

ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for June 2, 2022.

Please let us ([Micaela Parker](#), [Catherine Cramer](#), [Brad Stenger](#), [Laura Norén](#)) 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](#).

ANNUAL DSCN READER SURVEY

Last year's annual survey helped us shape the DSCN. We listened and adjusted. We dropped the Tweet of the Week because readers stopped liking it. We send out every other week; that's the sweet spot for desired frequency. We dialed up our editorializing – people liked a little opinion to situate the stories – and subsequently dialed it back – but not **that** much. We are planning a remodeling of the look-and-feel this summer. We want to hear from you as we plan changes and upgrades for the next 12 months. Have we hit the goldilocks zone?

[Take the survey!](#)

SPONSORED CONTENT



ON LIBRARY CLOSURES

The **University of Illinois'** "quirky" library for undergrads **has just closed** after 53 years. Losing old libraries is, for me, like losing an old church might be for a more religious person. The architecture bears the spirits and thoughts of so many who came before that to tear it down is symbolic violence, no matter what comes next. At UIUC, what comes next is a revamp of the campus's Main Library with more mixed uses in mind. Am I the only one who wishes there were a way to preserve the reading, writing, and printed matter storage component of libraries then build all the new, cool mixed-use facilities, too? Keep the old libraries as they are and simply build new resource centers for all the collaboration rooms, maker studios, reservable classrooms, and whatever else comes next for sharing knowledge?

CT SCAN OF A FRUIT FLY -> SUPER-COOL DEEP LEARNING

Imagine taking a CT scan of a fruit fly. Combine that with video data of fruit flies in motion. Feed it all into a deep learning model to **create a simulated fruit fly**. That's what **Pavan Ramdya, Auke Ijspeert, Jonathan Arreguit, Victor Lobato Ríos, Shravan Tata Ramalingasetty, and Gizem Özdil** at EPFL have done. Why? Accurately modeling a living organism from biomechanics and video data is a super-useful deep learning holy grail, one that has received far too little coverage, including by us. Imagine all that we can know if we are able to simulate organisms based on that type of data?

Linking behavior and biomechanics computationally suggests potentially

unsavory applications such as being able to identify individuals from streams of passersby based on gait or other gestures. Employers might want to calculate productivity in packing/shipping warehouses and for delivery drivers, baristas, food prep workers. Some **grocery stores** are already using gesture-prediction models to detect people hiding things like meat, cheese, and medicine to shoplift those items. This application definitely feels wrong – if people can't afford food and medicine, maybe they should be fed and provided basic health care, not kicked to the curb.

CLEARVIEW AI GETS TORCHED BY THE BRITS

The deeply unethical **Clearview AI** company that has scraped pictures of just about everyone off social media sites without our consent, then trained a classifier to identify all of us lovelies with facial recognition, and sold the ability to look up anyone's identity with a photo to law enforcement (and anyone else willing to pay and stoop that low) has been **fined** \$9 million in the UK. The Brits have also demanded that Clearview delete all images of British citizens and halt any practice that would allow Clearview to collect new images of British citizens. **Oscar Wilde** popularized the aphorism "imitation is the sincerest form of flattery," which was **originally penned** by the Brit **Charles Caleb Colton**. Perhaps other countries may wish to flatter the Brits and trounce Clearview AI with the same gusto as if trolloping Clearview AI had been their idea all along.

SPONSORED CONTENT



The image is a promotional banner for NJIT's Institute for Data Science. At the top left is the NJIT logo, and to its right is the text "THE INSTITUTE FOR DATA SCIENCE". Below this, a large blue rounded rectangle contains the text "Transforming the vision of technology to solve real-world problems in the modern data-driven economy" in a white, italicized serif font. At the bottom of this rectangle, a dark blue bar contains the text "Learn More: datascience.njit.edu" in white sans-serif font. The background of the banner shows a city skyline over water.

FACULTY LEAVING THE ACADEMY

According to **Karen Kelsky**, a career coach who has been working almost

exclusively with academics for a dozen years, "The overwhelming narrative is that people are happier once they leave academia." Closing out this particular academic year we have seen **a lot of career updates**. Industry compensation is generally higher than compensation in the academy for fewer hours per week. Even tenured faculty are moving to industry positions. Some institutions are laying professors off, others are seeing faculty unwilling to continue to shoulder increasingly heavy workloads while their compensation is cut. At **Goldsmiths**, part of the **University of London**, 20 positions were eliminated and pensions were cut by 38% for those who remained. Unfortunately, this reshuffling will most likely *not* result in openings for young scholars to move into tenure-track careers. Instead, there may be openings for contingent teaching and research positions. Those who remain working in academia will shoulder even heavier service commitments, having the knock-on result of pushing more to take industry positions, especially in computer science, data science, and computational social science where it's relatively easy to make the jump.

REMOTE WORK POLICY MATTERS

Speaking of people changing jobs, **Ian Goodfellow**, who oversaw machine learning and AI at **Apple** for a few years, **has quit** over the return to office policy. He would have been forced to work Mondays, Tuesdays, and Thursdays on an Apple campus. Instead, Goodfellow has rejoined Alphabet's **DeepMind** where he can apply to work remotely as part of Alphabet's more open-minded organizational structure. [Apple has since delayed its enforcement of the 3-day per week return to office expectation.]

Pew finds that **workers** are still clocking at least 30% of their time at home in 39 US states. And a new study by **Stanford University** researcher **Nicholas Bloom** suggests that workers who work from home are about **9% more productive**.

INFLATION HITS HARD; YOUNG ALUMS HAVE REGRETS

Average tuition for the US's top 20 schools **will go up** 3.7% to about \$58,396 which doesn't include room and board. It makes **President Biden's current plan** to cancel \$10,000 of existing student loan holder's debt (if they make less than \$150,000) look like barely enough to tread water. Of course, that 3.7% won't keep pace with inflation and it puts a lot of pressure on faculty who are seeing small or non-existent wage increases and higher service loads due to the continuation of hiring freezes.

A new **crowdsourced data set** suggests problems are even worse for grad students: "just 2% of the 178 institutions and departments in the data set guaranteed graduate students salaries that exceed the cost of living." PhD students are **demanding wage increases** to assuage their financial precarity.

Meanwhile **only 56%** of recent graduates (alums under 30 years old) feel that the benefits associated with getting their degrees surpassed the costs. It's difficult to raise tuition substantially in this environment, so penny pinching is expected to continue for campus employees.

Featured Event



Audio Data Across Domains (AudioXD) workshop for researchers who use audio data to share methods, discuss common challenges, and form teams to pursue grants together. Hosted by the **Kitzes Lab, University of Pittsburgh**. Free, in-person. Register by June 15, 2022. Event: Aug 1-3, 2022.

DATA 8 AT UC BERKELEY IS HUGELY POPULAR

The Data 8 "Foundations of Data Science" course at **University of California, Berkeley** **launched** in 2015 with an enrollment of 100. Now it's one of the most popular courses on campus with an enrollment of 1600+. The course has been copied on other campuses with **help from course instructors**. It is a testament to how quickly undergraduate can pivot into data science if supported with excellent introductory coursework. It's also a testament to the way public education is supposed to work. The introductory class is open to Berkeley students from any major. An upcoming instructors' **workshop** will help other colleges to develop curricula.

SAFEST CITIES ARE THOSE WITH MORE PEDESTRIANS

In a study designed to investigate the impact of urban design, transit mix (pedestrians, bikes, cars), and rates of crashes, those cities and areas with more pedestrians had a **calming effect** on all types of crashes, even those

that didn't involve pedestrians. Cars are the most dangerous modality, though the study pointed out that collisions between bikes and people are often under-reported so it was unable to truly estimate the relative danger of each modality. I'll never forget the time we invited an esteemed German researcher to give an urban studies talk. As she wrapped up her week visiting with our students and faculty, she stepped into the one-way bike lane and raised her arm to hail a taxi. A delivery biker riding the wrong way in the bike lane crashed into her, shattered her ankle, extended her stay by weeks, and exposed her to the US medical system. Bikes can damage peds.

U.S. STATE DEPT ALLOWS UNIVERSITIES TO SPONSOR REFUGEES

In a policy shift, the **U.S. Department of State** will **allow** colleges, universities, and other private organizations to directly sponsor refugee scholars.

Featured Job

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



Data Science Facilitator I. University of Wisconsin-Madison American Family Insurance Data Science Institute. Madison, WI (Some remote work possible).

LITTMAN STARTS NSF ROTATION

Computer science professor **Michael Littman**, of **Brown University**, **started** a rotation at the **National Science Foundation (NSF)** in the areas of AI, ML, and Human-Computer Interaction research. He is practically inviting our readers to explain to him what NSF needs to do to

better engage the data science community: "I'll be learning about what the research community needs, what NSF provides, and how to bring the two closer together!"

WILDFIRES IN THE WEST WILL BE BAD, BUT NOT AS BAD, THIS YEAR

Ronnie Abolafia-Rosenzweig and a team at **National Center for Atmospheric Research** (NCAR) have **trained a machine learning model** on data about the US west since 1984 (when satellite data was first available). The model predicts the 8th worst wildfire season on record for summer/fall 2022, which is bad, but not as bad as the last two years. The model predicts that "fires this summer will burn 1.9-5.3 million acres in the West, with 3.8 million acres being the most likely total," utilizing data that accounts for about 83% of the variability.

The **Boulder Daily Camera** newspaper thinks **a new wildfire model** from **University of Colorado** and **University Corporation for Atmospheric Research** (co-located with NCAR) will better inform the community's emergency response in case of wildfire.

PAST PEAK AGRICULTURAL LAND USE

The Earth has **passed the peak** in terms of the amount of land dedicated to agriculture. Closely tied to societal advancement, the amount of land used for agricultural purposes was a sign of human progress for millennia. Now, however, some agricultural land has been exhausted and the downsides are apparent: "expanding agriculture has been the biggest driver of the destruction of the world's wilderness," and "the wild mammal biomass has declined by 85%." Climate change will hasten this change. In India's Uttar Pradesh region, the heat wave has **slashed** mango yields by 80%.

COMMUNITY NORMS FOR RELEASING FOUNDATIONAL MODELS

Researchers at **Stanford University's Institute for Human-Centered Artificial Intelligence** (Stanford HAI) are **calling** for the development of community norms that can act as guardrails to gate how "foundational models" are released. There is currently a spectrum of release models – from totally open sourced to releasing none of the underlying training data, model parameters or code but instead releasing a descriptive paper. The team finds it "problematic for each foundation model developer to

determine its release policy independently. There are two reasons for this. First, a single actor releasing an unsafe, powerful technology could cause significant harm to individuals and society [...]. Second, precisely because the severity of the risks of foundation models are unclear, foundation model developers would benefit greatly from sharing best practices, rather than each organization reinventing the wheel and incurring the economic and social costs of rediscovering certain harms." It is difficult to argue with the first point – there is a significant risk that a bad actor could do harm with a powerful model. Further, the set of researchers who are sophisticated enough to do beneficial, compelling work is concentrated enough that other release mechanisms could be employed that would not dramatically diminish the value associated with open source releases.

We know this position paper will be controversial because it doesn't question the idea that there are foundational models. It also implies that open source is not the right solution for every scientific project. Hit reply and share your thoughts!

RESEARCH SOFTWARE ENGINEERS – GROWTH SPOT IN ACADEMIA

The Research Software Engineer role is **becoming** increasingly common at top research universities in the US, UK, and Australia and is a rare growth area in academia. An RSE builds software for research. Some top computational labs can afford to hire dedicated full-time RSEs, but "most research groups need — or can afford — an RSE only on an occasional basis. If their university has an RSE group, they can hire an in-house engineer for one day a week, or for a month at a time." Part of the reason universities are investing in high quality code and the humans it takes to produce it stems from the FAIR principles: "In the past, researchers weren't always incentivized to use or create good software [...]. Many journals now require authors to publish code, and that code has to be FAIR: findable, accessible, interoperable and reproducible." In other words, as the field of data science for science has professionalized and adopted higher standards, the number of professionals required to meet these standards has increased. This makes science more accessible, robust, and, yes, reproducible.

AI AND PATENT/COPYRIGHT LAW

There is no real debate about whether an AI can generate material that is patentable. A great deal of material generated by AI could be patentable or

copyright-able. The question is whether an AI can **hold a patent**. At present, the answer could be yes, no, or maybe, depending on the circumstances. This question is further complicated if the model had been trained on material **under copyright**, for which the model developers did not receive permission to store the copyrighted data in a training set. The ongoing court cases around AI has huge implications for how models are trained, who gets to train them, who gets to deploy them, and what becomes of their outputs.

RIGHT TO REPAIR LAW FAILS IN CALIFORNIA COMMITTEE

Broad public support was **not enough** to pass the Right to Repair bill in **California**, the state where it had the best chance of becoming law. Let's see what happens next legislative season.

ETHICAL WEB PRINCIPLES

The **World Wide Web Consortium** (W3C) has **published** Ethical Web Principles that are brevity at its haiku-like finest.

NEW PROGRAMS, FOLLOW THE MONEY

Click through to access **a structured spreadsheet** of New Programs and money moving around in academic data science.

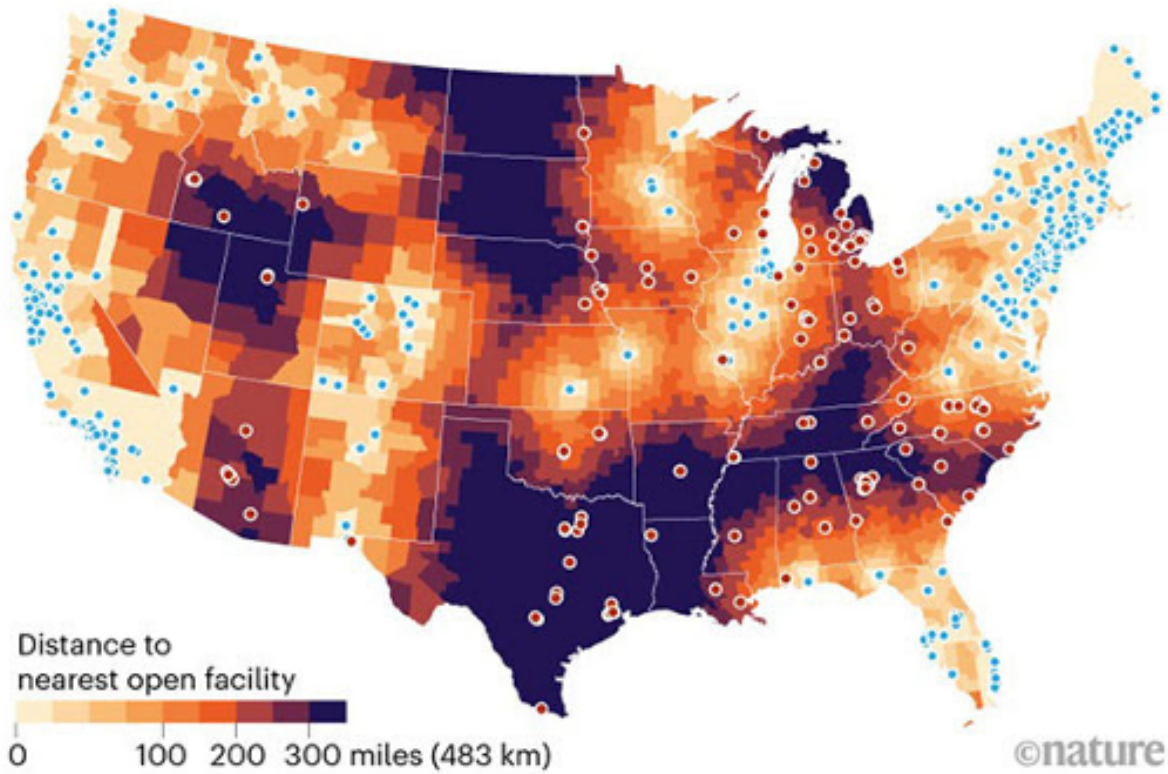
DATA VISUALIZATION OF THE WEEK

by Nature, Editorial from May 6, 2022

IF ROE IS OVERTURNED

In addition to 13 states with 'trigger bans,' 12 others are expected to enact new, restrictive policies. In this scenario, economists estimate that about 18 million women in the affected states would need to travel more than 200 miles to have the procedure, which would prevent around 100,000 women from reaching an abortion provider each year due to economic and other hardships.

- Abortion facility likely to close
- Abortion facility likely to remain open



Deadlines

Studies/Surveys

[Do you teach introductory statistics to undergrads?](#)

"This project is conducting research to inform the design, development, and evaluation of a new approach to teaching intro stats. Please complete the brief survey!"

Conferences

[Python at Scale is coming to Austin, TX this fall and CFP is OPEN!](#)

"Join us for news about big data solutions for #python devs, talks, tutorials, hackathons, workshops, and more ????" Deadline for submissions is

August 1.

Education Opportunities

Microsoft Research PhD Fellowships application deadline 7 June!

"For students enrolled in global PhDs"

Contests/Award

We will be running a competition track @satml_conf 2023.

We solicit competition proposals on topics that advance the science of trustworthy ML. Deadline for organizers to submit proposals for competitions: July 29 2022

Tools & Resources

What is #dataviz animation good for? A thread with examples.

Twitter, Joey Cherdarchuk from May 13, 2022

"By far the most common and intuitive use is to show time"

World Population Data — In Detail — Released To Public

CleanTechnica, U.S. Department of Energy from May 26, 2022

"Researchers at **Oak Ridge National Laboratory** have spent more than two decades working to solve this challenge through the development of cutting-edge population distribution models. Now their suite of LandScan datasets is available online to the global public under a new open-source creative commons license. "

Data Science in Context: Foundations, Challenges, Opportunities

Alfred Spector, Peter Norvig, Chris Wiggins, Jeannette M. Wing from May 1, 2022

"This website provides the Authors' Manuscript for Data Science in Context: Foundations, Challenges, Opportunities. This book will be published by **Cambridge University Press**. This pre-publication draft is free to view and download for personal use only."

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](#). Our archive of newsletters is at <https://academicdatascience.org/resources/newsletter>. Our mailing address is 1037 NE 65th St #316; Seattle, WA 98115.