

ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for May 14, 2021.

Please let us ([Micaela Parker](#), [Steve Van Tuyl](#), [Catherine Cramer](#), [Brad Stenger](#)) know if you have something to add to next week's newsletter. We are grateful for the generous financial support from the [Academic Data Science Alliance](#).

Thank you for helping DSCN with our first-ever reader survey. We look forward to sharing the results and incorporating your recommendations.

Academic Data Science News

Universities have been experimenting with data technologies, whether by necessity or curiosity. Results are mixed. Starting at the bottom rung, online proctoring appears to be [a "raging dumpster fire"](#) to be avoided at all costs. Here's [138 additional thumbs-down](#). Research on the effectiveness of video lectures [gets a slightly better grade](#) of "unlikely to be detrimental and usually improve student learning." (Is that a C+?) [Predictive models appear to still have some value](#) in understanding university retention rates, though mainly as a backdrop for reality, and as a "before" snapshot against which to develop improved models that address algorithmic bias. The bottom line: [AI is used by universities for admissions, grading, gauging mental health and more](#). "But universities would be making a mistake if they thought that automating decisions somehow relieved them of their ethical and legal obligations," said **Kirsten Martin**, a professor of technology ethics at the **University of Notre Dame**. And on the A+ side is the [increasing use of computational notebooks](#) as scholarly objects. Says **Henri Drake**, a grad student at MIT, "These kinds of notebooks can really get people excited and expose people to concepts that are otherwise kind of impenetrable."

University of Wisconsin-Madison researchers have been busy. Over the last three years they've written 21,500 articles, presented at over 1,500 conferences, earned almost 500 patents and generated nearly \$2 billion in grant funding. Fortunately now there is [a searchable database for all of this output](#), called [Research at UW-Madison](#). Researchers can add their own work, such as software, performances and exhibitions, while additional data comes from an outside service, [Academic Analytics](#), which is updated nightly and which populates the scholar profiles from available information in [CrossRef](#). Academic Analytics' services to universities are multi-faceted, and sometimes not transparent. **University of Missouri** chancellor **Mun Choi** is [taking heat](#) for using Academic Analytics to make tenure decisions. While he assured faculty that the data is not used as a "deciding factor," he did agree that faculty should not need to ask their department chair for access to their own data.

University of North Carolina-Chapel Hill is [soon to launch a data science initiative](#), good timing considering [the announcement](#) of **Apple's** new billion-dollar campus focused in machine learning, artificial intelligence, and software engineering, to be located in Research Triangle Park near campus. The announcement – coming on the heels of **Google's** [similar move](#) announced in March – underscores how North Carolina universities' moves into data science are making the area [attractive to tech companies](#). Also benefiting is the NC tech workforce, which – between Apple and Google by themselves – will have close to 5000 new jobs on offer. The Apple-university partnership

is [equally strong in San Diego](#), where **University of California-San Diego** is building two new engineering research buildings as Apple simultaneously announced it plans to bring over 4000 new jobs to town. **San Diego State University** and **University of San Diego** are similarly becoming well-poised for dancing with Apple.

Harvard Data Science Review, the academic publication launched by the **Harvard Data Science Initiative** in 2019 and published by **MIT Press**, devoted [the most recent issue](#) to an in-depth look at data science education and research at the University of California-Berkeley. **Michael Jordan** has an article in the issue, and when [asked why students go for data science](#) by **The Daily Californian** student newspaper, he said that "this is not just the thing that will make [them] money; it's the thing that will allow [them] to have an impact on the world."

Esports – organized video gaming competitions played for spectators – are [the backbone in a new degree minor](#) offered by the **University of Michigan School of Information**, thanks to a gift from **Activision Blizzard**, whose CEO, **Robert Kotick**, is an alum. The School of Information's program will address what dean **Thomas Finholt** called "critical social and behavioral questions surrounding games and game use, and the design of games."

Two **University of Minnesota** researchers used the Linux kernel community as a testbed. Without asking. And then they wrote and published a paper about it. The result? The entire university is [banned from submitting to the Linux kernel](#). In addition, the **Linux Foundation** asked the researchers to withdraw their paper, which they have.

High school students with vision impairment and other learning disabilities [will be able to access data science learning materials and tools](#) more readily thanks to an **NSF-supported** project led by **University of Nevada-Las Vegas** and including several other university partners. The team will develop multisensory content, combining visual, auditory, touch-based and natural language stimuli to make statistics, graphical representations and other information accessible.

The **U.S. Department of Defense** is [establishing centers of excellence](#) at two historically Black schools — **Howard University**, which will have a center focused on AI machine learning, and **Delaware State University**, whose DoD-backed center will focus on quantum sensing.

University of Wisconsin-Eau Claire will soon be [benefiting from a new Bose Supercomputer Cluster](#) thanks to a gift from nearby **Hewlett Packard**, which sees the enhanced computing capabilities as helping drive research as well as preparing students for careers at HP.

UCLA received a \$5 million grant from **Mellon Foundation** to establish [a social justice curriculum](#) that focuses on data. Plans include a new master's degree program in data and society, according to **Darnell Hunt**, the school's dean of social sciences.

The **Klingler College of Arts and Sciences at Marquette University** [began accepting applications](#) for a new Master of Science in Data Science degree program that will start in fall 2021. Marquette undergraduates have an accelerated option to complete the two-year program in one year along with their four-year undergraduate degree.

Oregon State University [plans to enroll](#) 40 graduate students as the College of Engineering's first cohort of data science master's and Ph.D. students. The program is highly interdisciplinary. Students work on core artificial intelligence topics while also choosing relevant courses from a range of disciplines across the university.

The **University of Michigan Ross School of Business** announced a [new Master of Business Analytics program](#). The ten-month residential program is a mix of data analytics and business courses.

The **University of Richmond** will increase its [data science offerings to undergraduates](#) with a new data science concentration for the school's computer science and math students, and a business analytics concentration for the school's business majors. This data science initiative is overseen by chemistry professor **Carol Parish**. **George Mason University** launched a new [bachelor of science in Health Informatics](#) degree program. **Arizona State** created a new [undergraduate neuroscience and brain science degree program](#).

Rebecca Nugent is the [new department head](#) of **Carnegie Mellon's Department of Statistics and Data Science**. Her appointment begins on June 15.

Susan Christoffersen [has been named dean](#) of the **University of Toronto Rotman School of Management**. Christoffersen is a finance professor and had been co-director of the [TD Management Data and Analytics Lab](#) at Rotman.

The **University of Missouri-St. Louis** [hired Will Werner](#) to direct UMSL's National Security and Community Policy Collaborative, a key post for collaborating with the **U.S. National Geospatial Intelligence Agency**, which is set to open its new St. Louis campus in 2025, and with the region's growing geospatial technology community.

Enormous gifts from **Boeing to Virginia Tech** (\$50 million) and **Bloomberg Philanthropies to Johns Hopkins University** (\$150 million) are earmarked for improving diversity in graduate technology education. The Boeing grant is [an investment](#) in Virginia Tech's new Innovation Campus in Alexandria, near Washington DC. (**Amazon announced plans** to increase staff at nearby HQ2 from 1,600 to 3,500.) The Bloomberg grant [will establish](#) the Vivien Thomas Scholars Initiative, a pathway for students from Historically Black Colleges and Universities and Minority Serving Institutions to pursue and earn PhDs in STEM fields. Six other institutions will also receive funds in association with the Initiative: **Howard University, Morehouse College, Morgan State University, Prairie View A&M University, Spelman College** and the **University of Maryland, Baltimore County**.

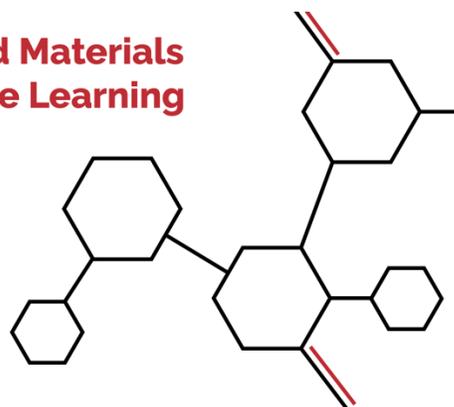
The **Heinz Endowments** [committed \\$30 million](#) to establish the **Center for Shared Prosperity at Carnegie Mellon**. The Pittsburgh-centric program is the largest grant in the history of The Endowments. CMU robotics professor **Illah Nourbakhsh** is the founding executive director.

Sponsored Content

Advancing Chemical and Materials Science through Machine Learning

Monday, June 14, 2021
10:30 AM – 5:00 PM EDT

Please register in advance on Eventbrite



Machine Learning for Chemistry & Materials Science Focused Research Program Symposium.

 Hariiri Institute for Computing

The *Advancing Chemical and Materials Science through Machine Learning* symposium looks to explore state of the art research in experimental and computational chemical and materials science and their interface with data sciences and machine learning through keynote speakers and focused panel discussions on challenges and opportunities in the fields. [Register in advance on Eventbrite by clicking here](#)

Editor's Picks

Researchers working on artificial intelligence and machine learning struggle to realistically project their work into the future. A [recent essay](#) by **Melanie Mitchell**, a **Santa Fe Institute** professor, discusses widespread errors in research assumptions that lead to overconfident prediction. "It's clear that to make and assess progress in AI more effectively, we will need to develop a better vocabulary for talking about what machines can do," Mitchell writes. "We will need a better scientific understanding of intelligence as it manifests in different systems in nature."

Technology transfer is another area where failing to include realism in communications hinders progress. Entrepreneur **Andrew Ng** participated in a recent **Stanford Institute for Human-Centered Artificial Intelligence Q&A**. He was asked why hospital decision-support tools perform no better than doctors, and said that it's one thing to write a research paper, and the challenge to make something clinically useful is far greater. The "gap between research and practice is not unique to medicine," Ng added.

Journalism catches flack as a major source of false expectations, but it's worthwhile to point out exemplars of realism. **The New Yorker** enlisted British science writer **Philip Ball** to [write about "animal translation."](#) He deftly navigates the animal intelligence showed by dolphins and octopus. Neither he nor his sources are confident that the research will yield understanding. Another New Yorker [article](#), by **Stephen Marche**, documents how GPT-3 has been a tech transfer success, and how a commercial implementation of GPT-3 changes how writers write.

Advancing the grounded, common language of AI should also help researchers who also struggle with ethics. **Virginia Dignum** from **Umea University** in Sweden points out, like Mitchell, the fallacies in researcher assumptions. She [calls AI-Fairness a myth](#). "Life isn't fair, and neither is AI," according to Dignum, "The point is not complete fairness, but the need to establish metrics and thresholds that ensure trust in AI systems."

Bruce Schneier, the computer security expert, gives [his assessment](#) of a world where AI does the malicious hacking that people do now. He makes sure to ask himself, "How realistic is AI hacking in the real world?" The inevitability isn't avoidable and it also isn't predictable, but it will be without precedent. When it comes to untrustworthy AI systems, Schneier foresees resilient governance capable of patching vulnerabilities rapidly, not the sclerotic leadership seen in many rule-making bodies.

Consider sports. Twenty-five years of Moneyball analytical hacks have removed athleticism from the game of baseball while adding dead time from pitching changes, pickoff attempts and strikeouts. **Major League Baseball** is scrambling to [bring in new rules](#) to reverse those trends and improve the product. The **National Football League** seeks to learn from baseball, embedding analysts in the league office, a group led by **Skidmore University** professor **Michael Lopez**. Each year Lopez gets hundreds of hackers to participate the [Big Data Bowl](#), the NFL's analytics contest put on with the help of **Amazon AWS** and **Kaggle**. This year's [winning team](#) included **Sam Walczak**, a student in the **University of Minnesota** data science master's program. Walczak [met teammates](#) **Marc Richards** and **Jack Werner** as undergraduates at **St. Olaf College**. **Wei Peng**, who studies statistics in the **University of Pittsburgh** Statistics Ph.D. program with Richards, completed the squad.

Research News

The crux of the argument by **Ryan Abernathy** (and his earth and atmospheric science co-authors) for [cloud-native repositories for big scientific data](#) is that "open architectures are the best path forward for big-data scientific infrastructure," and it is important to decouple "data storage from the computing platform without sacrificing performance." That paper in *IEEE Computing in Science and Engineering* is one of [three papers](#) that outline the vision that Abernathy and **Pangeo** project colleagues have for making effective shared infrastructure available to

data-intensive researchers. The timing is good for these talking points as efforts to create a National Research Cloud, led by **Stanford Institute for Human-Centered Artificial Intelligence** (HAI) [ramp up](#) after the U.S. federal government [passed](#) the *National Artificial Intelligence Initiative Act* in January. The law calls for [a task force](#) to investigate what a National Research Cloud should look like and how it should work. The assumption is that the new Cloud will [embrace and extend CloudBank](#), a **National Science Foundation** clearinghouse for cloud resources available to NSF grantees. CloudBank and its developer, **San Diego Supercomputer Center**, recently [teamed with European experts in cloud service procurements](#), namely **Strategic Blue**, a UK-based consultancy, and **CERN**, the famed particle physics laboratory.

The **University of California-Davis** partnered with **Maurice Gallagher**, a UC-Davis alum and entrepreneurship supporter, to [develop a new rapid COVID-19 test](#). The technology uses machine learning algorithms for detecting infectious disease from biological samples tested for chemical composition using mass spectrometry, an automated machine learned platform called MILO (Machine Intelligence Learning Optimizer) invented at the **UC-Davis School of Medicine** by professors **Hooman Rashidi** and **Samer Albahra**. The university and Gallagher will spin out a company, **SpectraPass**, to commercialize the test and the testing technology.

The versatility of Natural Language Processing is on display as researchers look for grammars that explain cellular protein products called biomolecular condensates, membrane-less cellular goo. **Cambridge University** researchers used the DeePhase NLP algorithm in "a hypothesis-free manner" to examine the conditions and constraints for how proteins form into biomolecular condensates during phase separation. The ability to learn about the basic chemistry and physics of a process without having to use pre-existing knowledge as inputs is a "fundamentally new way of approaching the problem," says researcher **Tuomas Knowles**.

University of California-San Diego is [inviting users of Oura Ring wearable technology](#) who have been pregnant to contribute their device data to an in-depth study of the physiological patterns of pregnancy and their related positive and negative health outcomes. **Oura** makes a sensor-laden ring that tracks temperature, heart rate and respiration rate and costs \$300. The Finnish startup recently also completed a \$100 million [venture funding round](#).

Artificial intelligence is being used to [digitally unfold and read](#) over 3,000 letters from the 17th Century. (Intricate folding patterns were the precursor to self-stick envelopes; unfolding 400-year-old paper is likely to damage it.) More ancient history, the Dead Sea Scrolls are now [thought to be written by two authors](#), thanks to AI-powered handwriting analysis.

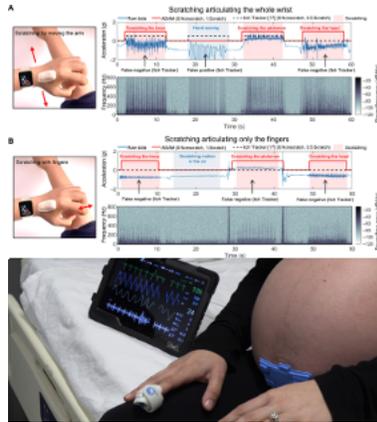
Tweet of the Week

Twitter, @KirkCodes from April 25, 2021



Data Visualization of the Week

Northwestern University, McCormick School of Engineering from May 10, and *ScienceAdvances*, John A. Rogers and Shuai Xu et al. from April 30, 2021



Events

See the [ADSA Events Page](#) for more details and more opportunities.

[Race and Technology: A Research Lecture Series](#)

Online May 26, starting at 10 a.m. Pacific. Speaker: **Sareeta Amrute**, **Data & Society Research Institute** and **University of Washington**. [registration required]

[Tech & Racial Equity Conference: Anti-Racist Technologies for a Just Future](#)

Online May 19-20. "Sponsored by the **Stanford Center for Comparative Studies in Race & Ethnicity (CCSRE)**, **Digital Civil Society Lab (DCSL)** at the **Stanford Center on Philanthropy and Civil Society**, and **Stanford Institute for Human-Centered Artificial Intelligence (HAI)**." [free, registration required]

Deadlines

Education Opportunities

[Virtual Public Health Modeling Course Offered at Yale](#)

"The **Yale School of Public Health** is offering a week-long virtual summer course in Public Health Modeling June 21-25." ... "The course is designed to provide researchers, clinicians, industry professionals and policymakers with the systems-based perspective and analytical tools they need to better understand and manage the complex forces that drive health outcomes." Deadline to apply is June 15.

RFPs

[Yale University and Boehringer Ingelheim Launch Biomedical Data Science Fellowship Program](#)

"The program will be based at the **Yale Center for Biomedical Data Science (CBDS)** in New Haven." Deadline for research proposals is June 15.

[Spell Open Research Grant for large-scale research projects in machine learning](#)

"We will be giving each winning project \$40,000 in cloud credits and free access to the **Spell** platform." ...

"Additionally, we will be rewarding smaller mini-grants of up to \$5,000 in cloud credits for smaller-scale research and ML-based art projects. " Deadline for research proposals is June 1.

Contests/Award

[Significance Magazine: Our writing competition for early-career statisticians and data scientists has just over a month left to run.](#)

"2018 winner @LetishaAudrey wrote 'Cooking up statistics: The science and the art'" ... "What will you write about?"
Deadline for submissions is May 31.

Tools & Resources

[Announcing the NASA Openscapes Framework](#)

Openscapes, NASA from Mar 10, 2021

"Together with the **National Center for Ecological Analysis and Synthesis (NCEAS)** at the **University of California Santa Barbara**, **University of Washington eScience Institute**, **Metadata Game Changers, 2i2c**, and **The Carpentries**, we are answering the call from **NASA** to support researchers from the **NASA Distributed Active Archive Centers (DAACs)** migrating workflows to the cloud."

[How to Change Gender Disparity Among Data Science Roles](#)

InformationWeek, Joao-Pierre S. Ruth from Apr 29, 2021

"**Tracy Ring**, managing director of cognitive, analytics and RPA alliances for **Deloitte Consulting** and founder of Deloitte's women in data science and analytics group, spoke to **InformationWeek** about her team's accomplishments and what organizations can do to improve gender parity especially in the data science workforce."

[We're excited to announce Observable Plot, a new open-source library for faster and easier data exploration on the web!](#)

Twitter, Observable from May 4, 2021

"Plot's concise API and thoughtful defaults are designed for a more joyful visualization process, try out the beta and let us know!"

Careers

See the [ADSA Jobs Page](#) for more opportunities.

About Us: The Data Science Community Newsletter was founded in 2015 in the Moore-Sloan Data Science Environment at NYU's Center for Data Science. We continue to be supported by the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation through the [Academic Data Science Alliance](#). The newsletter is written and the content is compiled by the Academic Data Science Alliance. Our archive of newsletters is at cds.nyu.edu/newsletter and is the process of transitioning to another, permanent location. Our mailing address is [1037 NE 65th St #316; Seattle, WA 98115](#).

OPT OUT: If you do not want to receive this newsletter, please email brad.stenger@gmail.com with the word 'unsubscribe' in the subject line.

OPT IN: Feel free to forward the Data Science newsletter to colleagues. They can sign up for the newsletter [using this web form](#).

