

## ADSA Data Science Community Newsletter

**Data Science Community Newsletter** features journalism, research papers and tools/software for August 11, 2022.

Please let us ([Micaela Parker](#), [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](#).

### IS THE PLANET ON THE BRINK OF SOCIETAL COLLAPSE?

"One can think of the nine planetary boundaries as credit cards, [with] six of those nine credit cards charged to the hilt to develop civilization as it now exists," writes **Bob Henson** for **Yale Climate Connections** (see also: [Data Viz of the Week](#)). Those are drought, war, sea-level rise + land subsidence, pandemics, collapse of Atlantic Meridional Ocean Current (AMOC), ocean acidification, and some as-yet unknown black swan climate surprise.

The linked [piece](#) details each scenario, but the sea level rise stood out: "the **Global Commission on Adaptation** estimated that sea level rise will lead to damages of more than \$1 trillion per year by 2050. Furthermore, sea-level rise, combined with other stressors, might bring about megacity collapse – a frightening possibility with infrastructure destruction, salinification of fresh water resources, and a real estate collapse potentially combining to create a mass exodus of people, reducing the tax base of the city to the point that it can no longer provide basic services." [Jakarta is winning](#) this particular race to the bottom. The halting of the AMOC also stood out, in this case because the collapse could be particularly imminent and would cause (more) drought across the Northern Hemisphere. A 2021 study "looked at eight independent measures of the AMOC, and found that all eight showed early warning signs that the ocean current system may be nearing collapse." Should the AMOC slow dramatically or stop altogether, the UK might see temperatures below freezing in July and August.

In the US west, the Great Salt Lake [has shrunk](#) from 3000 acres to 1000 acres, on pace to [disappear entirely](#) within the next decades, leaving that heavily populated area in drought, devastating a crucial migratory bird route, and releasing arsenic and other toxic heavy metals left over from mining operations that had been sealed in by the water. Further south, Lake Powell and Lake Mead are also [dropping precipitously](#), threatening agriculture, the [hydroelectric power supply](#), and household water supplies. Lake Powell currently only holds 27% of its capacity. In the Great Lakes, invasive species brought into the ecosystem by international shippers have [largely supplanted](#) native fishes.

In a glimmer of good news, the **U.S. Congress** is poised to pass a [piece of climate legislation](#) that may cut carbon emissions by 40%. For the researchers among us, another interesting bit of climate news out of the federal government is that the **Office of Financial Research** with the **Department of Treasury** has [launched](#) the **Climate Data and Analytics Hub**. This hub will initially only be available to the **Federal Reserve Bank of New York** and is designed to bring climate data into models that can

estimate the financial impact of the massive changes facing the planet.

Last week we noted that 1.9% of the land use in the US is dedicated to turf grass (e.g. lawns) which is such a water-guzzling way to live that lawns have been banned in many southwestern US communities.

## WILDFIRE RISK MAP IN OREGON IS WITHDRAWN FOR POLITICAL REASONS

After devastating wildfires hit Oregon in 2020, the legislature called for new wildfire risk maps that the **Oregon Department of Forestry** created with **Oregon State University**. When those maps were released earlier this year, they identified 80,000 lots at extreme risk of wildfire impact, and provided guidelines for roofing materials and defensible space that those property owners would have to follow should they undertake new construction or major upgrades.

A **huge uproar ensued** — property owners were concerned insurance rates could go up or become unavailable altogether. This is more or less the path we've predicted about how climate change will be rolled out to property owners – as a wake up call in the form of higher insurance rates (or policy non-renewal). What we didn't predict was the role universities and governments would play in producing risk maps, or that those maps would be withdrawn. Frankly, as far as the insurance industry is concerned, a map once produced cannot be unseen, so the home owners' fears about insurance rates may be substantiated whether or not local governments are able to enforce updated building codes. In fact, insurers may be even more reluctant to insure in an area that rejects risk mitigation efforts, but that will remain to be seen.

SPONSORED CONTENT



## FACULTY AT UF EYEING THE EXITS

A **survey** sent to 2000 **University of Florida** faculty (response rate =31%) reveals that recent erosions of academic freedom by **Governor Ron DeSantis** are still having a strong negative impact on morale. Nearly half have concerns about their academic freedom. Almost three-quarters of the respondents (74%) expressed concerns over whether the UF board of trustees "ensures that the university is free from undue political influence." As reported previously, UF is currently searching for a new president. Yet 70% of faculty "expressed a lack of confidence that the board will select a new president who 'prioritizes academic, scholarly and faculty interests'." It's tough to watch this from afar and undoubtedly tougher to be directly impacted.

## DATA SCIENCE AND THE BODY

Two major stories worth mentioning. In the first, a team from **MIT** led by **Xuanhe Zhao** has **created** a

small ultrasound sticker (e.g. like a band-aid) to provide continuous ultrasound imaging outside the hospital context. The main scientific breakthrough was in creating the right elastomeric substances to facilitate the recording of sound waves and to get the sticker to transmit properly. In terms of applications, the proof of concept focused mostly on diet and fitness considerations such as the changing diameter of major blood vessels and the heart muscle during exercise, stomach distention during consumption, and muscle tissue microtears during weight lifting.

The second piece of news about data science and the body also has an MIT connection and is more of an update. A team of scientists led by MIT biologist **Aviv Regev** at the **Broad Institute** in Cambridge, MA, have endeavored to **catalog all the cell types** in the human body. The idea to create a cellular atlas started in 2016. It followed in the wake of the Human Genome Project, another Broad Institute-led effort, and has an exceptionally successful parallel in the proteomics world. On the protein atlas, phase 1 has largely met its goal. **Alpha-fold AI** is capable of producing the structural simulation of every known protein in the human body.

One last related update that isn't exactly about the body, but *is* about health data: **Norway, Sweden, Finland, Iceland, and Denmark** have adopted a digital health data common standard. Called the **Nordic Digital Health and Evaluation Criteria** (NordDEC) programme, it is a world-first programme to unify digital health standards across multiple countries. Extracting meaningful findings out of standardized data is generally easier than the alternative, but we've seen such lengthy delays and derailments of electronic health record promises that we are waiting to see what this announcement amounts to. If history is a guide, we will be waiting for years.

## DO CENTERS/INSTITUTES FOR DATA SCIENCE NEED ETHICS BOARDS?

**Northeastern University** has **assembled** 20 experts — a mix of internal and external AI practitioners — to form an AI ethics board chaired by **Cansu Canca** that will provide guidance for AI research activities. This type of oversight is not completely novel in university research. Institutional Review Boards (IRBs) are a decades-old compliance mainstay, but they typically don't have the expertise to contemplate the specific considerations of AI research and application. And Canca is quick to point out that the AI Ethics Board will actively partner at the researcher and center-level to, "enhance the technology with all these multidisciplinary capabilities that we have."

## MEN MORE LIKELY TO SELF-CITE THAN WOMEN

In **a study** authored by **Vincent Traag** of citation practices across disciplines represented by journals on **JSTOR**, researchers found that men are more than 50% more likely to self-cite than women. However, **a similar study** looking at **Authority** (a **PubMed** source) with different controls found that the gender difference was largely a factor of age and career continuity, "papers by authors with short, disrupted, or diverse careers miss out on the initial boost in visibility gained from self-citations. Our data further suggest that this disproportionately affects women because of attrition and not because of disciplinary under-specialization."

But the real headline in self-citation over the past couple weeks was...unfit to summarize here in the DSCN due to the material's off-the-charts ick-factor (**link**, second warning: not for those who are already upset with the world). I haven't seen sociologists so quick to distance themselves from an anthropologist since the reaction to **Bronislaw Malinowski's** research with Trobriand islanders.

## RESEARCHERS TO CENSUS: STOP USING DIFFERENTIAL PRIVACY

A **letter** signed by an unknown number of researchers to **U.S. Census Bureau** director **Robert Santos** has **requested** that the agency **stop using differential privacy** on releases of annual population estimates and the **American Community Survey**. Their main complaints are that applying differential privacy techniques introduces significantly impactful delays; removes certain key variables (such as naturally small N variables like occupants per household and fine-grained racial data); there is too little known, understood, and currently understandable about statistical error rates that likely present equity issues; and a lack of "intelligible guidance" available regarding the use and underlying properties of data with DP applied.

This issue is not without controversy. Census reps have consistently upheld their adoption of differential

privacy (DP) as the only available, statistically robust strategy for distributing data without leaking private information either directly or through linkage with third-party data (of which there is plenty). Social scientists who are interested in studying naturally small N populations are particularly frustrated by DP. But these researchers also tend to realize that numerically small groups are have the most to lose if their personal information is deidentified.

## TRAINING SELF-DRIVING CARS WITH ROAD-PRINTING

**Cornell University** self-driving car researchers operate in a part of the country that gets a lot of snow. The snow changes the way roadways look, even in somewhat structural ways (a giant snow bank is a structural difference *and* a visual difference). A model that may have identified curb-like entities as entities that shouldn't be driven over may stop on snowy roads with 6" or more of snow plow troughs.

The team led by **Kilian Weinberger** have [started to incorporate](#) the ability of hyperparameter models to imprint specific features, in this case specific features of particular roadways. In this imprinting instance, the model will rely on its experience with a specific roadway for navigation rather than its summation of roadway features collected across all roadways, all times. I suppose this might work in combination with some kind of modified federated learning in which each vehicle can imprint on the typical traversals it takes while being blocked from updating the federated model in terms of those specific imprints. Drivers are habitual — commutes, errands, kid drop-off all traverse the same routes over and over again within a given car. Curious to see where this will go. And I'd note that imprinting is typically seen as a challenge for privacy. If that federated model example sent the imprinted routes back up to the main model, they would reveal a great deal about that particular car's daily location.

## REANIMATING DEAD SPIDERS AS ROBOTS

Researchers at **Rice University** have [applied nanorobots to dead spiders](#) to turn the spiders into pincer-like claws. Why? Dead spiders have a "perfect architecture for small scale, naturally-derived grippers." We have coined the term Zom-bots to describe this unholy union. They use the term necro-bots. Please provide alternative coinages!

## TIKTOK IS EXPOSING AN API TO "SELECT" RESEARCHERS

The infinite scroll of short, often humorous videos created by millions of average folk (and some celebrities) known as **TikTok** has exposed an [API for research partners](#). Thanks to our favorite TikTok creator (see her [TikTok playlist](#)), professor **Casey Fiesler**, for the update.

## HYBRID WORK — IT WORKS

A [new study](#) on hybrid work among knowledge workers found that those randomly selected into a hybrid work pattern had substantially lower attrition rates and moderately improved productivity compared to those who didn't consistently work from home during the study period. The results were strong enough and positive enough that hybrid work was extended to the entire company.

## MOST US FACULTY SUPPORT OPEN ACCESS PUBLISHING

In a [new survey](#) of faculty support for open access publishing, "Nearly three-quarters of faculty members under age 44 supported shifting to OA publishing, but older respondents were less enthusiastic. Sixty-three per cent of those aged 44–54 said they were in favour of such a move. Only 57% of those over 65 supported an OA model." The [survey](#) was conducted by **Ithaka S+R**.

## FEATURED AUDIO: TOPOL INTERVIEWS HASSABIS

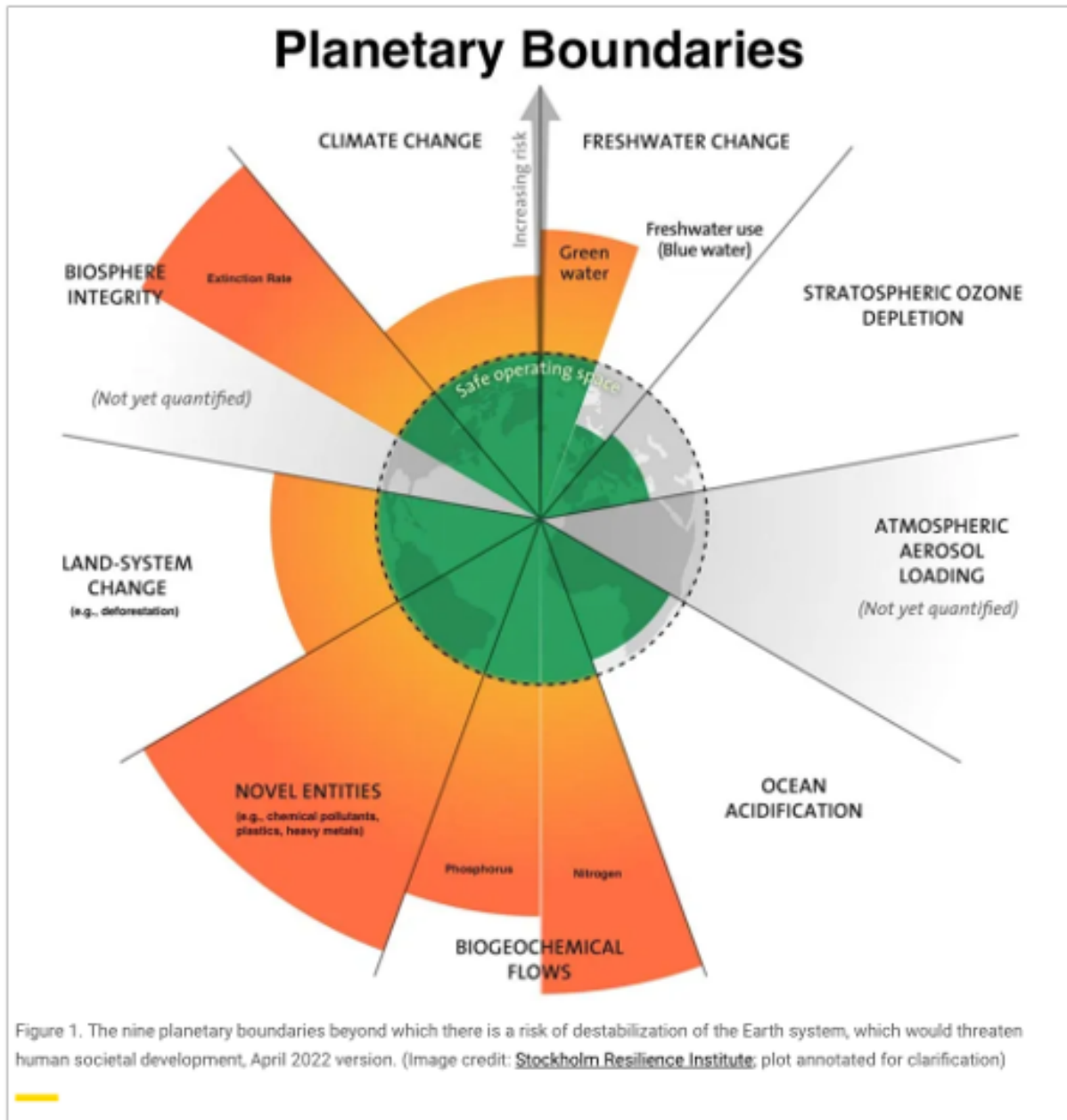
In our [featured audio for the week](#) — we're trying a new featured element! — **Eric Topol**, head of **Scripps Research Institute** in San Diego interviews **Demis Hassabis** of **DeepMind**.

## 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

by Stockholm Resilience Institute via Yale Climate Connections from June 22, 2022



### Deadlines

#### Studies/Surveys

##### Data Labeling for ML: Survey

"About 45% of the time in data science projects is consumed by processing and labeling data. It's fair to say that data labeling is one of the most expensive tasks of any machine learning project. How to work with data properly when preparing it? What are the best labeling methods and tools to use in machine learning solutions today?"

### Education Opportunities

[Paging all science & tech journalists!](#)

"The Computational Journalism Lab @NorthwesternU is looking for participants for a study about helping science journalists uncover potentially exciting leads from the latest research."

## Contests/Award

### U.S./U.K. Launch PETs Innovation Prize Challenges

"First announced at **President Biden's** Summit for Democracy, the U.S./U.K. privacy-enhancing technologies prize challenges, are now open for registration." Deadline to register and submit abstract is September 4.

## Tools & Resources

### AI Book Recs: Add These to Your Reading List

*Stanford University, Stanford Institute for Human-Centered Artificial Intelligence, Shana Lynch* from August 3, 2022

"Our HAI community offered up the best books in AI that they're reading. "

### Six tips for better spreadsheets

*Nature, Technology Feature, Jeffrey M. Perkel* from August 2, 2022

"Microsoft Excel and Google Sheets are powerful and widely used. But there's a right way and a wrong way to use them, data scientists say."

### Turn Your Adventures Into Scientific Expeditions

*Adventure Journal, The Revelator, Tara Lohan* from August 5, 2022

"I loved being out there, using my outdoor skills and actually helping — feeling like I was making a difference." ... "Every project we do is designed in partnership with a scientist or multiple. It's them saying, 'We need these data to solve this problem or to address this issue.' We couldn't do this work without incredible scientists who are trying to solve really big issues."

**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.