ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for December 16, 2021.

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.

DATA SCIENCE FOR SOCIAL IMPACT — ADSA'S FIRST VIRTUAL SUMMIT SESSION

As university data science programs have proliferated, so have "data for good" initiatives and projects. How are data science institutes running successful social impact programs?

Full-session Video
Neatly summarized Write Up

CAN WE GO ON LIKE THIS?

Universities and professors are taking stock of the changes they've made to accommodate the pandemic. In a **recent survey** of 3200 UK-based scientists, those working in industry reported more job satisfaction than those in academia, a departure from 2016 results. Many are questioning how they can go on like this. Some professors have **quit**. At a macro scale, universities like **Stanford** that allowed a fair amount of remote work and teaching are **now expecting** students, faculty, and most staff to be on-campus, in-person after the holiday break, regardless of Omicron. **Cornell**, on the other hand, is going on **alert level red** and sending students home after a big COVID outbreak despite their 97% vaccination rate. **NYU** is also seeing a big case load spike so they sent an all-call email this morning announcing faculty must make all finals remote and canceling all non-essential gatherings effective immediately.

Professors like **Courtney Thompson** who have "leaned hard into the 'pedagogy of compassion' mindset" and "cut readings, reduced workloads, recorded every lecture, and was generous with extensions" are rethinking because "the more flexible I try to be, the less happy my students are." **Paul Vermehen** **noted a similar pattern**, "I gave generous deadline extensions on two essays. Only about 60% were submitted. Last year, over the same period, about 92-93% of essays were submitted." Both report being exhausted; they also sound demoralized.

The question of "should we go back to the way it was?" is one area where there is a fair amount of divergence between academic and corporate life. Corporations have announced more permanent changes to work life, with less office-time and some allowing permanent remote work. Academia tends to look a lot more like Stanford, trying to get things back to the way they were. That

said, **Shannon Mattern** is teaching a **"Redesigning the Academy"** course next spring and many schools will **leave SAT admissions requirements** on the far side of COVID, though the anti-standardized-test trend started pre-pandemic. During the pandemic there was less of an effort to roll out the equivalent of a "pedagogy of compassion" in corporate America, but where flexible and remote work has been allowed, corporations likely cannot and **will not** return workers to the 5-day/week office life they had before. Universities, however, seem to be operating on the assumption that it is not only possible, but preferable to return to the way things were. Recall that the way things were involved a fair amount of union organizing among grad students and adjuncts pre-pandemic. That resistance has already led to a successful effort to **unionize** **Northwestern's** librarians (Dec 3, 2021) and an ongoing months-long **student worker strike** at **Columbia**. Additional academic labor organizing could pick up again in 2022.

The answer, then, is that the academy as a whole is likely to try to continue as it was pre-pandemic, largely because tuition-driven schools need to offer an in-person college experience to justify the cost of attendance. But as the corporate world changes and leaving academia slowly loses its stigma, younger faculty, staff and grad students may leave. This is particularly true for computer science, data science, and many bench science faculty who can get paid 2-3 times as much in industry. The Great Resignation is not a tsunami, it's more like a slow leak that will unfold over years.

INTERNATIONAL STUDENT ENROLLMENT DOWN 15%

International student enrollment fell 15% at US universities during 2020-2021 school year, a new **Pew Research Report** finds. Pew attributes this fall off primarily to the pandemic, but in the years leading up to the 2020-21 school year, the U.S. federal government showed an appetite for limiting student visas and restricting travel to the US by students from countries including Iran, Iraq, Libya, Somalia, Sudan, Syria, and Yemen (see **old coverage**). The **Trump administration** outright canceled the visas of thousands of students from China in December 2020; the policy has **been upheld** by **the Biden administration**. China normally sends the largest number of students to the US (~35% of foreign students) which is still true even though China had the second largest decline in the number of students last academic year. Foreign students pay the highest rates for tuition, so they are especially attractive to state schools and smaller private schools. Chinese students alone “paid an estimated $15 billion in tuition' in the 2018-19 year. These are non-trivial amounts for universities facing lower enrollments as some domestic students avoided from enrolling during the pandemic and demographic changes resulting from fewer babies born during the 2008 recession will hit in academic year 2026.

SPONSORED CONTENT

Subscribe to data.org for news and opportunities.

WHICH JOURNALS ARE GOOD FOR TECH ETHICS?

Try publishing in FaaCT, SIGCHI, CSCW, and Big Data & Society if you're doing tech ethics or data ethics research. **Casey Fiesler** has been teaching and researching in the computing, data, and tech ethics space for years. It hasn't always been easy because there aren't dedicated journals nor is publishing in a dedicated journal always a great career move. She **pulled together** a list of tech, computing, and or data ethics journals/conferences to demystify the landscape..

DC AG PLANS TO BAN ALGORITHMIC DISCRIMINATION

Washington DC Attorney General **Karl Racine** **introduced a bill** to ban “algorithmic discrimination" in decisions related to housing, loan making (closely tied to housing), and job seeking. Looks like gender, race, sexual orientation, family status, political beliefs, national origin, disability status, and quite a few other characteristics would be covered, following DC's Office of Human Rights, **Protected Traits guidance**. This ratchets up the bans some cities like Berkeley and Somerville, MA have placed on uses of facial recognition by regulating technology actively in use.

AI CONTROL COUNCIL? GLOBAL ETHICAL HACKER TEAM? WHAT SAY YOU?

A more interesting policy development is the proposal by **Anton Korinek** with **Brookings Institute** to **create** an AI Control Council to govern two classes of problems: the technical problem of direct control of AI and the societal problem of social control of AI. Direct control may refer to things like getting a handle on killer robots, self-driving cars, and even discriminatory modeling to make sure people aren't being killed or harmed by technology due to technical design flaws. The social control element investigates situations in which the model may be technically performing quite well — for instance, **Facebook** designed an algorithm that was very good at engaging viewers on their News Feed. However, the engagement tended to amplify misinformation, disinformation, and use patterns on Instagram that were bad for society overall. In this case, the overall impact of the technology was negative, even though the model delivered the outcomes it was designed to produce really well. Recent appointments at the **U.S. Federal Trade Commission**, including **Meredith Whittaker** **hired** into the new role of Senior Advisor on Artificial Intelligence, suggest that there is growing appetite in the US for creating oversight specific to AI. The UK has seen more regulatory efforts in this area (see Issue #230) and researchers at **University of Cambridge** are now **calling** for a loosely organized global group of ethical hackers to specifically target AI implementations to reveal their flaws before the flaws reveal themselves. This echoes the concept behind **Twitter's** algorithmic bias bug bounty that **Rumman Chowdhury** adapted from the InfoSec community and **announced** at **DefCon** last year. More corporations need to offer bug bounties tailored to AI — many would only need to expand existing programs to encompass equal impact and social impact concerns — to provide enough reward to motivate ethical AI hackers. At this point, it's fair to say that both bug bounty-style ecosystems and additional regulatory oversight can help innovations that use AI deliver the goods without the bads.

Featured Jobs

See the ADSA Jobs Page for more opportunities.

Assistant Professor - Statistics, Data Science, UC Berkeley, Statistics, Berkeley, CA
Assistant Teaching Professor - Probability, Statistics, and Data Science, UC Berkeley, Statistics, Berkeley, CA.

Postdoctoral Scholar University of Chicago - Data Science Institute. Chicago, IL.

TIMNIT GEBRU COMES OUT OF STEALTH WITH DAIR

Timnit Gebru, the AI researcher who was famously dismissed from Alphabet about a year ago has **come out of stealth** with her new research institute, Distributed AI Research (DAIR). DAIR is funded by a laundry list of private foundations - **Ford**, **MacArthur**, **Kapor Center**, **Rockefeller** and **Open Society**. Researchers, most of whom are yet to be hired, will likely come from marginalized groups and countries other than the US. In addition to research activities, the group will also offer consultative advice to other groups looking to investigate or mitigate bias and negative outcomes stemming from implementations of AI.

FLOATING AIMFULLY

Open water robots can now float their way from point A to point B **on their own**. **Caltech** mechanical engineering professor **John Dabiri** has spent his career investigating the unique dynamics of **jellyfish propulsion**. Recently, he and grad student **Peter Gunnarson** took inexpensive submersible robots with only onboard gyroscope and accelerometer inputs, and installed reinforcement learning algorithms. There is enough data to interpret the water's flow direction and use it for navigation, minimizing the power needed by an open water robot and, more important, improving its accuracy, especially in deeper water. With propelled robots, "it's almost impossible to control them with a joystick," says Dabiri.

Featured Events

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

UTSA
The University of Texas at San Antonio™

Draper Data Science Business Plan Competition. Open to students anywhere in the world. Info Session: 18 Jan 2022. Submission deadline: 31 Jan 2022. Event: April 13-15, 2022. 1st prize - $50,000 investment.

SOUTH KOREA PRESENTS PRIVACY PROBLEMS

South Korea will **use facial recognition** deployed against video feeds from an extensive array of CCTV cameras in Bucheon (suburb of Seoul) to perform contact tracing and mask wearing adherence to monitor the spread of the coronavirus. **Reuters** reports that South Korea already "has an aggressive, high-tech contact tracing system that harvests credit card records, cellphone location data and CCTV footage, among other personal information". People who support civil liberties and a balance of power are rightfully upset. I, for one, am curious about how accurate facial recognition is when everyone is wearing masks.

NON-PROFITS PRESENT PRIVACY PROBLEMS

Non-profits are exempt from most of the toughest data privacy laws, the ones enacted by California and Virginia, and some related marketing laws. In a desire to keep identifying and targeting new donors, the sites employ cookies and even keystroke/mouse position trackers to understand what visitors to their sites are interested in. **Planned Parenthood**, which is both a charitable organization and a community health clinic, has "28 ad trackers and 40 third-party cookies tracking visitors, in addition to so-called 'session recorders' according to **new research** by **Alfred Ng** and **Maddy Varner** of the **The Markup**. Many other non-profits like **Save the Children** and sites looking to attract patients to their addiction treatment centers also use about the same number of cookies and trackers as a typical commercial site. The difference is that commercial sites are required to be more responsive to user's requests to understand what is being collected and limit use and onward transfer of their data than non-profits are. The only exception to this situation is Colorado, where non-profits are expected to uphold that state's data privacy legislation.

A NOTE ON HUMANITIES VS. ENGINEERING

The culture wars have recently pitted engineering, computer science, math and other so-called "hard" majors against the humanities. A **new study** out of the Center of Law, Business, and Economics at **Northwestern University's Pritzker School of Law** that looks at college selectivity, choice of major, and earnings finds that short of going into a specific vocation like nursing or accounting, there's less difference in earnings between math-focused and writing-focused majors. "Although those with writing focused jobs didn't earn quite as much as those with math-focused jobs, the finding counters the stereotype that humanities majors are" unemployable or underpaid. "Indeed, the report found that a humanities major has a similar relationship to selectivity and earnings to majors" stereotypically perceived to be high earning, like engineering.

CELL DEATH INDICATOR, SOUNDS LIKE A METAL BAND

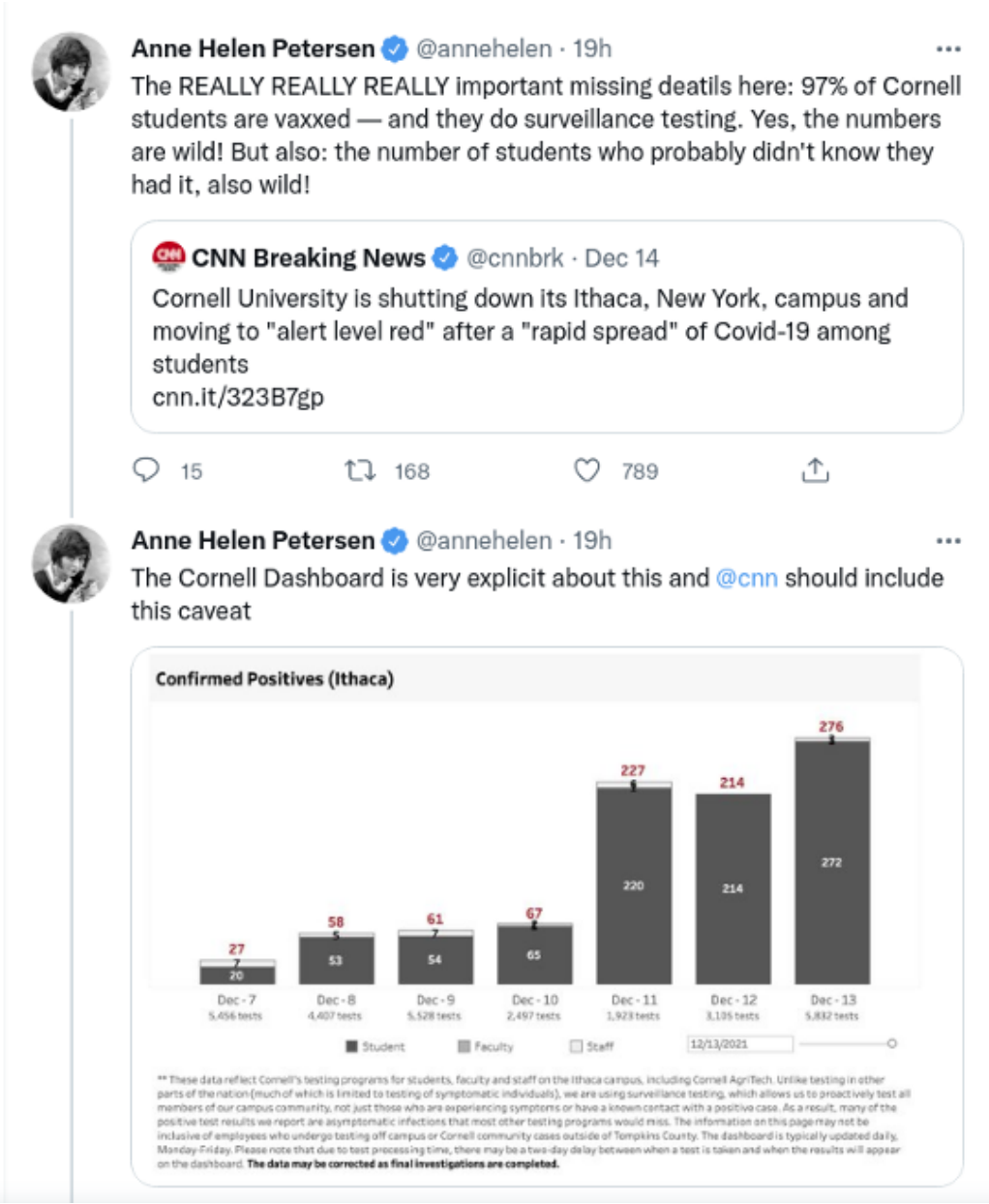
Before neurons die they degrade, which looks a lot like death, and really complicates any investigation into neurodegenerative disease. So **Jeremy Linsley**, a program leader in **Steve Finkbeiner's** lab at the **Gladstone Institutes** in San Francisco invented a calcium ion sensor, the morbidly named **genetically encoded death indicator** (GEDI), which lights up when a cell has passed the point of no return. Impressive, but the key to making the insight scalable came from his brother, **Drew Linsley**, a bioinformatics and computer vision professor at **Brown University**. Team Linsley went about training machine learning algorithms to differentiate between the not-so dead and the certifiably dead cells shown in neurological flourescent images. Their computer model, called BO-CNN (far less morbid, far less descriptive), is hitting 96% accuracy at the dead cell identification task, and at high-throughput rates and volumes. The advances are certain to dramatically improve researchers' understanding of neurological cell death.

NEW PROGRAMS, FOLLOW THE MONEY

Click through to access **a structured spreadsheet** of New Programs and money moving around in academic data science. It's university board meeting season. Lots of new programs on the list.

DATA VISUALIZATION OF THE WEEK

Twitter, Anne Helen Peterson, from December 14, 2021



Deadlines

Conferences

[ACM FAccT 2022 Tutorials Call for Proposals]

"ACM FAccT solicits proposals for tutorials to be presented at the 2022 Conference, which will be held both in-person and online. The in-person conference will take place in Seoul, South Korea on June 21-24 2022." Deadline for proposals is January 28, 2022.

Contests/Award

[US and UK to Partner on Prize Challenges to Advance Privacy-Enhancing Technologies]

"The US and United Kingdom (UK) will collaborate on a series of innovation prize challenges to catalyze research and advancements related to privacy-enhancing technologies (PETs). These technologies give the user greater control over the data being processed to protect personal information and intellectual property. The aim of the prize challenge is to bring together the top

minds in both countries to encourage and facilitate the adoption of PETs." Challenges begin in Spring, 2022.

Tools & Resources

Understanding User Interfaces with Screen Parsing

Carnegie Mellon University, Machine Learning at Carnegie Mellon, Jason Wu from December 10, 2021

"Recent efforts have focused on predicting the presence of an app's on-screen elements and semantic regions solely from its visual appearance. These have enabled many useful applications: such as allowing assistive technology to work with inaccessible apps and example-based search for UI designers. However, they constitute only a surface-level understanding of UIs, as they primarily focus on extracting what elements are on a screen and where they appear spatially. To further advance the UI understanding capabilities of machines and perform more valuable tasks, we focus on modeling the higher-level relationships by predicting UI structure."

How to Train your Decision-Making AIs

The Gradient, Ruohan Zhang and Dhruva Bansal from December 10, 2021

"My colleagues and I reviewed five types of human guidance to train AIs: evaluation, preference, goals, attention, and demonstrations without action labels. They don't replace imitation or reinforcement learning methods, but rather work with them to widen the communication pipeline between humans and learning agents."

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.