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

Many readers with Gmail addresses had DSCN #229 blocked by the service when we did our November 19 send out. We are resending the newsletter to you and, at the same time, working on solving the problem. Thank you for your help.

NEW BOOK READING GROUP FOR “OUT OF OFFICE”
Do you want to join a Slack-based reading group for the new book by Charlie Warzel and Anne Helen Petersen — Out of Office — about working from home (or somewhere other than the office)? Vox has a nice review: “The book isn't really about remote work — it's about work. And not just what it has meant and could mean, but also why the status quo isn't sustainable, for anyone”.

For context, the World Economic Forum investigated what's driving the Great Resignation and found lack of employer support for burnout and other mental health issues to be right up there.

Sign up for book co-reading here

DATA GOVERNANCE AT POPULATION-LEVEL
The vast majority of data governance regulations focus on informing and offering control to individual data subjects. Protecting the privacy of individuals, however, does little to mitigate the harms of misinformation campaigns or the concentration of capital derived from the information economy to a small number of businesses and governments. In a new Yale Law Journal piece, Salome Viljoen argues that, “data’s relationality results in widespread population-level interests in data collection and use that are irreducible to individual legal interests….and that enact or amplify social inequality." Because the law typically focuses on the rights of individual data subjects, there is a vexing aporia that leaves certain harms associated with the networked information economy to proliferate unchecked. Two examples, quickly: the refusal to vaccinate quickly by a numerical minority in the U.S. and the EU is linked to misinformation campaigns as is the erosion of democratic processes in the U.S. Individuals who exercise their rights to opt out of social media have not opted out of the broader impacts of misinformation.

In the final section of the piece, Viljoen lays out a theoretical framework that new democratic data governance law could follow: "personal choices over data sharing should reflect the effects these
choices have on others, not only because of the political and moral benefits of considering others, but also because under current conditions of datafication, individuals already directly relay information relating to others, which is used to predict and influence the behavior of others.

While I assume that most of our readers will not read the entire piece, it's worth noting that Viljoen's perspective emerges from a highly fertile discourse about the future of data governance which could look significantly different than the focus on individual privacy and consumer protection that underlies much of the existing data governance in place across the Americas and the EU. I wouldn't expect these changes soon, but the intellectual momentum is building.
policies promote the appropriation of intellectual property in ways that are considered counterproductive to U.S. national security interests, targeting individual professors has contributed to anti-Chinese and anti-Asian campus climates and produced a chilling effect for Chinese-origin scholars that is antithetical to the principles of academic freedom and international scientific cooperation. As we stare into a future marked by climate change and an ongoing global pandemic, international scientific collaboration is crucially important. Let's stay united as scientists, yes?

**OHIO STATE AIMS TO GRADUATE EVERY STUDENT DEBT-FREE**

Public state schools have historically existed to educate broad swaths of the public in order to foster economic development, good citizenship (democracy tends to break down unless citizens are educated), and scientific advances with broad benefit. This compact has been slowly eroding since the ~1960s when disinvestments in public higher education started to move the cost burden of higher education away from taxpayers and onto students, crescendoing in a massive amount of student debt for recent graduates. The Ohio State University system announced a "10-year plan to raise $800 million to eliminate all loans from financial aid packages given to undergraduates." With an average of $27,000 in debt at graduation, students are unable to do things like go to graduate school, buy a house, or choose jobs that would serve their communities, but pay less than positions that would make it easier for them to service their debt. Hopefully, other state schools can find a way to replicate this program.

**UK RELEASES 200,000 WHOLE GENOMES TO RESEARCHERS**

The UK Biobank (UKBB) has released the whole genomes and medical record data for 200,000 UK residents to scientific researchers. Five hundred thousand middle-aged and elderly people were enrolled in the project between 2006 and 2010. Full exomes (the regions of DNA that encode proteins) from this sample had been released in 2019. The new data release will "make it possible to explore the influence of noncoding DNA, which controls when genes are turned off or on, and of gene rearrangements, as well as missing, repeated, or extra stretches of DNA in genes.” Reading this in line with our next story suggests that it is not only possible, but good, to be a leader both in data sharing and in regulatory work to ensure good intentions are upheld (though releasing data is not the same as writing or analyzing an algorithm, the two are often related).

**UK HAS A PUBLIC SECTOR ALGORITHMIC ACCOUNTABILITY STANDARD**

The UK has issued a nationwide standard for the use of algorithmic decision making in the public sector. The guidelines will move into a pilot study phase with several public sector organizations before being finalized. Ongoing research and convening by the Center for Data Ethics and Innovation, a government expert body (see Review Into Bias In Algorithmic Decision Making), and a recent study by the Open Government Partnership, Ada Lovelace Institute, and AI Now Institute informed the new framework. The UK has emerged as a leader in AI policymaking, a situation made easier by Brexit.
NYC’S (WEAK) LAW MANDATING AUDITS FOR AI HIRING ALGOS
There’s no federal algorithmic accountability standard in the U.S., but New York City has been a leader in implementing AI accountability standards. A new law looks likely to go into effect in 2022 that would require employers who use algorithms in their hiring processes to audit those algorithms for fairness with respect to race and gender. DSCN reader Julia Stoyanovich of NYU is pleased that employers will have to admit to using algorithms in hiring, but noted the “rudimentary set of requirements" in the policy "are very easy to meet" and likely will not address broader concerns related to agism, ablism, and considerations for hiring outside of type.

APPLE ALLOWS CONSUMERS THE RIGHT TO REPAIR, SORT OF
Apple has a new program in which its customers can submit the serial number of their broken devices and then order genuine Apple parts with which to repair those devices. Apple already lets independent repair shops order genuine replacement parts, as of a couple years ago. This new announcement is nice for what I imagine to be a very small group of people who want to try to repair their own iPhones and Macs. I hope to see community education classes like "How to replace your Mac's battery and your iPhone's screen" coming to a library near me (and you).

Featured Jobs
See the ADSA Jobs Page for more opportunities.

OMICRON ORIGINS
Omicron is still going through the lab-based scientific exploration that precedes public health
guidance. It should be no surprise that University of Washington epidemiologist and 2021 MacArthur genius grant winner, Trevor Bedford is back on the scientific expert circuit. He gave an interview to NPR yesterday about the likely genetic origins of omicron...way back in 2020 possibly mutating for months in the body of someone with untreated HIV. Richard Lessells, part of the team that found omicron in South Africa, is the leading proponent of the HIV+ incubation origin for omicron. The scientific mysteries of this virus continue to deepen.

NO SIR
Time is not well accounted for in the "Susceptible-Infected-Resolved" (SIR) models use by epidemiologists to predict disease outbreaks. Brown University mathematicians led by George Karniadakis used physics-informed neural networks (PINNs) where physical laws that govern a system inform the algorithm to improve accuracy in nine different SIR Covid-19 models. “The computational time needed for re-training PINNs with new data is relatively short compared to the time-scale of pandemic evolution,” according to collaborator Ehsan Kharazmi. The “wrong” in wrong predictions compounds as predictions' trajectory fails to track reality. We're near-post-Covid. Keep up or keep out.

THE ROBOTIC LIFE OF FROG CELLS (VERY COOL - DO NOT SKIP)
Published a year ago by Josh Bongard at the University of Vermont and Michael Levin at Tufts, but rising to our attention now: algorithmically programmed frog egg cells can mush into other frog cells to become a little organism called a xenobot. Xenobots never grow into frogs or behave like frogs, but they can move around in fresh water on their own, can grow by smushing into and taking on other cells a la Pac Man, and can give rise to a second generation (no grandchildren, yet). Aside from fascinating questions about what it means to spawn life — what if the life spawned by an organism looks nothing like that organism? — these little xenobots may prove useful. Ongoing research is exploring whether they could be used to smush into and gather up microplastics in freshwater.

LONGER HUGS ARE BETTER
Please hug your loved ones. Do it for more than 1 second. With arms around the back, not the neck or waist. Cite: data viz of hugs.
For maximum impact, please hug your friends/family/children for at least 5 seconds.

HOLD ME LONGER

Longer hugs (5-10 seconds) are more pleasurable than 1 second hugs

1 SECS  5 SECS  10 SECS

12:00 PM · Nov 29, 2021 · Twitter Web App

307 Retweets  32 Quote Tweets  653 Likes
DATA VISUALIZATION OF THE WEEK
Twitter, Mark Lewis, from November 21, 2021

Mark Lewis
@marklewismd

I’ve never seen a better graphical depiction of grief

5:16 PM · Nov 21, 2021 · Twitter for iPhone

14.2K Retweets 589 Quote Tweets 73.3K Likes

Deadlines

Studies/Surveys
Submit your research for coverage in The Economist
"The Economist’s data team is looking for excellent research to cover in our Graphic Detail section, both online and in our print edition. One of our data-journalism team’s goals is to keep our global audience apprised of exciting and important new studies. If that sounds like your work, please send it to us."

Education Opportunities
Open AI Residency
"As part of our effort to support and develop AI talent, we’re excited to announce the OpenAI Residency. This new program offers a pathway to a full-time role at OpenAI for researchers and
engineers who don’t currently focus on artificial intelligence. The Residency will focus on recruiting from underrepresented groups in technology." Deadline for applications is January 14, 2022

**Tools & Resources**

**Cache warming: Leveraging EBS for moving petabytes of data**  
Netflix Technology Blog; Sailes Mukil, Prudhviraj Karumanchi, Tharanga Gamaethige, Shashi Madappa from November 26, 2021

"We had previously given an overview of how we perform cache warming on our EVCache clusters and why it’s needed. The architecture mentioned in the original cache warming article worked great for a vast majority of our use cases. However, as Netflix’s subscriber base grew over the years, the footprint of the data stored in EVCache has increased multiple fold and will continue to increase to meet new and additional demand. As the scale and sensitivity of the clusters increased, the architecture needed to evolve to keep up."

**The fastest way to share your notebooks - announcing NotebookSharing.space**  
Yuvi Panda from November 16, 2021

Just upload your notebook, get the link, and share it however you want. Notebook links are permalinks - once uploaded, a notebook can not be changed. You can only upload a new notebook and get a new link.

**Human-Centered Data Science**  
The MIT Press; Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller and Gina Neff from November 26, 2021

Coming March 2022. "Best practices for addressing the bias and inequality that may result from the automated collection, analysis, and distribution of large datasets."

**Events**

See the [ADSA Events Page](https://academicdatascience.org/resources/newsletter) for more details and more opportunities.

**MIT Technology Review - TR 35 Festival**  
*Online* December 8, starting at 11 a.m. Eastern. "MIT Technology Review welcomes you to our Innovators Under 35 Festival, where you’ll hear first-hand from our global community of honorees about their journey to tech success." [free registration required]

**2022 WiDS Datathon**  
*Online* Begins January 2022. "Aims to inspire women worldwide to learn more about data science, and to create a supportive environment for women to connect with others in their community who share their interests." [registration required]

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