Data Science Community Newsletter features journalism, research papers and tools/software for April 21, 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.

ADSA TALK DISTILLED: HOW TO *BE* NON-ABLEIST IN DATA SCIENCE

How can data science be made more inclusive of both practitioners and consumers of all abilities? **Anat Caspi**, Director of the **Taskar Center for Accessible TechnologiesTaskar Center for Accessible Technologies** discusses the challenges that dominant approaches impose on the non-ableist community, and the means by which technology mediates disabled people's experience in the world.

Full write-up.

COST OF FREE LABOR? [PULSE POLL]

The conversation that has been percolating among academics around water coolers and after seminars since at least 2015 is the subject of our next pulse poll: **Cost of Free Labor?**

What does a fair distribution of responsibility for "service work" in academia look like? How can adjuncts, postdocs, and grad students be equitably included in this important work without being overburdened by it? And the real kicker — which incentives are effective for service work? Are there places for honoraria and bonuses or should any service work be part of the standard expectations for any compensated role in higher education?

Take the poll.

SPONSORED CONTENT

Data Scientist positions at the University of Rochester — Apply!

ONE HOSPITAL'S MODEL OFTEN FAILS AT ANOTHER HOSPITAL
H Singh, **V Mhasawade**, and **Rumi Chunara** have a **new paper** in **PLOS** showing that predictive models for mortality developed using data from one hospital often have lower accuracy when applied to data from a different hospital. In other words, model generalizability is not robust. They find that part of the problem lies in differences in patient populations — racial differences in hospital-to-hospital populations stood out: "Findings demonstrate that predictive models can exhibit disparities in performance across racial groups, even while performing well in terms of average population-wide metrics." Another key problem was rooted in the consistency and availability of data; "distribution of all variable types (demography, vitals, and labs) differed significantly across hospitals and regions." The team concluded that algorithmic fairness alone would not be enough to identify and mitigate all of the root cause issues stemming from inconsistencies in the underlying data. The two takeaways are that predictive models in medical settings are getting more mature, but still aren't fully reliable, and that anyone building predictive models might want to think hard about their approach to training data. In this hospital example, if the training data were selected from within the hospital's population of patients, testing the within-hospital model might have worked as expected, but differences between hospitals would be undetectable.

CYBER-FLASHING WILL BE ILLEGAL IN THE UK
With the **new** **UK Online Safety Bill**, it will become illegal to "cyber-flash" on social media platforms. (The provision will not be enforced for text, email, or dating apps.) Cyber-flashing is sending unsolicited nudes. The most immediate enforcement case that comes to mind is **Twitter**. The platform currently allows pornography and it is unclear if the company will need to prevent UK users from viewing that content outright or if Twitter can continue to display pornographic content in the UK as long as they take measures to determine that users desire the content. The other interesting wrinkle is, of course, **Elon Musk's** effort to buy Twitter, take it private, and encourage "freer speech" on the platform.

Featured Job
See the ADSA Jobs Page for more opportunities.

Research Data Scientist, School of Law, University of Virginia, Arthur J. Morris Law Library, Legal Data Lab. Charlottesville, VA.

CHANCELLOR RESIGNS; FREE SPEECH SURVEY ON HOLD IN WISCONSIN

The **State University System in Wisconsin** has been roiled by administrative churn and political division for years. At present, a planned survey about free speech across the UW system has been **paused** over concern about its funding sources, ultimate goals, and the autonomy of individual campuses with respect to survey participation. The survey was funded by a prominent GOP donor, **John Menard**, through the **Menard Center** which was initially founded in 2017 by an agreement with the **Charles Koch Foundation** and an anonymous donor. Koch is a "proponent of conservative and libertarian causes". When a system-wide decision was made to allow the survey to proceed over concerns raised at individual campuses, including **University of Wisconsin-Whitewater**, the interim chancellor of UW-Whitewater, **Jim Henderson**, resigned. Now the survey has been put on hold again, until fall 2022.

PHILADELPHIA ORCHESTRA ACCOMPANIED BY GIANT AI VISUALIZATION

Last issue's data visualization was gorgeous, but inscrutable out of context. It was a still (with video upon click-through) of an 8.2 million pixel "thinking paintbrush" generated by an AI trained to produce a visual performance of Beethoven. The "thinking paintbrush" is the creation of artist **Refik Anadol**. He trained the model to visualize Beethoven's music by feeding it images of Renaissance art and architecture, things that Beethoven would have seen. In an **interview** with **WHYY public radio**, Anadol asserted that this his is the most ethical application of AI: "AI is a myth that people think of in the most negative way, but actually it can be a tool for humanity, for the next journey. Like any technology, it has a 50% chance of being used for good reasons. I'm not saying that it will never be bad or used for wrong, but the question is: what else can we do with it, such as creating art, beautiful experiences, and positive applications?"

Anadol earned his MFA from **University of California, Los Angeles**, **studying** in the esteemed **UCLA Digital Media Arts department** (graduating 2014), where he is now a lecturer. UCLA debuted Anadol's latest major work, titled *Moment of Reflection*, on April 19.

Featured Event

The Data Science Institute (DSI) and Institute for Mathematical and Statistical Innovation (IMSI) at the University of Chicago are jointly hosting an AI + Science summer school, August 8-12, 2022, on the UChicago campus. Register here by April 25.

ELECTRIC FUNGI — COMMUNICATING VIA THE HYPHAE NETWORK?

Ah, the wonderful world of fungus, one of the most mysterious forms of eukaryotic life, existing mostly as thread-like mycelium in soil. Professor **Andrew Adamatzky** at the **University of the West of England** runs an "unconventional" computing lab in Bristol. His team inserted tiny microelectrodes into fungal hyphae threads that make up the branching mycelia to investigate their potential ability to communicate along the hyphae network (which also happens to connect to trees and some other plants). He found that the electrical spikes detected in the network appeared as up to 50 different types, and **surmised** that these different spikes may operate like words that allow fungi to send messages, most likely to maintain the overall integrity of the fungus across its mycelia.

Other researchers are not so sure there's enough evidence to connect the distinct electrical spikes to the assertion that fungi use the spikes to communicate. **Dan Bebber** of the **University of Exeter** noted with a characteristically British delivery, "Though interesting, the interpretation as language seems somewhat overenthusiastic, and would require far more research and testing of critical hypotheses before we see 'Fungus' on Google Translate."

HOW DID US UNIVERSITIES SPEND FEDERAL PANDEMIC AID?

Funds from three different tranches of pandemic aid were earmarked for institutions of higher education. A **new report** released by the **Association of Public and Land-grant Universities** (e.g. many state college and university systems) explain how their $20 billion cut was allocated. The big blocks were "$8.5 billion for emergency grant aid, $10.5 billion for institutional relief, and an additional $1.7 billion for...Historically Black Colleges and Universities (HBCUs), Minority-Serving Institutions (MSIs), and other eligible institutions." Eighty percent of the grant aid to undergraduate

and graduate students has already been allocated to cover expenses such as "rent and food, as well as course materials, transportation costs, medical expenses, and other unforeseen costs." One of the most interesting funding gaps is for the improvement of HVAC systems and other physical plant changes to reduce classroom transmission rates as masks come off.

One new pandemic struggle is still emerging — recent high school graduates are showing up at college **unprepared** for the coursework. Long-time **University of Texas** math professor, **Uri Treisman** "nationally known for his techniques and philosophies for teaching calculus," said fall 2021 was the most difficult he's had in his 50-year career.

ANOTHER ARGUMENT FOR A "NEW" PUBLISHING MODEL IN ACADEMIA

Writing in **The Guardian**, **Stuart Ritchie**, **argues** that it's time to abolish the scientific paper in two ways. First, most people look at scientific papers on screens so no need to keep printing journals. Finally, the promise of the paperless office has arrived! Second, moving away from paper unlocks all sorts of things like the ability to attach data and code, update data, correct errors, and include reviewers' comments. This second take is not exactly new to the open science community, a community that Ritchie does not mention. Ritchie's most provocative (and least developed) argument is that scientists should stop listing their publications on their CVs because long CVs don't necessarily correlate with better science and may, in fact, correlate with the production of shoddy science. He doesn't explain what might take the place of a scientist's publication record or why digital publishing would lead to the end of the publication record. But! He placed this rather academic-insider piece in The Guardian. Wonder if he will list it on his CV.

NATIONAL CENTER FOR EDUCATIONAL STATISTICS: NEW STRATEGY

The **National Center for Education Statistics** (NCES) **published** its strategic report that outlines how the agency intends to address challenges and identify key goals. Of particular interest to our readers is NCES's emphasis on measuring progress toward diversity, equity, inclusion, and accessibility goals "throughout its organization and work" as well as by looking for opportunities to relink its data that doesn't have demographic info with data sets from other agencies that do. In order to do this, NCES will be "reconsider(ing) the consent language and planned usage of all primary collections, to support ongoing uses for statistical activities." Because the agency has limited funds they also want to try to expand access to additional researchers: "For mutual benefit, we recommend NCES establish a joint statistical research program for external researchers and fellows. NCES would also benefit from expanding and modernizing its data-licensing program, to further increase responsible data access for evidence building." If you have a graduate student in statistics, education, or social science, this might be the perfect time to start working closely with the NCES.

1-DAY PER WEEK IN THE OFFICE IS THE SWEET SPOT

A new study published as a **working paper** in **Harvard Business School** that was able to randomly assign workers to come in less than one day per week, about 1 day per week, and 2 or more days per week over a nine-week period found that those in the 1 day per week group had higher quantity of email communications and produced more "novel" work (novelty is a good thing in this case). Because the participants were randomly assigned, the study replicates what would likely happen if individuals — not managers — determined which days they should come in. It's a provocative finding that comes at a time when companies like **Target** and **Meta** have decided to follow **Twitter** and let employees work remotely as much as they want while others — like most big banks and **Google** — are expecting most or all employees to be in at least 3 days per week.

TWITTER AS A RESEARCH TOOL

There is a **new paper** on **arXiv** for researchers using Twitter data in their investigations. **Twitter** has long provided data to researchers via API, though not without criticism. Twitter's newest API for Academic Research allows scholars to "access Twitter's real-time and historical public data with additional features and functionality that support collecting more precise, complete, and unbiased

datasets," with access to the full historical archive of past Tweets. One of the friction points between platforms, researchers, and individual users is that users may expect platforms to reflect their intention to delete content, even removing that content from archives. Researchers may want the full archive. Platforms may reasonably choose to prioritize the choices of their users over the desires of the research community.

Platforms are also reluctant to share the algorithms they use to build and operate their recommendation algorithms, partially for privacy reasons (it is possible for neural networks to imprint on data in a way that is privacy revealing), but mostly because those algorithms are the company's intellectual property (IP). Sharing algorithms publicly would be a little like sharing the recipe for Coca-Cola publicly.

Josh Tucker (**NYU**) and **Surya Mattu** (investigative journalist) and others in the research community might **like to see more sharing** because consuming too much Coke or too much Twitter could be bad for some populations, but the companies are well-protected from divulging their trade secrets and other IP. Meanwhile, **Travis Greene**, writing in *Nature Machine Intelligence*, sees **an emerging void** in social science as digital platforms, like Twitter, undertake algorithmic behavior modification at global scale and don't allow researchers opportunities to investigate the behavior interventions.

The new Twitter API for Academic Research is a net positive, but it obviously doesn't allow researchers to fully understand how misinformation spreads because it's only one platform, and not the most widely viewed.
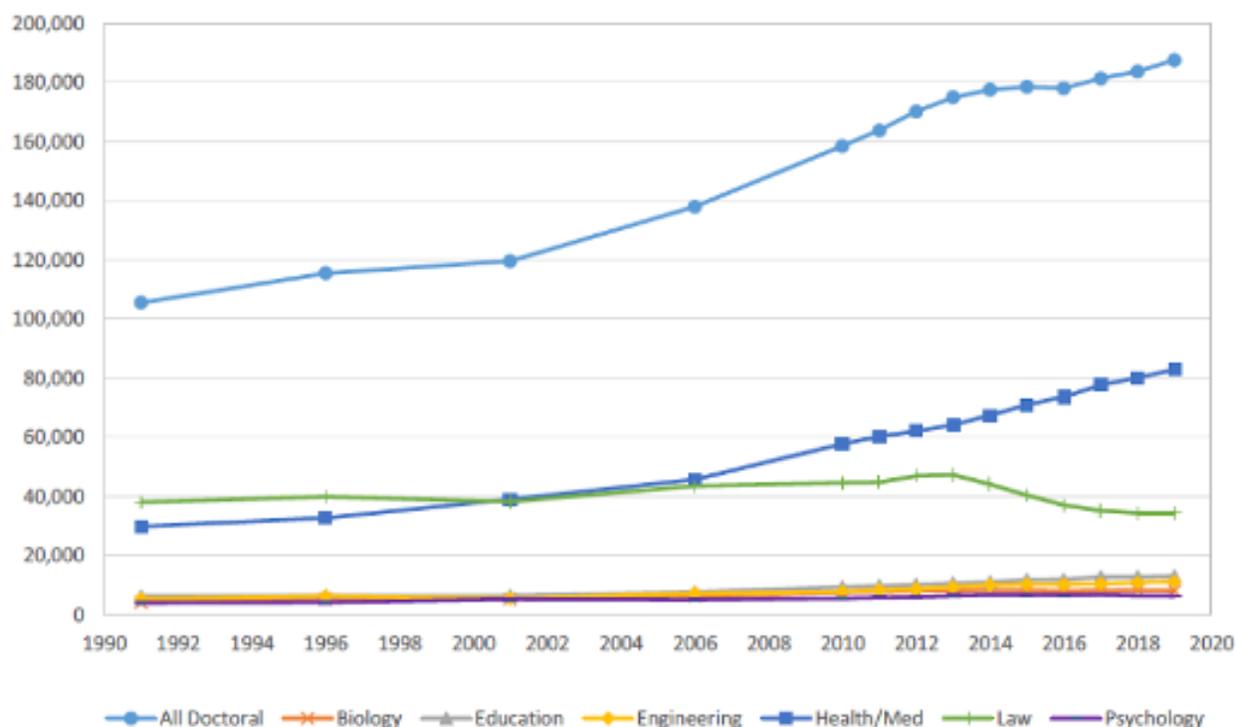
NEW PROGRAMS, FOLLOW THE MONEY
Click through to access **a structured spreadsheet** of New Programs and money moving around in academic data science.

DATA VISUALIZATION OF THE WEEK
NBER Working Papers, Data: Digest of Education Statistics by Michael F. Lovenheim and Jonathan Smith from April 7, 2022

Panel B. Doctoral Degrees Conferred

Deadlines

Conferences
Doing something cool in #ontology or #taxonomy? Join @hhedden, @uscholdm, @bolerio and others.
"Propose a talk for @DataDayTexas"

Contests/Award
The path to the #BellLabsPrize22 has never been simpler…
04/22 – Proposal deadline 06/03 – Selection announced 11/14 & 15 – Judging & ceremony

RFPs
NSF Convergence Accelerator Funding Opportunities
"The **NSF Convergence Accelerator** has released a new funding opportunity for three new research tracks: Enhancing Opportunities for Persons with Disabilities (Track H), Sustainable Materials for Global Challenges (Track I), and Food & Nutrition Security (Track J)." Deadline for letters of intent is May 31.

Tools & Resources
Fantastic writeup on the software architecture of Twitter.
Twitter, Peter Wang from April 19, 2022
"However, I think it's worth reflecting on how much of this complexity is driven by the single req that all writes to data structures (post/like/follow/etc) must synch within 0.5s." [thread]

[**ANES Announcement: Joint Study Linking General Social Survey and American National Election Studies Now Available**](#)
American National Election Studies from April 15, 2022
"Data released today make combining the General Social Survey (GSS) and American National Election Studies (ANES) possible for the first time."