



ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for March 1, 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](#).

Monday March 8 is the date for the annual [Women in Data Science conference](#) organized by Stanford University (WiDS). This year the event is virtual, global and will run for 24 hours straight, incorporating keynote presentations, speakers and panels from around the world. The [4th Annual WiDS Datathon](#) is currently open on Kaggle and will run until March 1. The first-time [WiDS Worldwide Workshops](#) offer 14 different skill-building sessions. Separate registrations (and in some cases, fees) are required for the conference, the datathon and the workshops.

Regional events taking place on March 8 are listed [at the Stanford conference website](#). WiDS events not planned for March 8 include [WiDS Charlotte](#) on March 4-5, [WiDS Pittsburgh@CMU](#) on March 10-12, [WiDS Cambridge](#) on March 11, and [WiDS Chicago](#) on March 12. Please check Google for WiDS events and dates in your area.

Academic Data Science News

The early post-vaccine phase of the pandemic is proving risky to universities. Spikes and clusters of COVID cases have created real scares [at the University of Maryland](#) and [at some colleges in Virginia](#). Lower key alerts are underway [at University of California-Berkeley](#) and [at New York University](#). **Dartmouth** has [shifted to grab-and-go dining](#) to temporarily replace student cafeteria dining. **Notre Dame** is [struggling to regulate students' risky off-campus behaviors](#). At this point everyone knows what the good (and bad) habits are, and campus danger levels are in-line with how well (or poorly) those habits have normalized. Emerging variants and spring breaks are wildcards, but there's a growing body of best practice literature (examples from [Science News](#), [Quanta Magazine](#)) for managing the situation.

Universities continue to drive COVID research. **Yale** scientists **William Jorgensen** and **Karen Anderson** use high-level computation and structural biology methods to [design promising COVID-19 therapeutic compounds](#). **University of Virginia Health** has new [software to monitor patients](#) with chronic disease as well as COVID-19. A team at **Notre Dame** are addressing [privacy concerns with mobile contact tracing apps](#) through funding from the **Office of the Director of National Intelligence**. **NIH** will be giving the **University of California-San Diego** \$33 million to [oversee five Rapid Acceleration of Diagnostics \(RADx\) projects](#) that address gaps in COVID-19 testing. Progress in the short run does not overshadow the [long run anxiety and mental consequences](#) experienced by college students. Hang in there everyone. Take care, stay safe, good habits.

Funding sources for computational social science are increasing their largesse. **Data.org** and the **Rockefeller Foundation** [spread \\$10 million across eight global projects](#). Academic groups from **Kansas State**, **University of California-Berkeley**, **Aalborg University** (Denmark), **EMPA** (Switzerland) and **University of Melbourne** (Australia) earned the [data.org Inclusive Growth and Recovery Challenge grants](#). **Cornell** researcher **J. Nathan Matias** [recently earned grants](#) totaling \$1.3 million from **MacArthur Foundation** and **Templeton World Charity Foundation** for his Citizens and Technology (CAT) Lab. **Twitch**, a user-generated live Internet entertainment company, filled its [second class of Research Fellows](#) with PhD students from **University of Washington**, **University of Maryland**, **University of Texas at Austin** and **Princeton**. Computational social science research has been fruitful of late,

making noteworthy progress in [cyber bullying](#), [social segregation in cities](#), and [police bias](#). Still, researchers can point to need gaps where data-informed social science would immediately help public policy, [risk data for natural disasters in rural areas](#) for example.

Duke University data science professor **Cynthia Rudin** [has a warning](#) for college students who are curious about "The Marriage Pact," a dating app that originated at **Stanford** and has exploded virally at other U.S. universities.

Humboldt State University in California is taking a [close look at becoming the State University system's third Polytechnic](#), joining **Cal Poly San Luis Obispo** and **Cal Poly Pomona**. Humboldt student enrollments have declined precipitously since 2015.

Laptop theft is a campus plague. Federal authorities recently [charged a 3-person crime ring](#) who allegedly stole at least 800 **Stanford University** laptops with a total value of \$2.2 million.

Pioneering nanotechnologist **Paul Alivisatos** is the new President of the **University of Chicago** and [will start the job this summer](#). Alivisatos graduated from Chicago in 1981 with his B.A. in Chemistry.

Yale University [announced](#) the **Wu Tsai Institute** for interdisciplinary neurocognition research. Psychology professor **Nicholas Turk-Browne** is the inaugural director. The institute targets two strategic priorities for Yale, neuroscience research and integrative data science.

The **University of Miami** (Florida) [received \\$12 million in grants](#) from the **Knight Foundation** and from **Phillip and Patricia Frost**, earmarked for the school's **Institute for Data Science and Computing**. The \$4.3 million Knight contribution will endow six chaired professorships at IDSC.

The **University of Pennsylvania** [showed architectural plans](#) for the school's \$117 million SEAS Data Science Building. It should be complete in Summer 2024.

The new Masters program in Data Science at the **University of Colorado** [is set to begin](#) in May for online students, and in the upcoming Fall term for residential

students. **DePaul University** in Chicago [announced](#) its new Masters in Data Science, and it has been taking students since the start of the year.

Government sources in the UK [are funding](#) the multi-university **UK Centre for Greening Finance & Investment** led by **Ben Caldecott**, finance professor at **University of Oxford**. The vision for CGFI, according to Caldecott, is "the greening of finance and the financing of green." Yup, lots of green.

The **University of Connecticut** will [open a data science technology incubator](#) at the school's Stamford branch campus.

Facebook is [donating \\$1.5 million](#) to the **University of California-Santa Barbara** to research data center efficiency. **Facebook AI Research** also [announced a formal affiliation](#) with **University College London's AI Center**. The planned four-year partnership is similar to relationships that FAIR has with other universities in France, Canada and the U.S.

NOAA and the **University of Southern Mississippi** have agreed to [a ten-year plan](#) to improve collection of environmental data with unmanned systems.

The **University of Wyoming** led a consortium that also included **University of Nevada-Reno** and **University of Montana** that received a \$6 million NSF grant to model ecological and biological systems in the Western U.S. The **West Big Data Innovation Hub** [helped to coordinate](#) the collaboration.

SPONSORED CONTENT

2i2c

Chris Holdgraf, Director of the **International Interactive Computing Collaboration (2i2c)**, presented to the **ADSA Data Science Education SIG** on opportunities for interactive computing in higher education. Chris covered the history and vision of 2i2c and answered participant questions about the organization and how it plans to build communities around interactive computing in academia. The one hour recorded session is [now available](#) on YouTube.

Editor's Picks

The New Yorker magazine has [a longread article](#) on AI ethics that raises lots of issues, and at times [lacks for sensitivity](#). The author, **Matthew Hutson**, provides a tourist's overview of the current state of ethical AI in practice, reporting from multiple conferences (NeurIPS, AAI, CHI, EMLNP) and interviewing leading researchers (**Michael Kearns, Alex Hanna, Katie Shilton, Eric Horvitz, Brent Hecht**). Hutson hits big questions (cultural stereotypes, biased outcomes, ethics training, peer review, institutional review, commercialization, personal impact statements, regulation). These are powerful forces at work, some pushing while others pull. Lowest common denominators (like training data sets, user populations, interface affordances) push the work forward broadly. The pull is toward what's exceptional in terms of what's good and what's bad, mainly through peer review and social backlash (and sometimes both). The hassle that comes with navigating the push-pull is discouraging researchers and disincentivizing government regulation, according to Hutson. It's also [putting off companies from developing products](#), according to another article in **Fortune**. Medical companies were hoping to incorporate AI into useful tools against COVID but the [real world impact has been marginal at best](#), according to **Mohana**

Ravindranath of **POLITICO**. It's healthy and it's also unpleasant to have all of AI ethics' dirty laundry on display, but let's keep hope alive. There's money from **National Science Foundation** and **Amazon** ([recent awards](#)) for [new research into Fair & Trustworthy systems](#). "We need to understand the different aspects of fairness," NSF program director **Henry Kautz** told **Nextgov**, "so that we can in turn understand how we can design our systems with 'fairness' built into them."

"[What is going on at Google AI?](#)" has become an increasingly urgent question as a second member of the Google AI ethics team – **Margaret Mitchell**, a computer scientist who studies algorithmic bias – has been fired. (The first, **Timnit Gebru**, who studies AI facial recognition bias, [was fired](#) in December 2020.) Why would Google create an ethics team to monitor its own substantial AI endeavors, and then fire two of the team leaders for doing exactly what they were hired to do – identifying potential internal ethics violations? (Hint: Might have something to do with Google being for-profit.) Adding to the substantial Google missteps is the apparent [mishandling of Tech Exchange](#), a program meant to address the "pipeline problem" by bringing Black college students to the Mountain View campus for exposure to tech careers. **Michaeli Graham** does an excellent job [connecting these threads](#) at **Afrotech**, an excellent Black in Tech news and community resource.

We're eyeballing intellectual property's role in data science since noticing that the **U.S. Patent and Trademark Office** has seen the number of patents for artificial intelligence [double in the last twenty years](#). Two large international banks, **Barclays** and **TD Bank Group**, [recently joined Open Invention Network](#), a consortium that legally defends open source code bases against abusive patent assertion lawsuits brought by patent trolls. Journal publishing is often the public disclosure trigger for the one-year patent filing clock in the U.S., as well as its own IP quagmire of copyright and license protections. **Public Library of Science** (PLOS) announced [an agreement](#) with **Uppsala University** to provide flat-fee top-tier journal publishing without article processing charges. PLOS has similar agreements in place with the **University of California System**, the **Big Ten Academic Alliance** and the **Canadian Research Knowledge Network**.

Research News

You know our experience of living with COVID has reached a new stage when the research turns to examining the research itself. **Global.health** is a [just-launched depository for anonymized COVID case data](#) funded by **Google** and the **Rockefeller Foundation**. There are studies about the [rise of remote clinical monitoring](#); the [effectiveness of governmental interventions](#) in the fight against COVID; a study of [contact tracing app effectiveness](#); how a COVID mortality model developed by a young man living in his parents' basement could [outperform all other models](#); opinion pieces on the failures [of COVID-19 data efforts in general](#) and [of the U.S. Centers for Disease Control in particular](#); and – perhaps surprising to some but not to sports fans – how the **NFL's** COVID data efforts [successfully addressed](#) those failures. New problems are also arising such as [data privacy issues with “vaccine passports”](#) as well as [increasing concern over virus variants](#) (from another trove of sports team data) plus ongoing [unanswered questions about zoonotic infections and the birth of COVID-19](#) and [whether or not schools can safely reopen](#).

Smartwatches, especially the ones with GPS and heart rate sensors, have become full-blown computers at this point with apps, customization and peripherals. Studies are underway using watch data to [identify worsening heart failure](#) and [diagnose COVID-19](#). Non-bracelet sources of health tracking are coming: patch-like wearables using [microneedle stabbers](#) and common public sampling devices like [smart hand sanitizers](#). Health tracking data issues loom around [personal privacy](#) and, potentially, [structural racism](#). Combining and using so much data is a new approach to medicine and **Stanford's Micheal Snyder** has taken to [calling it "multiomics."](#) In practicality, new business models and services are arriving to help health consumers to literally [navigate the American health landscape](#), providing needed patient advocacy in a system with more and more touchpoints, options and care coordination.

England's **Royal Society** has issued [a report](#) charting the path to carbon emission reduction through the use of existing digital technology, both to quantify progress and by using data to monitor the sector's own energy consumption. Here in the U.S. our own state of **Texas** could make good use of this kind of approach, as last week Texans learned the hard way [how systems will be tested](#) as the climate changes. A simple way to visualize the change is [comparing the number of record temperature highs to record lows](#) – the former now double the number of the latter, as **Gavin Schmidt** demonstrates.

The 20th Anniversary of the Human Genome Project [harkened back to a simpler time](#) for computational biology, before machine learning became essential for predictive advances in systems biology ([this](#) from researchers at **Stowers Institute for Medical Research** and **Stanford**) and structural biology ([this](#) from **UK Research and Innovation** and **DeepMind**). Machine learning continues to drive advances in basic life science like [classifying biochemical reactions](#) and [decoding cellular signals](#). Applied life science is also benefitting with [realtime tissue identification](#) under microscopes and solutions to [bacteria responsible for hospital infections](#). While the computational life science has been glorious, [data sharing and maintenance have suffered post-HGP](#); it's proven difficult to build easy access systems for mega-scale scientific data that are subject to evolving types, norms and profit motives.

Prague's **Svanda Theatre** and London's **Czech Centre** are premiering a [play written by artificial intelligence](#). (Actually, it's [eight scenes](#) written separately.) *AI: When a robot writes a play debuts* in Prague with live streaming on February 26. *Star Wars Episode X* is [another new 100% AI script](#). The stream of the live table read, filmed by **Blue Bantha Milk Co.**, is [online](#). "Weirdly Entertaining," says **Adam Bentz**, critic at **Screen Rant**.

Rodney Brooks shares the [origin story](#) for Robot Arm Programming Languages, some of it from his grad student years at MIT.

Precision farming via drones, digital imaging, machine learning and cloud computing is a path to global food security. New systems developed by researchers **Manchester Metropolitan University** (UK) and **University of Malaya** (Malaysia) are [improving small farms' rice cultivation](#) in developing countries. Also, first world developers at UK consultancy **Monolith AI** have found the perfect recipe for consistently [fluffy pancakes](#).

Tweet of the Week

Twitter, Sarita Schoenebeck from Feb 20, 2021

← Tweet



Sarita Schoenebeck
@syardi



MIT just announced guaranteed transitional funding for graduate students who wish to leave an unhealthy advising relationship. This is a great idea.
rise4mit.medium.com/rise-update-on...

2:48 PM · Feb 20, 2021 · Twitter Web App

912 Retweets 209 Quote Tweets 6,614 Likes



m @MPoirier1976 · Feb 20
Replying to @syardi



Isn't that an indicator of disease in the system?

1



21



Data Visualization of the Week

Twitter, The Institute for Ethical AI & Machine Learning, and the Linux Foundation, from Feb 16, 2021

The 8 LFAI Principles for Trusted AI – (R)REPEATS

Reproducibility

Robustness

Equitability

Privacy

Explainability

Accountability

Transparency

Security

Events

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

7th annual Analytics Frontiers Conference, University of North Carolina-Charlotte

Online April 22-23. "The Analytics Frontiers Conference is the largest data science conference in the region." [registration required]

The Learning Salon, a forum in which we explore bridges and contentions in biological and artificial learning.

Online March 5 starting at 4 p.m. Eastern. Speaker: **Randall O'Reilly**. [rsvp required]

HAI's 2021 Spring Conference Intelligence Augmentation: AI Empowering People to Solve Global Challenges

Online March 25, starting at 9 a.m. Pacific. [registration required]

Deadlines

Conferences

2021 NBER Economics of AI Conference: Call for Papers — Economics of AI

September 23-24. Deadline for submissions is May 31.

Tools & Resources

Designing for Voice User Interfaces — Principles and Best Practices

Snappymob blog, Zoe Chin from Jan 18, 2021

"How are voice-first interfaces designed differently from screen-first interfaces? We explore the ins-and-outs of UX in voice design."

Uncovering Unknown Unknowns in Machine Learning

Google AI Blog, Lora Aroyo and Praveen Paritosh from Feb 11, 2021

"To address the problem of [ambiguities] in ML models, we recently launched the Crowdsourcing Adverse Test Sets for Machine Learning (CATS4ML) Data

Challenge at HCOMP 2020 (open until 30 April, 2021 to researchers and developers worldwide). The goal of the challenge is to raise the bar in ML evaluation sets and to find as many examples as possible that are confusing or otherwise problematic for algorithms to process. CATS4ML relies on people's abilities and intuition to spot new data examples about which machine learning is confident, but actually misclassifies."

The second edition of "Modern Data Science with R" is coming out in March

Twitter, Nicholas Horton from Feb 15, 2021

"the bookdown version is available now!"

Careers

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

Krishna A. Bharat Chair in Computational Journalism, Georgia Institute of Technology

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.