Data Science Community Newsletter features journalism, research papers and tools/software for November 4, 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.

NEW INDICES FOR ACADEMIC ARTICLES
Carl Malamud, a collaborator of the late Aaron Swartz, has released an indexed, searchable (sort of) database of "107,233,728 journal articles" with signatures of support from Vint Cerf, Ian Foster, James Evans, and Tim O'Reilly, among 96 others. [See also: Nature article]

There's another new index – Lacuna – that keeps track of predatory publishers who extract publishing fees from authors but perform zero peer review such as OMICS, or those that scrape work published elsewhere and rebrand it as their publication.

ARXIV WORKING AS INTENDED
What a month for one of the original and most prominent pre-print services, arXiv. A National Academy of Science member and Professor emeritus at University of Texas at Austin, John Kormendy, posted a paper proposing to guide hiring and tenure decisions based on "citations of refereed papers, citations of all publications normalized by the numbers of co-authors, and citations of all first-author papers," techniques that have widely been seen as locking out talented, underrepresented people with arguments that sound like meritocracy but actually perpetuate systems that deliver rewards to people who had the most advantages to begin with. Kormendy offered an apology, though some think he doesn't understand why he should apologize.

Then there was this paper – Delphi: Towards Machine Ethics and Norms – that is a tragedy of comedic proportions and (hopefully) was meant to be an object lesson in why we should not delegate ethical decision making to machines. The basic creation here is a “prediction” of whether or not a proposed behavior is ethical, using a Reddit corpus as training data, with a user interface similar to a virtual Magic 8 Ball - ask a question, hope to receive permission to proceed. Unfortunately, the Commonsense Norm Bank engine recommends behaviors such as deceive, hate crimes, or killing someone if the question is phrased correctly. FACEPALM AWARD WINNER! And the competition is tough lately.

On a positive note! Please see Carnegie Mellon University faculty member Rayid Ghani’s work on how to develop fair and equitable machine learning. He argues that, "you don't have to sacrifice accuracy to build systems that are fair and equitable. But it does require you to deliberately design systems to be fair and equitable. Off-the-shelf systems won't work." Ghani has been working on fair and equitable machine learning for years and it shows. Rather than developing an oracular
project, Ghani has been deeply engaged with city agencies and community groups on local resource allocation problems.

And a tweet reminder from Thomas Dietterich at Oregon State: @arxiv is not @github. Don't submit drafts to arXiv; you are wasting your colleagues' time.

ACADEMIC FREEDOM UNDER FIRE
The University of Florida denied four professors the ability to sign an amicus brief in a voting rights case. Now a pediatrics professor, Jeffrey Goldhagen, had his request to provide expert testimony about mask mandates in K-12 schools denied. Florida Governor Ron DeSantis banned mask mandates for K-12 students. It is quite typical for professors to sign amicus briefs and act as expert witnesses in trials. The state accredits is now investigating whether UF professors are able to exercise their academic freedom independent of political pressure.

Up at MIT there will be campus-wide meetings about the state of academic freedom and freedom of speech more generally. This follows the case we reported last issue in which geophysicist Dorian Abbot had his invitation to give the large, public Carlson lecture down-graded to an invitation to give a less prestigious, less public talk on campus. Abbot had published opinions questioning the premise of Diversity, Equity, and Inclusion initiatives in Newsweek.

At the University of Oregon, there is now a new provostial office dedicated to academic freedom and advising faculty about how to exercise it.

MINNESOTA’S TRIBAL MEMBERS GET BIG TUITION BREAK
The University of Minnesota has announced it will offer reduced or free tuition at all four of the state's public universities to members of the state's eleven native tribes. As far as we know, it is the first university system in the United States to create a program like this.

KYLE CRANMER MOVES TO WISONSIN
Physicist and long-time emotional supporter of the DSCN since its infancy, Kyle Cranmer, is moving from New York University to University of Wisconsin-Madison. He will lead the American Family Insurance Data Science Institute. The upper midwest is so hot right now. ;)

SPONSORED CONTENT
The Data Science Institute at Columbia University trains the next generation of data scientists, develops innovative technology, fosters collaborations to interpret data and address pressing societal problems, and works closely with industry to bring promising ideas to market.

RIGHT TO REPAIR INCHES FORWARD
The US Copyright Act revamped a section of its code to add exemptions that expand the types of devices that can be diagnosed, maintained, and repaired without violating copyright. Even medical devices that are not implantable are now exempted from certain repairs, a move the medical device industry has lobbied to prevent. The Biden Administration, several legislators, and the U.S. Federal Trade Commission are also working to advance right to repair. We have previously written about the momentum building behind Right to Repair in our longer From the Desk of… section.

FACEBOOK QUITS FACIAL RECOGNITION
Facebook is shutting down its facial recognition technology, though it will not be deleting the DeepFace algorithm at its core. As a reminder to people who stopped using Facebook a while back, the company was one of the first to have decently accurate facial recognition models. One of the primary use cases was automatically tagging people who appeared in photos uploaded to the site. With over 1 billion “facial templates” in its data stores, the company noted that uncertainty about the regulatory fate of facial recognition contributed to the decision. Readers, pay attention to voice recognition, a key feature of voice-activated technologies. “Voice prints” are just as identifying as facial recognition but less biased in the ways facial recognition has been biased. As this newsletter has previously asked: is the main tech ethics concern surrounding facial recognition about pervasive, persistent, almost inescapable surveillance or about the accuracy of that kind of surveillance across skin tones? For those who are more concerned about the surveillance than the skin tone bias (which skilled technologists can reduce), voice recognition is an interesting application. Less biased, it is also generally designed to “listen” only when explicitly called into active listener status by a key command – “Hey, Siri” or “Alexa” or “Ok, Google”. The ability for people to control active listening is a critical design feature, an area where technology companies can work to build user trust. [Comedic Reference: memorable SNL skit “Amazon Echo”]

META
The ADSA editorial board considered commenting on Facebook’s new name, Meta, but will not be mentioning it.

PAY FOR VAX DOESN'T WORK
New evidence confirms what social scientists recommended at the outset – paying people to get vaccines does not work. Whatever people are being paid, whether that’s $100 or $500, it is super low relative to the value someone puts on their own life. And it has the effect of making it appear as though authorities do not value their constituents, further degrading trust. This has definitely been an issue full of situations in which scientists and policy makers did not read the social science literature before proceeding. Might I interest anyone in the annals of Administrative Science Quarterly? New studies on the impact of AI in the latest issue. Or how about American Sociological Review articles that confirm physical attractiveness (with racialized stereotypes) matters in contexts where we might wish it didn't? TL;DR being physically attractive matters more than race or gender when it comes to earnings. Oy.
PARENTS AND PARENTS TO BE – JUICY TARGETS

Hit reply to this email and let us know how you feel about menstruation, fertility, and pregnancy tracking apps. After the early, high-profile fail at Target in which a pregnant teen's father was alerted to her pregnancy by all the diaper and baby marketing material from the retailer that started arriving at the home, I have personally been skeptical that there will be a solid, impenetrable wall between users and the ads, promos, features, and other personalizations that may target them about pregnancy.

Marketers know that the birth of a child is the rare opportunity not only to sell adults a bunch of new things for the baby, but also to change their own habits. Adults rarely change their brand loyalties as easily as they do when they are about to become parents, so knowing someone might be about to become a parent is extremely valuable to marketers. That's why I'm skeptical that any fertility, menstruation, and pregnancy trackers will resist the sirens (marketers) call. But maybe I am being too skeptical. Are parents-to-be happy to have new brand relationships? Are app makers motivated to keep user data totally walled off from marketing? Please share thoughts. Note that your thoughts won't be anonymous if they come from your email address, but we will ask before we share them with anyone other than DSCN editorial staff.

COVID’S LONG-LASTING IMPACT ON SCIENTIFIC PRODUCTIVITY

Scientific productivity took a hit from COVID. Even though "researchers' productivity levels have mostly returned to pre-pandemic highs, scientists who did not pursue COVID-19-related research initiated 36% fewer new projects in 2020 compared to 2019." Lead author Dashun Wang notes that the perceived productivity is coming from scientists, "busy working on established topics, writing up existing research, reviving legacy projects or revisiting old data...no fields were immune to the reduced number of projects." Unsurprisingly, the impact was higher for parents and those in other caretaking roles.

BIG QUESTIONS ABOUT OLD GAMES

Can quantum computing solve chess? Elisabeth Paehtz is the top women's chess player in Germany, taught by her father, a grandmaster. She asked the quantum question to a chess programmer at Chessbase, a leading chess software maker and publisher in Hamburg, and "he said it's absolutely irrelevant." The number of possible chess positions is galactically big, ten to the power of forty. "The problem is the data saving," she told a Dusseldorf audience. It's "more data than could be saved on the planet." A simpler question is how long does it take ordinary people to get good at chess? Joseph Wong, a Boeing engineer and casual chess player mined the lichess database for an answer. The online chess website lichess.org freely publishes data from all of the chess matches played there. Wong found that beginners improve quickly at first but an experienced player typically needs 3-4 years to increase their lichess skill rating by 100 more points. There are games older than chess, some of which exist in historical fragments that have otherwise been lost to time. Cameron Browne and colleagues at Masstricht University are three years into a project that uses AI to reconstruct ancient games that were popular in millennia BC. Like now "anyone who had time to kill would often teach those around them the games they knew," says Browne.
Define Data Scientist

Have you worked in data science for 5+ years?

Help the US Dept of Labor define Data Scientist for a database used by millions of employers, workers, educators, and students.

VOLUNTEER TO PARTICIPATE

The [Occupational Information Network](https://www.onetonline.org) (O*NET) is sponsored by the US Department of Labor. Participants must be currently active in the field and based in the US.

**FOLLOW THE MONEY**

$125,000,000 National Institutes of Health -> 16 teams to study the science of senescent (aging) cells.

$100,000,000 Mark and Lynne Benioff -> [1t.org](https://1t.org) a fund to plant 1 trillion trees to mitigate the effects of climate change.

$85,000,000 Investors led by Tiger Global Management -> OctoML, a University of Washington spinout led by Luis Ceze, that "uses ML to optimize ML."

$74,500,000 National Institutes of Health -> 19 institutions for data science initiatives in Africa.

$43,300,000 Baylor University and Harvard University -> The Global Flourishing Study, a five-year project that will "measure the factors of human flourishing."

$25,000,000 Scott Smith, founder of Qualtrics -> Utah Valley University for a new engineering building that, upon completion in 2024, should double the number of computer science and engineering graduates at the school

$18,000,000 Robert and Donna Manning -> UMass - Amherst for the Robert and Donna Manning College of Information and Computer Sciences.

$15,000,000 John and Marcia Price -> University of Utah to build a new home for computing which will likely bear the Price name.
$13,200,000 National Institute on Drug Abuse -> University of Michigan to develop precision health care for drug abuse, HIV and chronic condition treatment.

$5,000,000 Army Corps of Engineers -> University of Nebraska-Lincoln and University of Nebraska at Omaha for research to expand the lifespan of bridges through new monitoring technology

$4,092,000 data.org -> London School of Hygiene and Tropical Medicine to develop epidemiological software tools.

$2,100,000 Richard and Carol Dean Hertzberg -> Anne Wallace, MD, at UC San Diego Health for to develop and maintain a breast cancer database.

$1,950,000 Air Force Office of Scientific Research -> University of Massachusetts Amherst and Brown University to develop “scientific machine learning” based on prior knowledge, not big data.

$1,450,000 National Science Foundation -> University of Arkansas to investigate the use of artificial intelligence in infrastructure applications.

$800,000 Amazon -> Tennessee State University to endow a Computer Science Professorship for 4 years at the Historically Black College/University (HBCU).

negative $40,000,000 Princeton Club of New York -> Sterling National Bank. The club is closed indefinitely while its mortgage is in default. Princeton Club has no financial relationship with Princeton University.

NEW PROGRAMS
Science Hub for Humanity and Artificial Intelligence a partnership that will be housed at UCLA in collaboration with Amazon.

Massachusetts AI and Technology Center for Connected Care in Aging and Alzheimer’s Disease based at the University of Massachusetts Amherst with research partners from Brigham and Women’s Hospital, Massachusetts General Hospital, Brandeis University and Northeastern University, also covered in The Boston Globe

Master of Science in Biomedical Image Computing at University of Illinois at Urbana-Champaign

IDEaS Cloud [Computing] Hub at Georgia Institute of Technology with support from Microsoft

Master of Science in Data Science at University of Houston-Victoria

DATA VISUALIZATION OF THE WEEK
Nature, Diana Kwon, from October 26, 2021
Deadlines

Conferences

Data & Society — The Social Life of Algorithmic Harms
New York City March 10-11, 2022. Deadline for applications to participate is November 29.

Education Opportunities

U.S. Centers for Disease Control Public Health Informatics Fellowship Program
"Help local, state, and international public health agencies solve complex #PublicHealth informatics challenges using #DataScience." Deadline to apply is December 1.

Tools & Resources

Did you know educators can launch Competitions for their students on Kaggle's InClass
platform for free?
Twitter, Meg Risdal from October 29, 2021
"Great video from one of our InClass hosts @jeffheaton walking through the setup from end-to-end!"

API attacks are both underdetected and underreported
Help Net Security, Akamai from October 28, 2021
"While DDoS attacks and ransomware are both major issues, attacks on APIs don’t receive the same level of attention, in large part because criminals use APIs in ways that lack the splash of a well-executed ransomware attack, but that doesn’t mean they should be ignored."

New UChicago initiative aims to improve health care algorithms for underrepresented groups
University of Chicago, UChicago News from November 3, 2021
"Research by Sendhil Mullainathan of Chicago Booth helped spur the founding of the Algorithmic Bias Initiative, which this summer released the Algorithmic Bias Playbook. The free guide offers a framework for identifying, correcting and preventing algorithmic bias."

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

NSF CISE Distinguished Lecture – From Seeing to Doing: Understanding and Interacting with the Real World
Online November 4, starting at 12 p.m. Eastern. Speaker: Fei-Fei Li of Stanford University.

Dr. Latanya Sweeney: How Technology Will Dictate Our Civic Future
Online November 5, starting at 12 p.m. Pacific. "This talk is part of the FIAT Justice Lecture Series coordinated by the Algorithmic Fairness and Opacity Group (AFOG), Cal NERDS, and the D-Lab and sponsored by the Public Interest Technologies University Network (PIT-UN)."

AI Festival - When Machines Dream the Future
Dresden, Germany, and Online November 12-14. "A series of debates, artworks, documentary films and workshops in which experts and creative artists investigate what "artificial intelligence" means, which problems it can solve, and how conflict can be intensified or even caused by AI." [registration required]

StreamWare: A Scalable Framework for Accelerating Streaming Data Science
Online November 16, starting at 12:15 Central, presenting at the Supercomputing 2021 conference. "This Birds-of-a-Feather session brings together a diverse community of interest around streaming data and the development of StreamWare, an open source framework supported in part by the NSF Principles and Practice of Scalable Systems (PPoSS) program." [registration required]

Featured Jobs
See the ADSA Jobs Page for more opportunities.

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.