Data Science Community Newsletter features journalism, research papers and tools/software for April 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.

Our first-ever audience survey is poised to hit inboxes within the week. Your input is a big help. Can't wait to get the responses in – we really want to know more about our readers. Thanks!

Academic Data Science News

Data journalists at the **Stanford Daily** student newspaper counted the available undergraduate tech courses and then they counted the undergraduate tech ethics courses. Charts display the historical rise of student interest in computing, and the simultaneous non-history of ethics coursework for them. The situation is improving. The article mentions an interdisciplinary course first offered in 2018 and created by political science and computer science faculty, "CS182: Ethics, Public Policy and Technological Change." At **Northwestern University**, **Mason Secky-Koebel** and **Bijal Mehta** recently founded **Responsible Artificial Intelligence Student Organization** (RAISO). The group has traction and plans to maintain a presence among STEM and non-STEM students on campus.

Erin Cech and **Cindy Finelli**, **University of Michigan** faculty members in the departments of sociology and electrical engineering and computer science, are designing a new master's-level course to address what they view as a hole in the education of engineering students in the area of professional responsibility. The project is funded by the **National Science Foundation** (NSF) and it's titled "Advancing Engineers' Ability to Recognize, Strategize About, and Act On Concerns Related to Public Welfare."

Students who have gained admission to top schools benefitted in many cases from their high SAT scores, a subject that **This American Life** recently addressed, using the pandemic's no-SAT-required application boom as its news peg. Troubling proof that the institutional bias is academic, not test-specific, comes from the **Stanford Center for Educational Policy Analysis**, which recently showed college applicants' essay content is strongly related to household income, just like SAT scores are strongly related to household income. But other new research by **Chris Bennett** of **Vanderbilt** shows student diversity gains from test-optional admissions policies.

Like students at top universities, tenure-track academics have benefitted from institutional bias in the process of obtaining their station. There's new research by **Allison Morgan** from **University of Colorado** where she surveyed 7000 tenure-track faculty. It showed that they come from higher-than-average income households and they are 25 times more likely than the rest of the population to have parents who earned a Ph.D. So, bias is pervasive, which doesn't make it any less important to call out, to

work on – and with hard work – to improve.

Princeton University and **WHYY**, the **NPR** outlet in Philadelphia, recently premiered [a new podcast](#) called "A.I. Nation." Hosted by **Ed Felten**, founding director of the **Center for Information Technology Policy at Princeton**, and **Malcolm Burnley**, new episodes will be released weekly during April. **Harvard Data Science Review** also has [a new podcast](#). Each monthly episode is a "case study" into how data is used today.

The **University of Wisconsin-Madison** says that Data Science is the [fastest growing major on campus](#). 314 undergrads have declared the major in the program's first six months. Note: 314 is 1% of the undergrad population at UW and Data Science is one out of 125 possible degrees. For comparison, 587 students earned Economics degrees, the most popular undergrad program according to College Factual's [most recent data](#). So, pretty good thus far, and with upside potential.

Downeast Magazine [interviews](#) **Chris Mallett** from **Roux Institute**, the **Northeastern University** artificial intelligence and data science outpost in Portland, Maine. Roux enrolled 200 grad students in its first year and trained another 200 individuals from corporate partner institutions. The Institute is succeeding at taking college students from **Colby** and **Bates** and adding to their tech skills.

Cornell will name the **College of Computing and Information Science** after **Ann S. Bowers**, a 1959 alumna and longtime Silicon Valley executive. Bowers' landmark gift to the school [will fund](#) construction of a new building for the college and hiring for up to 15 new faculty.

Clarkson University renamed its Masters Degree program in Data Science. The [new MS degree](#) is in Applied Data Science. The 36-credit program can be completed in one or two years, on-line or on-campus. Very convenient.

The **School of Information Studies at Syracuse University** is set to launch [two new undergraduate degree programs](#): "Innovation, Society and Technology" and "Applied Data Analytics." These two new majors will complement the school's other degree program in Information Management and Technology. **St. Lawrence University** faculty [recently approved](#) a new undergraduate major in Data Science. Northern New York State is teeming with trees and lakes and soon, young data scientists.

Miami Dade College started the Data Science 4 All, DS4A/Empowerment [partnership](#) with **SoftBank** and **Correlation One** to create career opportunities in data science for qualifying community college students.

The **State University System of Florida** [approved](#) a new Master of Science in Applied Data Science at **New College** in Sarasota.

Microsoft [gifted](#) $5 million to **The Institute for Protein Design (IPD) at the University of Washington School of Medicine** in order to advance research on biotechnological tools that may be used against future pandemics.

Cleveland Clinic and **IBM** agreed to [a ten-year partnership](#) to launch the Discovery Accelerator that will focus on comprehensive clinical and research computing infrastructure, including quantum systems. One beneficiary will be Cleveland Clinic's new **Global Center for Pathogen Research and Human Health**.

The **Lilly Endowment** granted **Purdue University** $10 million [to establish Indiana Digital Crossroads](#), a statewide data science education initiative that evolves The Data Mine education program established by **Mark Daniel Ward**. The grant is part of a [larger $70 million effort](#) by Lilly Endowment that funded new programs at 16 Indiana colleges and universities.

The **Koch Institute for Integrative Cancer Research at MIT** [named](#) biology professor **Matthew Vander Heiden** as director and successor to **Tyler Jacks**. Vander Heiden "expects that one key area of focus in

the coming years will be applying the power of machine learning and artificial intelligence to understanding cancer."

The **Broad Institute**, a computational biology collaboration between **Harvard** and **MIT**, launched the **Eric and Wendy Schmidt Center**. Research at the center will integrate machine learning and the life sciences. The Schmidt Center has a $150 million endowment and is co-directed by **Anthony Philippakis**, chief data officer at Broad, and **Caroline Uhler**, associate professor of Electrical Engineering and Computer Science at MIT.

The **NYU Center for Data Science** is the focal point for New York University's university-wide efforts in Data Science. CDS offers rigorous, top-ranked Data Science education programs to PhD, Master's, and Undergraduate students, and is home to 17 joint faculty and a prestigious postdoctoral research program. For more information on academic admissions, visit the **CDS website** and email **datascience-group@nyu.edu** with any questions. Open faculty and postdoctoral positions are posted to **cds.nyu.edu/jobs**.

Editor's Picks

High-quality AI services, basic business (or life) functions that rely on Artificial Intelligence in some form or fashion, are a sign that we are on our way to figuring out how to do the human-machine collaborations that AI enables. Two forces are at work with these services. Is there onerous work for the computer to do? And more interestingly, is there a decision outcome that suits blameless computing rather than blame-able humans? (New research from **University of Georgia** suggests that people trust computers more than humans.) You can see both forces at work with the AI that researchers are using to marshal submission reviews for academic conferences.

Manufacturers have long used AI services like optical inspection systems. The current state of the art is a steel weld inspection process developed by **John Deere** and **Intel**. Robots do the welding and now robots do the welding inspection too. The ideas are trickling down to small business. **FloWaste** is a startup from the **Trinity College Dublin** incubator that puts cameras on commercial kitchen wastebaskets to capture data and perform analytics on what's being wasted. Service industries need AI services too. Insurance relies on catastrophic risk modelers like **Risk Management Solutions** (RMS). RMS is launching new climate change models this June which will affect predictions for future hurricanes and floods.

Government, the provider of many basic services, is benefitting from new AI services. A new and in theory more fair approach to election redistricting is being tried at the **MGGG Redistricting Lab at Tufts**, which collects election and Census geodata, researches redistricting and election reform and focuses on building free, open-source tools like Districtr, an open mapping tool for creating district and community maps. Low quality services in poor communities is nothing new. **University of Iowa** researchers are accessing thousands of emails pertaining to the **Flint MI** water crisis to analyze, annotate, and categorize individual files in the dataset, and in the process find out how things got so bad there. AI may also be able to correct for the long-standing lack of energy grid resilience, a persistent problem impacting poor communities and developing nations. Nashville will soon get state-of-the-art traffic management infrastructure courtesy of a collaboration between **Southwest Research Institute** and **Vanderbilt University**. Goodbye I-24 gridlock.

Consumers might be the biggest beneficiaries of new AI services. AI is a crucial element in how **The**

Simpsons is produced. At first [AI wrote](#) simple investment reports and sports stories, and now it [writes decent longform](#) articles

A UK-based project called **Lost Tapes of the 27 Project** is [pioneering new music](#) in the style of famous, but dead, musicians. AI will [augment your appreciation](#) for the emotion in paintings. [New improvements](#) to **Google Maps** guide you on your way in the most pollution-free way possible. Self-driving robots provide [contactless food deliveries](#) from Ann Arbor startup **Refraction AI**. Another startup, **Marble**, founded by **Harvard** data science student **Blake Bullwinkel**, has a [smartphone app](#) that estimates the carbon footprint of supermarket products.

Equity, fairness and usability in medical data have been getting close looks lately. One area is in medical devices, [still being sold and used](#) even after being labeled as biased. Whether the bias is physical, computational or interpreted, it will show up in data gathered from these devices, such as spirometers and pulse oximeters. In the use of Electronic Health Records (HER) research indicates that there are [gender differences in how EHR data are used](#) – raising fairness and bias flags. The lack of coordination between healthcare services and social services negatively affects those who need them both the most. The Leveraging Integrated Networks in Communities (LINC) to Address Social Needs Act (S.509) [will award grants](#) to states to develop or enhance collaboration between health care organizations and social service sectors, translating into integrated and interoperable data.

Reid Blackman, writing at **Harvard Business Review**, [recommends Institutional Review Boards](#) for every company that uses AI. Painful, but perhaps pain is what's needed?

Research News

For those of us who rely on **NSF** funding – and anyone else who might appreciate greater support for science – it comes as excellent news that now both houses of **Congress** are working on [bills to double the NSF budget](#) and create a new technology directorate. Fingers crossed on that, and also on getting [an accurate 2020 Census count](#) – not so easy, based on reporting from Reveal and their use of crowdsourced data. AI is top of mind throughout governmental offices. The **National Security Commission on Artificial Intelligence** (NSCAI) released its Final Report and Recommendations and there are [assumptions galore about AI](#), among them: the US and China are in an "AI arms race"; autonomous weapons systems are inevitable; and AI itself has powers and promise far outside its actual use. The EU seems to be doing a far better job in regulating AI-enhanced household appliances and consumer electronics. Here in the US the **Consumer Product Safety Commission** (CPSC) is starting to take a [closer look](#) at consumer product safety in the Age of AI – and hopefully will come up with something that has teeth. Or legs. Or both.

Ubiquitous terrestrial and extra-terrestrial sensing translates to useful ecosystem-wide information but [can have unintended consequences](#). The plethora of data-gathering and data-bouncing satellites already in orbit – including the 1000 launched by **SpaceX Starlink**, not including the tens of thousands to be launched in the coming years – add about 10% more diffuse light everywhere in the night sky, negatively impacting sensitive astronomical measurements. "As space gets more crowded, the magnitude of this effect will only be more, not less," says **John Barentine**, director of public policy at the **International Dark-Sky Association**.

Quantified self no longer applies exclusively to humans. **Jay Levine** at **North Carolina State** [has added sensing](#) know-how to mussels' innate environmental scrubbing capability. When mussels thrive so do coastlines. Researchers at **Washington State University's Bear Research, Education, and Conservation Center** [trained full-grown bears to walk on treadmills](#), comparing their walk patterns to human hiking patterns (a/k/a trails). Heads up: Bears like the same flat easy paths we do. So, for a bear-free picnic, choose the steep rocky trail.

Xenobots! [Xenobots](#) are frog embryo cells that develop non-frog structure and function over time and they began as an idea, a test case performed in silico, using algorithms created by **Josh Bongard** and **Sam Kriegman** of the **University of Vermont**. **Stanford University** biologist **Dominique Bergmann** used single-

cell RNA sequencing technologies to [track genetic activity](#) in nearly 20,000 plant cells as they formed an Arabidopsis leaf. She found considerable ambiguity, especially during early development, in the paths cells traversed in reaching their biological outcome.

Brian Cunningham at the **University of Illinois** recently invented a new optical microscope capable of [counting the individual virus and small protein material](#) in a sample at room temperature. The device opens new doors to virus testing and studying protein interactions, and it is about to be prototyped for analyzing blood and breath samples.

A collaboration between **MIT**, **J. Craig Venter Institute** and the **National Institute of Standards and Technology** (NIST) [added seven genes](#) to what had been the simplest synthetic organism created using a human-selected genome. The orirginal synth blob lacked for consistent reproductive characteristics but the updated synth, now 480 genes, reproduces into neat, separate, uniform orbs. Not so blobby. "Life is still a black box," NIST's **Elizabeth Strychalski** said, but "scientists are getting a good look at what's going on inside."

Best practices for using machine learning for materials discovery are still to be determined, according to [this overview](#) of benchmarking studies by **Chemical & Engineering News**. Chemists who use machine learning to guide new materials engineering are as likely as computer vision researchers to see out-of-left-field errors. Frequently they are unaware of their technical options for machine learning tools, often working off uninformed recommendations that the benchmarks hope to prevent. Still the list of potential tools grows. An ["atom-mapping tool"](#) invented at **IBM Zurich** claims to extract an "organic chemistry grammar from unsupervised learning of chemical reactions." **Chemistry World** magazine has an [in-depth feature](#) on another **IBM** AI + materials effort, this one to discover new anti-microbial peptides.

the **University of Vermont**. **Stanford University** biologist **Dominique Bergmann** used single-cell RNA sequencing technologies to [track genetic activity](#) in nearly 20,000 plant cells as they formed an Arabidopsis leaf. She found considerable ambiguity, especially during early development, in the paths cells traversed in reaching their biological outcome.

Data Visualization of the Week
Twitter, Ben Lorica from Apr 7, 2021
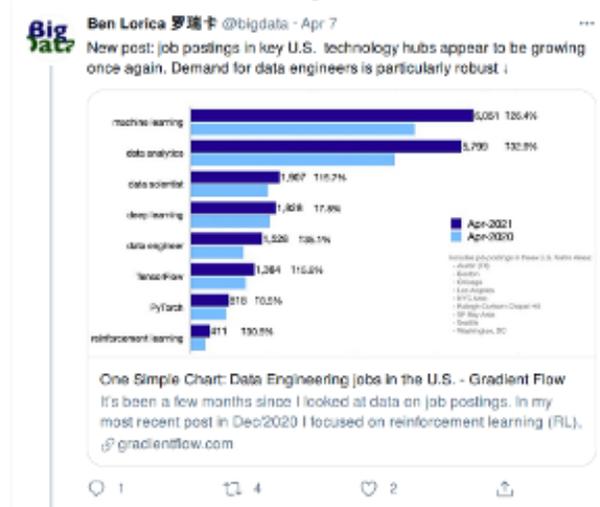
Tweet of the Week
Twitter, Hidden Brain from Apr 12, 2021

Hidden Brain ✔
@HiddenBrain

New research finds that when we come up with ways to improve on something, whether it's a device or an idea, we tend to focus on what we can add, not on what we can subtract.

And so "people systematically overlook opportunities to change things through subtraction."

2:25 PM · Apr 12, 2021 · Twitter Web App

68 Retweets   18 Quote Tweets   285 Likes

Events

See the ADSA Events Page for more details and more opportunities.

Smart Cities Miami Conference
Online April 15, starting at 1 p.m. Eastern. "This year's conference theme is health and well-being in the post-pandemic city" [registration required]

Policymaking and artificial intelligence: A conversation with John R. Allen and Darrell M. West
Online April 21, starting at 2:30 p.m., organized by **The Brookings Institution**. "A fireside chat on their book, *Turning Point: Policymaking in the Era of Artificial Intelligence*." [registration required]

Mark your calendars for the 3rd midyear event of ASA's Organizations, Occupations & Work section: Race and Racism in Research on OOW
Online April 21 starting at 1 p.m. Eastern. "Featuring @phdlt @blackcapitol @victorerikray and more." [registration required]

Deadlines

Conferences
New Languages for NLP Building Linguistic Diversity in the Digital Humanities
Workshop-I is June 21-25. Deadline for submissions is April 30.

Education Opportunities
Registration for our Data Science for Everyone course is now open!
"Data Science for Everyone is the flagship undergraduate course of the **NYU Center for Data Science**." Deadline is May 27 for NYU students and May 12 for non-NYU students.

Tools & Resources
Using Machine Learning and Qualitative Interviews to Design a Five-Question Women's Agency Index

National Bureau of Economic Research, Working Paper; Seema Jayachandran, Monica Biradavolu & Jan Cooper from Apr 04, 2021

"We propose a new method to design a short survey measure of a complex concept such as women's agency. The approach combines mixed-methods data collection and machine learning. We select the best survey questions based on how strongly correlated they are with a 'gold standard' measure of the concept derived from qualitative interviews."

[Shedding light on fairness in AI with a new data set](#)

Facebook AI from Apr 8, 2021

"A new, unique data set called Casual Conversations, consisting of 45,186 videos of participants having nonscripted conversations. It serves as a tool for AI researchers to surface useful signals that may help them evaluate the fairness of their computer vision and audio models across subgroups of age, gender, apparent skin tone, and ambient lighting."

[Denison faculty publish new data science textbook](#)

Denison University, News from Apr 7, 2021

"Mathematics and Computer Science faculty **Thomas Bressoud** and **David White** have written a new data science textbook, titled *Introduction to Data Systems* and published by **Springer Publishing**."

Careers

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

[Director, American Family Insurance Data Science Institute](#), University of Wisconsin; Madison, WI.

[Assistant Professor - Critical Cartography](#), Oregon State University, College of Earth, Ocean and Atmospheric Sciences; Corvallis OR. **Apply by May 1.**