining steadily and suggested that a sharp decline in the retention of newcomers was the cause. This lack of retention was due to several changes that the Wikipedia community had made to manage quality and consistency in the face of the massive growth in participation. Specifically, the restrictiveness of the encyclopedia's primary quality control mechanism and the algorithmic tools used to reject contributions appeared to be the key causes of decreased newcomer retention. So for the future they are looking at three key areas: (1) gaps in representation and content; (2) growth in newcomer retention; and (3) third-party re-use.

With regard to representation and content, current contributors are primarily male, mostly from North America and Western Europe, the majority of which cite English Language sources [63]. They are using Artificial Intelligence to fill-in the gaps, especially in under-developed content areas [64]. With regard to the growth in newcomer retention, they are using AI to improve quality control and identify the most serious acts of vandalism. The system can direct humans to review the most damaging edits and determine the caliber of mistakes. Thus "rookie" mistakes can be treated more appropriately as innocent [65].

With regard to third-party reuse he demonstrated how a search on Google will display Wikipedia information in Google's Knowledge Panel. Just do a Google search on the question "How long do goats live" and you can see the response [66]. They have found that Google users are fifty percent less likely to click a Wikipedia link and five times as likely to assume that the information in the Knowledge Panel "comes from" Google [67]. He added that the same information feeds Alexa and that the problem with Wikipedia's "openness" is that they have difficulty tracking usage and that their "brand" is diluted.

In closing Halfaker provided some links that may be of interest:

1. Information on their Open Access policy and related links: https://upload.wikimedia.org/wikipedia/commons/d/d0/Wikimedia_Public_Research_Resources.pdf

2. Their beta software to run SQL queries against Wikipedia & other databases from your browser: https://quarry.wmflabs.org

3. The Wikipedia Community Survey form: https://paws.wmflabs.org

4. The Wikipedia Research GitHub site: https://github.com/wikimedia-research/iwsc-2017-workshop

5. Information on the new AI system used to detect vandalism (ORES): https://upload.wikimedia.org/wikipedia/commons/4/48/ORES_-_Facilitating_re-mediation_of_Wikipedia%27s_socio-technical_problems.pdf

Halfaker's slides are available on the NFAIS website.

16. Conclusion

The goal of this conference was to explore the application and implication of Artificial Intelligence (AI) across all sectors of scholarship and, as you know by now, the topics covered were quite diverse. As

I noted at the beginning of this overview, the conference was not for the AI computer-savvy, rather it was a conference for non-techies who wanted to learn how Machine Learning (ML) and AI were being used within the Information Community - by publishers, librarians, and vendors - and from my perspective the goal was met. My first take-away from the conference is the following:

> "Academic publishing is right now poised to substantially leverage technologies led by artificial intelligence [68]".

The speakers made it very clear that today Artificial Intelligence (AI) and Machine Learning (ML) are significantly advancing the information industry. Why? Because effective use of these tools allow publishers to serve as alchemists - transforming their content from "raw data" into invaluable "knowledge". This "transformation" improves the user search experience as users do not have to invest a lot of time and effort in distilling and making sense of search results. The outcome of effective use of AI and ML are knowledge-based information services that are becoming the gold standard in scholarly communication.

The corollary to this message is the following:

> "The playing field is poised to become a lot more competitive, and businesses that don't deploy AI and data to help them innovate in everything they do will be at a disadvantage [69]".

It was made quite clear that if publishers are not already utilizing AI and ML they need to do so sooner rather than later in order to remain relevant to their users. Certainly, speakers such as Michael Puscar and Sam Herbert, were adamant in their belief that publishers need to take action *now*. And the publishers who spoke at the conference (Elsevier, CAS, Clarivate Analytics, ProQuest) echoed their message - these publishers saw the writing on the wall and have adopted AI and ML as part of their technology arsenal.

I should mention that this very same message was voiced earlier this year at the 2019 NFAIS Annual Conference. At that meeting Vincent Cassidy of the Institution of Engineering and Technology (IET) talked about how they have added value to the Inspec database via the use of ML to develop concept relationships and Knowledge Graphs, which has resulted in user and revenue growth [70]. Travis Hicks from the American Society of Clinical Oncology (ASCO) talked about increasing the discovery of their digital content using ML to build a semantic search model [71]. And Bob Kasenchak (at the time with Access Innovations, Inc.) also talked about the importance of building concept relationships and semantic search capabilities [72]. It is quite clear that the use of AI and ML is becoming the norm for today's publishers and Elsevier is a great example of a well-entrenched, traditional publisher who has embraced change and is reinventing themselves by moving from a document-centric approach (where everything revolves around books and journals) to a knowledge-centric organization using AI and ML.

If your organization is change-adverse, I suggest that you have key staff read Puscar's and Herbert's "rah rah" articles followed by that of Peter Ciufetti who was actually able to take their advice and use the current AI/ML tools offered by the Big Tech companies to build a ML Model for the development of ProQuest's video recommendation system - even though he is not a Data Scientist. (*Note*: if your organization does decide to move forward, make sure that the technical staff read Daniel Vasicek's article on the challenges to building ML models!)

But in retrospect, I believe that there was an even more important take-away message that came out of the conference - and one that I did not expect to emerge from a meeting that was primarily publisher-focused. The message was broader and far more philosophical, and it was a call for educators and librarians to better prepare students - indeed, Society in general - for a future where AI is increasingly imbedded in all that we do. The questions raised by James Vileta in his presentation are critical: How is AI applied to professions, commerce, and industry? How is it applied to careers and personal life planning? How will

it impact our social, political and economic world? Should young people, in particular, be more aware of its potential impact on their lives? These are questions that we all need to think about as we move forward in an increasingly automated world and this is the message that most resonated with me.

As someone who has worked in publishing for many years I am aware that technology has always been ahead of the curve in our industry - but policy (e.g. copyright) has always lagged behind. This has been true for the music business as well, but should not be the case for education. In 2018, the Pew Foundation asked almost one thousand technology pioneers, innovators, developers, business and policy leaders, researchers, and activists about the impact that AI will have on society by the year 2030 [73]. Amy Webb, founder of the Future Today Institute and professor of strategic foresight at New York University, commented, "As AI matures, we will need a responsive workforce, capable of adapting to new processes, systems and tools every few years. The need for these fields will arise faster than our labor departments, schools and universities are acknowledging". (Take a look at the report - it is interesting).

Since the goal of institutions of higher education is to allow each of us to increase our knowledge and maximize our intelligence - AI definitely needs to be on their radar. Consider the following quote:

> "I think what makes AI different from other technologies is that it's going to bring humans and machines closer together. AI is sometimes incorrectly framed as machines replacing humans. It's not about machines replacing humans, but machines augmenting humans. Humans and machines have different relative strengths and weaknesses, and it's about the combination of these two that will allow human intents and business process to scale 10x, 100x, and beyond that in the coming years [74]".

The truth us that advances in AI are *not* totally computer-dependent - they are, in fact, an extension of human intelligence. As of today all Machine Learning techniques still require human-based, hand-programmed components. We are the future of AI, so in my opinion, the second message of the conference may actually be the most important.

This unexpected message is what has made the NFAIS conferences so interesting and valuable over the years. They have provided a neutral venue in which all information industry stakeholders could get together for open and productive discussions. Thus it was with mixed feelings that I read that on June 30, 2019 that NFAIS merged with the National Information Standards Organization (NISO) [75].

That said, *MARK YOUR CALENDAR!!!* The long-standing NFAIS Annual Conference will continue, albeit with a new format and a new name - the NISO Plus Conference. Scheduled to take place February 23–25, 2020 in Baltimore, MD, the goal of the meeting is "to bring all of the constituencies in the information economy together to generate discussion, seek common ground, identify areas of opportunity and provide a space and resources to work towards solving common problems [76]". Indeed, plan to attend the meeting. It appears that the neutral venue in which publishers, librarians, policy-makers, and vendors can discuss new and potentially controversial issues productively and with respect for differing opinions will continue!!!

I will be there - will you?

17. Recommended Additional Reading

Hlava, M. M. K., "The Data you have…Tomorrow's Information Business," *Information Services and Use,* Vol. 36, No. 1–2, pp. 119–125, 2016, available at: https://content.iospress.com/download/information-services-and-use/isu799?id=information-services-and-use%2Fisu799, accessed October 12, 2016.

Bryson, J. J., "The Past Decade and Future of AI's Impact of Society," *OpenMind*, 2019, available at: https://www.bbvaopenmind.com/wp-content/uploads/2019/02/BBVA-OpenMind-Joanna-J-Bryson-The-Past-Decade-and-Future-of-AI-Impact-on-Society.pdf, accessed October 21, 2019. (*Article from the book Towards a New Enlightenment? A Transcendent Decade*, see: https://www.bbvaopenmind.com/en/books/towards-a-new-enlightenment-a-transcendent-decade/, accessed October 12, 2019).

Kaplan, A., Haenlein, N., "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons*, Vol. 61, Issue 1, pp. 15–25, January-February 2019, available at: https://www.sciencedirect.com/science/article/pii/S0007681318301393?via%3Dihub, accessed October 12 2019.

Garbade, M. J., "Clearing the Confusion: AI vs. Machine Learning vs. Deep Learning Differences," *Towards Data Science*, September 14, 2018, available at: https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb, accessed October 12, 2019.

"Intelligent Economies: AI's Transformation of Industry and Society," A Report from the Economist Intelligence Unit, available at: https://eiuperspectives.economist.com/sites/default/files/EIU_Microsoft%20-%20Intelligent%20Economies_AI%27s%20transformation%20of%20industries%20and%20society.pdf, accessed October 12, 2019.

About the Author

Bonnie Lawlor served from 2002–2013 as the Executive Director of the National Federation of Advanced Information Services (NFAIS), an international membership organization comprised of the world's leading content and information technology providers. She is currently an NFAIS Honorary Fellow. Prior to NFAIS, Bonnie was Senior Vice President and General Manager of ProQuest's Library Division where she was responsible for the development and worldwide sales and marketing of their products to academic, public, and government libraries. Before ProQuest, Bonnie was Executive Vice President, Database Publishing at the Institute for Scientific Information (ISI - now Clarivate Analytics) where she was responsible for product development, production, publisher relations, editorial content, and worldwide sales and marketing of all of ISI's products and services. She is a Fellow and active member of the American Chemical Society and a member of the Bureau of the International Union of Pure and Applied Chemistry for which she chairs their Publications and Cheminformatics Data Standards Committee. She is also on the Board of the Philosopher's Information Center, the producer of the *Philosopher's Index,* and she serves as a member of the Editorial Advisory Board for *Information Services and Use*. She has served as a Board and Executive Committee Member of the former Information Industry Association (IIA), as a Board Member of the American Society for Information Science & Technology (ASIS&T), and as a Board member of LYRASIS, one of the major library consortia in the Unites States.

Ms. Lawlor earned a B.S. in Chemistry from Chestnut Hill College (Philadelphia), an M.S. in chemistry from St. Joseph's University (Philadelphia), and an MBA from the Wharton School, (University of Pennsylvania), with subsequent studies at INSEAD in Fontainebleau, France. Contact: chescot@aol.com.

About NFAIS

Founded in 1958, the National Federation of Advanced Information Services (NFAIS™) acted as a global, non-profit, volunteer-powered membership organization that served the information community; i.e., all those who create, aggregate, organize, and otherwise provide ease-of-access to and effective navigation and use of authoritative, credible information.

Member organizations represented a cross-section of content and technology providers, including database creators, publishers, libraries, host systems, information technology developers, content management providers, and other related groups. They embodied a true partnership of commercial, nonprofit, and government organizations that embraces a common mission - to build the world's knowledgebase through enabling research and managing the flow of scholarly communication. NFAIS existed to promote the success of its members and for sixty-one years it provided a forum in which to address common interests through education and advocacy.

About NISO

NISO, the National Information Standards Organization, a non-profit association accredited by the American National Standards Institute (http://www.ansi.org/ANSI), identifies, develops, maintains, and publishes technical standards to manage information in today's continually changing digital environment. NISO standards apply to both traditional and new technologies and to information across its whole lifecycle, from creation through documentation, use, repurposing, storage, metadata, and preservation.

Founded in 1939, incorporated as a not-for-profit education association in 1983, and assuming its current name the following year, NISO draws its support from the communities it serves. The leaders of over 70 organizations in the fields of publishing, libraries, IT, and media serve as its voting members. Many of the experts and practitioners serve on NISO working groups, committees, and as officers of the association. NISO is where content publishers, libraries, and software developers turn for information industry standards that allow them to work together. Through NISO, all of these communities are able to collaborate on mutually accepted standards - solutions that enhance their operations today and form a foundation for the future.

References

[1] S. Lynch, Andrew Ng: why AI is the new electricity, *Stanford Business Insights* (2017), available at: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity, accessed October 12, 2019.

[2] A.M. Turing, Computing machinery and intelligence, *Mind* **49**: (1950), 433–460, available at: https://www.csee.umbc.edu/courses/471/papers/turing.pdf, accessed October 12, 2019.

[3] Logic Theorist, *History-Computer*, available at: https://history-computer.com/ModernComputer/Software/LogicTheorist.html, accessed October 12, 2019.

[4] J. Moor, The Dartmouth College AI conference: the next 50 years, *AI Magazine* **27**(4) (2006), 87–91, available at: https://www.aaai.org/ojs/index.php/aimagazine/article/view/1911, accessed October 12, 2019.

[5] R. Anyoha, The History of Artificial Intelligence, *SITN*, Harvard University, available at: http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/, accessed October 12, 2019.

[6] R. Iriondo, Machine Learning vs. AI, Important Differences Between Them, *Data Driven Investor*, October 15, 2018, available at: https://medium.com/datadriveninvestor/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6, accessed October 12, 2019.

[7] T. Mitchell, *Machine Learning* McGraw Hill, 1997, p. 414, ISBN 0070428077, for more information see: http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html, accessed October 12, 2019.

[8] Expert Panel, 13 Industries Soon to be Revolutionized by Artificial intelligence, Forbes Technology Council, available at: https://www.forbes.com/sites/forbestechcouncil/2019/01/16/13-industries-soon-to-be-revolutionized-by-artificial-intelligence/#61cd7b973dc1, accessed October 12, 2019.

[9] S. Abu-Mostafa, M. Magnon-Ismail and H. Lin, *Learning form Data-a Short Course* AML Book, 2012, ISBN:1600490069 9781600490064. *Note* a MOOC based on this book and the actual course taught at Caltech is available at: https://work.caltech.edu/telecourse.html, accessed September 19, 2019.

[10] B. Marr, How much data do we create every day? The mind-blowing stats that everyone should read, *Forbes* (2018), available at: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#3ceb60ee60ba, access September 20, 2019.

[11] Apache Hadoop, *Wikipedia,* see: https://en.wikipedia.org/wiki/Apache_Hadoop, accessed October 15, 2019.

[12] NoSQL, *Wikipedia,* see: https://en.wikipedia.org/wiki/NoSQL, accessed October 15, 2019.

[13] Barbosky listed a Stuart G. Walesh as the author, see: https://www.goodreads.com/author/quotes/331712.Stuart_G_Walesh, accessed September 20, 2019. Elsewhere a Leo Cherne is cited as the author, see: https://ghostcodebuilds.wordpress.com/2017/05/30/the-computer-is-incredibly-fast-accurate-and-stupid/, accessed September 20, 2019.

[14] Edmond de Belamy, *Wikipedia,* available at: https://en.wikipedia.org/wiki/Edmond_de_Belamycite_note-2, accessed September 20, 2019.

[15] V. LeCorvec, J. Jeffrey and J. HuntHow, How Gauss Determined the Orbit of Ceres, presented on December 7, 2007 and available at: https://math.berkeley.edu/~mgu/MA221/Ceres_Presentation.pdf, accessed September 11, 2019.

[16] For more information, go to: https://www.tensorflow.org/, accessed September 11, 2019.

[17] *Artificial Intelligence: How Knowledge is Created, Transferred, and Used: Trends in China, Europe, and the United States*, Elsevier, 2018, full report available at: https://www.elsevier.com/__data/assets/pdf_file/0011/906779/ACAD-RL-AS-RE-ai-report-WEB.pdf, accessed September 11, 2019.

[18] *Artificial Intelligence: How Knowledge is Created, Transferred, and Used: Trends in China, Europe, and the United States*, Elsevier, 2018, Executive Summary available at: https://www.elsevier.com/__data/assets/pdf_file/0007/827872/ACAD_RL_RE_AI-Exec_Summary_WEB.pdf, accessed September 11, 2019.

[19] J. Vincent, Google's Sergey Brin warns of the threat from AI in today's "technology renaissance", *The Verge* (2018), available at: https://www.theverge.com/2018/4/28/17295064/google-ai-threat-sergey-brin-founders-letter-technology-renaissance, accessed September 12, 2019.

[20] B. Lawlor, An overview of the NFAIS 2019 annual conference: creating strategic solutions in a technology-driven marketplace, *Information Services and Use* **39**(3) (2019), in press.

[21] C. Zandonella, New software allows for 'decoding digital brain data', *Science News* (2017), accessible at: https://www.sciencedaily.com/releases/2017/02/170224133923.htm, accessed September 12, 2019.

[22] Case Study: *Content Matchmaker at SAGE Publishing,* available at: https://www.67bricks.com/wp-content/uploads/2019/07/67-Brick-CS_Content-Matchmaker-v3.pdf, accessed September 12, 2019.

[23] M. Ware, Peer review: recent experience and future directions, *New Review of Information Networking* **16**(1) (2011), 23–53.

[24] M. Ware and M. Mabe, *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*. International Association of Scientific, Technical, and Medical Publishers, Oxford, UK, 2012.

[25] T. Reuters, Global Publishing: Changes in submission trends and the impact on scholarly publishers, April 2012.

[26] Peer Review: How We Found 15 Million Hours of Lost Time, AJE Scholar, available at: https://www.aje.com/en/arc/peer-review-process-15-million-hours-lost-time/, accessed September 12, 2019.

[27] Augusta Ada King, *Wikipedia,* see: https://en.wikipedia.org/wiki/Ada_Lovelace, accessed September 12, 2019.

[28] P. Thrower, *Eight Reasons I Rejected Your Article* Elsevier, 2012, available at: https://www.elsevier.com/connect/8-reasons-i-rejected-your-article, accessed September 12, 2019.

[29] AI-enhanced peer review: Frontiers launches next generation of efficient, high-quality peer review, available at: https://blog.frontiersin.org/2018/12/14/artificial-intelligence-peer-review-assistant-aira/, accessed September 12, 2019.

[30] Based upon the Defense Advanced Research Projects Agency (DARPA "AI Next Campaign", see: https://www.darpa.mil/work-with-us/ai-next-campaign, accessed September 13, 2019.

[31] For an interesting article on this topic see: Extance, A., How AI Technology can Tame the Scientific Literature, *Nature*, September 10, 2018, available at: https://www.nature.com/articles/d41586-018-06617-5, accessed September 13, 2019.

[32] Science Direct, *Wikipedia,* https://en.wikipedia.org/wiki/ScienceDirect, accessed September 19, 2019.

[33] Scopus, *Wikipedia,* https://en.wikipedia.org/wiki/Scopus, accessed September 19, 2019.

[34] *Topic Prominence in Science*, available at: https://www.elsevier.com/solutions/scival/releases/topic-prominence-in-science, accessed September 19, 2019.

[35] *ArtificiaI Intelligence: How Knowledge is Created, Transferred, and Used: Trends in China, Europe, and the United States*, available at: https://www.elsevier.com/__data/assets/pdf_file/0011/906779/ACAD-RL-AS-RE-ai-report-WEB.pdf, accessed September 19, 2019.

[36] R. Eveleth, Academics write papers arguing over how many people read (and cite) their papers, *Smithsonian Magazine* (2014), accessible at: http://www.smithsonianmag.com/smart-news/half-academic-studies-are-never-read-more-three-people-180950222/, accessed September 14, 2019.

[37] L. Meho, The rise and rise of citation analysis, *Physics World* **20**(1) (2007), 32, available at: https://arxiv.org/ftp/physics/papers/0701/0701012.pdf, accessed September 14, 2019.

[38] Y. Gardner and S. Inger, *How Readers Discover Content in Scholarly Publications*, 2018, available at: http://oca.unc.edu.ar/files/How-Readers-Discover-Content-2018-Published-1-2.pdf, accessed September 14, 2019.

[39] See: https://www.ncbi.nlm.nih.gov/books/NBK3827/, accessed September 14, 2019.

[40] M. Cockerill, Case Study: "What Now? TrendMD Guides Readers to the Most Relevant Further Reading," available at: https://www.trendmd.com/blog/what-now-trendmd-guides-readers-to-the-most-relevant-further-reading/, accessed September 14, 2019.

[41] See: https://www.trendmd.com/blog/how-to-increase-views-of-your-scholarly-articles-by-87/, accessed September 14, 2019.

[42] P. Paul Kudlow, M. Cockerill, D. Toccalino, D. Bissky Dziadyk, A. Rutledge, A. Shachak, R.S. McIntyre, A. Ravindran and G. Eysenbach, Online distribution channel increases article usage on mendeley: a randomized controlled trial, *Scientometrics* **112**(3) (2018), 1537–1556, available at: https://link.springer.com/article/10.1007/s11192-017-2438-3, accessed September 14, 2019.

[43] *Framework for Information Literacy for Higher Education,* ACRL, January 11, 2016, available at: http://www.ala.org/acrl/standards/ilframework, accessed September 14, 2019.

[44] Statista, available at: https://www.statista.com/statistics/967402/united-states-smart-speakers-in-households/, accessed September 14, 2019.

[45] B. Lawlor, An overview of the NFAIS 2017 Annual Conference: the big pivot: re-engineering scholarly communication, *Information Services and Use* **37** (2017), 290, accessible at: https://content.iospress.com/download/information-services-and-use/isu854?id=information-services-and-use%2Fisu854, accessed September 14, 2019.

[46] B. Lawlor, An overview of the NFAIS 2018 annual conference: information transformation: open, global, collaborative, *Information Services and Use* **38** (2018), 19, available at: https://content.iospress.com/download/information-services-and-use/isu180013?id=information-services-and-use%2Fisu180013, accessed September 14, 2019.

[47] J.J. Selingo, Are colleges preparing students for the automated future of work?, *The Washington Post* (2017), available at: https://www.washingtonpost.com/news/grade-point/wp/2017/11/17/are-colleges-preparing-students-for-the-automated-future-of-work/, accessed September 15, 2019.

[48] Basic Income, *Wikipedia*, accessible at: https://en.wikipedia.org/wiki/Basic_income, accessed September 14, 2019.

[49] R. Susskind and D. Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts* Oxford University Press, 2016, ISBN: 9780198713395.

[50] *Information Services and Use*, 39, 3, 2019, in press.

[51] URI to Launch Artificial Intelligence Lab, Press Release, December 20, 2017, available at: https://today.uri.edu/news/uri-to-launch-artificial-intelligence-lab/, accessed September 14, 2019.

[52] C. Baraniuk, Inside Finland's plan to become an AI powerhouse, *Wired* (2019), available at: https://www.wired.co.uk/article/finland-artificial-intelligence-online-course, accessed September 15, 2019.

[53] B. Lawlor, An overview of the NFAIS conference: blockchain for scholarly publishing, *Information Services and Use* **38**(3) (2018), 114, available at: https://content.iospress.com/download/information-services-and-use/isu180015?id=information-services-and-use%2Fisu180015, accessed September 15, 2019.

[54] See: https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html, accessed September 19, 2019.

[55] S. Rendle, Factorization machines, in: *Proceedings of the 2010 IEEE International Conference on Data Mining, December 13–17,* 2010, pp. 995–1000.available at: https://cseweb.ucsd.edu/classes/fa17/cse291-b/reading/Rendle2010FM.pdf, accessed September 19, 2019.

[56] See; https://en.wikipedia.org/wiki/List_of_Wikipedias, accessed September 20, 2019.

[57] Dunbar's Number, *Wikipedia*, available at: https://en.wikipedia.org/wiki/Dunbar%27s_number, accessed September 21, 2019.

[58] Linus' Law, *Wikipedia*, available at: https://en.wikipedia.org/wiki/Linus%27s_Law, accessed September 21, 2019.

[59] Wikipedia: Policies and Guidelines, *Wikipedia*, available at: https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines, accessed September 21, 2019.

[60] R.S. Geiger and A. Halfaker, When the Levee breaks: without bots, what happens to wikipedia's quality control processes?, 2012, preprint available at: http://files.grouplens.org/papers/geiger13levee-preprint.pdf, accessed September 21, 2019.

[61] Ibid.

[62] B. Schwartz, Google to close Google+ after 7 years: A look back at the impact it once had on Google Search, *Search Engine Land*, available at: https://searchengineland.com/google-to-close-google-after-7-years-a-look-back-at-the-impact-it-once-had-on-google-search-306360, accessed September 21, 2019.

[63] S.W. Sen, H. Ford, D.R. Musicant, M. Graham, O.S. Keyes and B. Hecht, Barriers to the localness of volunteered geographic information, in: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, April 2015*:pp. 197–206., available at: https://www-users.cs.umn.edu/~bhecht/publications/localnessgeography_CHI2015.pdf, accessed September 21, 2019.

[64] T. Simonite, Artificial intelligence aims to make wikipedia friendlier and better, *MIT Technology Review* (2015), available at: https://www.technologyreview.com/s/544036/artificial-intelligence-aims-to-make-wikipedia-friendlier-and-better/, accessed September 21, 2019.

[65] B. Marr, The amazing ways how wikipedia uses artificial intelligence, *Forbes* (2018), available at: https://www.forbes.com/sites/bernardmarr/2018/08/17/the-amazing-ways-how-wikipedia-uses-artificial-intelligence/#51d984422b9d, accessed September 21, 2019.

[66] See: https://www.google.com/search?q=hoe+long+do+goats+live&oq=hoe+long+do+goats+live&aqs=chrome..69i57j0l5.4719j1j8&sourceid=chrome&ie=UTF-8, accessed September 2019.

[67] C. McMahon, I. Johnson and B. Hecht, The substantial interdependence of wikipedia and google: a case study on the relationship between peer production communities and information technologies, in: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* 2017, pp. 142–151.ISBN: 9781577357889.

[68] Artificial Intelligence Leads the Change in Academic Publishing, Impelsys *Blog*, available at: https://www.impelsys.com/blog/artificial-intelligence-leading-change-academic-publishing/, accessed October 12, 2019.

[69] P. Daugherty, What's the future of work in the age of AI? *Salesforce Blog*, available at: https://www.salesforce.com/blog/2019/03/future-of-work-in-the-age-of-ai.html, accessed October 12, 2019.

[70] B. Lawlor, An overview of the NFAIS 2019 annual conference: creating strategic solutions in a technology-driven marketplace, *Information Services and Use* **39**(3) (2019), in press.

[71] bid.

[72] B. Kasenchak, What is semantic search? and why is it important?, *Information Services and Use* **39**(3) (2019), in press.

[73] J. Anderson and L. Rainie, Artificial intelligence and the future of humans, Report from the Pew research Center, available at: https://www.pewinternet.org/2018/12/10/artificial-intelligence-and-the-future-of-humans/, accessed October 9, 2019.

[74] R. Bordoli, Chief Executive Officer, Figure Eight, available at: https://www.salesforce.com/video/1718054/, accessed October 12, 2019.

[75] Merger of Major Information Industry Associations Finalized, NISO press releease, July 1, 2019, https://www.niso.org/press-releases/2019/07/merger-major-information-industry-associations-finalized, accessed July 16, 2019.

[76] The NISO Plus Conference, NISO Events, available at: https://www.niso.org/events/2020/02/niso-plus-conference, accessed October 10, 2019.