ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for March 10, 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.

APPLICANTS ARE ASKING: IS GRAD SCHOOL "WORTH IT"? [SURVEY]

Readers with Masters, JDs, MDs, PhDs — how are you advising former undergraduates who are now thinking of getting Masters or PhDs (or MDs/JDs)? We've been hearing more trepidation about the *value* of grad school both as an investment in future earnings and as a possible challenge to life and relationships outside of grad school. This is a little different than previous cycles when questions focused on selecting the correct program to match specific interests and on how to prepare a successful application. **Take Survey**

UNDERGRAD IS DEFINITELY WORTH IT

Georgetown University's First Try at ROI project ranked 4500 colleges on their 40-year return on investment to undergraduates. Looks like undergrad definitely pays. Readers of this particular newsletter may be surprised to see that ROI *is* higher for private schools compared to public schools but *is not* higher for engineering schools compared to liberal arts colleges. From the abstract, “the median ROI of liberal arts colleges is nearly $200,000 higher than the median for all colleges. Further, the 40-year median ROI of liberal arts institutions ($918,000) is close to those of four-year engineering and technology-related schools ($917,000), and four-year business and management schools ($913,000)." Hop on over and use the tool to see how much your undergraduate degree might yield by the time you're ~65.

New efforts to make college **tuition free** at some public colleges, community colleges, and a handful of privates — either for all students, or for students whose families are not in the upper class — should also improve the ROI.

FACIAL RECOGNITION FOR ARCTIC ICE WEDGES

Researchers at the **University of Connecticut**, the **National Center for Supercomputing Applications**, and the **Woodwell Climate Research Center** are using **newly available** commercial imaging of the Arctic and deep learning algorithms originally developed for facial recognition applications to measure changes in the permafrost. Human faces and arctic ice wedges can both be understood as a set of polygons. **Chandi Witharana**, one of the project leads, said the new technique will give real-time evidence to "see how climate change is happening."

In a much warmer island climate, Hawaii will now benefit from **a single portal** with which to access

research, data, and modeling tools. The project will be run by the **University of Hawaii**.

This research and coordination comes as a **new IPCC report** shows that temperatures on Earth are already "1.2℃ warmer than in preindustrial times, some ecosystems are nearing a hard limit on their ability to adapt, including warm water coral reefs, coastal wetlands and rainforests, and the frigid mountain and polar realms." The range of maladies befalling humans has already begun and will accelerate: "up to 3 billion people could face water scarcity. Snowmelt for irrigation could decline by 20% in many river basins." The report notes that governments are moving too slowly to stop these outcomes and must dramatically accelerate their mitigation efforts.

BENGIO AND TEAM INTRODUCE NEURAL PRODUCTION SYSTEM (NPS)

In a **new paper** led by **Yoshua Bengio**, his team from **Mila** in Montreal, and researchers from **DeepMind** and **Waverly**, a departure from "equivariant" Graph Neural Networks (GNNs) is demonstrated. The new approach should be better adapted to sparse learning conditions and situations in which predictions are tied closely to properties of a given entity. In this case, NPS was trained on video data where it "comprises a set of rule templates applied by binding placeholder variables in the rules to specific entities, serving to factorize entity-specific and rule-based information in rich visual environments."

WEARABLES FOR PATIENT DIAGNOSIS

Niraj Jha, **Shayan Hassantabar**, and their teammates from **Princeton** have **developed** a bracelet that can predict whether someone has COVID within minutes. Future applications may include predicting a wide array of conditions from "depression, bipolar disorder, schizophrenia, diabetes and sickle cell disease" to "cardiovascular disorders and sepsis infections."

Meanwhile, **Stanford Health Care** has appointed a Chief Data Scientist, **Nigam Shah**, to **operationalize** uses of data science at patient scale. Shah's vision also includes using wearables to facilitate diagnosis, likely much more cost effective than over-reliance on imaging (which can also subject patients to radiation, depending on the imaging technique).

The **Wu Tsai Human Performance Alliance** based at **Stanford** was founded last year on the principle that studying peak human performance will help everyone to thrive, regardless of their daily step count. The new **Female Athlete Science and Translational Research (FASTR) program** is led by **Emily Kraus** (physician, runner, cyclist) and **Megan Roche** (epidemiologist, pro trail runner) and it focuses on the gender gap in human-performance research. When performance studies on male subjects are generalized to women, misconceptions follow, leading to misguided notions of appearance, training and nutrition that can be mentally and physically detrimental.

Down the coast in Los Angeles, **Cedars-Sinai Medical Center** has appointed **Sumeet Chugh** to head the new division of **Artificial Intelligence in Medicine** and **University of Southern California, Viterbi School of Engineering** researchers are also **looking into wearables** for diagnosis.

If your lab happens to be working on wearable prototypes, you probably want to check out this **novel small-scale prototype technique** developed at **University of California, Berkeley** by **Renxiao Xu** that can replace expensive photolithography fabrication with a $200 vinyl cutter. Not only is it cheaper, the technique also cuts the modification cycle time by 90%. Manufacturing the specific materials sensors required to implement all the various biosensing that has been tested in academic environments is one of the key challenges **Shana O. Kelley** **enumerates** as impediments to bringing sensor technology to market.

CORRELATION, CAUSATION & NEW "VIBRATING" VARIABLES TECHNIQUE

Chirag Patel, **Aleksandra Kostic**, and **Braden Tierney** have developed a **new technique** for

testing the robustness of models derived from typically smallish clinical data sets by perturbing their findings by inserting data from unobserved cases to see if the correlations in the original model hold. In other words, a study about the link between gut microbiome and a particular type of cancer controlled for age, smoking behavior, family risk, gender, and local pollution exposure. But unobserved variables may also be strongly correlated, an age-old problem in experimental design that can lead researchers astray.

In a recent computational paper, the authors tested their "vibration of effects" tool to show that "a full one-third of 581 reported microbe-disease associations were inconsistent, with outcomes changing depending on how the design was tweaked and which other variables were included in the analysis." Diabetes research was particularly inaccurate — “more than 90 percent of the research findings of studies exploring the link between gut microbes and type 1 and type 2 diabetes were inconsistent." At present, the approach is computationally brutal: "the team ran more than 6 million statistical modeling strategies on the findings of previous studies, adding and subtracting variables and testing different combinations of variables." (An earlier application of the vibration of effects model debunked the claim that Vitamin D ingestion has a therapeutic effect on COVID infection.)

ERIC TOPOL WARNS AGAINST COVID COMPLACENCY

As many in the epidemiological community believe, we are at a lull in the COVID storm rather than the dawning of a new COVID-free era. Topol is the founder and director of the **Scripps Research Translational Institute** and professor of Molecular Medicine. "Any proclamation that the pandemic is over ignores the potential recrudescence of a new variant with high transmission and immune escape. We will still benefit from using masks for many situations including protecting immunocompromised and vulnerable people," he **wrote** on his personal Substack. "The metrics do not lend any support to the mission of containment accomplished." He goes on to note that the US has been particularly sluggish about getting 2nd and 3rd shots compared to countries in Asia and Europe: "the United States ranks 55th in the world for booster shots and 67th for 2-shot vaccinated." The conclusion: “Premature cessation of mitigation measures, particularly in many regions of the country with high levels of circulating virus, leaves our immunocompromised population susceptible" and "being entranced with a false sense of security, will just add to a historic big miss."

One of the most promising early warning systems has come from wastewater analysis, which has been able to return results for community-level virus load faster and more reliably than testing. Wastewater analysis emerged from campuses so it is no surprise that a university research group at **University of California, Berkeley** has recently published a how-to **paper**, explaining how to set up a “wastewater testing laboratory including standard operating procedures, laboratory buildout and workflow, and a quality assurance plan". Hopefully, public health utilities, more campuses, and other agencies can rapidly implement this plan to make up for the growing gaps in testing surveillance as people turn to at-home tests that are much more convenient, but do not get officially recorded.

ELSEVIER VS. RESEARCHGATE

In 2017 **Elsevier** **sued** **ResearchGate** in a German court, claiming that ResearchGate violated Elsevier's copyrights by hosting 50 research papers on the ResearchGate website. The court has **finally ruled**, though both sides plan to appeal. ResearchGate was found to be in breach of copyright by hosting the papers, which had already been taken down, even though the papers were uploaded to the site by the papers' authors. No damages were paid and the court opened a question about whether a corresponding authors' agreement to copyright terms applies to all authors. In the meantime, two major competitors to Elsevier have decided to partner with ResearchGate instead of suing, Elsevier has added **trackable fingerprints** to all pdfs downloaded from its site (to better capture copyright violators), and various countries have **moved toward** mandates that all research funded by national bodies be publicly available for free.

TEACHING RESOURCE — RESPONSIBLE TECH DEVELOPMENT

MIT has added to its collection of OpenCourseWare (OCW) with a **new set** of "original active learning projects, homework assignments, in-class demonstrations, and other resources and tools" designed to give students the skills and mental models to address the social impact of technology.

RESEARCH TIES TO RUSSIA — MIT UNDER INTENSE PRESSURE

Longstanding ties — including a board seat for a Russian oligarch that was quietly retired — and several donations of $200 million, $300 million, and recurrent deposits of $50 million from Russian-backed sources to **MIT** have resulted in a **public accusation** in *Fortune* penned by researchers at **Yale** this week. **Steven Tian**, **Anjani Jain**, and **Jeffrey Sonenfeld**, all at Yale, had been writing directly to MIT about the national security implications of their ties to Russia for years. The authors explain, "MIT's collaboration with Russian institutions includes sensitive domains of national security. Moreover, there has been no public acknowledgement by MIT of hundreds of millions of dollars of opaque payments the institution is reported to have received from Russia for engaging in these collaborations." Notably, MIT was also in similar trouble when it turned out **Jeffrey Epstein**, after his first conviction for sex crimes, had donated money to the **MIT Media Lab**. (Full disclosure: This writer, an MIT alum, called for the school to provide an audit of all non-governmental funding at that time.)

NATURE OPENED THE DOOR ON PEER REVIEW

For about a year **Nature** has been allowing authors to make their peer reviews public; peer reviewers can choose to remain anonymous or reveal their names. In this inaugural period, 46% of papers published are accompanied by peer reviews and author exchanges. Some disciplines have even larger adoption rates. Genetics and genomics had 66% of its papers published alongside peer reviews. More authors are expected to publish reviews in 2022 which has Nature's editorial board **exuberantly pleased**: "the benefits to research are huge. Opening up peer review promotes more transparency."

WHY DO BATTERIES DIE?

William Chueh and **Haitao "Dean" Deng** at **Stanford** already knew that batteries wear out because of mechanical deformations that accrue during charging. Batteries physically degrade over time, but nobody quite knew what this looked like at the atomic level. With high-resolution microscopy they were able to image the atomic interactions. That data was trained to inform an AI application to look specifically at lithium-ion phosphate batteries. The **eventual goal** is to find battery compositions that can withstand charge/discharge cycles with less mechanical degradation.

HOW TO IMPROVE THE COMMENTS SECTION, LESSONS FROM 13 MILLION EXAMPLES

Researchers **Yixue Wang** and **Nick Diakopoulos** **analyzed 13 million comments** left on articles at **The New York Times** website. The dataset has a nice built-in feature — some of the comments were selected as high-quality by NYTimes editorial staff. That training set was paired with a similar set of comments that raters deemed as equally good to determine the impact of being selected on the same commenter's future propensity to comment, and to do so in a high quality way. Researchers found that getting selected by the NYTimes led commenters to comment with higher quality than they might have, at least for a while before they sink back to mediocrity. The lesson for product designers is fairly clear; place higher quality comments at the top of the feed. Let commenters know they have met a high bar for quality comments after their first high quality comment, then re-up that encouragement somewhat later.

ROBOTS AND SENSORS FOR INFRASTRUCTURE

A team of young people at the **Carnegie Mellon** spinout company **Mach 9 Robotics** have designed a rig that can detect signs of imminent bridge failure in Pittsburgh. The truck is equipped with sensors that are similar to an x-ray for concrete and rebar structures. Resulting imaging combined with visual inspection allows the team to diagnose a bridge's condition much the way a radiologist and ER doctor might diagnose a patient. With the imaging rig, results eliminate a lot of guesswork.

The **University of New Mexico** has a **robot** that can be sent out to tap rocks along the roadside so that inspectors don't have to risk their lives figuring out which boulders are about to come rumbling down. This is helpful at a time when a large bolus of infrastructure nears the end of its lifespan and more mudslides and other weather events destabilize hillsides.

Cornell researchers trained an AI to **assist** the construction of new hydroelectric dams in the Amazon basin and to reduce their impact on the surrounding environment.

Seoul Robotics, a startup funded by the South Korean government, **adapted** its LIDAR computer vision system developed to steer robot cars for infrastructure use. Touted as a giant leap from video cameras, the system tracks identified and unidentified objects with four centimeter accuracy — good for assembly lines, vehicle traffic, retail customers, super-spreading undergrad Spring breakers, whatever moves.

NEW PROGRAMS, FOLLOW THE MONEY

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

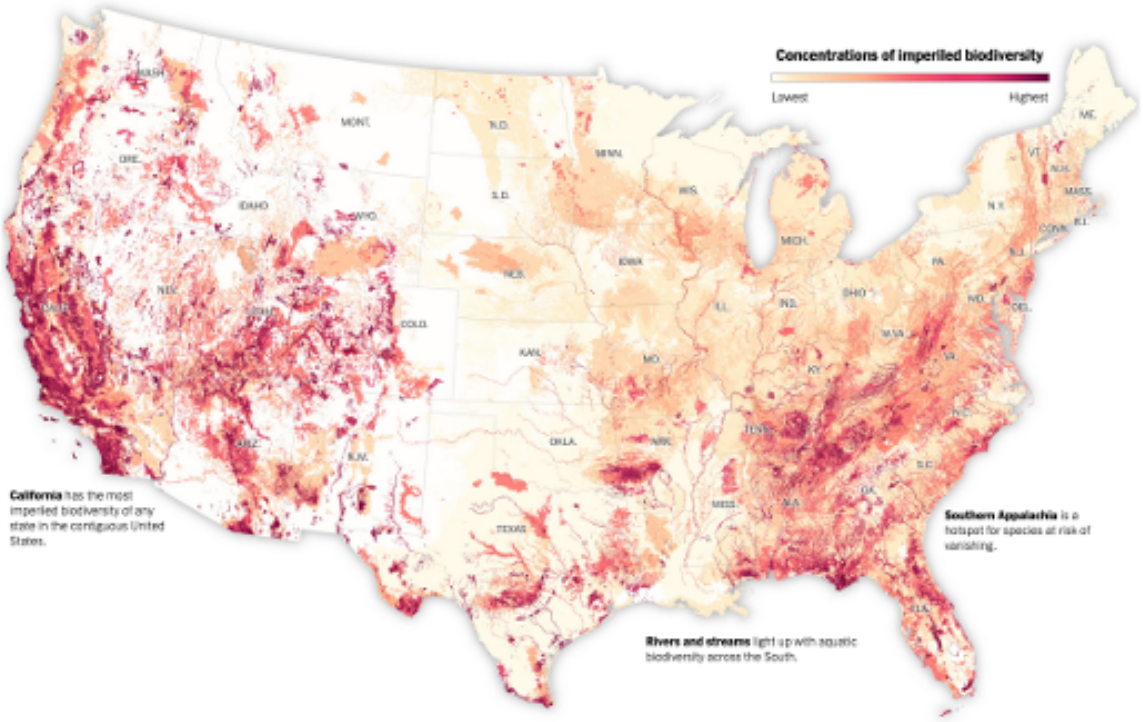
DATA VISUALIZATION OF THE WEEK

The New York Times, Catrin Einhorn and Nadja Popovich from March 3, 2021

This Map Shows Where Biodiversity Is Most at Risk in America

By Catrin Einhorn and Nadja Popovich March 3, 2022

Let your eyes wander to the areas of this map that deepen into red. They are the places in the lower 48 United States most likely to have plants and animals at high risk of global extinction.



[Full-text Original paper]: "Increasing taxonomic diversity and spatial resolution clarifies opportunities for protecting US imperiled species" in *ESA Ecological Applications* journal online, January 19, 2022.

Deadlines

Studies/Surveys

Attention researchers who are studying health misinformation, this is a great opportunity to inform policy making in this space.

"Today, I'm [**Dr. Vivek Murthy**, U.S. Surgeon General] making an official request to the general public for stories and research on #HealthMisinformation. This includes a public call to tech companies to transparently share what they know about the impact of misinfo on their platforms."

Conferences

NeurIPS 2021 Call for Papers

"Abstract submission deadline: Monday, May 16"

Education Opportunities

Apply to participate in NeuroHackademy 2022

"Applications are due by April 4th."

Contests/Award

The Information is Beautiful Awards are Back!

"During our State of the Industry panel at Outlier, we welcomed a video address from a special guest: **David McCandless**. Founder of the Information is Beautiful (IIB) Awards. He shared big news with our community: after a two-year hiatus during the pandemic, the Information is Beautiful Awards will return, hosted by the **Data Visualization Society**." Deadline tbd.

Tools & Resources

This is one of the most interesting GNN results I've seen. Insightful, general, and useful!

Twitter, Aaron Clauset, Hamed Hassani from March 8, 2022

"Can GNNs compute the shortest path? Or min-cut? Or more generally, are GNNs aligned with dynamic programs? We provide an answer in [https://arxiv.org/abs/2202.08833]; joint work with Mohammad Fereydounian, @JDadashkarimi, and @aminkarbasi "

Announcing AI Blueprints Public Preview

The Next Web, Ben Dickson from March 6, 2022

"An alternative to manual design is "neural architecture search" (NAS), a series of machine learning techniques that can help discover optimal neural networks for a given problem. Neural architecture search is a big area of research and holds a lot of promise for future applications of deep learning."

You're using email wrong

Arne Bahlo from March 6, 2022

"As controversial as the company behind HEY is, the concept that they introduced with their mail app has fundamentally changed how I think about email. I adapted the part that resonated with me to my Fastmail account like this: Inbox, Papertrail, Newsfeed. "

Featured Events

See the ADSA Events Page for more details and 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.