

Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for December 7, 2020

Please let us ([Micaela Parker](#), [Steve Van Tuyl](#), [Laura Noren](#), [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](#) .

Academic Data Science News

It's college testing season. U.S. college prep is [mostly the same](#) (despite a growing number of [test center closings and cancellations](#)) but so much of college admissions and decision making has changed. The SAT is this weekend; ACT next weekend. Most colleges aren't asking for test scores on applications. (The [exception](#): Florida's public universities.) Still, most pre-college students plan to take the test, believing good scores will help with admissions and scholarships.

The sturdy norm of American teenage achievement testing runs counter to this past summer's awakening to systematic inequality, according to **The Brookings Institution**, which recently studied 2020 SAT data and found that the test [reinforced, rather than reduced, race gaps](#) in college admissions.

College applicants are preferring to stay closer to home with their school choices. Forty percent fewer foreign applicants will compete for slots at U.S. colleges. A growing number of American higher education institutions are outspending revenues and/or losing enrollment, and there is [no indication that a safety net](#) will be available for students who attend a college that closes. Universities aren't helping, choosing to obscure financial difficulties, and so far government regulators have failed to demand greater transparency.

The current reality is that a large number of incoming college students [will major in Computer Science](#). At some schools that is a ticket to perfectly fine jobs in Web development and IT administration. At **Yale** it might bring about [momentary insecurity](#), when you see that the CS department is ranked #124. At top-tier research universities that also emphasize data science, formal computer studies are a pathway to research. And for them, **Computing Research Association** is offering a [Best Practices Report for Scaling Undergraduate CS Research Opportunities](#). CRA also has posted an [analysis of Future Computer Science Needs](#), adopted from recent faculty search advertisements.

Another report by **Ithaka SR** consultancy tallied the inventory of [research data services offered by U.S. colleges and universities](#). Among the findings, libraries are an essential service provider, mostly via generalist consultation. Libraries, IT and academic departments also sometimes provide specialized and/or focused data services.

Universities are in the spotlight as they ramp up efforts in COVID testing and research, both for their own students and the population in general. Arizona universities [have taken several approaches](#): **Arizona State University** developed a saliva-based test and aims to monitor the spread through frequent mass testing and **University of Arizona** is focusing on wastewater testing (more in Research News), while **Northern Arizona University** is using a hybrid testing scenario. Duke University has [found success](#) using a technique called [pooled testing](#).

Universities are not keeping their coronavirus research to themselves. **Purdue University** and **National Science Foundation** started the [Records Evaluation for COVID-19 Emergency Research \(RECOVER\) initiative](#) to mine multicenter patient datasets for insights into COVID comorbidities and alternative treatments. The **University of Texas at Austin COVID-19 Modeling Consortium** has developed a free, interactive dashboard to help families and school staff anywhere in the US estimate how many infected people are likely to show up on their campus on a given day, and **Oregon State** received a \$2 million grant from the **David and Lucile Packard Foundation** to [create a national TRACE Center](#) that will expand the OSU's COVID-19 public health project to other states. **University of Maryland** and **Carnegie Mellon** are teaming up with **Facebook** to [analyze over 30 million survey results](#) from around the world that will help to track COVID spread, **Northeastern University** is [analyzing COVID test swabs](#) donated by students and faculty, while **University of Colorado** researchers are [gathering data on COVID survivors](#). If there were such a thing as a list of silver linings to the pandemic, the opportunity for our substantial university research capacity to shine would be one of them.

New York University will now offer a joint Bachelor's degree in Data Science and Mathematics. [The program](#) is a collaboration between **NYU Center for Data Science** and **NYU Courant Institute's Department of Mathematics**, and it will begin next semester.

NYU Center for Data Science has also [partnered](#) with the **National Society of Black Physicists** to offer CURP, the NYU CDS Undergraduate Research Program.

MIT Schwarzman College of Computing is [adding research centers](#) to its research centers. The Center for Exascale Simulation of Materials in Extreme Environments (CESMIX) will be based at the college's Center for Computational Science and Engineering (CCSE).

The **University of Vermont** [will be eliminating](#) 12 undergraduate majors, 11 minors and four Master's programs, including the school's program in Vermont Studies. Students looking to focus their interdisciplinary studies on cheese, snow, dirt roads and craft beer will have to look elsewhere.

Tufts University School of Engineering, aided by **Noodle Partners**, will launch [online graduate programs](#) in data science and in computer science.

Montana State University, the **University of California-San Diego**, **Drexel University** and the **OCLC** library cooperative [launched](#) the LIS Education And Data Science Integrated Network Group (LEADING) to build data science capacity among university libraries.

The **University of California-Santa Barbara** is reporting [a substantial gift](#) from **Amazon** that will support researchers and graduate students in the Department of Computer Science, the Department of Statistics and Applied Probability and the new **UCSB Center for Responsible Machine Learning**.

The **State of Illinois** [has released \\$142 million](#) for a range of university research projects under the state's **Illinois Innovation Network**. \$98 million goes to **University of Illinois at Chicago** for its computer design research and learning center. \$23.5 million goes to build the **University of Illinois Discovery Partners Institute** headquarters in Chicago's South Loop. \$14 million goes to design the new data science center at **University of Illinois** in Champaign.

University of Toronto's Schwartz Reisman Institute and **AI Global** kicked off [an effort to establish globally recognized safety and security standards](#) for trustworthy artificial intelligence technologies, starting with a certification mark that indicates responsible and trusted use of AI.

Pacific University in Stockton, California, is introducing a [hybrid \(in-person and online\) Master of Science degree](#) in Computer Science.

Editor's Picks

The trends, predictions and evidence for sea level rise are, in a word, [worrisome](#), and set to grow more so with the [recent launch of Sentinel-6](#)

[Michael Freilich](#), a high-precision Earth observation satellite, by **NASA** and the **European Space Agency**.

Tara Houska—Zhaabowekwe, a Couchiching First Nation Ojibwe and founder of **Giniw Collective**, recently wrote in **Vogue** magazine, [What the Climate Movement Can Learn From Indigenous Values](#). Her essay is a counter-narrative to "The story of our self-destruction and what to do has been mostly told in cold, statistical analysis recited by a handful of mostly male, mostly non-POC, almost entirely non-Indigenous voices." The essay brings to mind the famous [1950 Nobel Prize Banquet Speech](#) by author **William Faulkner** who surfaced human values – compassion, sacrifice, endurance – as touchpoints for writers and poets seeking evidence that humans will prevail despite the threat of inevitable thermonuclear conflict hanging in the air. Houska's touchpoint comes from "the language of the land" and is more industrious than uplifting, "I listened to the softly rustling rice and lifted up my cedar knockers, as my ancestors have for millennia. There's work to be done." The [history of climate action](#) is a timeline of rising CO₂ and of high-level global climate threat testimonies, conferences and agreements. "The line going up doesn't mean we've failed to make progress," **Simon Evans**, an environmental journalist, tweeted, "It just means we haven't done enough." No need to fear environmental Armageddon. We just need to do the work.

AI may be notable for many things, but it is infamous for the ongoing instances of inherent bias that have been built into too many AI-driven tools. At least now some attention is being paid. [Research is being conducted](#) on fairness and bias by some of the same entities guilty of AI-ethics transgressions such as **Google**, **Amazon** and **Microsoft**, and a recent feature in **Nature** described many of these instances that [haunt facial-recognition research](#), presenting results from a survey of facial-recognition researchers on ethical questions. While we seem to have only gotten as far as requiring authors to attach statements addressing ethical concerns and potential negative outcomes from their work, on the plus side the needle is moving (albeit slowly) in the right direction as far as diversity in the data science field. Supporting evidence can be found in [9 Black Women That You Should Know in Data Science](#). Breaking news: On the evening of Wednesday, December 2, Timnit Gebru, the co-lead of Google's ethical AI team, [announced](#) on **Twitter** that the company had forced her out, ostensibly because of her disregard for internal publication review protocols. Several other **Google** researchers have [gone on record](#) disputing this reasoning. As of this writing, it [remains unclear](#) what Google's actual reason for terminating her was and what their next steps will be. Sadly, the aforementioned moving needle seems to have gotten stuck.

Research News

Since March, the **University of Arizona Water & Energy Sustainable Technology (WEST) Center** has been using the implementation of wastewater-based epidemiology (WBE) to help communities stop the spread of SARS-CoV-2 by testing their wastewater. In August, [the WEST Center team started testing on campus as students began to return](#), quickly stopping at least one dorm outbreak and becoming the first US campus to use this valuable tool to keep students and staff safe. **University of California-San Diego** was one of the first universities to announce [they would be using wastewater testing](#) as a way to reduce COVID-19 spread and have since greatly expanded their use of wastewater testing on campus.

Researchers at **Arizona State** have been doing [pioneering work on wastewater-based epidemiology](#). (An ASU spinout, **AquaVitas**, recently [signed a contract](#) with **Federal Health and Human Services** for national-scale testing.) By analyzing wastewater, researchers are able to detect whether SARS-CoV-2 RNA is present in a sample and, importantly, at what concentration. Because infected individuals shed the virus prior to showing symptoms, testing wastewater provides early detection of the disease. Monitoring wastewater gives public health officials and facility managers [four to 10 days of advance notice](#), while infected individuals are still asymptomatic, to take steps to reduce the spread of COVID-19. And it can detect the genetic signal of the virus in sewer water generated by a discrete segment of the community, such as a university, public school or factory. At the **University of Nevada Las Vegas**, researchers are using wastewater testing to look for [new strains of the virus](#).

The proliferation of wastewater testing has also produced many opportunities for collaborations between government and universities. ASU's **Rolf Haden** and AquaVitas are [advising New York City](#) on testing. The **University of Wisconsin Laboratory of Hygiene** working with the **Wisconsin Department of Health Services**. Wisconsin is one of five founding states in [a national surveillance system](#) funded by the **Center for Disease Control and Prevention**. **Michigan Department of Environment, Great Lakes, and Energy (EGLE)** is providing \$800,000 to expand [an ongoing COVID-19 detection program in wastewater](#), a partnership with **Michigan State**; **University of Kansas School of Engineering** has been [taking weekly sewage samples at both of the City of Lawrence's wastewater treatment plants](#) to test for components of the virus, **North Dakota State University** undergrads are conducting wastewater testing for local communities and several Canadian universities are preparing to test wastewater from long-term care homes in Ottawa, Toronto and Edmonton.

Some cities in North Dakota have [declined to participate](#) in wastewater testing. A heads-up, the **National Institutes of Health** currently is [seeks bids from private companies](#) to test up to 30% of the country's municipal wastewater treatment plants, and expects to award a total of US\$19 million.

High resolution imaging combined with machine learning to monitor, inspect or otherwise scrutinize [industrial](#), [environmental](#) and [social](#) phenomena needs its own word, a shorthand for identifying what computer vision is up to that isn't too jargon-y. A human intelligence that visually recognizes baseball pitches in a strike zone is called an umpire. Umpires are sometimes good, sometimes bad at their job, but we don't have to always explain what umpires do when they come up as a subject. One thought, the "covpire" – short for computer-vision-umpire. In a sentence: Deploying the widget assembly covpire will save the company billions!

If Earth-bound fatigue has set in, it may be energizing to look at [precise maps of millions of bright quasars](#), which apparently show our place in the cosmos as never before, as well as the new Gaia data revealing the "[best-yet map of the Milky Way](#)." One of those stars – TYC 2597-735-1 to be precise – and its surrounding blue ring has been a mystery – until now. Astronomers believe they have solved the 16-year-old mystery of a [star surrounded by an enigmatic ring of ultraviolet light](#). "Blue Ring Nebula" offers a glimpse of the transition from a merged pair of stars to the weird stars seen millions of years later and is considered the "Rosetta Stone" of that process. The star and its ring were first discovered by the Galaxy Evolution Explorer (GALEX), a **NASA** mission that ended in 2013. Back to the calamitous Earth, where [the announcement](#) came from **National Science Foundation** that the instrument platform of the 305-meter telescope at **Arecibo Observatory** in Puerto Rico fell, resulting in damage to the dish and surrounding facilities. An on-site 12-meter radio telescope will be repaired and the **University of Central Florida** will continue as research manager, but the giant telescope will be decommissioned, a loss to geospace researchers and, less well known, an irreplaceable loss to [terrestrial defenses against collisions](#) with giant space rocks.

The **AlphaFold group at DeepMind** have realized a landmark achievement in computational biology, [predicting how proteins fold](#) in 3-dimensions based on the amino acid sequences coded in DNA.

Tweet of the Week

Twitter, Katie Siek from December 03, 2020



Katie Siek @katiesiek · Dec 3

...

Thank you @timnitGebru for continually doing the right thing and forcing tech to look in the mirror and re-examine where we need to do better. Thank you for ensuring underrepresented groups have a voice and we attempt to decrease data violence.



Google's Co-Head of Ethical AI Says She Was Fired for Email

Timnit Gebru, a co-leader of the Ethical Artificial Intelligence team at Google, said she was fired for sending an email that management

[bloomberg.com](https://www.bloomberg.com)

48

94

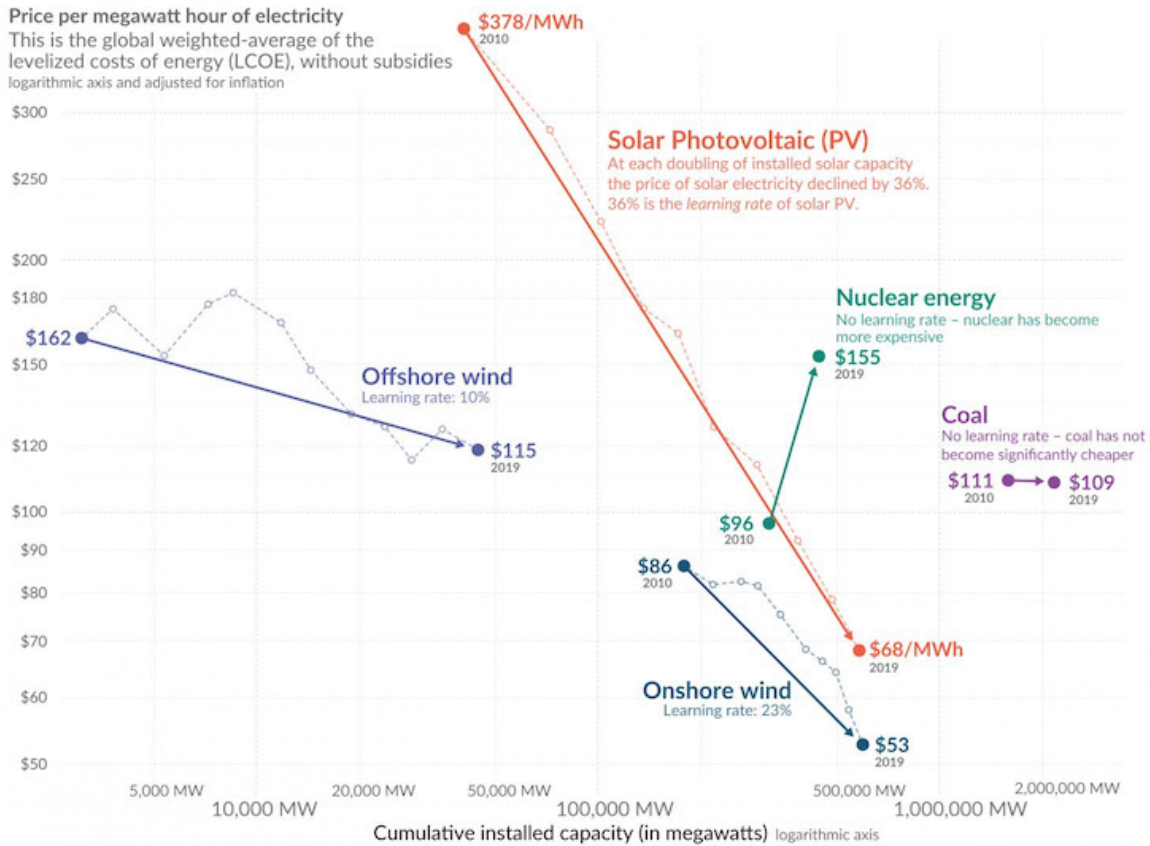
253



Data Visualization of the Week

Our World in Data, Max Roser, and Twitter, Hannah Ritchie from December 01, 2020

Electricity from renewables became cheaper as we increased capacity – electricity from nuclear and coal did not



Source: IRENA 2020 for all data on renewable sources; Lazard for the price of electricity from nuclear and coal – IAEA for nuclear capacity and Global Energy Monitor for coal capacity. Gas is not shown because the price between gas peaker and combined cycles differs significantly, and global data on the capacity of each of these sources is not available. The price of electricity from gas has fallen over this decade, but over the longer run it is not following a learning curve.

OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY
by the author Max Roser

Events

2020 John Jay Iselin Memorial Lecture: The Atlantic's Ed Yong

Online December 7, starting at 5 p.m. "COVID-19 has caused a staggering 1.26 million deaths worldwide. Nearly a year to the date of when the virus was first reported internationally, science journalist **Ed Yong** speaks about the global pandemic as part of the 2020 John Jay Iselin Memorial Lecture." [registration required]

The Future of COVID-19 Epidemiology

Online December 8, starting at 4 p.m., organized by **Radcliffe Institute for Advanced Study at Harvard University**. "To what extent is our future with COVID-19 knowable? As new information about the transmission, demographics, and treatment of COVID-19 emerge, epidemiologists continue to address complex data and generate new predictive models to better understand the dynamics of the virus. Join leading epidemiologists for a panel

discussion as they assess the current and future state of the epidemic." [registration required]

Stanford Big Earth Hackathon Takes on Wildfires

"As severe as the wildfires have been this year, they might have been even worse were it not for an amazing event at **Stanford**: the 2020 Big Earth Hackathon. Now in its third year, the annual hackathon has teams of 1–4 Stanford students compete to develop data-driven solutions for important environmental use cases. Projects are judged on several characteristics, including defining the specific problem to be addressed, project impact, degree of completion, and quality of presentation. The core of the Big Earth Hackathon is the creativity and innovation of the solution." Deadline for applications to participate is December 19.

COVID19 Precision Immunology App-a-thon

"**precisionFDA** calls upon the scientific and analytics communities to develop innovative and user-friendly approaches to improve the understanding of the relationship between personalized immune repertoires and COVID disease-relevant phenomena and their associated external factors by accurate interpretation of rapidly shared datasets in public databases, such as iRECEPTOR Gateway. This precision immunology app-a-thon phase I will launch on Nov. 30, 2020 and end on Jan. 29, 2021." [sign up required]

Deadlines

Contests/Award

Riiid AIEd Challenge

The inaugural AIEd Challenge is "inviting leading global AI talent to compete on Google's Kaggle competition platform to create deep learning algorithms for education." Winning teams will share \$100,000 in prize money. Deadline for entries is December 31, h/t **Craig Smith** from [Eye on AI](#) podcast.

Education Opportunities

DATA STEWARDSHIP - Developing a Data Reuse Strategy for Solving Public Problems

"This program will help data stewards—responsible data leaders in the public and private sector—develop a data re-use strategy to solve public problems. Noting the ways data resources can inform their day-to-day and strategic decision-making, we teach participants ways they can use data to improve how they operate and pursue goals in the public's interests. By working differently—using agile methods and data analytics—public sector and private sector leaders can further open data and reduce data access inequities in ways that advance their institution's goals." Deadline for applications is December 17.

The FINAL ROUND of applications for hackNY's 2021 Summer Fellowship starts today!

"College students - get your applications in before December 30th to be considered for your summer of community, learning, and fun."

RFP

NSF Launches Minority-Serving Institutions Research Expansion Program

"The Computer and Information Science and Engineering Minority-Serving Institutions Research Expansion Program (CISE-MSI Program) is now open for solicitation. With this solicitation, the **National Science Foundation's (NSF) Directorate for Computer and Information Science and Engineering (CISE)** is launching a new, focused program to support research expansion for Minority-Serving Institutions (MSIs). The goal of the CISE-MSI program is to broaden participation by increasing the number of CISE-funded research projects from MSIs." Deadline for proposals is April 15, 2021.

Tools & Resources

NEW from a big collaboration at Google: Underspecification Presents Challenges for Credibility in Modern Machine Learning

Twitter, Alexander D'Amour from November 09, 2020

"Explores a common failure mode when applying ML to real-world problems."
Thread 1/14

We've released 1.2 million reviews in 6 languages in the Multilingual Amazon Reviews Corpus (MARC)!

Twitter, Phillip Keung from November 24, 2020

"w/ **Y. Lu G. Szarvas @nlpnoah** <http://arxiv.org/abs/2010.02573>"

We are pleased to announce the first official release of CAMEL Tools

Twitter, CAMEL Lab from November 16, 2020

"An open-source Python Arabic NLP toolkit for pre-processing, morphological modeling, dialect identification, named entity recognition, and sentiment analysis."

Faceted Search with Solr

LucidWorks, Yonik Seeley from November 27, 2020

"Faceted search has become a critical feature for enhancing findability and the user search experience for all types of search applications. This article gives you an introduction to faceted search with Apache Solr."

Careers

Tenured and tenure track faculty positions

Tenure-Track Faculty Position in Computation, Equity, and the Health of the Planet

Massachusetts Institute of Technology, Department of Urban Studies and Planning; Cambridge, MA

Click [here](#) to receive the Data Science Community Newsletter and/or to have us follow your twitter feed so that our data science twitter bot can easily grab links from your tweets.

Data Science Community Newsletter Issue 208